



RESEARCH ARTICLE

When standard network measures fail to rank journals: A theoretical and empirical analysis

Giacomo Vaccario  and Luca Verginer 

ETH Zurich, Zurich, Switzerland

an open access  journal



Citation: Vaccario, G., & Verginer, L. (2022). When standard network measures fail to rank journals: A theoretical and empirical analysis. *Quantitative Science Studies*, 3(4), 1040–1053. https://doi.org/10.1162/qss_a_00225

DOI: https://doi.org/10.1162/qss_a_00225

Peer Review: https://publons.com/publon/10.1162/qss_a_00225

Received: 5 July 2022
Accepted: 27 September 2022

Corresponding Author:
Giacomo Vaccario
gvaccario@ethz.ch

Handling Editor:
Ludo Waltman

Copyright: © 2022 Giacomo Vaccario and Luca Verginer. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.



Keywords: citation network, citation paths, journal rankings, ranking bias, PageRank

ABSTRACT

Journal rankings are widely used and are often based on citation data in combination with a network approach. We argue that some of these network-based rankings can produce misleading results. From a theoretical point of view, we show that the standard network modeling approach of citation data at the journal level (i.e., the projection of paper citations onto journals) introduces fictitious relations among journals. To overcome this problem, we propose a citation path approach, and empirically show that rankings based on the network and the citation path approach are very different. Specifically we use MEDLINE, the largest open-access bibliometric data set, listing 24,135 journals, 26,759,399 papers, and 323,356,788 citations. We focus on PageRank, an established and well-known network metric. Based on our theoretical and empirical analysis, we highlight the limitations of standard network metrics and propose a method to overcome them.

1. INTRODUCTION

Bibliometricians and scientometricians often use citation-based indicators to rank and evaluate articles, journals, and authors in academic publishing (Hicks, Wouters et al., 2015; Owens, 2013). The impact factor and *h*-index are among the most widely used indicators to assess journals (Braun, Glänzel, & Schubert, 2006; Garfield, 1964; Hirsch, 2005). These indicators are *local* in the sense that they are based on the number of citations received by a given article, author, or journal within a given period. More sophisticated indicators have been developed using citation data and network analysis, such as the journal influence measure by Pinski and Narin (1976), a precursor to PageRank (Brin & Page, 1998), the Eigenfactor metric (Bergstrom, West, & Wiseman, 2008), and the SClmago Journal Rank (SJR) indicator (Guerrero-Bote & Moya-Aneón, 2012). SJR and the Eigenfactor are widely accessible indicators as they are reported on Scopus and the Journal Citation Report (Waltman, 2016), two of the largest commercial providers of bibliometric data. These indicators are based on eigenvector centralities and rely on *nonlocal* information. The rationale for using nonlocal information is to give more weight to citations from well-cited papers.

The assumption at the core of both local and nonlocal indicators is that the citing paper is influenced by the cited one. This assumption is motivated in two ways, namely by knowledge flow and the allocation of scientific credit. Specifically, it is assumed that knowledge flows in the opposite direction to citations. Thus, a paper receiving many citations contains knowledge that is often reused to create new knowledge (i.e., new papers). Similarly, authors endorse each other by citing their works, and hence, citations proxy credit allocation. Nonlocal

indicators also rely on the path transitivity assumption (i.e., given a network, all sequences of links represent a possible path). For example, given two paper citations ($c \rightarrow b$) and ($b \rightarrow a$), the transitivity assumption implies that there is a path ($c \rightarrow b \rightarrow a$), and hence, paper c may influence paper a via b . In other words, there is a possible causal connection between the three papers. We argue that the projection of citations among papers onto journals violates this transitivity assumption and that the causal connection is lost. We show that this violation affects journal rankings derived from nonlocal indicators.

The citation paths implied by the path transitivity assumption at the journal level might not match the empirical paper-to-paper citation paths for two reasons. First, the journal aggregation of the citation links may violate the path transitivity assumption. Given two consecutive links between journals A , B , and C , we do not know if the paper in B cited by the paper in A is also the paper citing the paper in C . Hence, we do not know if there was any influence from A to C via B . Path transitivity would instead incorrectly imply the presence of a path between A and C . Second, the time aggregation of citation links also violates path transitivity because we lose the ordering of citation events. In other words, when aggregating citations of papers published at different times, one erroneously assumes that younger papers can influence older ones.

In the present work, we study the effect of violating the path transitivity assumption in general. Note that our argumentation is valid for the knowledge flow and the scientific credit allocation perspectives. For this reason, we will use the term *fictitious influence* to refer to both.

The remainder of this paper is structured as follows. In Section 2, we briefly review the usage of journal rankings and recent findings in network science, highlighting the importance of the path transitivity assumption. Section 3 clarifies the pitfalls in projecting paper citations onto journals. In Section 4, we show empirically how journal rankings are biased by fictitious influence. Finally, in Section 5, we summarize and discuss our results.

2. LITERATURE REVIEW

Scientometricians and bibliometricians traditionally use citation analysis to develop quantitative indicators. These indicators are obtained by identifying the properties of documents through their cross-referencing. One example is the commonly used impact factor (Garfield, 1964). This captures the influence of journals by computing the average number of citations received by papers published in them. More sophisticated indicators have been developed by combining citation analysis with network analysis. Specifically, practitioners have used this analysis by constructing a citation network at the journal level. In this network, journals are nodes, and links are citations among papers published in them. Network measures, such as eigenvector and betweenness centralities, have been proposed as indicators to determine journal influence (Guerrero-Bote & Moya-Anegón, 2012; Pinski & Narin, 1976) and their interdisciplinarity (Leydesdorff, 2007; Leydesdorff, Wagner, & Bornmann, 2018). Moreover, such measures have been used to quantify the influence of authors (Radicchi, Fortunato et al., 2009) and papers (Chen, Xie et al., 2007; Zhou, Zeng et al., 2016).

As mentioned in the introduction, the use of citation data is motivated by the credit allocation mechanism. In other words, we assume that when an author cites a paper, they endorse the authors of the cited paper. When projecting citations onto journals, we implicitly assume the same, namely that citation links among journals capture credit allocation from one journal to the other. Additionally, most network measures rely on the *path transitivity assumption*. When inferring (from data) the existence of links from A to B and B to C , we automatically

permit a path of length two from A to C via B . Specifically, practitioners rely implicitly on this assumption to construct paths from citation links at the journal level. These paths represent possible flows of knowledge between journals and have been used to compute journals' similarity (Small & Koenig, 1977), journal influence (Pinski & Narin, 1976), and journal interdisciplinarity (Leydesdorff, 2007; Leydesdorff et al., 2018).

Despite the proliferation and wide usage of citation-based indicators, they are also criticized. A first concern arises from the fact that the citation practices vary across scientific fields (Bornmann & Daniel, 2008; Radicchi, Fortunato, & Castellano, 2008; Schubert & Braun, 1986). These differences introduce biases in citation-based indicators that cannot be easily overcome (Albarrán, Crespo et al., 2011; Vaccario, Medo et al., 2017; Waltman, van Eck, & van Raan, 2012). A second concern relates to the fact that publications are increasingly written by multiple coauthors. Various works have shown that coauthorship and the number of citations are deeply intertwined (Parolo, Pan et al., 2015; Persson, Glänzel, & Danell, 2004; Sarigöl, Pfitzner et al., 2014; Zingg, Nanumyan, & Schweitzer, 2020). Further concerns about using citation and bibliographic data come from the results on how editorial biases relate to social factors, such as previous coauthorship (Dondio, Casnici et al., 2019; Sarigöl, Garcia et al., 2017) and citation remuneration (Petersen, 2019). These findings, with many others, questioned the objectivity of citation-based indicators.

Recent advances in network theory have also raised concerns about the naive applications of network analytic tools to complex data (Borgatti & Everett, 2020; Butts, 2009; Zweig, 2011). In particular, Butts (2009) stresses the importance of correctly matching the unit and purpose of the analysis with the appropriate network representation. These concerns, we argue, are also valid when one applies network measures to rank journals using paper citations. To do this, one moves the unit of analysis from papers to journals without fully understanding the implications. Moreover, Mariani, Medo, and Zhang (2015) show how PageRank fails to identify significant nodes in time-evolving networks. This problem particularly applies to citation networks, which are continuously growing with the publication of new papers. Finally, Scholtes, Wider et al. (2014) and Vaccario, Verginer, and Schweitzer (2020) identify temporal properties in the dynamics of real-world systems, which violate the path transitivity assumption. These results raise concerns about correctly modeling dynamic processes on networks, such as scientific credit diffusion and knowledge flow.

To address the problem introduced by the violation of the path transitivity assumption, Lambiotte, Rosvall, and Scholtes (2019), Rosvall, Esquivel et al. (2014), and Scholtes et al. (2014) propose novel network models based on path abstraction. In this abstraction, instead of analyzing dyads, one looks at the time-ordered path sequences between nodes. Specifically, in citations, instead of concentrating on individual citation links, one should consider consecutive citations between articles to obtain *citation paths*. In our work, we use precisely this notion of citation paths to address the violation of path transitivity and its effect on journal rankings.

3. CITATION PATHS AND THE VIOLATION OF PATH TRANSITIVITY

In citation data, we usually have a set of documents $\mathcal{D} = \{p_1, p_2, \dots, p_N\}$, and a set of citation edges among them $\mathcal{E} = \{(p_2, p_1), (\dots), \dots\}$ where (p_j, p_i) represents a citation from document p_j to p_i with $i < j$. Note that the subscript of documents represents their publication order. So for example, p_1 is older than p_2 and p_2 is older than p_3 and so on and so forth.

We restrict our attention to the case where the documents in \mathcal{D} are scientific *papers* published in *journals*. From the sets of papers, \mathcal{D} , and of citations, \mathcal{E} , we can build a citation

network at the *paper* level, where nodes are the papers and links are the citations. One could argue that to investigate the citation network at the *journal* level, we could define a new network where nodes are journals that contain the papers, and links are the citations projected at the journal level. Even though the first part is correct, the second step discards information required to quantify indirect interjournal influence. To understand why this is the case, consider the example illustrated in Figure 1:

- (a) we have four papers $\mathcal{D} = \{p_1, p_2, p_3, p_4\}$ and three journals $\mathcal{J} = \{A, B, C\}$. The younger paper, p_4 , belongs to journal A , the second and third papers, p_2 and p_3 , belong to journal B and the older paper, p_1 , belongs to journal C . Additionally, we have the following citations $\mathcal{E} = \{(p_4, p_3), (p_3, p_1)\}$.
- (b) we have the exact same setting as before, but we change one citation link: instead of (p_3, p_1) , we have (p_2, p_1) , i.e., $\mathcal{E}' = \{(p_4, p_3), (p_2, p_1)\}$.

In Figure 1, we build the citation network at the *journal* level for both examples by aggregating and projecting the citations from the papers onto journals. Here, we find that the citation networks at the journal level are the same. However, the two citation networks at the *paper* level are not the same (i.e., $\mathcal{E} \neq \mathcal{E}'$). What do we miss by looking at the citation network at the journal level? In the first case, Figure 1(a), we see that information, knowledge, and influence can propagate from journal C to journal A via journal B thanks to the citation links. In the second case (see top of Figure 1(b)), this is impossible as neither citations nor citation paths connect papers in the journals A and C . When looking at the citation network at the journal level, we cannot detect such a difference.

The standard projection of paper citations onto journals implies the existence of citation paths among journals that do not exist. As illustrated in Figure 1(b), the projection implies the existence of a citation link from p_3 to p_2 just because they are published in the same journal. In other words, the projection introduces relations between journals that do not exist. As mentioned in Section 1, we refer to this problem as *fictitious influence*.

On a higher level, one can understand the problem of fictitious influence by comparing the topology of paper and journal citation networks. In paper citation networks, younger papers only cite older ones, and hence, we have *direct acyclic graphs* (see Figure 2(a)). On such

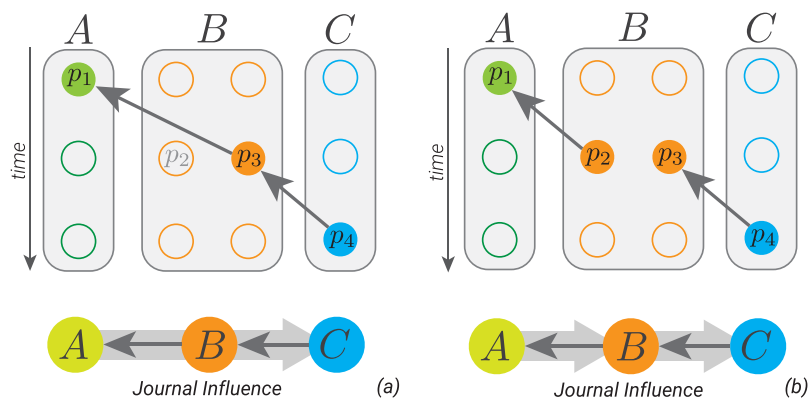


Figure 1. The citation projection from the paper to the journal level. In (a), we illustrate the case where journal A may influence journal C via journal B through citation links. The citation network at the journal level correctly captures this feature. However, in (b), A cannot influence C via B because no uninterrupted citation path connects the three journals. This fact is not captured by the citation network at the journal level.

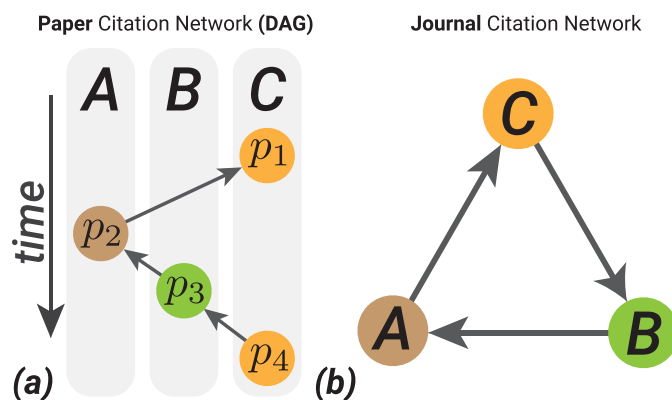


Figure 2. (a) An example of a paper citation network where four journals cite each other, creating a direct acyclic graph (DAG). (b) The projection of the acyclic structure depicted in (a) onto journals: A cycle is created.

topologies, we can define causal paths between papers: Younger papers reuse knowledge and information from older papers and cite them; in other words, older papers are a possible cause for the existence of younger ones. This statement about a causal link does not have to be true, as citations serve different purposes (Bornmann & Daniel, 2008). However, the acyclic topology of the citation network is a necessary (but not sufficient) condition for a causal connection to exist. When one projects the citations at the journal level, one creates a network with many cycles and breaks the possible causal structure captured by the acyclic topology (see Figure 2(b)). In other words, one cannot define causal paths between journals but only possible correlations between them. Hence, when using nonlocal indicators on the journal citation network, one neglects the causal structure and introduces a fictitious influence between journals. The following section shows that this fictitious influence affects journal rankings.

4. AN EMPIRICAL INVESTIGATION

We perform an empirical investigation to quantify the importance of fictitious influence on journal rankings. To be precise, we construct the citation network for papers and their projection at the journal level. Then, we use these two networks to derive two journal rankings based on PageRank. Other measures could have been used to discuss the problem of fictitious influence, but we chose PageRank as it is a prominent centrality measure widely used in networks science and scientometrics.

We compute the first ranking on the journal citation network, and the fictitious influence will affect this ranking. Then, we compute the second ranking using citation paths extracted from the paper citation networks, and the fictitious influence will not affect this second ranking. A stark discrepancy in the two rankings would indicate that fictitious influence is not innocuous. We choose PageRank because it is a prototypical nonlocal indicator used to rank journals (Guerrero-Bote & Moya-Anegón, 2012) in addition to websites (Brin & Page, 1998).

4.1. Data

We use citation data from MEDLINE obtained from the Torvik Group by combining various publicly available sources, including the MAG, AMiner, and PubMed Central. The data contains detailed information about 26,759,399 papers published between 1940 and 2016 with more than 460 Mio citations. To link the papers to the journals, we use a dump of PubMed. We find that these papers belong to 24,135 journals and have 323,356,788 citations. Note that

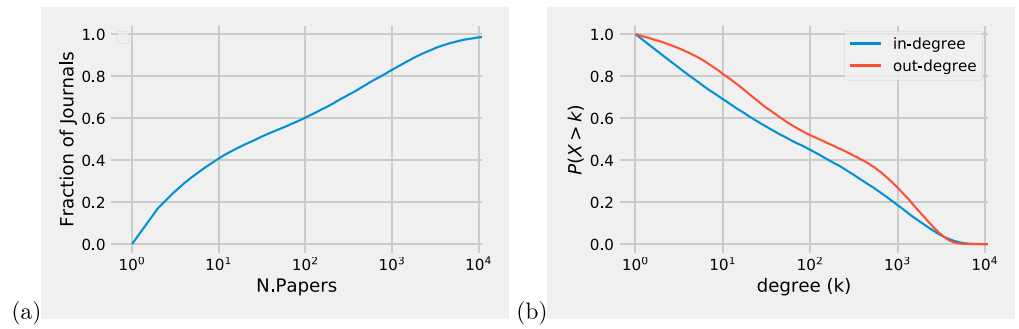


Figure 3. Number of papers and citations per journal. (a) Fraction of journals with at least a given number of papers. (b) Counter cumulative distribution of incoming (in-degree, blue) and outgoing citations (out-degree, red) per journal.

more than 50% of the journals have at least 20 papers, 50 incoming citations, and 100 outgoing citations (see Figure 3).

4.2. Methods

To rank journals according to PageRank, the standard approach is to project citations from papers onto journals to obtain a journal citation network. On such a network, it is then possible to compute PageRank centrality according to

$$PR = \left(\frac{1-d}{n} E + dT \right) PR \quad (1)$$

where PR is the vector containing the PageRank scores of the nodes, d the damping factor¹, E is an $n \times n$ matrix of 1s, and T is the transition matrix of the journal citation network (Brin & Page, 1998). As discussed in the previous section, this standard approach introduces fictitious influence between journals.

To understand how to avoid the fictitious influence when computing PageRank scores, let us recall the dynamic process captured by this centrality. In this process, we have a random walker placed on a node. The walker can either follow a link or “teleport” to a random node in the network from this node. Then, the PageRank score of a node is its visitation probability (i.e., how likely it is to find the walker on that node (Brin & Page, 1998)). For a detailed discussion of random walks and diffusion on networks see Masuda, Porter, and Lambiotte (2017).

The simplest way to address the fictitious influence problem is to unfold the random walk on the paper citation network instead of the journal citation network. Indeed, on the paper citation network, the random walker can *only* follow the empirical citation paths. To rank journals according to PageRank computed on these paths, we (a) place the random walker on a journal, (b) move the walker on a random paper belonging to the journal, and (c) let the walker follow the citation paths (i.e., on the paper citation network), or “teleport” to a random journal. Note that the teleportation occurs at the journal level as we want to capture journal importance using PageRank. After teleporting to a random journal, we are back to step (a).

Depending on the visitation probability of papers, we obtain the paper PageRank scores. By summing the PageRank scores of papers belonging to the same journal, we obtain the overall journal PageRank scores PR_C .

¹ We choose $d = 0.5$ as proposed by Chen et al. (2007).

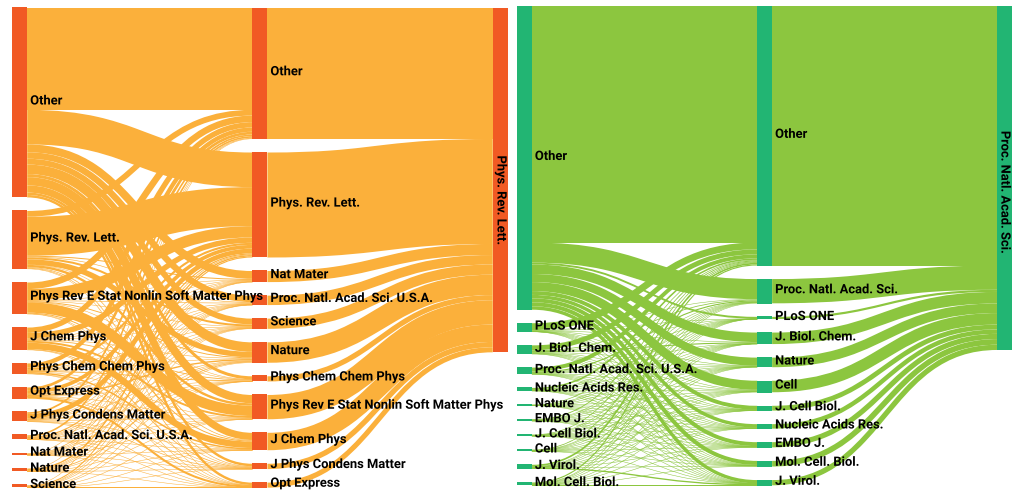


Figure 4. Citation paths of length two ending at papers published in “Phys. Rev. Lett.” (left) and “Proc. Natl. Acad. Sci.” (right). In each alluvial diagram, the rightmost column represents the focal final journal. The first and second columns represent the intermediate journals that are traversed to reach the focal journal. We show the top 10 journals contributing the most to the citation paths ending in the focal journal. All other journals are aggregated in the “Other” label. The size of a journal in the first/second column is proportional to the number of citation paths of length 2 that have this journal as the starting/middle journal. For example, “Phys. Rev. Lett.” is at the end of 13,143,749 citation paths, and 16% of these paths start from “Phys. Rev. Lett.” itself.

A straightforward implementation of such a process is to compute a *personalized* PageRank on the paper citation network. In other words, when the random walker teleports to a random paper/node, this paper is not chosen uniformly at random. If one chooses the paper uniformly at random, then journals with more papers are more likely to be the starting points of the random walk. This preference in the starting point would create biases in the random walk and the final ranking.

To ensure that the journals are uniformly sampled as starting points, we consider journal sizes (i.e., the total number of papers belonging to each of them). In particular, we define the size of a journal to which a paper i belongs to be S_i . Then, the probability to teleport to a paper i is inversely proportional to the size of the journal (to which the paper i belongs) S_i and the total number of journals analyzed n^2 . Formally, we write this as:

$$\widetilde{\text{PR}} = \left(\frac{1-d}{n} \widetilde{E} + d\widetilde{T} \right) \widetilde{\text{PR}} \quad (2)$$

where \widetilde{T} is the transition matrix of the paper citation network, \widetilde{E} is an $N \times N$ matrix with N equal to the number of papers, and each element $(\widetilde{E})_{ij} = 1/S_i$ with S_i being the size of the journal to which the paper (at row) i belongs. Then for each journal, we sum the scores $\widetilde{\text{PR}}$ of its papers and obtain PR_C .

4.3. Results

The number of unique citation paths of length 2 observed in the data set is 1,095,968,097. In Figure 4 we show an example of the extracted citation paths reaching “Proceedings of the National Academy of Sciences of the United States of America” (PNAS) and “Physical Review

² The i th element of the personalization vector is $\left(\frac{1}{S_i n}\right)$ and it is normalized as $\sum_i^N \frac{1}{S_i n} = \sum_k^n S_k \frac{1}{S_i n} = 1$, where the first equality comes from changing the summation index from papers to journals.

Letters” (PRL) in two steps. In this representation, we distinguish the top 10 journals, citing more often the respective focal journals. The figures show the variety and distribution of citations paths leading to PRL and PNAS. When projecting the paper citations at the journal level, the number of implied paths is 340,997,180,016. By projecting the citations, we introduce more than 300 billion paths that are never observed in the data. These are the paths that may give rise to fictitious influence. Note that if we consider even longer paths (i.e., longer than 2), the problem becomes even more pronounced.

In Table 1, we report the rankings of the top 20 journals according to PageRank computed with the standard network approach (PR). Additionally, we also report the rank position of these top 20 journals according to PageRank computed using the empirical citation paths (PR_C). This table shows that several journals have changed their position within the ranking. For example, we find that the ranking of journals such as “Proc. Natl. Acad. Sci. U.S.A.,”

Table 1. Top 20 journals according to PR. The first column (PR) contains the rank of the journal as computed on the journal citation network. The second column (PR_C) contains the rank of the journal as computed using the citation paths. The Change column contains an arrow pointing downwards when the journal loses positions in the PR_C ranking, an upward arrow if the journal gains positions, and an equal sign if the rank is the same

PR	PR _C	Change	Journal name
1	4	-3↓	Science
2	2	=	Proc. Natl. Acad. Sci. U.S.A.
3	3	=	Nature
4	1	+3↑	J. Biol. Chem.
5	5	=	N. Engl. J. Med.
6	6	=	Lancet
7	9	-2↓	JAMA
8	10	-2↓	Cell
9	7	+2↑	Circulation
10	14	-4↓	J. Clin. Invest.
11	13	-2↓	J. Immunol.
12	12	=	Cancer Res.
13	15	-2↓	Blood
14	19	-5↓	BMJ
15	20	-5↓	Nucleic Acids Res.
16	34	-8↓	J. Neurosci.
17	36	-9↓	Pediatrics
18	106	-88↓	Am J Public Health
19	22	-3↓	J. Exp. Med.
20	29	-9↓	Ann. Intern. Med.

Downloaded from http://direct.mit.edu/qss/article-pdf/3/4/1040/2070729/qss_a_00225.pdf by guest on 12 June 2024

“Nature” and “Lancet” are not affected. In contrast, journals such as “Science,” “J. Neurosci.” and “Am J Public Health” lose several positions. One extreme example is “Am J Public Health,” which moves from 18th to 106th position in the ranking. For other journals, such as “Pj. Biol. Chem.,” we see an improvement in their ranking position. In Table 2, we report the top 20 journals according to PR_C for completeness.

To quantify the difference between the two rankings, we first compute the overlap between the rankings. To be precise, we calculate the Jaccard similarity between two sets of journals listed among the top k journals according to the two approaches. In Figure 5, we report this similarity for different values of k . We see that for small values of k (i.e., when considering the top positions) we have about 80% overlap, indicating that the rankings share the same 80% of journals in these top positions. However, when comparing a larger fraction of the rankings, the

Table 2. Top 20 journals according to PR_C . The first column (PR) contains the rank of the journal computed on the journal citation network. The second column (PR_C) contains the rank of the journal computed using the citation paths. The Change column contains an arrow pointing downwards when the journal loses positions in the PR_C ranking, an upward arrow if the journal gains positions, and an equal sign if the rank is the same

PR	PR_C	Change	Journal name
4	1	+3↑	J. Biol. Chem.
2	2	=	Proc. Natl. Acad. Sci. U.S.A.
3	3	=	Nature
1	4	-3↓	Science
5	5	=	N. Engl. J. Med.
6	6	=	Lancet
9	7	-2↓	Circulation
27	8	+19↑	Phys. Rev. Lett.
7	9	-2↓	JAMA
8	10	-2↓	Cell
22	11	+11↑	Biochim. Biophys. Acta
12	12	=	Cancer Res.
11	13	-2↓	J. Immunol.
10	14	-4↓	J. Clin. Invest.
13	15	-2↓	Blood
31	16	-15↑	Biochem. J.
24	17	+7↑	Biochemistry
25	18	+7↑	Cancer
14	19	-5↓	BMJ
15	20	-5↓	Nucleic Acids Res.

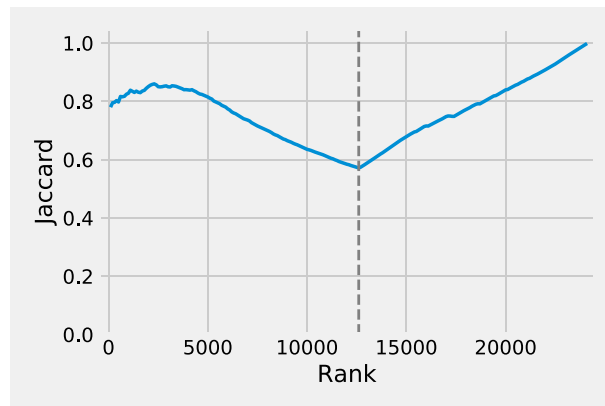


Figure 5. We plot the overlap between the two rankings (i.e., we consider the top k ranked journals according to PR and compute the intersection with the top k ranked journals according to PR_C).

intersection decreases to 60%. In other words, almost half of the journals listed in the two rankings are different. This indicates that the two rankings are extremely different. For larger values, the intersection increases linearly to the value 1. This result is expected as the complete rankings contain the same journals, and their similarity is trivially 1.

To further quantify the difference between the rankings coming from PR and PR_C , we compute the Kendall τ coefficient (KT) (Kendall, 1945). When considering the full ranking, we obtain a low value of around 0.5. Similar to before, we also compute the KT coefficient by considering the top k journals according to PR for different values of k . In Figure 6, we report how the KT coefficient changes with k . We find that it increases when considering the first $\approx 12,500$ ranked journal, and then we have a sharp decrease. First, note that the increase of the KT coefficient does not imply that the ranking is similar, as only less than 60% of the journals are the same. It only means that the relative positions of these 60% of journals are correlated.

Second, the sharp decrease of the KT coefficient marks the point where PR fails in ranking the journals. Indeed, PR assigns to many journals identical scores for a position lower than 12,500. In contrast, PR_C , which uses the empirical citation paths, assigns unique PageRank

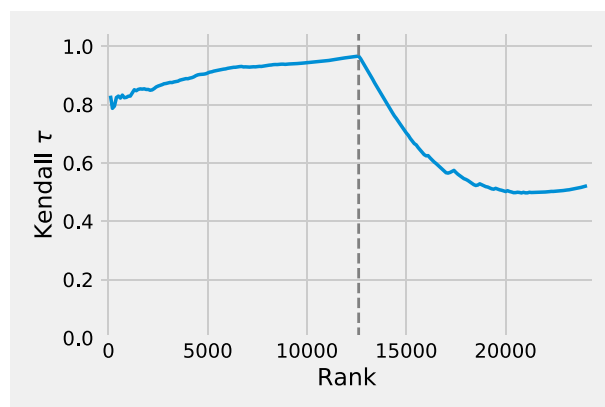


Figure 6. Comparison between the two rankings. We plot the Kendall's τ coefficient between the top k ranked journal according to the PageRank computed on the journal citation network and the corresponding relative rankings of these k journals according to the PageRank computed using the empirical citation paths.

scores also to these less central journals. Note that to rank these journals, PR_C relies on fewer assumptions (i.e., we have relaxed the transitivity assumption).

The rankings created with and without correcting for fictitious influence are substantially different. In other words, the discrepancy in the rankings indicates that computing the network measure on the journal citation network yields wrong and possibly misleading results.

5. DISCUSSION

Increasing attention has been given to data to guide science and research policy (Hicks et al., 2015). This usage has produced the need to develop new and more sophisticated measures to quantify scientific performance. In particular, several measures have been constructed by combining bibliometric and network methods. However, even though numerous measures have been proposed to rank journals, there is no ground truth (i.e., a ranking that is universally accepted). Focusing on measures for journal impact, we have shown how a naive combination of these methods may lead to misleading or even wrong results. Specifically, we have argued that a standard projection of paper citations onto journals may introduce nonexistent relations, which we call fictitious influence.

First, we have explained how fictitious influence arises from the transitivity assumption, which is a common and central assumption in many standard network methods. In particular, we have identified two ways in which fictitious influence may arise: the time and the journal aggregation of citation links. By time-aggregating citations, one loses the ordering of citations between journals. By aggregating citations inside journals, one mixes the incoming and outgoing citations of papers belonging to the same journal. These aggregations introduce relations between journals that do not respect the empirical citation patterns among papers.

Second, we have shown that the fictitious influence is not an innocuous effect when computing rankings of journals. To do this, we have used real-world citation data from MEDLINE, the largest open-access bibliometric data set in the life sciences. We have first computed the number of paths of length 2 on the paper citation and the journal citation network with this data. The former represents the empirically observed paths, whereas the latter represents implied paths after projecting paper citations onto journals. We find that only 0.3% of the implied citation paths are present in the data set. This discrepancy highlights that the projection introduces many wrong citation paths, allowing for fictitious influence. Then we have computed two journal rankings using the standard journal citation network and the paper citation network. We have computed the PageRank scores of journals biased by the fictitious influence on the former network. On the latter network, we have computed the unbiased PageRank scores. Among the top 2,500 journals, we have found that the overlap between the rankings is relatively high (≈ 0.85) with a low Kendall's τ 0.70. These results indicate that even though the same journals belong to the top of the rankings, they occupy different positions. When considering the top 12,500 journals, we have found that the overlap between the rankings decreases to approximately 0.60, and Kendall's τ increases. This indicates that the two rankings become extremely different, as they share less than 60% of the journals, but the relative positions of these journals are consistent across rankings. Overall, our results indicate that the fictitious influence significantly affects the reliability of PageRank as a measure for journal ranking.

One may find it strange that the large difference between the observed and implied citation paths (only a 0.3% overlap) still results in a Kendall τ of ~ 0.50 . This result can be understood by considering the definition of PageRank (Eq. 1). PageRank is influenced by paths of any length. These paths are weighted by the powers of the damping factors, d^l , where d is the

damping factor and l is the path length. We have set $d = 0.5$, and hence the paths of length 2 have a weight of 0.25 (and paths of length 3 have a weight of 0.125, etc.). In other words, paths of length 2 have a 25% probability of being used by the random walker before teleporting. Hence, their influence on the visitation probabilities of PageRank is limited.

To overcome the problem of fictitious influence, one could argue that higher order networks are possible solutions. On the one hand, these network models could help because centrality measures computed on them correlate with measures computed on the original sequence data (Scholtes, 2017). On the other hand, they assume that there are temporal correlations in the data allowing us to summarize them. For an overview and applications of these models to various data, see Lambiotte et al. (2019). Higher order networks allow the use of network measures while addressing the technical issue of fictitious influence. However, the development of adequate scientometric indicators is a very complex task. For example, the Leiden Manifesto suggests balancing an indicator's complexity with its transparency (point 4 in Hicks et al., 2015). Using well-known network measures could increase transparency; at the same time, the added complexity of the higher order networks could obscure their meaning. Hence, the viability of these methods will depend on the intended usage.

This work has the following primary limitations. We used only citation data from MEDLINE, a database with a primary focus on bibliometric information in the life sciences. Hence, we have analyzed a biased sample of bibliographic data. This bias limits the reliability of the obtained rankings. However, the discrepancies found between the rankings highlight the fundamental problem of fictitious influence. The second limitation is that we only considered one possible nonlocal indicator, PageRank. There are many other nonlocal network indicators, and for each of them, the effect of fictitious influence could be different. Future work can replicate our analysis on a larger citation data set and consider other nonlocal indicators to address these limitations.

To conclude, we have shown that journal rankings based on nonlocal journal indicators may be wrong. This problem arises because a naive projection of paper citations onto journals introduces fictitious relations. To address this problem, we propose to adopt a path-based perspective. With this work, we have highlighted the shortcomings of the standard network approach to create journals rankings. Also, we propose a new perspective to perform citation analysis at the journal level. The path perspective supports research evaluators and administrators in the challenging tasks of assessing scientific performance.

ACKNOWLEDGMENTS

We thank Frank Schweitzer for helpful discussions. Also, we thank Ingo Scholtes for his many critiques and suggestions, which improved the manuscript.

AUTHOR CONTRIBUTIONS

Giacomo Vaccario: Conceptualization, Formal analysis, Investigation, Methodology, Writing—original draft, Writing—review & editing. Luca Verginer: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing—original draft, Writing—review & editing.

COMPETING INTERESTS

The authors have no competing interests.

DATA AVAILABILITY

We use citation data from MEDLINE obtained from the Torvik Group by combining various publicly available sources, including the MAG, AMiner, and PubMed Central. Access to this data was obtained by getting in contact with the Torvik group: <https://abel.lis.illinois.edu/>. To link the papers to the journals, we use a dump of PubMed: https://www.nlm.nih.gov/databases/download/pubmed_medline.html.

FUNDING INFORMATION

No funding was received for this research.

REFERENCES

- Albarrán, P., Crespo, J. A., Ortuño, I., & Ruiz-Castillo, J. (2011). The skewness of science in 219 sub-fields and a number of aggregates. *Scientometrics*, *88*(2), 385–397. <https://doi.org/10.1007/s11192-011-0407-9>
- Bergstrom, C. T., West, J. D., & Wiseman, M. A. (2008). The Eigenfactor™ metrics. *Journal of Neuroscience*, *28*(45), 11433–11434. <https://doi.org/10.1523/JNEUROSCI.0003-08.2008>, PubMed: 18987179
- Borgatti, S. P., & Everett, M. G. (2020). Three perspectives on centrality. In *The Oxford handbook of social networks* (pp. 334–351). <https://doi.org/10.1093/oxfordhb/9780190251765.013.22>
- Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80. <https://doi.org/10.1108/00220410810844150>
- Braun, T., Glänzel, W., & Schubert, A. (2006). A Hirsch-type index for journals. *Scientometrics*, *69*(1), 169–173. <https://doi.org/10.1007/s11192-006-0147-4>
- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, *30*(1), 107–117. [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)
- Butts, C. T. (2009). Revisiting the foundations of network analysis. *Science*, *325*(5939), 414–416. <https://doi.org/10.1126/science.1171022>, PubMed: 19628855
- Chen, P., Xie, H., Maslov, S., & Redner, S. (2007). Finding scientific gems with Google's PageRank algorithm. *Journal of Informetrics*, *1*(1), 8–15. <https://doi.org/10.1016/j.joi.2006.06.001>
- Dondio, P., Casnici, N., Grimaldo, F., Gilbert, N., & Squazzoni, F. (2019). The “invisible hand” of peer review: The implications of author-referee networks on peer review in a scholarly journal. *Journal of Informetrics*, *13*(2), 708–716. <https://doi.org/10.1016/j.joi.2019.03.018>
- Garfield, E. (1964). “Science Citation Index”—A new dimension in indexing. *Science*, *144*(3619), 649–654. <https://doi.org/10.1126/science.144.3619.649>, PubMed: 17806988
- Guerrero-Bote, V. P., & Moya-Aneón, F. (2012). A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *Journal of Informetrics*, *6*(4), 674–688. <https://doi.org/10.1016/j.joi.2012.07.001>
- Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). Bibliometrics: The Leiden Manifesto for research metrics. *Nature*, *520*, 429–431. <https://doi.org/10.1038/520429a>, PubMed: 25903611
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, *102*(46), 16569–16572. <https://doi.org/10.1073/pnas.0507655102>, PubMed: 16275915
- Kendall, M. G. (1945). The treatment of ties in ranking problems. *Biometrika*, *33*, 239–251. <https://doi.org/10.1093/biomet/33.3.239>, PubMed: 21006841
- Lambiotte, R., Rosvall, M., & Scholtes, I. (2019). Understanding complex systems: From networks to optimal higher-order models. *Nature Physics*, *15*(4), 313–320. <https://doi.org/10.1038/s41567-019-0459-y>, PubMed: 30956684
- Leydesdorff, L. (2007). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology*, *58*(9), 1303–1319. <https://doi.org/10.1002/asi.20614>
- Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2018). Betweenness and diversity in journal citation networks as measures of interdisciplinarity—A tribute to Eugene Garfield. *Scientometrics*, *114*(2), 567–592. <https://doi.org/10.1007/s11192-017-2528-2>, PubMed: 29449751
- Mariani, M. S., Medo, M., & Zhang, Y.-C. (2015). Ranking nodes in growing networks: When PageRank fails. *Scientific Reports*, *5*, 16181. <https://doi.org/10.1038/srep16181>, PubMed: 26553630
- Masuda, N., Porter, M. A., & Lambiotte, R. (2017). Random walks and diffusion on networks. *Physics Reports*, *716*, 1–58. <https://doi.org/10.1016/j.physrep.2017.07.007>
- Owens, B. (2013). Research assessments: Judgement day. *Nature*, *502*(7471), 288–290. <https://doi.org/10.1038/502288a>, PubMed: 24132272
- Parolo, P. D. B., Pan, R. K., Ghosh, R., Huberman, B. A., Kaski, K., & Fortunato, S. (2015). Attention decay in science. *Journal of Informetrics*, *9*(4), 734–745. <https://doi.org/10.1016/j.joi.2015.07.006>
- Persson, O., Glänzel, W., & Danell, R. (2004). Inflationary bibliometric values: The role of scientific collaboration and the need for relative indicators in evaluative studies. *Scientometrics*, *60*(3), 421–432. <https://doi.org/10.1023/B:SCIE.0000034384.35498.7d>
- 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>
- Pinski, G., & Narin, F. (1976). Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics. *Information Processing & Management*, *12*(5), 297–312. [https://doi.org/10.1016/0306-4573\(76\)90048-0](https://doi.org/10.1016/0306-4573(76)90048-0)
- Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences*, *105*(45), 17268–17272. <https://doi.org/10.1073/pnas.0806977105>, PubMed: 18978030
- Radicchi, F., Fortunato, S., Markines, B., & Vespignani, A. (2009). Diffusion of scientific credits and the ranking of scientists.

- Physical Review E*, 80(5), 056103. <https://doi.org/10.1103/PhysRevE.80.056103>, PubMed: 20365039
- Rosvall, M., Esquivel, A. V., Lancichinetti, A., West, J. D., & Lambiotte, R. (2014). Memory in network flows and its effects on spreading dynamics and community detection. *Nature Communications*, 5, 4630. <https://doi.org/10.1038/ncomms5630>, PubMed: 25109694
- Sarigöl, E., Garcia, D., Scholtes, I., & Schweitzer, F. (2017). Quantifying the effect of editor-author relations on manuscript handling times. *Scientometrics*, 113(1), 609–631. <https://doi.org/10.1007/s11192-017-2309-y>, PubMed: 29056793
- Sarigöl, E., Pfitzner, R., Scholtes, I., Garas, A., & Schweitzer, F. (2014). Predicting scientific success based on coauthorship networks. *EPJ Data Science*, 3(1), 9. <https://doi.org/10.1140/epjds/s13688-014-0009-x>
- Scholtes, I. (2017). When is a network a network?: Multi-order graphical model selection in pathways and temporal networks. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1037–1046). ACM. <https://doi.org/10.1145/3097983.3098145>
- Scholtes, I., Wider, N., Pfitzner, R., Garas, A., Tessone, C. J., & Schweitzer, F. (2014). Causality-driven slow-down and speed-up of diffusion in non-Markovian temporal networks. *Nature Communications*, 5, 5024. <https://doi.org/10.1038/ncomms6024>, PubMed: 25248462
- Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics*, 9(5–6), 281–291. <https://doi.org/10.1007/BF02017249>
- Small, H. G., & Koenig, M. E. (1977). Journal clustering using a bibliographic coupling method. *Information Processing & Management*, 13(5), 277–288. [https://doi.org/10.1016/0306-4573\(77\)90017-6](https://doi.org/10.1016/0306-4573(77)90017-6)
- Vaccario, G., Medo, M., Wider, N., & Mariani, M. S. (2017). Quantifying and suppressing ranking bias in a large citation network. *Journal of Informetrics*, 11(3), 766–782. <https://doi.org/10.1016/j.joi.2017.05.014>
- Vaccario, G., Verginer, L., & Schweitzer, F. (2020). The mobility network of scientists: Analyzing temporal correlations in scientific careers. *Applied Network Science*, 5(1), 36. <https://doi.org/10.1007/s41109-020-00279-x>
- Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365–391. <https://doi.org/10.1016/j.joi.2016.02.007>
- Waltman, L., van Eck, N. J., & van Raan, A. F. (2012). Universality of citation distributions revisited. *Journal of the American Society for Information Science and Technology*, 63(1), 72–77. <https://doi.org/10.1002/asi.21671>
- Zhou, J., Zeng, A., Fan, Y., & Di, Z. (2016). Ranking scientific publications with similarity-preferential mechanism. *Scientometrics*, 106(2), 805–816. <https://doi.org/10.1007/s11192-015-1805-1>
- Zingg, C., Nanumyan, V., & Schweitzer, F. (2020). Citations driven by social connections? A multi-layer representation of coauthorship networks. *Quantitative Science Studies*, 1(4), 1493–1509. https://doi.org/10.1162/qss_a_00092
- Zweig, K. A. (2011). Good versus optimal: Why network analytic methods need more systematic evaluation. *Central European Journal of Computer Science*, 1(1), 137–153. <https://doi.org/10.2478/s13537-011-0009-x>