The MIT Press

Are link-based and citation-based journal metrics correlated? An Open Access megapublisher case study

Enrique Orduña-Malea[1] (iD) and Isidro F. Aguillo[2] (iD)

[1]Department of Audiovisual Communication, Documentation and History of Art, Universitat Politècnica de València, Valencia, Spain
[2]Cybermetrics Lab, Institute of Public Goods and Policies (IPP), Spanish National Research Council (CSIC), Madrid, Spain

ABSTRACT

The current value of link counts as supplementary measures of the formal quality and impact of journals is analyzed, considering an open access megapublisher (MDPI) as a case study. We analyzed 352 journals through 21 citation-based and link-based journal-level indicators, using Scopus (523,935 publications) and Majestic (567,900 links) as data sources. Given the statistically significant strong positive Spearman correlations achieved, it is concluded that link-based indicators mainly reflect the quality (indexed in Scopus), size (publication output), and impact (citations received) of MDPI's journals. In addition, link data are significantly greater for those MDPI journals covering many subjects (generalist journals). However, nonstatistically significant differences are found between subject categories, which can be partially attributed to the "series title profile" effect of MDPI. Further research is necessary to test whether link-based indicators can be used as informative measures of journals' current research impact beyond the specific characteristics of MDPI.

1. INTRODUCTION

The advent of the Worldwide Web (Berners-Lee, Cailliau et al., 1992) facilitated scientific journals in experiencing a digital transformation, shifting from the Gutenberg galaxy to the internet galaxy (Castells, 2002). The creation of journal websites not only allowed publishers to create new scholarly communication channels for readers but also facilitated metaresearchers to capture a wide range of online metrics related to both on-site (e.g., visits, downloads, reads) and off-site (e.g., mentions, links) events (Orduña-Malea & Alonso-Arroyo, 2017). The chance of measuring (massively) new journal–reader interactions led the scientific community to investigate the role of journal websites in the access and dissemination of scientific research (Vaughan & Thelwall, 2003), and to design and test new web-based journal-level metrics to complement citation-based metrics in the research assessment of scientific impact.

An example was the case of the Usage Impact Factor, an indicator aimed to mimic the operation of the Journal Impact Factor by using server web log data (on-site metrics) instead of citations (Bollen & Van de Sompel, 2008). The negative correlation found between usage data and the Journal Impact Factor helped to spread a multidimensional notion of scholarly impact (Bollen, Van de Sompel et al., 2009).

The application of web usage data at large scale was, however, limited by several technical aspects that jeopardized its use, especially data accessibility (i.e., permissions are needed from webmasters), data coverage (i.e., limited number of journals systematically collecting data), and data accuracy (i.e., fair comparisons were compromised). Although practical standards for reporting and transmitting usage statistics recorded by scholarly publishers were proposed, such as SUSHI (Chandler & Jewell, 2006; NISO, 2014) or COUNTER (Shepherd, 2006)[1], the problems already pointed out are still valid.

Link-based metrics (off-site metrics) have been also examined as potential signals of scientific journals' impact, constituting the basis on which this study is based. As with usage data, early studies did not yield positive correlations between link data and Journal Impact Factors (Harter & Ford, 2000; Smith, 1999). However, subsequent works evidenced a significant correlation (Vaughan & Hysen, 2002; Vaughan & Thelwall, 2002), probably due to the evolution of the Web and the improvement of the available link data sources. The size, age, and discipline(s) covered by the journals were found to be variables determining the number of links received by journal websites (Vaughan & Thelwall, 2003).

However, counting the number of links received showed both general and specific practical limitations.

As regards the general limitations we can highlight the proper interpretation of the motivations to create links (Bar-Ilan, 2005; Thelwall, 2003), link obsolescence, spam, and the dependence on link data providers (Thelwall & Kousha, 2015).

With regard to the specific limitations, online access from subscription-based services limits obtaining links (Thelwall, 2012). In addition, the use of journal management services favors the creation of long unfriendly URLs, which hinders link discoverability. Moreover, the creation of different web domains (e.g., a web domain to host the Open Journal System platform and another one to host the official journal website) scatters the web impact, making the measurement of links received difficult (Orduña-Malea, 2019).

Nevertheless, it is the use of the Digital Object Identifier (DOI) which introduces the major practical limitation, as journal articles are commonly linked to through DOI URLs. Despite some journals creating customized DOI URL versions[2], the pure DOI URL version belongs to an independent web domain (doi.org)[3], generating a remarkable loss of links received by the journal websites. For example, the URL path "doi.org/10.3390," which belongs to the MDPI publisher, receives near 18 million links to their publications according to Majestic's historic index (as of January 4, 2022).

Because the DOI was adopted as an international standard in 2012 (ISO, 2012), increasing its use massively since then, those pioneering studies on journal websites did not face this current web visibility problem.

Journal websites constitute online research objects that provide users with scientific results along with other informative content. Therefore, their web design, published contents, and web dissemination can exert an influence on the number of users who discover, access, and consume the available content (Codina & Morales-Vargas, 2021).

The creation of a link from any webpage to a journal website implies not only the potential interest of the webmaster in making the journal website visible to users but also the possibility

---

[1] COUNTER v. 5.0.2 was published on September 28, 2021. https://cop5.projectcounter.org/en/5.0.2/index.html.

[2] For example, https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/epi.2015.sep.08.

[3] For example, the following URL (https://doi.org/10.3145/epi.2015.sep.08) counts for the "doi.org" web domain, not to "profesionaldelainformacion.com," the web domain of the corresponding journal.

of driving users (web visitors) to the site (Thelwall, 2012). This might turn into article downloads, reads, and eventually, citations. In addition, the number of links received by websites, especially from trusted websites, is used by search engines' algorithms to determine the position of the linked websites in the search engine results pages (Ledford, 2008, p. 11), thereby increasing the chances of being clicked and accessed.

Even though links may have been generated for nonacademic reasons, such as promotions or gratuitous links (Thelwall, 2003), those links from the academic web (other journals, universities, research societies, research blogs, etc.) or highly reputable websites (government entities, large companies, media, informational resources) might acquire great value and significance to evidence nonscholarly uses of research (Thelwall, 2012). For that reason, counting links to a journal website, especially those from reputable web domains, provides signals about the journal website's impact and influence.

The relation of link-based indicators with citation-based indicators at the journal-level constitutes the main objective of this work. The absence of correlation between these two types of indicators would imply that link-based indicators do not yield signs of scientific impact, providing distinct information in relation to the impact of the scientific content published by the journal. However, a strong correlation might imply that link-based indicators might bear evidence of scientific impact.

Because link-based metrics operate at higher orders of magnitude than citations, are generated (and can be collected) almost instantaneously, and provide information about the wider impact of academic research (Thelwall, 2012), their calculation and monitoring could serve to provide complementary evidence of scientific journals' impact.

Give the evolution of the Web during the last 15–20 years, the increasingly complexity of the journal websites, and the emergence of the DOI (and other article IDs), it is deemed necessary to revisit journal websites studies to determine the current value of link counts as supplementary measures of journals' research impact.

To accomplish this objective, the following research questions are drawn:

- (RQ1) Are link-based metrics related to the formal quality of journals?
- (RQ2) Are link-based metrics related to the discipline covered by the journals?
- (RQ3) Are link-based and bibliometric-based journal metrics correlated?
- (RQ4) Where do links to journal websites come from?

To carry out this study, the Multidisciplinary Digital Publishing Institute (MDPI) publishing house will be used as a case study.

2. METHODS

2.1. MDPI as an Open Access Megapublisher Case Study

Based in Basel (Switzerland), MDPI (originally Molecular Diversity Preservation International) was launched in 1996 as a nonprofit institute for the promotion and preservation of the diversity of chemical compounds, evolving into an open access publisher in 2010 under a new name (Multidisciplinary Digital Publishing Institute).[4]

The MDPI publishing portfolio covers all research disciplines, comprising 352 peer-reviewed journals and near 600,000 articles published (as of July 2021), being one of the

---

[4] https://www.mdpi.com/about/history.

major open access commercial publishers, along with BioMed Central, Frontiers in … and Hindawi (Rodrigues, Abadal, & de Araújo, 2020).

The use of MDPI as a baseline for the journal-level link analysis is supported by the following considerations.

First, all journals are created using the same web template, including—with slight variations—the same journal sections and information architecture (Codina & Morales-Vargas, 2021), and sharing the same URL syntax (e.g., mdpi.com/journal/agriculture), which avoids variability due to web quality features.

Second, the number of journals available (352) is large enough for comparative purposes, being also able to filter by journal age, discipline, and formal quality (i.e., whether the journal is indexed in prestigious bibliographic databases).

Third, all the articles published by MDPI are made immediately available worldwide under an open access license, favoring the obtaining of links.

2.2. Data Collection

The bibliographic data related to all journals published by MDPI (name, ISSN, release year, total number of articles published, website URL) were collected from the publisher's official website[5] as of July 18, 2021, yielding 352 journals. Proceedings-based journals were excluded to keep the sample as homogeneous as possible[6].

The thematic classification of journals was established through the 10 subject categories established and assigned by MDPI. Subsequently, each journal was labeled as specialized (assigned to only one subject category), multidisciplinary (two, three, or four categories), or generalist (assigned to five or more subject categories), as Table 1 shows.

Although the cutoff between multidisciplinary and generalist journals is rather loose (five categories, the 50% of all categories used by MDPI), it helps in distinguishing between those journals admitting publications from many disciplines on the one hand, and those journals accepting publications from few disciplines without being purely specialized journals on the other.

The Majestic database was used as a source for link-based data. Each link from a website (hereinafter referred to as source URLs) to each of the 352 MDPI journal websites (hereinafter referred to as *target URLs*) were gathered through the historic index[7] as of July 17–18, 2021, which yielded a total of 1,084,805 raw links[8].

A data cleaning process was necessary to solve inconsistencies, such as robot failures due to crawling loops[9], name changes of journals[10], or web redirections[11]. Finally, all links

---

[5] https://www.mdpi.com/about/journals.

[6] https://www.mdpi.com/about/proceedings.

[7] This database covers more than 3,580 billion URLs since 2006.

[8] The source URL generates outlinks, and the target URL receives inlinks.

[9] For example, the following source webpage is due to a loop, and consequently, was removed. https://www.easn.net/newsletters/issues/taxonomy/term/44/all/feed/feed/feed.

[10] The journal *Microarrays* changed to *High-throughput*, and finally, to *Biotech*. These three journals were merged for link purposes.

[11] "clinicsandpractice.org" redirects to the MDPI journal *Clinics and Practice*; "current-oncology.com" redirects to the MDPI journal *Current Oncology*; "scipharm.at" redirects to the MDPI journal *Scientia Pharmaceutica*; and "tomography.org" redirects to the MDPI journal *Tomography*. All links from these websites to their corresponding journals were considered self-links, and consequently were removed.

Table 1. Subject categorization of MDPI journals

Profile	LABEL	SUBLABEL	Discipline	N1	N2	N3
Specialized—strictly one discipline	A	A1	Biology & Life Sciences	23	132	63
		A2	Business & Economics	6	30	13
		A3	Chemistry & Materials Science	18	97	47
		A4	Computer Science & Mathematics	13	59	36
		A5	Engineering	12	100	52
		A6	Environmental & Earth Sciences	5	69	35
		A7	Medicine & Pharmacology	6	90	36
		A8	Physical Sciences	8	52	24
		A9	Public Health & Healthcare	6	84	32
		A10	Social Sciences, Arts and Humanities	9	39	10
			SUBTOTAL	106		
Multidisciplinary Two, three, or four disciplines	B			230		
Generalist More than four disciplines	C			10		
Not defined				6		
TOTAL				352		

N1: number of journals that are assigned only to the corresponding category; N2: number of journals that are assigned at least to the corresponding category; N3: number of journals that are at least assigned to the corresponding category and are also indexed in Scopus. Source: https://www.mdpi.com/about/journals.

received by a target from one specific source webpage were considered as one, to avoid artificial link inflation. After this debugging process, the final set was reduced to 567,900 links from source webpages to MDPI journal websites.

The next step consisted of obtaining link-based indicators related both to the target URLs (each MDPI journal website) and the source URLs (each external domain name holding webpages linking to MDPI journal websites).

All the indicators related to target URLs are journal-level metrics (i.e., the domain name covers the journal in its entirety) and reflect the web impact achieved by each MDPI journal website.

Web visibility indicators (link counts and referring domain counts) were collected as basic link-based metrics in terms of web impact (Björneborn & Ingwersen, 2004). In addition, the Citation Flow and Trust Flow scores (referred to as *flow metrics*) of each target URL were collected from Majestic. These are normalized indicators that allow measuring the influence of target URLs based on the quantity of links received and the quality of the websites generating those links, respectively (Orduña-Malea, 2019), thus minimizing the effects of link inflation.

The number of links from sites with a minimum Trust Flow value (referred to as *Links counts TF25*) is introduced to test whether counting links only from trusted websites might change the

relation of link-based indicators with citation-based indicators. This parameter excludes links from poor-quality and fraudulent websites, most of them with low Trust Flow scores.

Finally, we calculated network indicators (Eigenvector centrality and PageRank) from Majestic's data via Gephi software[12] to determine whether the connectivity between the source and target URLs do influence the journals' citation-based impact.

Web visibility metrics, flow metrics, and network metrics jointly allow us to have broader information about the target URLs' web impact.

The characteristics of the source URLs were also analyzed. These indicators are aimed to measure the characteristics of those websites linking to the MDPI journal websites. The underlying rationale is that links from webpages with few external outlinks (links to other sites) generally reflect genuine interest on the target URL linked, and links from webpages with many external outlinks might reflect unnatural or shallow linking behaviors.

This way, the number of external outlinks and outdomains were collected for each source URL. These indicators were aggregated to the journal-level through median values.

Likewise, the context in which each link is generated is informative. Links placed near many other outlinks denote less importance (e.g., the outlinks can be placed in navigation menus). The link density is an indicator that measures the percentage of outlinks near the outlink targeted to each MDPI journal website. This parameter is calculated by Majestic for each source-URL/target-URL combination. To the best of our knowledge, this is the attempt to measure link density scores related to journal websites.

A detailed description for each web indicator is available in Appendix A. Additional information can also be found at the official Majestic Glossary[13].

Scopus was used as a source of bibliometric data. Given that Scopus follows an indexing procedure based on the fulfilment of a set of quality criteria[14], the inclusion of the journals in this database was also used as a control group to determine whether the journals' web impacts vary depending on their formal quality. All publications from MDPI journals indexed in Scopus (523,935 publications from 159 journals, which corresponds to 89% of all MDPI publications and 45.2% of all the journals, respectively) were collected as of July 2021.

The number of citations received at the journal level (aggregating the number of citations received by each publication) constitutes the central metric to be collected. As this metric is size dependent, the journal age and the number of publications (all publications, indexed publications, recent publications, and cited publications) per journal were also collected to check whether the journal age or size influence the correlation between links and citations.

Unlike citation-based indicators, link-based metrics are not cumulative (Ingwersen & Björneborn, 2004). Links can disappear as the source websites change or are deleted. Therefore, link counts are not necessarily correlated to size-dependent indicators, such as citations counts. For this reason, other citation-based metrics were collected to check whether link-based metrics are sensitive to them. To this end, relative (Citescore), weighted (SJR), and normalized (SNIP) indicators were collected for each journal (2021 values)[15]. A detailed description for each web indicator is available in Appendix B.

---

[12] https://gephi.org.

[13] https://majestic.com/help/glossary.

[14] https://www.elsevier.com/solutions/scopus/how-scopus-works/content/content-policy-and-selection.

[15] https://www.scopus.com/sources.

2.3. Data Analysis

Due to the skewed distribution of link-based metrics, a Kruskal-Wallis median test (Kruskal & Wallis, 1952) was used to determine whether being indexed in Scopus generates statistically significant differences between the journal websites' online impact (RQ1). In addition, the potential effect of the journal disciplinary profile (RQ2) was determined. Spearman's rho correlations (Spearman, 1904) were used to measure the strength of association between these metrics (RQ3), and descriptive statistics were applied to find out the most important linking websites (RQ4). All statistical tests were carried out through XLSTAT 2021.1.1[16].

3. RESULTS

3.1. Are Link-Based Metrics Related to the Formal Quality of Journals? (RQ1)

The formal quality of journals has been operationalized as being indexed in Scopus. The nonindexed journals have also been divided into two subcategories: new journals (those less than 3 years old, and therefore with no time to be indexed in Scopus) and established journals (those 3 or more years old). The comparison of median values of link-based metrics shows that journals indexed in Scopus have attracted a statistically higher number of links and referring domains than both new and established nonindexed journals.

The number of links received by new nonindexed journals is slightly overrepresented due to the link behavior of *Encyclopedia*, a journal that receives 113,186 links (mainly from trusted websites). The *International Journal of Environmental Research and Public Health*, a new nonindexed journal occupying second position according to the number of links received, only attracts 5,019 links.

Otherwise, those source websites linking to the indexed journals generate statistically significant lower numbers of external outlinks and outdomains, whereas they achieve a lower link density score, evidencing a more selective linking behavior. However, when the number of links is normalized by the number of journals' publications, the new nonindexed journals achieve significantly higher averages (Table 2).

3.2. Are Link-Based Metrics Related to the Discipline Covered by the Journals? (RQ2)

Generalist journals attract a statistically significant higher number of links than the specialized journals (Table 3), and these links come from a larger number of referring domains. These results might be explained due to the greater number of publications published by generalist journals (median = 483.5) than the specialized journals (median = 123). In contrast, the number of links received per publication does not show significant differences.

Link-based indicators do not show significant statistical differences by subject categories for those journals indexed in Scopus. However, the boxplots performed for a few specific link-based metrics reveal noteworthy behaviors (Figure 1). For example, the Social Sciences, Arts and Humanities journals (A10) show better performance when links are selective and when they are normalized by the number of publications. Physical Sciences journals (A8) attract a great number of links, but from a limited number of referring domains.

---

[16] https://www.xlstat.com.

Table 2. Link-based metric values according to whether a journal is indexed in Scopus

Link-based metrics	Nonindexed journals (N = 193)				Indexed journals (N = 158)		p-value
	New journals (N = 141)		Established journals (N = 52)				
	Mean	Median	Mean	Median	Mean	Median	
Target—Links counts (T)	896.6	34.0	391.9	188.5	2,665.2	503.0	< 0.0001*
Target—Links counts (TF25)	854.9	3.0	98.7	47.0	306.6	140.0	< 0.0001*
Target—Links counts (T)/Publications	22.3	2.5	1.0	0.8	1.3	0.5	< 0.0001*
Target—Links counts (TF25)/Publications	18.2	0.3	0.3	0.2	0.2	0.1	0.001*
Target—Referring domains counts	11.0	8.0	68.1	44.0	121.8	93.0	< 0.0001*
Target—Trust Flow score	22.7	23.0	25.1	24.0	28.1	25.0	< 0.0001*
Target—Citation Flow score	28.6	29.0	31.9	31.0	34.9	35.0	< 0.0001*
Source—Link density score	27.1	25.5	19.0	16.0	13.3	7.5	< 0.0001*
Source—External outlink counts	560.2	445.0	128.7	54.0	38.6	25.5	< 0.0001*
Source—External outdomain counts	10.0	8.0	25.0	11.3	13.5	10.0	< 0.0001*

*p-value is lower than alpha-value (0.05). Kruskal-Wallis test.

The median values for the link-based metrics are shown in Table 4, where a lack of disciplinary pattern is evidenced. However, a pairwise comparison reveals noteworthy differences between metrics. For example, Environmental & Earth Sciences has a median referral domains count value of 129.5, but Physical Sciences has 73.5.

Table 3. Link-based metric values according to the journals' coverage profile

Link-based metrics (median values)	Subject category			p-value
	Specialized (N = 106)	Multidisciplinary (N = 229)	Generalist (N = 10)	
Target—Links counts (T)	124.5	218.0	437.0	0.017*
Target—Links counts (TF25)	30.5	60.0	135.5	0.066
Target—Links counts (T)/Publications	1.4	1.0	0.9	0.090
Target—Links counts (TF25)/Publications	0.2	0.1	0.2	0.174
Target—Referring domains counts	23.5	45.0	99.5	0.035*
Target—Trust Flow score	23.0	24.0	26.0	< 0.0001*
Target—Citation Flow score	30.0	30.0	35.5	0.034*
Source—Link density score	19.8	10.0	13.0	0.044*
Source—External outlink counts	36.0	40.0	32.0	0.726
Source—External outdomain counts	8.0	8.0	9.8	0.231

*p-value is lower than alpha-value (0.05). Kruskal-Wallis test. *Note*: all metrics are totals for the journal.
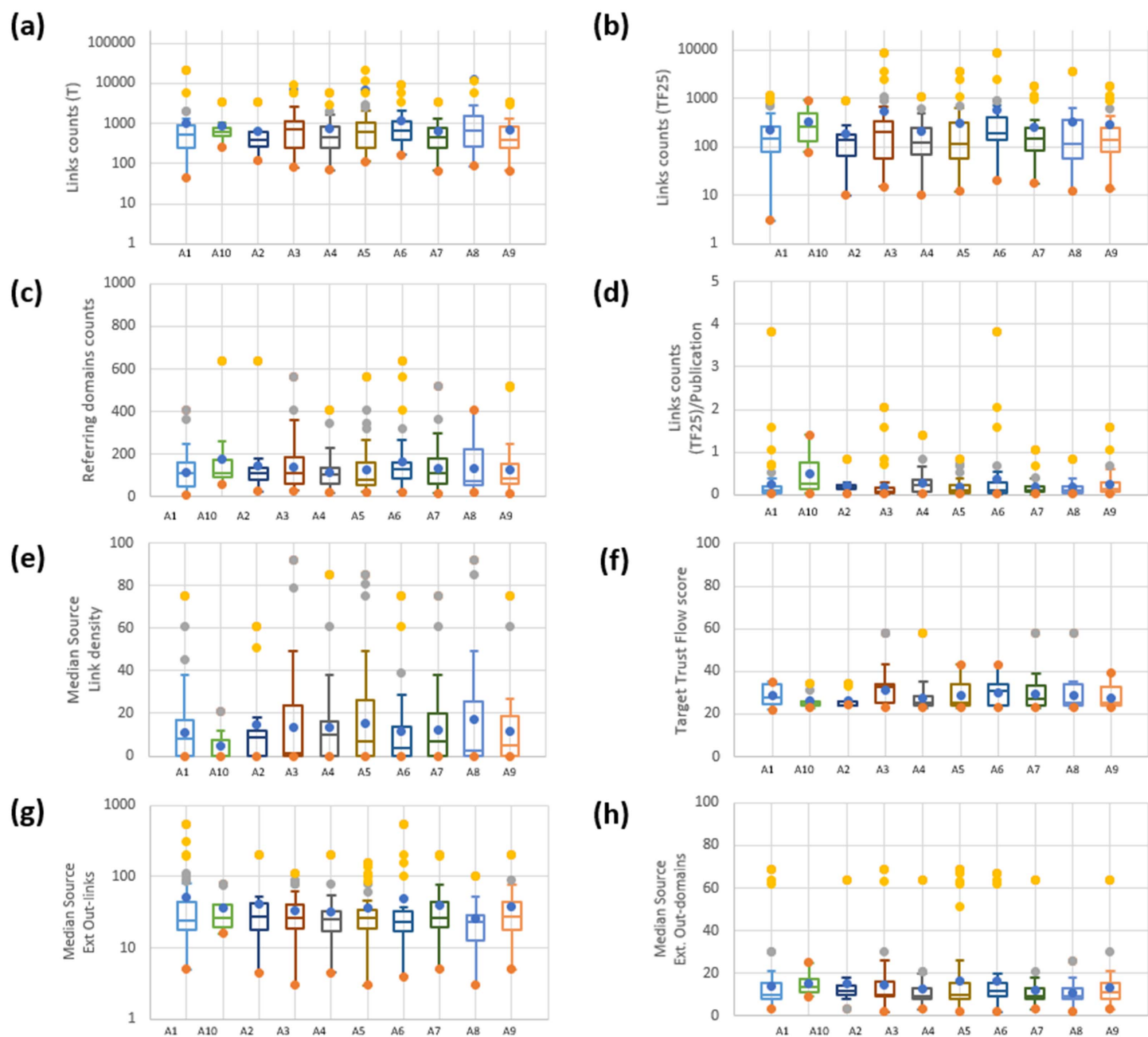
Figure 1. Link-based metrics for journal subject categories. (a) Target—Links counts; (b) Target—Links counts (TF25); (c) Target—Referring domain counts; (d) Target—Links counts (TF25)/publication; (e) Source—Source link density; (f) Target—Trust Flow score; (g) Source—External outlink counts; (h) Source—External outdomain counts. A1: Biology & Life Sciences (*N* = 63); A2: Business & Economics (*N* = 13); A3: Chemistry & Materials Science (*N* = 47); A4: Computer Science & Mathematics (*N* = 36); A5: Engineering (*N* = 52); A6: Environmental & Earth Sciences (*N* = 35); A7: Medicine & Pharmacology (*N* = 36); A8: Physical Sciences (*N* = 24); A9: Public Health & Healthcare (*N* = 32); A10: Social Sciences, Arts and Humanities (*N* = 10). Note: One journal can appear in more than one subject category.

In general terms, Chemistry & Materials Science's journals have the highest median values for Links counts, Trust Flow, and Citation Flow scores, receiving links from low link density areas. Conversely, Business and Economics' journals attract lower number of links, generated in areas of higher link density. Environmental & Earth Sciences' journals receive links from a greater number of referring domains.

Table 4. Link-based metric values according to the subject categories

Variables (median values)	Subject categories									
	Biology & Life Sciences	Business & Economics	Chemistry & Materials Science	Computer Science & Mathematics	Engineering	Environmental & Earth Sciences	Medicine & Pharmacology	Physical Sciences	Public Health & Healthcare	Social Sciences, Arts and Humanities
Target—Links counts (T)	525.0	398.0	690.0	455.0	603.0	640.5	455.5	677.0	400.0	601.5
Target—Links counts (TF25)	151.0	142.0	200.0	126.0	114.0	198.5	146.0	119.0	139.5	265.0
Target—Links counts (T)/ Publication	0.4	0.5	0.3	0.8	0.5	0.3	0.3	0.6	0.4	0.8
Target—Links counts (TF25)/ Publication	0.1	0.2	0.1	0.2	0.1	0.1	0.1	0.1	0.1	0.3
Target—Referral domains counts	109.0	113.0	109.0	103.0	80.0	129.5	109.5	73.5	89.0	111.0
Target—Trust Flow score	28.0	24.0	33.0	25.0	25.0	31.0	27.0	25.0	25.0	25.0
Target—Citation Flow score	35.0	35.0	36.0	35.0	35.0	36.0	35.5	35.0	34.0	35.5
Source—Link density score	8.0	9.0	1.0	10.0	7.0	4.0	7.0	2.5	4.8	0.0
Source—External outlink counts	24.0	27.0	26.0	25.0	26.0	23.0	26.8	23.5	27.0	26.5
Source—External outdomain counts	10.0	12.0	10.0	9.5	10.0	11.5	9.5	9.5	11.0	13.5

Note. All metrics are totals for the journal.

3.3. Are Link-Based and Bibliometric-Based Journal Metrics Correlated? (RQ3)

The number of publications (whether total, indexed, recent, or cited publications) achieves strong positive and statistically significant correlations with link-based metrics (Figure 2). Specifically, the total number of publications published by the journals strongly correlates with the total number of links received by the corresponding journal website ($R_s = 0.83$) and with the number of referring domains ($R_s = 0.83$). It is also worth noting the strong correlation achieved between the number of citations received and the number of referring domains ($R_s = 0.74$), which is even larger for recent citations ($R_s = 0.78$). The size-dependent nature of both citations and links received might explain these strong correlations. Network measures (Eigenvector and Page Rank) also evidence a strong correlation between web connectivity and the number of citations received, especially for recent publications.

A lack of correlation has been found between the link-based indicators and the impact-related journal indicators, whether normalized (SNIP), relative (CiteScore), or weighted (SJR) indicators. Only the SNIP indicator achieves significant correlation with the number of referring domains ($R_s = 0.36$). These results are aligned to the low number of links per publication previously obtained (see Table 2), probably due to the fact that these journal-level impact measures follow a similar rationale (citations per publication).

Source-related web metrics evidence a lack of correlation (outdomains counts) or even negative correlation (link density score and external outdomains counts) with the number of citations received. As these metrics are related to the link behavior of the source webpages, these results suggest that links from webpages that generate many links (e.g., directories) might reflect publishers' promotion instead of scholarly journal impact.

BIBLIOMETRIC INDICATORS	LINK-BASED INDICATORS									
	Target Eigenvector	Target Page Rank	Target Links Counts (all)	Target Links Counts (TF25)	Target Ref. Domains counts	Target Trust Flow	Target Citation Flow	Source Link Density	Source Outlinks counts	Source Outdomains counts
Journal age	*0.57	*0.58	*0.56	*0.51	*0.57	*0.55	*0.58	*-0.30	*-0.30	0.08
Total publications	*0.83	*0.79	*0.73	*0.77	*0.83	*0.80	*0.79	*-0.30	*-0.50	0.05
Indexed publications	*0.80	*0.76	*0.70	*0.74	*0.8	*0.77	*0.76	*-0.30	*-0.50	0.02
Recent publications	*0.80	*0.76	*0.72	*0.78	*0.80	*0.73	*0.75	*-0.20	*-0.50	0.02
Cited publications	*0.77	*0.73	*0.69	*0.74	*0.77	*0.74	*0.74	*-0.30	*-0.40	0.03
Cited recent publications	*0.78	*0.74	*0.71	*0.77	*0.78	*0.69	*0.73	-0.20	*-0.50	0.01
SNIP	*0.36	*0.34	0.20	*0.27	*0.36	0.20	0.15	*-0.30	-0.10	0.05
SJR	*0.25	0.17	0.08	0.19	*0.25	*0.34	0.12	-0.20	0	-0.10
Citescore	0.08	0.10	0.07	0.08	0.08	0	0.04	0	0.01	0
All citations	*0.74	*0.71	*0.66	*0.72	*0.74	*0.69	*0.72	*-0.30	*-0.40	0
Recent citations	*0.78	*0.74	*0.70	*0.78	*0.78	*0.71	*0.70	*-0.20	*-0.50	0

Figure 2. Spearman correlation between bibliometric-based and link-based metrics. *Note*: All link-based metrics and journal age are totals for the journal; all the remaining bibliometric metrics are aggregated article-level metrics. * Values different from 0 with alpha significance level = 0.05.

The age of journals shows a statistically significant correlation with the number of links received ($R_s = 0.56$), links from trusted webpages ($R_s = 0.51$), and Trust Flow values ($R_s = 0.55$). The moderate values obtained show that age is significant but not critical for link attraction.

When the correlation values are disaggregated by subjects, we find similar patterns (Figure 3).

First, we can observe a strong positive correlation between the link-based metrics and publication-based metrics, with no significant variations depending on the type of publications considered (total, indexed, recent, or cited publications). The number of referring domains is strongly correlated to the total number of publications for all 10 disciplines, especially Chemistry & Materials Science ($R_s = 0.91$) and Social Sciences, Arts & Humanities ($R_s = 0.98$).

The journal-level impact indicators (SNIP, SJR, Citescore) achieve weak correlations with link-based indicators in most disciplines, even negative correlations in the case of Computer Science & Mathematics, and Social Sciences, Arts & Humanities. However, we find significant positive strong correlations between the SNIP indicator and the number of referring domains for Medicine & Pharmacology ($R_s = 0.63$) and Public Health & Healthcare ($R_s = 0.61$), which evidence disciplinary differences.

The number of citations received by journals and the number of referring domains are also strongly correlated in all 10 disciplines. Journals from Business & Economics are those reflecting weaker correlations (stronger for recent citations, $R_s = 0.83$). However, the low number of journals indexed in this subject category (13) prompts us to consider the values obtained with caution.

Otherwise, weak correlation values have been obtained between the Trust Flow scores and all the bibliometrics-based indicators for Business & Economics journals, an aspect that does

BIBLIOMETRIC INDICATORS	LINK-BASED INDICATORS																													
	Biology & Life Sciences			Business & Economics			Chemistry & Materials Science			Computer Science & Mathematics			Engineering			Environmental & Earth Sciences			Medicine & Pharmacology			Physical Sciences			Public Health & Healthcare			Social Sciences Arts & Humanities		
	TF25	RDC	TTF	TF25	RDC	TTF	TF25	RDC	TTF	TF25	RDC	TTF	TF25	RDC	TTF	TF25	RDC	TTF	TF25	RDC	TTF	TF25	RDC	TTF	TF25	RDC	TTF	TF25	RDC	TTF
Journal age	*0.40	*0.43	*0.37	0.03	0.23	0.22	*0.55	*0.73	*0.67	*0.41	*0.56	*0.63	*0.54	*0.66	*0.68	*0.58	*0.75	*0.69	*0.53	*0.48	*0.56	*0.60	*0.74	*0.68	*0.40	0.23	*0.38	0.63	*0.84	*0.85
Total publications	*0.61	*0.86	*0.83	*0.58	*0.73	0.08	*0.73	*0.91	*0.85	*0.70	*0.87	*0.75	*0.71	*0.84	*0.87	*0.57	*0.86	*0.80	*0.74	*0.88	*0.83	*0.75	*0.85	*0.88	*0.69	*0.83	*0.86	*0.83	*0.98	0.65
Indexed publications	*0.60	*0.84	*0.80	*0.60	*0.69	0.08	*0.74	*0.91	*0.84	*0.71	*0.87	*0.74	*0.72	*0.85	*0.86	*0.56	*0.86	*0.79	*0.73	*0.85	*0.83	*0.74	*0.85	*0.90	*0.68	*0.83	*0.86	*0.79	*0.95	0.73
Recent publications	*0.63	*0.85	*0.82	*0.74	*0.80	0.04	*0.73	*0.87	*0.79	*0.77	*0.86	*0.65	*0.74	*0.84	*0.85	*0.57	*0.86	*0.79	*0.76	*0.87	*0.81	*0.77	*0.84	*0.87	*0.67	*0.85	*0.86	*0.83	*0.93	0.65
Cited publications	*0.61	*0.84	*0.78	*0.57	*0.69	0.15	*0.73	*0.91	*0.08	*0.72	*0.86	*0.76	*0.71	*0.84	*0.85	*0.59	*0.86	*0.79	*0.72	*0.85	*0.82	*0.75	*0.85	*0.89	*0.70	*0.83	*0.85	*0.79	*0.95	0.73
Cited recent publications	*0.62	*0.84	*0.79	*0.80	*0.84	0.03	*0.74	*0.88	*0.75	*0.77	*0.87	*0.64	*0.74	*0.84	*0.86	*0.57	*0.85	*0.78	*0.73	*0.86	*0.79	*0.78	*0.86	*0.87	*0.67	*0.83	*0.84	*0.88	*0.95	0.58
SNIP	0.13	*0.33	*0.37	-0.07	-0.08	0.24	0.27	*0.45	0.20	-0.27	-0.30	-0.32	0.12	0.19	0.16	0.00	0.22	0.15	0.32	*0.63	*0.54	0.20	0.34	0.16	0.34	*0.61	0.45	-0.60	-0.36	0.28
SJR	0.01	*0.33	*0.26	-0.01	0.02	0.33	0.21	0.28	0.18	-*0.36	-0.17	0.03	0.09	0.10	*0.31	0.10	0.24	0.15	0.33	*0.59	*0.59	-0.08	-0.08	0.15	0.32	*0.67	*0.61	-0.60	-0.60	0.05
Citescore	-0.07	0.09	-0.01	0.34	0.22	0.10	0.26	*0.35	0.17	-0.14	-0.19	-0.23	0.09	0.10	0.16	-0.08	-0.01	-0.12	0.29	*0.58	*0.37	0.11	0.30	0.18	*0.40	*0.62	0.40	-0.22	-0.30	0.09
All citations	*0.57	*0.81	*0.73	0.44	0.53	0.34	*0.71	*0.91	*0.74	*0.62	*0.76	*0.76	*0.70	*0.82	*0.87	*0.58	*0.85	*0.78	*0.69	*0.83	*0.80	*0.71	*0.86	*0.90	*0.68	*0.83	*0.85	0.69	*0.88	0.73
Recent citations	*0.56	*0.83	*0.78	*0.80	*0.83	0.06	*0.72	*0.89	*0.75	*0.73	*0.80	*0.61	*0.73	*0.84	*0.88	*0.57	*0.85	*0.79	*0.68	*0.86	*0.80	*0.74	*0.86	*0.88	*0.68	*0.85	*0.84	0.52	0.67	0.43

Figure 3. Spearman correlation between the bibliometric-based and the link-based metrics according to subject categories. Note 1: TF25: Number of links received from webpages with a TF ≥ 25; RDC: Number of referring domains; TTF: Target Trust Flow. Note 2: all link-based metrics and journal age are totals for the journal; all the remaining bibliometric metrics are aggregated article-level metrics. * Values are different from 0 with alpha significance level = 0.05.

Table 5. Top referring domains providing links to MDPI journals: all referring domains, universities, and organizations

All domains			Universities			UK universities			Organizations		
Domain	**No. Links**	**No. Journals**	**Domain**	**No. Links**	**No. Journals**	**Domain**	**No. Links**	**No. Journals**	**Domain**	**No. Links**	**No. Journals**
4m-net.org	130,778	1	cf.ac.uk	3,334	4	cf.ac.uk	3,334	4	4m-net.org	130,778	1
4m-association .org	130,512	1	lsmuni.lt	1,045	1	strath.ac.uk	225	55	4m-association.org	130,512	1
encyclopedia .pub	113,585	28	unios.hr	974	3	salford.ac .uk	183	53	isbe-online.org	20,494	1
isbe-online.org	20,494	1	ualg.pt	701	1	ncl.ac.uk	36	16	metaconferences.org	15,858	2
metaconferences .org	15,858	2	universitaspertamina .ac.id	687	4	abdn.ac.uk	36	8	scimatic.org	4,018	6
mdpi.cn	8,784	351	ntnu.edu	623	60	mdx.ac.uk	33	12	doaj.org	1,452	215
iao.ru	8,620	2	icbms.fr	527	1	lse.ac.uk	31	17	fen.org.es	1,424	3
mpg.de	6,116	141	uio.no	490	57	warwick.ac .uk	27	13	observatorioeconomiasocial .org	1,187	1
journaltocs.ac .uk	5,309	167	vu.lt	471	7	lancs.ac.uk	23	13	iccsa.org	817	2
tanger.cz	4,443	5	hmu.gr	366	3	ox.ac.uk	22	16	qoam.org	669	172

Note. All link-based metrics are totals for the journal.

not occur in any other category. The low number of links received by this discipline (see Table 4) might explain this issue.

3.4. Where Do Links to Journal Websites Come From? (RQ4)

All 352 MDPI journals have received 567,900 links from 9,568 unique referring domains, showing a highly skewed distribution of links per referring domain (three referring domains provide the 66% of all links received by the MDPI journals). The following categories of referring domains can be pointed out:

- *Self-promotion*: MDPI journals receive links from other MDPI sites (e.g., 8,784 links from mdpi.cn; 1,777 from mdpi.rs; 1,395 from mdpi.es). These websites provide links to many MDPI journals.
- *Bibliographic data*: MDPI journals receive links from websites dedicated to providing journals' bibliographic data, such as JournalTOCs (5,309), DOAJ (1,452), Quality Open Access Marker (675 links from qoam.eu, and 669 links from qoam.org), SHERPA (419 links), Hypotheses.org (412 links), Observatory of International Research (344 links), Research4Life (294 links), or Scimago Journal & Country Rank (255 links). Generally, these informational websites provide links to many MDPI journals.
- *Universities*: More than 10% of all referring domains belong to higher education institutions, where the United Kingdom (136 referring domains), stands out. Generally, these academic websites provide links to few MDPI journals.
- *Events*: Conference websites held by academic-related associations generate a significant amount of the total number of links targeted to the MDPI journals, where the 4M Association (130,778 links from m-net.org, and 130,512 from 4m-association.org) stand out. Generally, event websites generate a high number of links to few journals. For example, 4M Association links to just one journal (*Micromachines*), and the International Society of Bionic Engineering (isbe-online.org) provides 20,494 links to only one journal (*Bioengineering*).

Table 5 includes the top 10 referring domains with most links to MDPI journals, as well as the number of journals each referring domain is linking to. Referring domains belonging to universities and organizations are also included by way of illustration.

4. DISCUSSION

This work provides evidence of strong positive correlation between citation-based and link-based journal-level metrics for the 159 open access journals published by MDPI and covered by Scopus. These results reinforce early studies on journal link analyses. However, direct comparisons cannot be carried out, as the sources for citations (Scopus) and links (Majestic) did not exist when those previous studies were published (Vaughan & Hysen, 2002; Vaughan & Thelwall, 2002, 2003). Moreover, the dynamics of the WWW as well as the implementation of the DOI as a permanent URL standard ID have also changed the analytical framework.

The results obtained should be treated cautiously due to the limitations of the sources used and should be circumscribed by the data sources used (MDPI, Scopus, and Majestic).

4.1. MDPI: The Journal Data Source

This study has analyzed all journals published by one unique publisher. This design allowed data comparisons, as all journals are governed by the same publication guidelines, with

identical website designs and marketing promotion. In fact, as an exponent of the series titles phenomenon (such as *BMC Series* or *Frontiers in*), MDPI might be viewed as a broad disciplinary scope journal (Spezi, Wakeling et al., 2017), diminishing the identity of each journal while enhancing the whole MDPI brand. This behavior could minimize differences between journals when measured through web data.

The results obtained could be different when analyzing other publishers, especially journals behind subscription paywalls. The characteristics of the publisher (the number of journals managed, topics covered, and the publication rate) might affect the results obtained. Specifically, the behavior of some megajournals can distort the results obtained, given their elevated annual publication output. The use of medians in the statistical tests carried out allowed us to minimize the effect of outliers.

Although the scientific community has expressed concerns related to megajournals in general (e.g., Björk, 2015, 2018; Björk & Catani, 2016; Borrego, 2018; Brainard, 2019; Heneberg, 2019; Petersen, 2019; Siler, Larivière, & Sugimoto, 2020; Spezi, Wakeling et al., 2017, 2018; Wakeling, Willett et al., 2016; Wakeling, Creaser et al., 2019; Wellen, 2013), and to MDPI in particular (Copiello, 2019; Oviedo-García, 2021; Repiso, Merino-Arribas, & Cabezas-Clavijo, 2021), we do not question MDPI's editorial practices, using its portfolio simply as a baseline for link studies.

4.2. Scopus: The Bibliometric Data Source

Scopus has been used to collect the number of citations received by journals as well as different impact-based journal indicators (SNIP, SJR, Citescore). We acknowledge that using other databases (e.g., Web of Science, Dimensions, or Google Scholar), with distinct coverage of both journals and citations (Martín-Martín, Thelwall et al., 2021; Mongeon & Paul-Hus, 2016; Singh, Singh et al., 2021; Visser, van Eck, & Waltman, 2021), could have yielded other results. Further studies should check whether the results vary depending on the bibliographic database used. Therefore, the results obtained should be restricted to Scopus.

Scopus has been used as a filter to determine the formal quality of journals (indexed vs. nonindexed). This decision might filter out quality journals that are not indexed in Scopus yet (especially new journals). To minimize this effect, the nonindexed journals were divided into new and established journals. Although Scopus evaluates the formal quality of journals in a particular way, this evaluation is considered good enough for the purposes of this work.

4.3. Majestic: The Link Data Source

Majestic's historic database has been used to collect the external links received by the journal websites. This link-intelligent tool has already been used successfully in webometric studies (Orduña-Malea, 2021). The analysis required a data cleaning process to avoid crawling errors. This process (which reduced the initial set of links collected by 52.4%) is deemed necessary to achieve reliable results, although it is time consuming. As with bibliographic databases, the use of other link sources (e.g., Ahrefs, Link Explorer) might produce different results as the link coverage can differ from one source to another. Therefore, the results obtained are limited to those obtained from Majestic.

Majestic calculates all link-based metrics related to each URL through a self-made search engine that crawls the entire Web. As a private company, the exact calculation of web metrics, especially the Flow Metrics, is not publicly disclosed due to industrial property rights. Therefore, further studies aimed at checking the accuracy of other web sources are advisable.

4.4. Worldwide Web: The Analytical Framework

Beyond the general features of the web data source used, the following aspects related to the web environment must be considered to contextualize the results obtained.

First, results collected at a fixed date should not be interpreted cumulatively (as are bibliometric indicators), but rather as the status of the source and target websites at that time. For example, a website redesign project could eventually generate misleading results. For this reason, longitudinal studies would be desirable to avoid potential data collection errors. In this sense, the Trust Flow score is useful, as it holds its value long enough to avoid ephemeral changes over time.

Second, the massive inflation of link counts does not necessarily reflect bad web practices, but natural web behavior. For example, this study has revealed cases where links appear in website navigation menus (e.g., the personal academic website "lluiscodina.com" links to the *Journalism and Media* journal, due to a link that appears in the footer of each webpage). Likewise, logos can link massively to one specific journal (e.g., links from "cytofluidix.com" to the *Micromachines* and *Fluids* journals). Related projects can also generate massive links to one specific journal. For example, the referring domain "encyclopedia.pub" generates 113,183 links to the journal *Encyclopedia*, as they are related[17].

To avoid these problems, counting referring domains instead of links is advisable, as this work has shown.

The answers to the specific research questions are given below.

4.5. Are Link-Based Metrics Related to the Formal Quality of Journals? (RQ1)

Those journals indexed in Scopus attain links from a greater number of trustworthy referring domains than the nonindexed journals. Considering the indexing of journals in Scopus as a quality filter, debugged link data provides evidence of the web influence acquired by the indexed scientific journals. However, these results are conditioned by the dependence of the link counts on the number of publications, significantly greater in the indexed journals (median = 761 publications) than in the nonindexed ones (median = 27).

4.6. Are Link-Based Metrics Related to the Discipline Covered by the Journals? (RQ2)

Those journals covering a greater number of subject categories (generalist journals) attract links from a greater number of trustworthy referring domains than those journals covering only one subject (specialized journals). Although covering a wide range of subjects could help generalist journals to generate the interest of a wider audience, their significantly greater volume of publications might explain the results obtained.

The differences found between all subject categories were not statistically significant, and no clear disciplinary patterns have been found, but there are differences in particular metrics. For example, Chemistry & Materials Science is the subject category with the greatest median Trust Flow score. Environmental & Earth Sciences holds the highest median referring domains count. Social Sciences achieves the highest median links TF25 count. Physical Sciences journals show the highest links scores and the lowest median referring domain counts.

---

[17] https://encyclopedia.pub/about.

A plausible explanation is that the exclusion of DOI-based URL citations might enhance the "series title" profile of the whole publisher (Spezi et al., 2017), diminishing differences between disciplines. Additionally, the low number of journals in some subjects can also affect the results obtained.

4.7. Are Link-Based and Bibliometric-Based Journal Metrics Correlated? (RQ3)

Link-based metrics (especially referring domain counts and Trust Flow scores) achieve a statistically significant strong positive correlation with both the size of the journals (number of publications) and their impact (number of citations received). These correlations are strong for all subjects, except for Business & Economics.

However, link-based metrics do not correlate with journal-level impact indicators (SJR, Citescore, SNIP). A plausible explanation is the different nature of these metrics, which do not index all the contents, use small citation temporal windows, and hold their value for a whole year. Conversely, link-based metrics represent the journals' status at the time of data collection, considering all links received for all contents created.

Another potential reason for the uncorrelated values obtained is the fact that these indicators are based on (estimated) averages of citations per document, a distorted metric because a few documents are responsible for most of the citations received (Larivière & Sugimoto, 2019). In fact, a similar circumstance occurs with the link counts per document obtained (see Tables 2 and 3), which generate completely different results from the remaining online metrics.

4.8. Where Do Links to Journal Websites Come From? (RQ4)

Although the motivations behind the creation of each link cannot be directly addressed (Bar-Ilan, 2005; Thelwall, 2003), the origin of links (referring domain categories) have pointed out the importance of navigational links from scientific information products. As links in those strategic and valued websites can potentially drive quality visitors (i.e., visitors with potential to submit articles or cite MDPI publications) to the MDPI websites, the coverage of the journal in those websites is taken as signals of certain web impact.

Links from conference websites reflect the sponsored activities of some MDPI journals, which collaborate and support academic events (most links come from banners on conference websites). Contributions originally submitted to these events can also potentially be submitted finally to specific special issues in those journals. In any case, this issue only affects statistically few journals.

Links from universities reflect authors' self-archiving activities, being authors depositing the author or final version of their papers in their institutional repository or personal websites. As each paper includes a link to the journal website, links from universities can be related to MDPI publication patterns of university staff.

Although these results help to contextualize the results found, the correlation between citation counts and link counts needs further research. A reasonable explanation is that uncited MDPI publications might have not been self-archived in university repositories or have been published in journals not covered in scientific-related information websites. In any case, an analysis at the article level (URLs to each MDPI publication, especially to DOI-based URLs) are deemed necessary to test this hypothesis.

5. CONCLUSIONS

Link-based indicators have been proved to be sensitive to the quality (being indexed in Scopus), size (number of publications), and impact (number of citations received) of MDPI journals. Therefore, we suggest that link-based indicators can be used cautiously as informative measures of the MDPI journals' current performance.

The number of referring domains, the number of links from trusted websites, and the Trust Flow achieved by journal websites should be highlighted as robust metrics. These metrics are selective (they depend on the existence of reliable, active websites generating links to each journal), stable over time (their variation is less volatile than the number of total links received), and not so easily manipulated.

The results obtained in this work can be useful for journal publishers, who can monitor these link-based indicators to obtain fresh information about the journals' web impact, and thus are able to design strategic decisions in advance for the optimal dissemination of the journals. Library catalogues and bibliographic databases offering information about scientific journals can also include these link-based metrics to add information to users.

Experts on science studies can also use these results to better understand the relation between science communication (journal website as an online channel) and scholarly communication, and to explore the nonscholarly impact of journals and publications. Likewise, experts in webometrics can better understand the nature of online indicators related to academic and scholarly online objects.

The links counted in this study were only those targeted to the journal websites (any webpage inside the official journal website), excluding DOI links to publications. For this reason, the link-based indicators obtained cannot be directly related to the research impact of the publications but to the journals' web impact.

To better understand the nature of web indicators and their relationship with the scientific impact of journals, it is necessary to carry out studies at the article level (using both the DOI and the different URLs created by the journals for each article). Further studies are also necessary to evaluate link-based metrics for journals under different publication policies.

AUTHOR CONTRIBUTIONS

Enrique Orduña-Malea: Conceptualization, Formal analysis, Methodology, Writing—Original draft. Isidro F. Aguillo: Methodology, Supervision, Writing—Review & editing.

COMPETING INTERESTS

The authors have no competing interests.

FUNDING INFORMATION

DATA AVAILABILITY

The raw data used in this study provides citation-based and web-based indicators for 352 journals and includes 567,900 typified links to them. The data is openly available (Orduña-Malea & Aguillo, 2022).

REFERENCES

Bar-Ilan, J. (2005). What do we know about links and linking? A framework for studying links in academic environments. *Information Processing and Management*, *41*(4), 973–986. https://doi.org/10.1016/j.ipm.2004.02.005

Berners-Lee, T., Cailliau, R., Groff, J., & Pollermann, B. (1992). World-Wide Web: The information universe. *Internet Research*, *2*(1), 52–58. https://doi.org/10.1108/eb047254

Björk, B.-C. (2015). Have the "mega-journals" reached the limits to growth? *PeerJ*, *3*, e981. https://doi.org/10.7717/peerj.981, PubMed: 26038735

Björk, B.-C. (2018). Publishing speed and acceptance rates of open access megajournals. *Online Information Review, 45*(2), 270–277. https://doi.org/10.1108/OIR-04-2018-0151

Björk, B.-C., & Catani, P. (2016). Peer review in megajournals compared with traditional scholarly journals: Does it make a difference? *Learned Publishing*, *29*(1), 9–12. https://doi.org/10.1002/leap.1007

Björneborn, L., & Ingwersen, P. (2004). Toward a basic framework for webometrics. *Journal of the American Society for Information Science and Technology*, *55*(14), 1216–1227. https://doi.org/10.1002/asi.20077

Bollen J., & Van de Sompel, H. (2008) Usage impact factor: The effects of sample characteristics on usage-based impact metrics. *Journal of the American Society for Information Science and Technology*, *59*(1), 136–149. https://doi.org/10.1002/asi.20746

Bollen, J., Van de Sompel, H., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PLOS ONE*, *4*(6), e6022. https://doi.org/10.1371/journal.pone.0006022, PubMed: 19562078

Borrego, A. (2018). Are mega-journals a publication outlet for lower quality research? A bibliometric analysis of Spanish authors in *PLOS ONE. Online Information Review, 45*(2), 261–269. https://doi.org/10.1108/OIR-04-2018-0136

Brainard, J. (2019). Open-access megajournals lose momentum. *Science*, *365*(6458), 1067. https://doi.org/10.1126/science.aaz4585, PubMed: 31515364

Castells, M. (2002). *The Internet galaxy: Reflections on the Internet, business, and society.* Oxford, UK: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199255771.001.0001

Chandler, A., & Jewell, T. (2006). Standards—Libraries, data providers and SUSHI: The Standardized Usage Statistics Harvesting Initiative. *Against the Grain*, *18*(2), 82–83. https://doi.org/10.7771/2380-176X.4669

Codina, L., & Morales-Vargas, A. (2021). Soluciones de arquitectura de la información en plataformas digitales editoriales: Revisión comparativa de Taylor and Francis Online, SAGE Journals, *PLOS One*, MDPI y Open Research Europe. *Anuario ThinkEPI*, *15*. https://doi.org/10.3145/thinkepi.2021.e15e01

Copiello, S. (2019). On the skewness of journal self-citations and publisher self-citations: Cues for discussion from a case study. *Learned Publishing*, *32*, 249–258. https://doi.org/10.1002/leap.1235

Harter, S., & Ford, C. (2000). Web-based analysis of e-journal impact: Approaches, problems, and issues. *Journal of the American Society for Information Science*, *51*(13), 1159–1176. https://doi.org/10.1002/1097-4571(2000)9999:9999<::AID-ASI1029>3.0.CO;2-P

Heneberg, P. (2019). The troubles of high-profile open access megajournals. *Scientometrics*, *120*(2), 733–746. https://doi.org/10.1007/s11192-019-03144-6

Ingwersen, P., & Björneborn, L. (2004). Methodological issues of webometric studies. In H. F. Moed, W. Glänzel, & U. Schmoch (Eds.). *Handbook of quantitative science and technology research* (pp. 339–369). Dordrecht: Springer. https://doi.org/10.1007/1-4020-2755-9_16

ISO. (2012). *ISO 26324:2012 Information and documentation—Digital object identifier system.* https://www.iso.org/standard/43506.html

Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, *47*(260), 583–621. https://doi.org/10.1080/01621459.1952.10483441

Larivière, V., & Sugimoto, C. R. (2019). The journal impact factor: A brief history, critique, and discussion of adverse effects. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 3–24). Cham: Springer. https://doi.org/10.1007/978-3-030-02511-3_1

Ledford, J. L. (2008). *Search engine optimization bible* (2nd ed.). Indianapolis, IN: Wiley.

Martín-Martín, A., Thelwall, M., Orduna-Malea, E., & Delgado López-Cózar, E. (2021). Google Scholar, Microsoft Academic, Scopus, Dimensions, Web of Science, and OpenCitations' COCI: A multidisciplinary comparison of coverage via citations. *Scientometrics*, *126*(1), 871–906. https://doi.org/10.1007/s11192-020-03690-4, PubMed: 32981987

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, *106*(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5

NISO. (2014). *ANSI/NISO Z39.93-2014 The Standardized Usage Statistics Harvesting Initiative (SUSHI) Protocol.* https://www.niso.org/publications/z3993-2014-sushi

Orduña-Malea, E. (2019). Rendimiento de las revistas científicas en la Web: El caso de Colombia. *4° Encuentro Regional de Editores de Revistas Académicas.* Medellín, 5–7 junio. https://doi.org/10.13140/RG.2.2.35924.24962

Orduña-Malea, E. (2021). Dot-science top level domain: Academic websites or dumpsites? *Scientometrics*, *126*(4), 3565–3591. https://doi.org/10.1007/s11192-020-03832-8

Orduña-Malea, E., & Aguillo, I. F. (2022). *The MDPI dataset: A link analysis* [Data set]. Universitat Politècnica de València. https://doi.org/10.4995/Dataset/10251/183269

Orduña-Malea, E., & Alonso-Arroyo, A. (2017). *Cybermetric techniques to evaluate organizations using web-based data.* Oxford, UK: Elsevier.

Oviedo-García, M. Ángeles. (2021). Journal citation reports and the definition of a predatory journal: The case of the Multidisciplinary Digital Publishing Institute (MDPI). *Research Evaluation*, *30*(3), 405–419. https://doi.org/10.1093/reseval/rvab020

Petersen, A. M. (2019). Megajournal mismanagement: Manuscript decision bias and anomalous editor activity at *PLOS ONE*. *Journal of Informetrics*, *13*(4), 100974. https://doi.org/10.1016/j.joi.2019.100974

Repiso, R., Merino-Arribas, A., & Cabezas-Clavijo, Á. (2021). El año que nos volvimos insostenibles: Análisis de la producción española en Sustainability (2020). *Profesional de la Información*, *30*(4). https://doi.org/10.3145/epi.2021.jul.09

Rodrigues, R. S., Abadal, E., & de Araújo, B. K. H. (2020). Open access publishers: The new players. *PLOS ONE*, *15*(6), e0233432, https://doi.org/10.1371/journal.pone.0233432, PubMed: 32502146

Ruhnau, B. (2000). Eigenvector-centrality—A node-centrality? *Social Networks*, *22*(4), 357–365. https://doi.org/10.1016/S0378-8733(00)00031-9

Shepherd, P. T. (2006). COUNTER: Usage statistics for performance measurement. *Performance Measurement and Metrics*, *7*(3), 142–152. https://doi.org/10.1108/14678040610713101

Siler, K., Larivière, V., & Sugimoto, C. R. (2020). The diverse niches of megajournals: Specialism within generalism. *Journal of the Association for Information Science and Technology*, *71*(7), 800–816. https://doi.org/10.1002/asi.24299

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, *126*(6), 5113–5142. https://doi.org/10.1007/s11192-021-03948-5

Smith, A. G. (1999). A tale of two web spaces: Comparing sites using Web Impact Factors. *Journal of Documentation*, *55*(5), 577–592.

Spearman C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*(1), 72–101. https://doi.org/10.2307/1412159

Spezi, V., Wakeling, S., Pinfield, S., Creaser, C., Fry, J., & Willett, P. (2017). Open-access megajournals: The future of scholarly communication or academic dumping ground? A review. *Journal of Documentation*, *73*(2), 263–283. https://doi.org/10.1108/JD-06-2016-0082

Spezi, V., Wakeling, S., Pinfield, S., Fry, J., Creaser, C., & Willett, P. (2018). "Let the community decide"? The vision and reality of soundness-only peer review in open-access mega-journals. *Journal of Documentation*, *74*(1), 137–161. https://doi.org/10.1108/JD-06-2017-0092

Thelwall, M. (2003). What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation. *Information Research*, *8*(3). https://informationr.net/ir/8-3/paper151.html?text=1

Thelwall, M. (2012). Journal impact evaluation: A webometric perspective. *Scientometrics*, *92*(2), 429–441. https://doi.org/10.1007/s11192-012-0669-x

Thelwall, M., & Kousha, K. (2015). Web indicators for research evaluation. Part 1: Citations and links to academic articles from the Web. *Profesional de la Información*, *24*(5), 587–606. https://doi.org/10.3145/epi.2015.sep.08

Vaughan, L., & Hysen, K. (2002). Relationship between links to journal Web sites and impact factors. *Aslib Proceedings*, *54*(6), 356–361. https://doi.org/10.1108/00012530210452555

Vaughan, L., & Thelwall, M. (2002). Web link counts correlate with ISI impact factors: Evidence from two disciplines. *Proceedings of the American Society for Information Science and Technology*, *39*(1), 436–443. https://doi.org/10.1002/meet.1450390148

Vaughan, L., & Thelwall, M. (2003). Scholarly use of the Web: What are the key inducers of links to journal Web sites? *Journal of the American Society for Information Science and Technology*, *54*(1), 29–38. https://doi.org/10.1002/asi.10184

Visser, M., van Eck, N. J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, *2*(1), 20–41. https://doi.org/10.1162/qss_a_00112

Wakeling, S., Willett, P., Creaser, C., Fry, J., Pinfield, S., & Spezi, V. (2016). Open-access megajournals: A bibliometric profile. *PLOS ONE*, *11*, e0165359. https://doi.org/10.1371/journal.pone.0165359, PubMed: 27861511

Wakeling, S., Creaser, C., Pinfield, S., Fry, J., Spezi, V., Willett, P., & Paramita, M. (2019). Motivations, understandings and experiences of open-access mega-journal authors: Results of a large-scale survey. *Journal of the Association for Information Science and Technology*, *70*(7), 754–768. https://doi.org/10.1002/asi.24154, PubMed: 31763360

Wellen, R. (2013). Open access, megajournals, and MOOCs: On the political economy of academic unbundling. *Sage Open*, *3*(4), 1–16. https://doi.org/10.1177/2158244013507271

APPENDIX A. LINK-BASED INDICATORS (MAJESTIC)

ID	Indicator	Scope	Level	Type
L1	Eigen centrality	Score that measures the prestige of a node (journal) if it is connected to many other nodes who themselves have high scores and vice versa (Ruhnau, 2000).	Journal-level	Weighted
L2	PageRank	Variant of eigen centrality, which also takes link direction and weight into account to measure the prestige of a node in a network.	Journal-level	Weighted
L3	Links counts (T)	Number of links received by the journal website from other external domains.	Journal-level	Size-dependent
L4	Links counts (TF25)	Number of links received by the journal website from other external domains, with a Source Trust Flow value of at least 25. It constitutes a selective link counts metrics.	Journal-level	Size-dependent
L5	Referring domains counts (RDC)	Number of web domains providing at least one link to the journal website.	Journal-level	Size-dependent
L6	Target-Trust Flow (TTF)	Score on a scale from 0 to 100 achieved by the journal website. It is based on the number of hyperlinks (and clicks on these links) from trusted seed sites that the journal website's URL receives. These seed sites have been manually curated by Majestic.	Journal-level	Weighted
L7	Target-Citation Flow (TCF)	Score on a scale from 0 to 100 achieved by the journal website, based on the number of hyperlinks it receives. It measures how often the journal website's URL is linked.	Journal-level	Weighted
L8	Source-Link Density	Percentage of surrounding links around the link to the journal website. Each linking webpage is divided into text segments. The number of links in the segment containing the link to the journal website is computed.	Journal-level	Relative
L9	Source-External outlink counts	Median of the total number of links from each journal website to other web domains.	Journal-level	Size-dependent
L10	Source-External outdomain counts	Median of the total number of web domains linked from each journal website.	Journal-level	Size-dependent

APPENDIX B. BIBLIOMETRIC-BASED INDICATORS

ID	Indicator	Scope	Source	Level	Type
B1	Age	Number of years since the journal release.	MDPI	Journal-level	–
B2	Publications (T)	Total number of publications published by the journal.	MDPI	Aggregated article-level	Size-dependent
B3	Publications (I)	Total number of publications published by a journal and indexed in Scopus.	SCOPUS	Aggregated article-level	Size-dependent
B4	Publications (R)	Number of publications published by the journal in the period 2017–2020 and indexed in Scopus.	SCOPUS	Aggregated article-level	Size-dependent
B5	Publications (C)	Number of publications published by a journal that has been cited.	SCOPUS	Aggregated article-level	Relative
B6	Publications (RC)	Number of publications published by a journal in the period 2017–2020 that have been cited in Scopus.	SCOPUS	Aggregated article-level	Relative
B7	SNIP	The number of citations given in the present year to publications in the past three years divided by the total number of publications in the past three years, normalized by discipline.	SCOPUS/ CWTS	Aggregated article-level	Normalized
B8	SJR	The average number of weighted citations received in the selected year by the documents published in the selected journal in the three previous years, excluding journal self-citations.	SCOPUS/ SCIMAGO	Aggregated article-level	Weighted
B9	Citescore	Citation counts to peer-reviewed documents published in a range of four calendar years, divided by the number of these documents in these same four years.	SCOPUS	Aggregated article-level	Relative
B10	Citations (T)	Total number of citations received by a journal indexed in Scopus.	SCOPUS	Aggregated article-level	Size-dependent
B11	Citations (R)	Total number of citations received by a journal in the period 2017–2020 in Scopus.	SCOPUS	Aggregated article-level	Size-dependent