Journal of Informetrics 17 (2023) 101456

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

The k-step *h-index* in citation networks at the paper, author, and institution levels

Alex Jie Yang [a,b,c,*], Linwei Wu [a,b,c], Qi Zhang [a,b,c], Hao Wang [a,b,c], Sanhong Deng [a,b,c,*]

[a] *School of Information Management, Nanjing University, Nanjing 210023, China*
[b] *Jiangsu Key Laboratory of Data Engineering and Knowledge Service, Nanjing 210023, China*
[c] *International Joint Informatics Laboratory, Nanjing University, Nanjing 210023, China*

ARTICLE INFO

Keywords:
h-index
Citation network
Science of science
Innovation

ABSTRACT

The evaluation of scientific impact plays a crucial role in assessing research contributions. In this study, we introduce the concept of the k-step *h-index* and investigate its applicability in citation networks at different levels, including papers, authors, and institutions. By incorporating higher generations of citation information, the k-step *h-index* provides a comprehensive and nuanced measure of scientific influence. It demonstrates exponential growth in k-step citations, capturing valuable information from the Hirsch core and tail. Through power law distribution analysis, we uncover the presence of highly influential entities coexisting with less influential ones, revealing the heterogeneity of impact within citation networks. To validate the effectiveness of the k-step *h-index*, we utilize a vast dataset from APS, conducting a thorough examination of its consistency and convergent validity. Our findings demonstrate strong correlations between the k-step *h-index* and conventional metrics, as well as alignment with measures of innovation. This confirms the reliability of the k-step *h-index* and its ability to capture innovative contributions. Notably, when compared to benchmarks, the k-step *h-index* outperforms in accurately ranking expert-selected items, including milestone papers, distinguished authors, and prestigious institutions. Higher values of the k-step *h-index* consistently exhibit superior performance, showcasing their predictive power in identifying prominent scientific entities. These findings hold significant implications for research evaluation, policy-making, and strategic planning, as they pave the way for a more holistic understanding of scholarly contributions.

1. Introduction

The science of science within the framework of complex networks has garnered significant attention across diverse academic disciplines in recent years (Fortunato et al., 2018; Jo et al., 2022; L. Liu et al., 2023; Lü et al., 2016; Park et al., 2023; Waltman, 2016; Wang & Uzzi, 2022; Zeng et al., 2017). A multitude of metrics that incorporate network analysis have emerged as valuable instruments for quantifying the impact of scientific papers (Hutchins et al., 2016; Park et al., 2023; Wang et al., 2013), evaluating the performance of authors (Ding & Cronin, 2011; Jo et al., 2022), and assessing the achievements of institutions (Dunaiski et al., 2018; Zhang & Wu, 2021).

* Corresponding authors.
E-mail addresses: alexjieyang@outlook.com (A.J. Yang), sanhong@nju.edu.cn (S. Deng).

https://doi.org/10.1016/j.joi.2023.101456
Received 12 February 2023; Received in revised form 12 August 2023; Accepted 12 September 2023
Available online 23 September 2023
1751-1577/© 2023 Elsevier Ltd. All rights reserved.

The introduction of the *h-index* by Jorge Hirsch (2005) sparked widespread interest in its properties, variations, and mathematical modeling within the field of information science (Alonso et al., 2009; Lü et al., 2016; Poirrier et al., 2021). In recent years, the *h-index* has been expanded to encompass other levels, including papers and institutions (Blumm et al., 2012; Iniguez et al., 2022), and has found broad applications. The *h-index* effectively strikes a balance between quality and quantity (Zeng et al., 2017) while demonstrating resilience against random outliers and citation inflation (Alonso et al., 2009; Waltman, 2016). However, despite its numerous advantages, the *h-index* overlooks citation distributions and the number of publications, potentially leading to inconsistencies under certain circumstances (Brito & Navarro, 2021; Waltman & Van Eck, 2012). Moreover, several other limitations of the *h-index* have been identified, such as inconsistencies across disciplinary standards, biases against early-career items, and a lack of discriminative power (Bornmann & Leydesdorff, 2018; Mohammed et al., 2020; Prathap, 2019). These limitations may, in part, arise from the disregard of information in the tail of the Hirsch index (Ye & Rousseau, 2010).

From the perspective of complex networks, the *h-index* is essentially a local metric that solely considers direct information from neighboring nodes (Lü et al., 2016). However, previous findings have demonstrated that global (high-generation) metrics, which take into account the entire network structure, generally outperform local metrics (Hu & Rousseau, 2017; Ullah et al., 2021). In fact, numerous global metrics have been employed to identify influential scientific entities (Maji, 2020; Rosvall et al., 2014). These global metrics overcome the limitations associated with local metrics and surpass the citation count in detecting groundbreaking research (Rosvall et al., 2014; Xu et al., 2020) and esteemed authors (Yang, Hu, et al., 2023).

Building upon these considerations, we put forth the proposition of introducing the k-step *h-index*. By incorporating a broader spectrum of global information derived from high-generation citation networks, the k-step *h-index* augments the depth and dimensionality of the original *h-index*, thus providing valuable insights for the development of talent-based policies that promote scientific advancement. With the establishment of k-step *h-indices* at the levels of papers, authors, and institutions, along with the definition of the k-step Hirsch core, this study aims to address the following Research Questions (RQs):

(**RQ1**) What are the definitions and characteristics of the k-step *h-indices* at the levels of papers, authors, and institutions? How does the methodology of k-step *h-indices* differ from other h-type indices?

(**RQ2**) To what extent does the k-step *h-index* enhance heterogeneity in rankings by incorporating citation information from both the k-step Hirsch core and Hirsch tail?

(**RQ3**) Do the k-step *h-indices* demonstrate consistency with existing metrics, particularly in relation to measures of innovation?

(**RQ4**) To what degree does the k-step *h-index* outperform the original indices in accurately identifying items selected by experts at the levels of papers, authors, and institutions?

By addressing these questions, our objective is to shed light on the distinct attributes and functionalities of the k-step *h-index*, thereby contributing to the advancement of our comprehension of the intricate dynamics underlying the assessment of scholarly impact. To accomplish this, we integrate two large-scale datasets and employ author and institution name disambiguation procedures. We employ multiple methodologies to examine the effectiveness of the k-step *h-index* at each level. The subsequent sections of this paper are organized as follows: In Section 2, we provide an extensive review of the existing literature. Section 3 introduces the definitions and properties of the k-step *h-indices*. Section 4 outlines the data and preprocessing procedures. In Section 5, we analyze the distributions, perform power law fitting, and investigate the k-step Hirsch core and Hirsch tail. Section 6 delves into the principal findings, discussing the consistency and convergent validity of the k-step *h-indices*. Finally, in Section 7, we offer a concise and insightful conclusion, summarizing the key contributions of our study and highlighting its broader implications.

2. Literature review

2.1. High generations in citation cascades

The incorporation of network analysis methods within the field of information science holds great potential for enhancing metrics and evaluation processes (Acemoglu et al., 2016; Barabási, 2016; Iniguez et al., 2022; Yang, Deng, et al., 2023; Yang, Hu, et al., 2023). Within the context of citation networks, the notion of citation generations was first introduced by Margolis (1967) and subsequently explored in greater depth by Rousseau (1987), who delved into the assessment of indirect influence based on citation relationships spanning multiple generations.

Expanding upon this foundational work, Min et al. (2021) defined and quantified citation cascades, incorporating focal papers, direct and indirect citations, kth citation generations, and the flow of knowledge as integral components. These metrics take into account the successive layers of citations initiated by a focal paper, whereby subsequent papers cite both the focal paper and its prior citations, resulting in a cascading effect (Yang, Deng, et al., 2023). Wang et al. (2021) conducted an analysis of the global structure of citation networks to assess the influence of publications, yielding results that surpassed those obtained through conventional metrics. Additionally, Hu and Rousseau (2017) demonstrated the effectiveness of indirect measures in identifying high-value research by investigating three generations of citation documents.

By employing network analysis techniques, scholars have gained novel insights into the intricate dynamics of citation networks, thereby enhancing the understanding of scholarly influence and knowledge dissemination processes (Wang et al., 2021; Zeng et al., 2017). These methods have proven to be a valuable addition to traditional evaluation approaches, as they uncover hidden connections and patterns that may go unnoticed through standard citation-based metrics (Mariani et al., 2019; Wolcott et al., 2016). Through the examination of citation cascades and multi-generation citation relationships, researchers have been able to identify influential papers and gauge their impact more accurately, leading to improved assessments of research quality and significance.

Table 1
A summary of the h-type indices.

Type	Description	Representative works
Normalization	Employing normalization techniques to address the limitations of the *h-index* by considering factors such as field, age, co-authorship, and other relevant factors. These approaches aim to provide a more accurate and fair evaluation of researchers' scientific impact.	h_α-index (Hirsch, 2019), generalized *h-index* (Radicchi et al., 2008), h_i-index (Bi, 2023) and h_a-index (Fassin, 2023)
Generalization to other levels	Expanding the application of the *h-index* beyond the individual author level to encompass the paper level and other aggregated levels	The paper-level *h-index* (Egghe, 2010; Schubert, 2009; Yan et al., 2016), the journal-level *h-index* (Braun et al., 2006), and the successive *h-index* (Schubert, 2007)
Network Node	Extending the *h-index* to complex networks to enable the evaluation of the significance of individual network nodes, offering insights into their importance within the network structure. This extension also seeks to establish a measure that can converge with the coreness of a node.	Key studies in this area include the works of Zhao et al. (2011), Lü et al. (2016), Wu et al. (2019) and Zhai et al. (2018, 2019)
High generation	Incorporating high generation information derived from citation cascades to expand the Hirsch core of the *h-index*.	To the best of our knowledge, there are currently no specific works within this area of research.

2.2. H-type indices

The original *h-index*, introduced by Hirsch, primarily aimed to evaluate the scientific impact of individual authors (Hirsch, 2005). However, in order to achieve more nuanced assessments of researchers, several h-type indices have been proposed to overcome the limitations of the original *h-index*. These include the g-index (Egghe, 2006), h_α-index (Hirsch, 2019), generalized *h-index* (Radicchi et al., 2008), h_i-index (Bi, 2023), h_a-index (Fassin, 2023), among others (Alonso et al., 2009; Y. Liu et al., 2023; Zeng et al., 2017). The emergence of h-type indices has had a significant impact on the field of information science in the past two decades (Alonso et al., 2009; Fortunato et al., 2018; Waltman, 2016; Zeng et al., 2017). These extensions primarily utilize normalization techniques to overcome the limitations of the *h-index* (see Table 1), but do not extend the citation information within the Hirsch core.

Beyond the individual author level, Schubert (2007) introduced the concept of successive *h-indices*, which was later formalized by Egghe (2008). By employing successive *h-indices*, it becomes possible to measure scientific entities at higher levels of aggregation, enabling the evaluation of an institution's influence and academic reputation through these successive *h-indices* (Arencibia-Jorge, 2009). Schubert (2009) also introduced the paper-level *h-index* for assessing individual publications and demonstrated its robustness across different fields and ages, solidifying its position as a valuable tool for paper-level evaluation. Egghe (2010) applied the Lotkaian model to analyze the single publication *h-index* of papers within the Hirsch core of a researcher, as well as the indirect *h-index*. Empirical analysis conducted by Yan et al. (2016) confirmed the coherence of utilizing the *h-index* and other h-type indices for evaluating individual publications.

In recent years, the integration of *h-indices* into network analysis has yielded fruitful results (Y. Liu et al., 2023). Zhao et al. (2011) introduced the concept of the h-degree in citation networks as a means to identify critical nodes. Lü et al. (2016) extended the notion of h-type metrics to nodes within networks and developed an iterative algorithm for their computation. Wu et al. (2019) calculated the *h-index* of each node on weighted networks by replacing the node's degree with the node's strength. Zhai et al. (2018, 2019) proposed bi-directional *h-indices* as a more comprehensive measure in citation networks, incorporating the h-degree as the initial value and leveraging additional information from evolving citation networks to update the significance of a node.

However, to the best of our knowledge, empirical studies investigating higher-generation *h-indices* for assessing the influence of publications, authors, or institutions are currently lacking. This research gap serves as a crucial motivation for the present study, where we aim to address this gap and explore the potential of higher-generation *h-indices* as evaluation tools.

3. Methodology

3.1. Paper-level k-step h-index

Citation-based metrics have gained significant traction as valuable tools for assessing the academic impact of publications, authors, and institutions (Blumm et al., 2012; Iniguez et al., 2022). Among these metrics, the *h-index* has emerged as a widely used citation-based indicator for evaluating academic impact (Koltun & Hafner, 2021). While traditionally employed for author-level assessment, the *h-index* has also been extended to encompass evaluations at the publication and institution levels.

The calculation of the *h-index* involves sorting a set of papers in descending order based on their citation counts. Starting from the highest-ranked paper to the lowest-ranked paper, the h^{th} ranked paper must have at least h citations (Brito & Navarro, 2021). To account for multiple citation generations, the analysis can be expanded to include citation documents that cite the citing publications of the focal paper, offering a more comprehensive understanding of their impact, particularly for highly cited articles (Wang et al., 2021). By incorporating multiple citation generations, we propose the k-step *h-index*, applicable at the paper, author, and institution levels.

In the context of paper-paper citation networks (as depicted in Fig. 1a), we begin by introducing the definition of the k-step *h-index* at the paper level. This index takes into consideration broad citation information of a focal paper, including indirect citations through multiple generations.

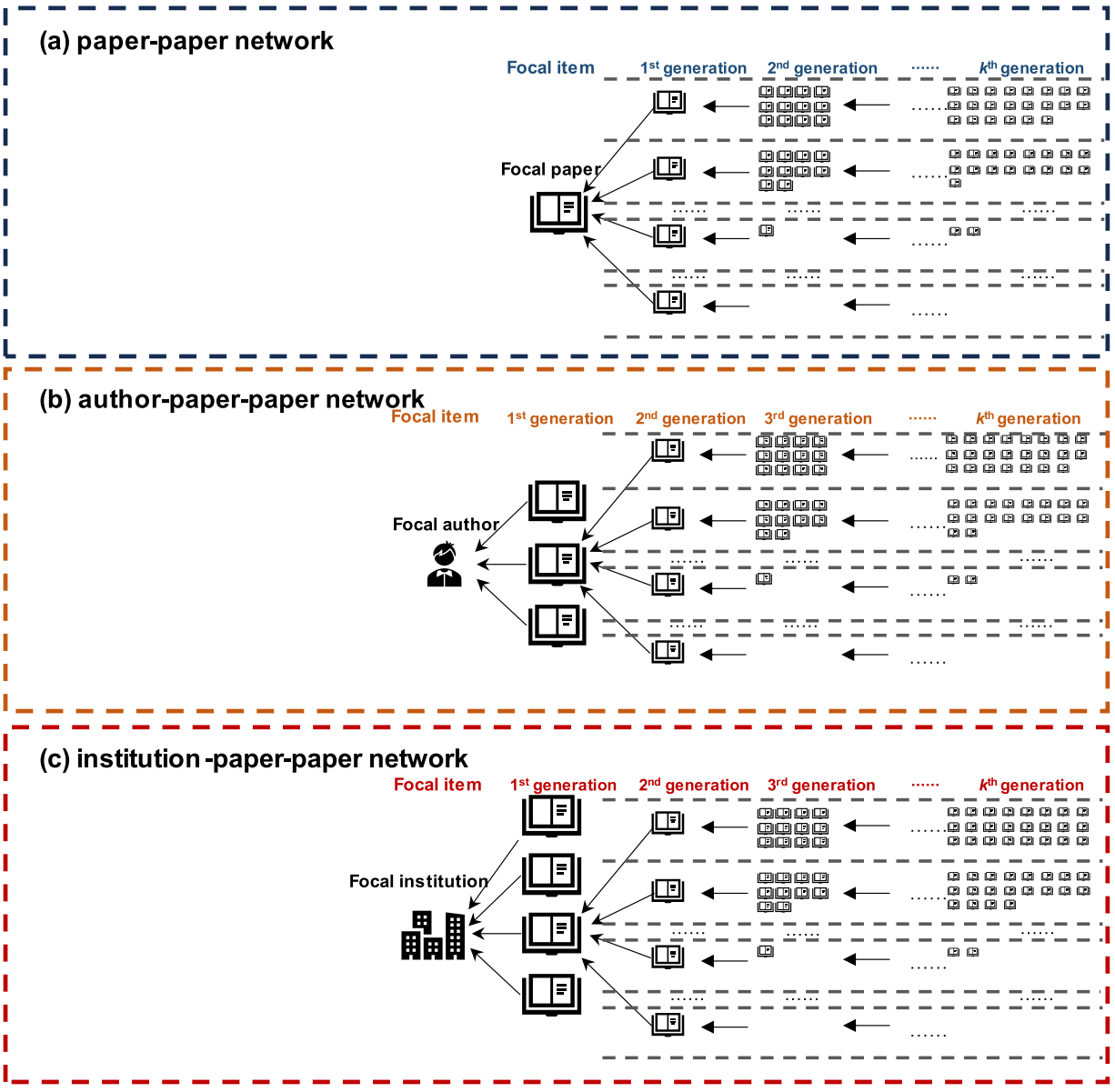


Fig. 1. Network visualizations for calculating k-step *h-indices* at different levels: Paper, Author, and Institution. (a) The paper-paper network of a focal publication. (b) The author-paper-paper network of a focal author. (c) The institution-paper-paper network of a focal institution. Each level considers citation documents from the first generation to the kth generation. The generation layout differs between (a) the paper-paper network and (b) the author-paper-paper network and (c) the institution-paper-paper network. Note that the k-step *h-index* utilizes k+1-step citation information at each level.

Definition 1. In the paper-paper network of a focal paper C, the k-step *h-index* of C, denoted as $h_k(C)$, is the highest integer value for which, in the kth-generation of C's citation cascades, there are at least $h_k(C)$ documents with a minimum citation count of $h_k(C)$.

3.2. Author-level k-step h-index

The assessment of author influence can be conducted within paper-paper citation networks or co-author networks. However, when aggregating impact scores from publications, the co-author network may encounter the issue of credit misallocation (An et al., 2022; Shen & Barabási, 2014). To measure author influence within paper-paper citation networks, we introduce an author-paper-paper network (as shown in Fig. 1b). In this network, we adopt the full-credit-counting method (Waltman, 2016), where the credits assigned to all authors in a single publication are considered equal and set to 1.

Definition 2. In the author-paper-paper network of a focal author A, the k-step *h-index* of A, denoted as $h_k(A)$, is the largest natural

number such that, in the k[th]-generation of A's citation cascades, there are at least $h_k(A)$ documents, each with a citation count of at least $h_k(A)$.

3.3. Institution-level k-step h-index

To follow the concept of successive *h-index* (Arencibia-Jorge, 2009; Egghe, 2008; Schubert, 2007), we aggregate the influence of publications within an institution by constructing an institution-paper-paper network. Similar to the author-level assessment, we employ the full-credit-counting method (Waltman, 2016) to address credit allocation.

Definition 3. In the institution's institution-paper-paper network, the k-step *h-index* of the focal institution I, denoted as $h_k(I)$, is the largest natural number such that, in the k[th]-generation of I's citation cascades, there are at least $h_k(I)$ documents, each with a citation count of at least $h_k(I)$.

Fig. 1 illustrates the process of calculating higher-step citation count and the k-step *h-index* within the citation network of a focal item. In the paper-paper network (Fig. 1a), the 1st citation generation encompasses all the articles that cite the focal paper. The 2nd generation includes the cited publications of each paper in the 1st generation, with duplication, and this pattern continues for subsequent generations, up to the kth generation. Similarly, in the author-paper-paper network (Fig. 1b) and the institution-paper-paper network (Fig. 1c), the 1st citation generation refers to publications authored by the focal author or institution, where they are listed as authors. The 2nd generation consists of the cited publications of each paper in the 1st generation. This process extends until the kth generation. It is important to note that the k-step *h-index* incorporates the citation information up to the k+1th generation at each level.

3.4. Comments

Higher-step metrics in citation networks are rooted in the fundamental concept that a focal paper initiates a cascade of citations. As the focal paper is cited by subsequent papers, those papers, in turn, become the subjects of citation by their successors, and so forth. The k-step *h-index* embodies this notion by delving deeper into the citation networks of the focal publication, thereby considering a broader spectrum of citation numbers for individual papers. It serves as a compromise between the raw citation counts of the first generation and the subsequent influences of papers within that generation. By incorporating information from up to k+1 generations within the citation network, the k-step *h-index* enables a comprehensive evaluation of the lasting impact of individual publications, authors, and institutions. It acknowledges that certain scientific entities may not receive a substantial number of direct citations in the first generation of the citation cascade. Nevertheless, they might have triggered a multitude of subsequent influential works, evident in the posterior k generations of citation documents within the cascade. Thus, the k-step *h-index* is devised to capture these latent scientific patterns and provide a holistic understanding of scientific development and technological progress.

However, it is important to acknowledge that the k-step *h-index* is not without its biases and limitations. Firstly, it does not incorporate age and field normalization methods (Fortunato et al., 2018; Radicchi & Castellano, 2011; Radicchi et al., 2008), which can introduce biases when comparing scientific entities across different disciplines or time periods. Additionally, the k-step *h-index* does not consider citation windows (Wang, 2013), which could impact the assessment of temporal influence and the dynamics of citation patterns over time.

Moreover, as the k-step *h-index* primarily relies on citation-based metrics, it may introduce other biases into the evaluation process (Hicks et al., 2015). For instance, it may not adequately account for excellence in locally relevant research conducted by non-English language scholars, papers, or institutions, thereby potentially marginalizing their contributions. Furthermore, the k-step *h-index* does not quantify the uncertainty and error associated with citation data, which can lead to misplaced concreteness and false precision in the evaluation of scientific impact. Therefore, it is essential to examine the properties and distributions of the k-step *h-indices* and to test the consistency and convergent validity of the k-step *h-indices* with qualitative and expert assessments to ensure a comprehensive and accurate evaluation of scientific entities.

4. Data and preprocessing

4.1. Citation data

In this study, we utilized two extensive citation datasets: the APS dataset and the MAG dataset. The APS dataset, obtained from the American Physical Society, comprises a total of 700,179 publications and 9,370,286 citation links spanning the years 1913 to 2021. On the other hand, the MAG dataset, provided by Microsoft, includes over 200 million publications and more than 1.7 billion citation links spanning the years 1800 to 2021. Notably, the MAG dataset contains comprehensive metadata for a significant portion of the papers within the APS dataset, including DOIs, author information, and institutional affiliations. Through meticulous matching and integration procedures, we obtained a total of 463,348 publications from 1913 to 2010 as our empirical data, resulting in an impressive matching rate of 96.9 %. To construct the paper-paper citation network, we employed the Python libraries igraph-0.10.2 and networkx-2.8.8. It is important to mention that while the innovation measures and control variables were calculated using the entire MAG dataset, the calculations of the k-step *h-indices* and network-based variables were confined to the APS dataset. In our analysis, the in-degree of a node represented its citation count, while the out-degree indicated its reference count. However, it should be noted that the actual citation counts are considerably higher since our local database solely encompasses articles from the APS dataset.

When calculating the k-step *h-indices*, it is possible for a publication to appear multiple times across different generations and in

various locations within a particular generation. To address this, we adhered to the criteria outlined by Min et al. (2021), which allow for a document to appear simultaneously in different generations and permit a document to appear repetitively within a single generation. By employing these criteria, we ensure the accurate calculation of the k-step *h-indices* and maintain the integrity of our analysis.

4.2. Name disambiguation

In our study, ensuring the accurate disambiguation of authors and institutions is crucial, as highlighted in previous research (Wang et al., 2013; Yin et al., 2019; Zeng et al., 2017). Within the APS dataset, a notable challenge arises from the presence of multiple names used to refer to the same author or institution. To address this issue, we adopted a disambiguation methodology inspired by the work of Sinatra et al. (2016) for the author name disambiguation process.

For the author-level name disambiguation procedure, we conducted a comprehensive comparison of all author pairs within the APS dataset. Our selection criteria involved evaluating the following conditions: (1) ensuring that the last names of the authors are identical, (2) verifying that the initials of their first and middle names are the same, and (3) confirming that the two authors have cited each other at least once, or they share common co-authors or affiliations. By implementing these rigorous criteria, we successfully disambiguated a total of 294,086 authors within the APS dataset, thereby enhancing the accuracy and integrity of our analysis.

At the institution level, we applied a similar disambiguation procedure, in addition to collecting institution name information from reliable sources such as Wikipedia and official websites. We meticulously combined different institution names manually, taking into account variations and aliases. Through these operations, we successfully identified 4,446 distinct institutions from 125 countries. This meticulous matching process ensures that our evaluation of institution-level metrics aligns with the esteemed institutions recognized by prestigious rankings such as the Shanghai Ranking (see S.M. Note 1).

During the data preprocessing stage, we performed calculations for the k-step *h-indices*, as well as other relevant metrics such as total citations and publication numbers, for the 463,348 publications, 294,086 authors, and 4,446 institutions included in our study. Due to computational and memory limitations, we focused on calculating the 2-step and 3-step *h-indices* at each level. By considering these specific k-step *h-indices*, we obtain valuable insights into the impact and influence of scientific entities within a reasonable computational scope.

5. Distribution and properties

5.1. RQ2: descriptive statistics

In order to provide a comprehensive overview, we present the descriptive statistics of the k-step citations (with duplication) and k-step *h-indices*. The results, as shown in Table 2, reveal that the k-step citation exhibits exponential growth as the value of k increases at the paper, author, and institution levels. This exponential growth can be attributed to the allowance for a document to appear simultaneously in different generations and the potential for a document to appear repetitively within a single generation. Notably, for certain highly-cited papers, the maximum 3-step citation count at the paper level has reached a staggering 27 billion, indicating a remarkably rich collection of information within the Hirsch core and Hirsch tail.

Conversely, the k-step *h-index* also experiences rapid growth as k increases, albeit in a linear manner rather than exponentially. This observation aligns with expectations, as higher-step *h-indices* incorporate aggregated citation counts from higher generations, leading

Table 2
Descriptive statistics of the k-step citations and k-step *h-indices*.

	Metrics	Observation	Mean	Std	Min	Median	Max	Skew	Kurt
Paper level	**1step-c**	463,348	142.4	584.5	0	17	1.1e+05	39.1	4,677
	2step-c	463,348	1,891	10,164	0	69	2.0e+06	45.1	5,756
	3step-c	463,348	975,415	5.6e+07	0	663	2.7e+10	359.2	154,666
	1step-h	463,348	3.6	4.6	0	2	1.2e+02	3.0	16.7
	2step-h	463,348	10.0	14.9	0	5	4.2e+02	3.3	20.7
	3step-h	463,348	52.0	132.6	0	11	4.3e+03	7.5	93.9
Author level	**1step-c**	234,086	61.2	234.1	0	9	1.6e+04	16.9	548.9
	2step-c	234,086	803.9	4,355	0	39	3.6e+05	24.8	1,130
	3step-c	234,086	10,247	70,825	0	162	6.9e+06	31.1	1,721
	1step-h	234,086	2.5	3.2	0	1	5.8e+01	3.5	19.6
	2step-h	234,086	7.0	10.4	0	3	2.3e+02	3.7	24.5
	3step-h	234,086	16.9	28.8	0	7	8.2e+02	4.8	44.7
Institution level	**1step-c**	4,446	1,527	7,535	0	27	1.5e+05	10.8	148.1
	2step-c	4,446	19,245	116,433	0	137	3.3e+06	13.8	254.4
	3step-c	4,446	233,956	1.7E+06	0	718.5	5.5e+07	16.7	380.0
	1step-h	4,446	8.0	13.5	0	3	1.5e+02	3.7	19.9
	2step-h	4,446	19.4	35.5	0	6	5.2e+02	4.5	30.9
	3step-h	4,446	45.5	96.3	0	12	1.4e+03	5.4	42.4

Note: The k-step citation refers to the total number of citations at the kth generation, with duplications. The 1st-step citation (1step-c) at the author level indicates the total number of citations for that specific author.

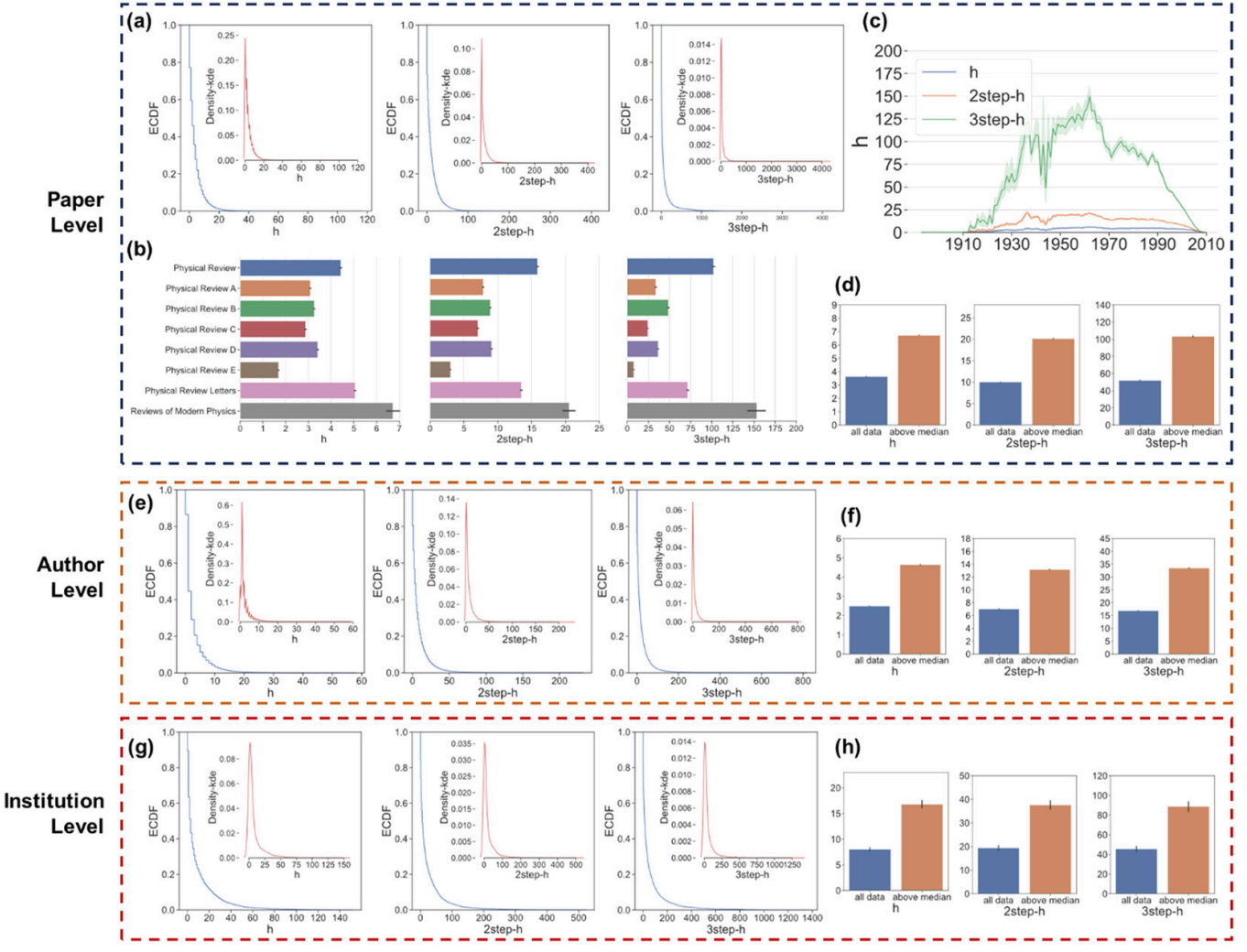


Fig. 2. Distribution of k-step *h-indices* at the paper, author, and institution levels in citation networks. (a) Empirical cumulative distribution function (ECDF) curves and kernel density estimation (KDE) insets showcasing the distributions of the k-step *h-indices*. (b) Bar graph presenting the average values of the k-step *h-indices* for different journals, with the black line indicating the standard deviation. (c) Graph illustrating the average values of the k-step *h-indices* for papers published in different years, with shaded areas representing 95 % confidence intervals. (d) Graph displaying the average values of the k-step *h-indices*, along with the average values above the median. (e) ECDF curves at the author level for the k-step *h-indices*, with KDE insets depicting the *h-index* distributions. (f) Bar graph showcasing the average values of the k-step *h-indices* for different journals, with the black line representing the standard deviation. (g, h) Curves and bars depicting the distributions at the institution level using the same methodology.

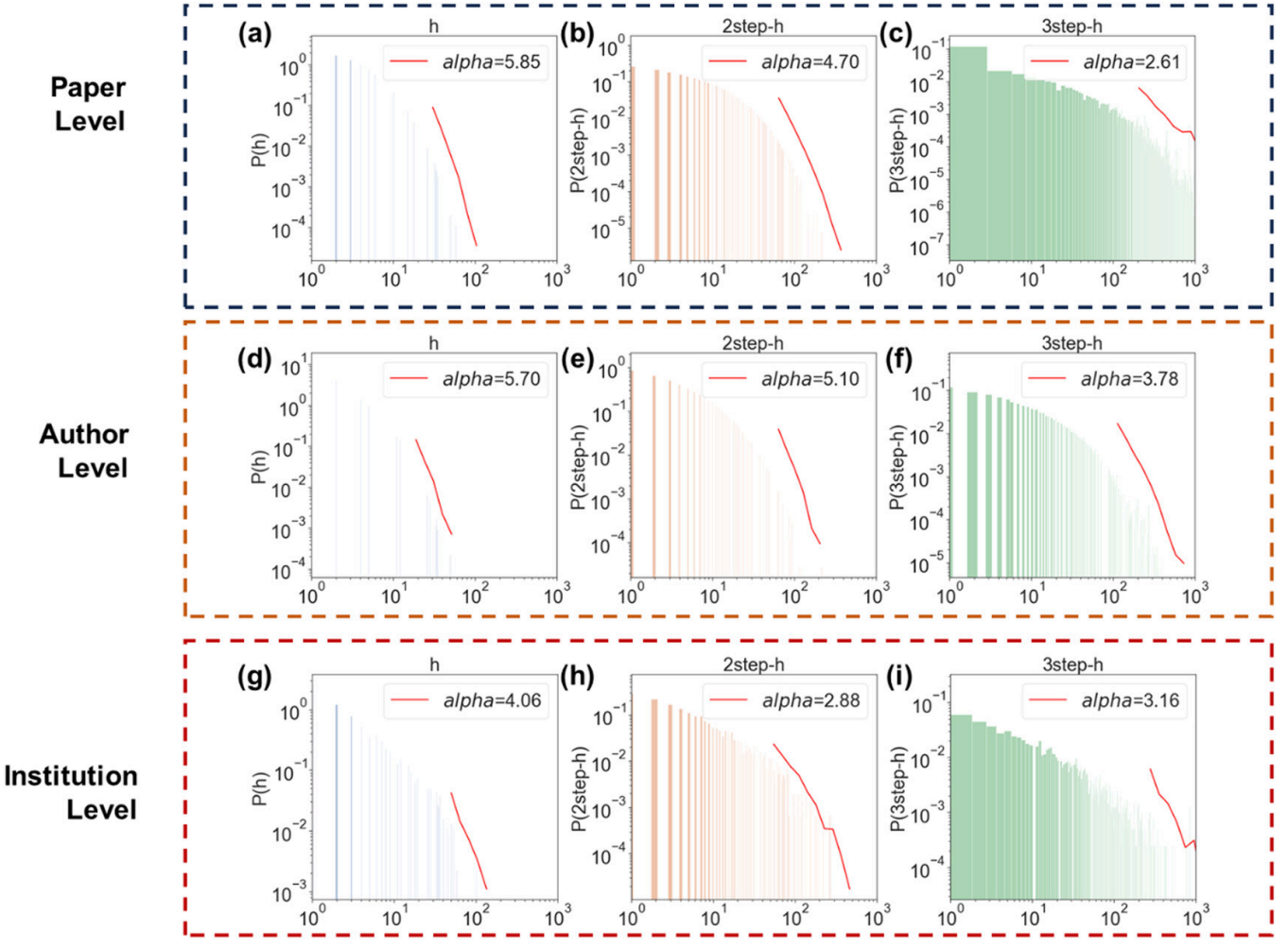


Fig. 3. Power law distribution fitting results for k-step *h-indices*. The histograms and fitting curves for each measure are displayed in (a) *h-index*, (b) 2-step *h-index*, and (c) 3-step *h-index* at the paper level, in (d) *h-index*, (e) 2-step *h-index*, and (f) 3-step *h-index* at the author level, in (g) *h-index*, (h) 2-step *h-index*, and (i) 3-step *h-index* at the institution level. All measures exhibit significant power law distributions.

to numerical expansion. However, it's worth noting that the rate of growth for k-step h-indices is notably more modest in comparison to k-step citations. To illustrate, while the average increase in 3-step citation exceeds 6,000-fold, the average 3-step *h-index* only experiences a growth factor of 14 compared to its 1-step equivalent. This divergence can be attributed in part to the nature of the k-step *h-index*, which seeks to strike a harmonious equilibrium between the quality and quantity of citations.

The distributions of the k-step *h-indices* at the paper, author, and institution levels are illustrated in Fig. 2. The empirical cumulative distribution function (ECDF) and kernel density estimation (KDE) curves presented in Fig. 2a, e, g reveal that the distribution patterns of the *h-indices* exhibit fat-tailed distributions, both at lower-step and higher-step values. It is worth noting that the steepness of the ECDF and KDE curves is more pronounced at the paper and author levels compared to the institution level, indicating a less pronounced long-tail effect as we transition to highly aggregated levels. This observation suggests that as we move from individual papers and authors to institution-level analysis, the concentration of influence becomes less skewed. The presence of the Matthew effect (Merton, 1968) becomes evident in the higher-step *h-index*, wherein prominent authors and institutions enjoy a greater advantage. This observation is further supported by the significant disparity between the average values of the entire sample and the above-median samples, which is particularly pronounced for the higher-step *h-index* as depicted in Fig. 2d, f, h.

Furthermore, an exploration of the distributions across different journals (Fig. 2b) and publication years (Fig. 2c) at the paper level sheds light on the uneven distribution of k-step *h-indices*. This finding aligns with previous research that indicates citation-based indicators are subject to biases influenced by publication date and field to some extent (Radicchi & Castellano, 2011; Radicchi et al., 2008). Notably, the degree of heterogeneity increases as the value of k grows. This suggests that higher-generation k-step *h-indices* have the potential to further enhance the measurement of already well-performing entities. These findings underscore the importance of incorporating field-normalized methods and considering citation windows (Waltman, 2016) in future studies and practical applications. By accounting for variations across fields and temporal dynamics, a more comprehensive and accurate assessment of scientific impact can be achieved.

5.2. RQ2: power law distribution

The power law distribution is a fundamental and widely observed phenomenon in various domains, including the field of science, as it signifies the presence of preferential attachment mechanisms and the Matthew effect (Barabasi & Albert, 1999; Peterson et al., 2010). In order to explore the power law distribution characteristics of the metrics under investigation, we employed the function $N(x) \sim x^{-\alpha}$ to fit the power law distributions of nodes and examined the power law distribution patterns of the k-step *h-indices* and benchmark measures. For this analysis, we utilized the powerlaw package developed by Alstott et al. (2014).

The power law distribution fitting results, as presented in Fig. 3 and S.M. Fig. S2 indicate that the k-step *h-indices* at the paper, author, and institution levels all exhibit a power law distribution. This finding suggests the presence of a limited number of highly influential entities alongside a large number of less influential ones. The alpha values obtained through the fitting process serve as indicators of the level of heterogeneity within the distribution of node importance measures, with smaller values indicating a more heterogeneous distribution. The outcomes reveal that the distributions of higher k-step *h-indices* demonstrate greater heterogeneity compared to the *h-index*, signifying that these metrics possess a higher capacity to capture the characteristics of highly disruptive entities. For instance, at the paper level (Fig. 3a-c), the fitted alpha value for the *h-index* is 5.85, for the 2-step *h-index* is 4.70, and for the 3-step *h-index* is 2.02. This finding aligns with the observations made in the preceding sections and partially addresses RQ2, thereby demonstrating that the k-step *h-index* enhances heterogeneity in the assessment of scholarly impact.

5.3. RQ2: K-step Hirsh core

To unravel the underlying essence and functioning of the k-step *h-index*, it is imperative to delve into the analysis of the k-step Hirsch core and Hirsch tail of the entities being evaluated. The Hirsch core and Hirsch tail, as introduced by Ye and Rousseau (2010), refer to the cumulative number of citations received by publications that exceed or fall below the *h-index* threshold. In the context of the k-step *h-index*, we define the k-step Hirsch core as the aggregate number of citations, at the kth generation of the focal paper's citation cascades (with duplication), that surpass or fall below the *h-index* value.

In order to identify the k-step Hirsch core, the publications at the kth generation are sorted in ascending order based on their citation counts. The total citations of the top h publications in this ranking are considered as the k-step Hirsch core, while the remaining publications constitute the k-step Hirsch tail. The outcomes of the Hirsch core and tail analysis for the k-step *h-indices* are depicted in Fig. 4. For each level (paper, author, and institution), we have selected representative items with the highest *h-indices* to showcase their respective k-step Hirsch cores and tails. Remarkably, we observe a striking similarity among these items across different levels. It is noteworthy that the numerical value of the k-step Hirsch core is equal to S1 + S2, representing the total citations within this region, while the Hirsch tail corresponds to S3. Through this analysis, we gain valuable insights into the k-step *h-index*. We observe that both the 2-step Hirsch core and 2-step Hirsch tail encompass significantly larger citation information compared to their 1-step counterparts, at the paper, author, and institution levels.

Furthermore, as the k-step *h-index* increases, the relative size of the Hirsch tail tends to grow, while the relative size of the Hirsch core diminishes. For instance, consider the author-level item P.W. Anderson, highlighted in orange in Fig. 4c. His 1-step Hirsch core is relatively large, while his 1-step Hirsch tail is comparatively small compared to his peers. However, the ranking results differ when examining the 2-step Hirsch tail and 2-step Hirsch core. This observation underscores how higher-step *h-indices* can yield distinct evaluations and rankings. Tentatively, we can conclude that the k-step *h-indices*, particularly those at higher steps, have the potential to offer a more comprehensive and profound assessment of the Hirsch core by effectively incorporating information from the Hirsch tail.

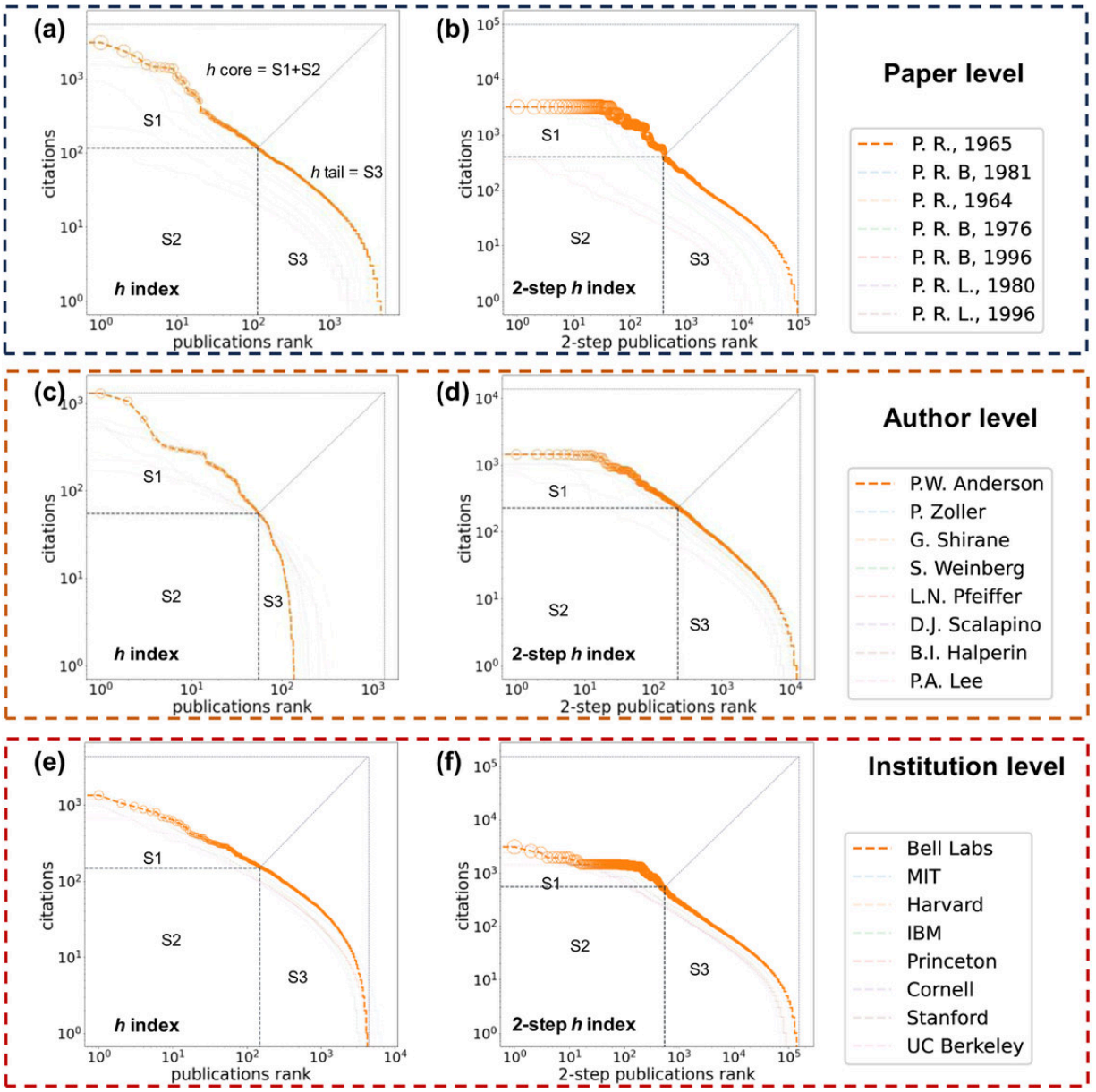


Fig. 4. The illustration of the Hirsh core and Hirsh tail of k-step *h-indices* at the paper, author and institution level, respectively. (a-f) The citations of each publication of a single (a-b) paper, (c-d) author, and (e-f) institution, with publications ranked from 1 to n (n refers to the k^{th}-generation cited publications). The sizes of circles are proportional to the citations of each publication. About 8 prestigious entities are chosen for comparison (shown in different colors) at each level, yet only the highest *h-index* ones (marked as the orange circles and lines) are chosen to illustrate the Hirsh core and Hirsh tail. The logarithmic scale (log10-log10 axes) is used. Note that the x-axis and y-axis differ in each subgraph, indicating that the 2-step Hirsch core and Hirsch tail are significantly larger than the 1-step counterparts.

6. Consistency and convergent validity

6.1. RQ3: correlation analysis

Analyzing the consistency between different evaluation metrics allows us to gain a deeper understanding of their performance and actual value (Lü et al., 2016; Zhai et al., 2018). To assess the relationship between the k-step *h-index* and conventional metrics, we employ Kendall's Tau correlation analysis (Kendall, 1938). Kendall's Tau is calculated as $\tau = \frac{n^c - n^d}{\sqrt{(n^c+n^d+n^x) \times (n^c+n^d+n^y)}}$, where n^c represents the number of groups with consistent sequence, n^d denotes the number of groups with inconsistent sequence, n^x indicates the number of groups with identical sequence, and n^y signifies the number of groups with the same ordering. A higher correlation

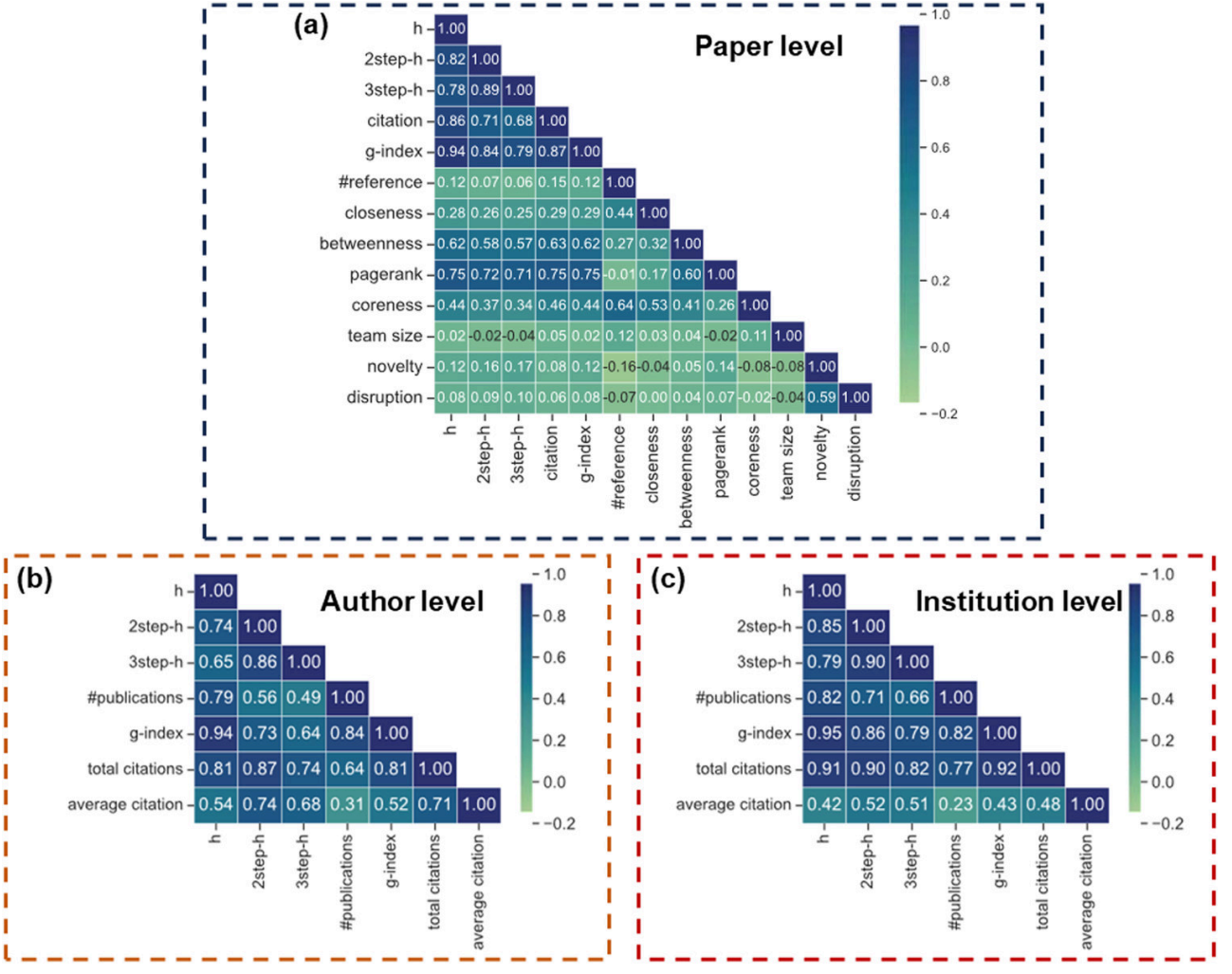


Fig. 5. The Kendall-tau correlations of metrics at the paper, author and institution level, respectively. (a) The Kendall-tau correlations of the k-step *h-indices* and benchmarks at the paper level. (b) The Kendall-tau correlations of the k-step *h-indices* and some benchmarks (publications, total citations and average citations) at the author level. (c) The Kendall-tau correlations of the k-step *h-indices* and some benchmarks (publications, total citations and average citations) at the institution level.

coefficient τ indicates a stronger relationship between the two sets of data, while a value of $\tau = 0$ signifies complete independence.

The results of the correlation analysis, illustrated in Fig. 5, reveal several important insights. Firstly, there is a strong correlation between lower-step and higher-step *h-index* values. However, the correlation coefficients vary across different levels, with coefficients above 0.78 for papers, above 0.65 for authors, and above 0.79 for institutions. To demonstrate the consistency and distinctiveness of the k-step *h-index*, we compare it with benchmark measures at the paper level (Fig. 5a). Coreness and PageRank are commonly used metrics to assess the impact of individual publications (Wang et al., 2021; Xu et al., 2020), whereas the *h-index* is not typically employed for paper evaluation. Notably, there are notable differences in the correlation coefficients among these metrics, highlighting the potential value of the k-step *h-index* in measuring the impact of papers.

Secondly, at the author and institution levels, there are noticeable differences in the correlations between the k-step *h-index* and publication numbers. Specifically, at the author level, the correlation coefficient between the *h-index* and publications is 0.79, whereas the correlation coefficient between the k-step *h-index* and publications drops to 0.49. Conversely, the k-step *h-index* maintains consistency in its relationship with total citations and average citations at the author and institution levels. These findings underscore the importance of considering higher generations of citation information to achieve a comprehensive and accurate evaluation of scientific entities. While the k-step *h-index* demonstrates strong correlations with certain conventional metrics, it also reveals discrepancies in correlation coefficients across different levels and measures, suggesting its unique contribution and potential for providing novel insights into scholarly impact assessment.

6.2. RQ3: consistency with innovation measures

Among the plethora of measures in the field of scientometrics (Wu et al., 2022), our particular focus lies in examining the consistency of the k-step *h-index* with innovation measures. To demonstrate this consistency, we employ two specific measures that capture the innovation of papers. Firstly, we utilize the level of atypical combinations of journal pairs (Uzzi et al., 2013) to signify novelty. Secondly, we calculate the level of paradigm shifting in citation networks, representing the disruption (Funk & Owen-Smith, 2017) (see S.M. Note 1 and Note 2 for detailed definitions and further information). It is worth noting that these two innovation measures primarily rely on the citation database (Ruan et al., 2021), necessitating the use of the entire MAG database for their calculation, rather than solely relying on the local APS database as outlined in Section 4.1.

Remarkably, in Fig. 5, we observe that the k-step *h-index*, particularly with higher generations, exhibits higher correlation coefficients with measures of novelty and disruption. However, it is important to consider confounding factors to ascertain the true levels of consistency. To address these confounding factors, we conducted a regression analysis using ordinary least squares (OLS) and controlled for several variables, including the number of authors, the number of institutions, citation gap (Chao Min et al., 2021), the number of journal pairs (Uzzi et al., 2013), and the number of references. Additionally, we included fixed effects for field and year. The regression results are presented in Table 3.

The results demonstrate that the k-step *h-indices*, particularly with higher generations, exhibit a positive and statistically significant

Table 3
OLS regression of the k-step *h-indices* and innovation measures.

Model	Dependent variable: Novelty				Dependent variable: Disruption			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
h	0.155***				0.0295***			
	(0.0151)				(0.0014)			
2step-h		0.1802***				0.0526***		
		(0.0133)				(0.0011)		
3step-h			0.2***				0.0533***	
			(0.0146)				(0.0012)	
g-index				0.1684***				0.0414***
				(0.0166)				(0.0012)
#Author	0.006***	0.006***	-0.0305***	-0.0264***	-0.0259***	-0.0288***	-0.0305***	-0.0264***
	(0.0019)	(0.0019)	(0.0025)	(0.0025)	(0.0025)	(0.0025)	(0.0025)	(0.0025)
#Institution	-0.0051***	-0.0051***	0.0337***	0.0285***	0.0276***	0.0314***	0.0337***	0.0285***
	(0.0017)	(0.0017)	(0.0023)	(0.0023)	(0.0023)	(0.0023)	(0.0023)	(0.0023)
Citation gap	0.0011	0.001	-0.0018	0.005***	0.0059***	0.0022	-0.0018	0.005***
	(0.0011)	(0.0011)	(0.0015)	(0.0015)	(0.0015)	(0.0015)	(0.0015)	(0.0015)
# Journal Pairs	0.4963***	0.4963***	0.0277***	0.0335***	0.0349***	0.0309***	0.0277***	0.0335***
	(0.0057)	(0.0057)	(0.0078)	(0.0077)	(0.0077)	(0.0077)	(0.0078)	(0.0077)
#Reference	-0.3447***	-0.3448***	-0.1278***	-0.1337***	-0.1351***	-0.1305***	-0.1278***	-0.1337***
	(0.0027)	(0.0027)	(0.0037)	(0.0037)	(0.0037)	(0.0037)	(0.0037)	(0.0037)
_cons	0.9901***	0.9901***	0.5212***	0.5211***	0.5211***	0.5212***	0.5212***	0.5211***
	(0.0043)	(0.0043)	(0.0059)	(0.0059)	(0.0059)	(0.0059)	(0.0059)	(0.0059)
Field FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R2	0.114	0.113	0.113	0.113	0.023	0.02	0.019	0.022
Observations	463348	463348	463348	463348	463348	463348	463348	463348

Note: *p < 0.1, **p < 0.05, ***p < 0.01. Robust standard errors are indicated in parentheses. The normalization methods were applied to both independent variables and control variables, scaling them to a range of 0-1.

relationship with measures of innovation. Specifically, for each unit increase in the k-step *h-index*, there is a corresponding increase in the measure of novelty and disruption, respectively. The coefficients for the 2-step *h-index* and 3-step *h-index* indicate a stronger relationship with novelty compared to the *h-index* alone or g-index. This suggests that the k-step *h-indices*, with their incorporation of higher generations of citation information, more effectively capture and reflect the innovative nature of papers. These findings provide substantial evidence supporting the consistency between higher generation k-step *h-indices* and measures of innovation, encompassing both novelty and disruption. The positive and statistically significant coefficients indicate that the k-step *h-indices*, particularly with higher generations, serve as robust indicators of innovation, capturing and reflecting the innovative and disruptive nature of papers. These findings suggest that the k-step *h-indices* have the potential to complement and enhance the insights provided by conventional metrics, acting as valuable tools for assessing and quantifying innovation in scientific research.

6.3. RQ4: ranking of expert-selected Items

A key aspect of our investigation centers around assessing the effectiveness of the k-step *h-indices* in identifying milestone papers, exceptional authors, and prestigious institutions. To address this question, we conducted an experiment to examine the performance and efficiency of the k-step *h-indices* through convergent validity analysis. Convergent validity refers to the extent to which the metrics accurately measure what they are intended to measure (Bornmann & Tekles, 2021; Rowlands, 2018). By utilizing expert-selected items, including Nobel Prize-winning papers and milestone papers identified by the American Physical Society at the paper level,

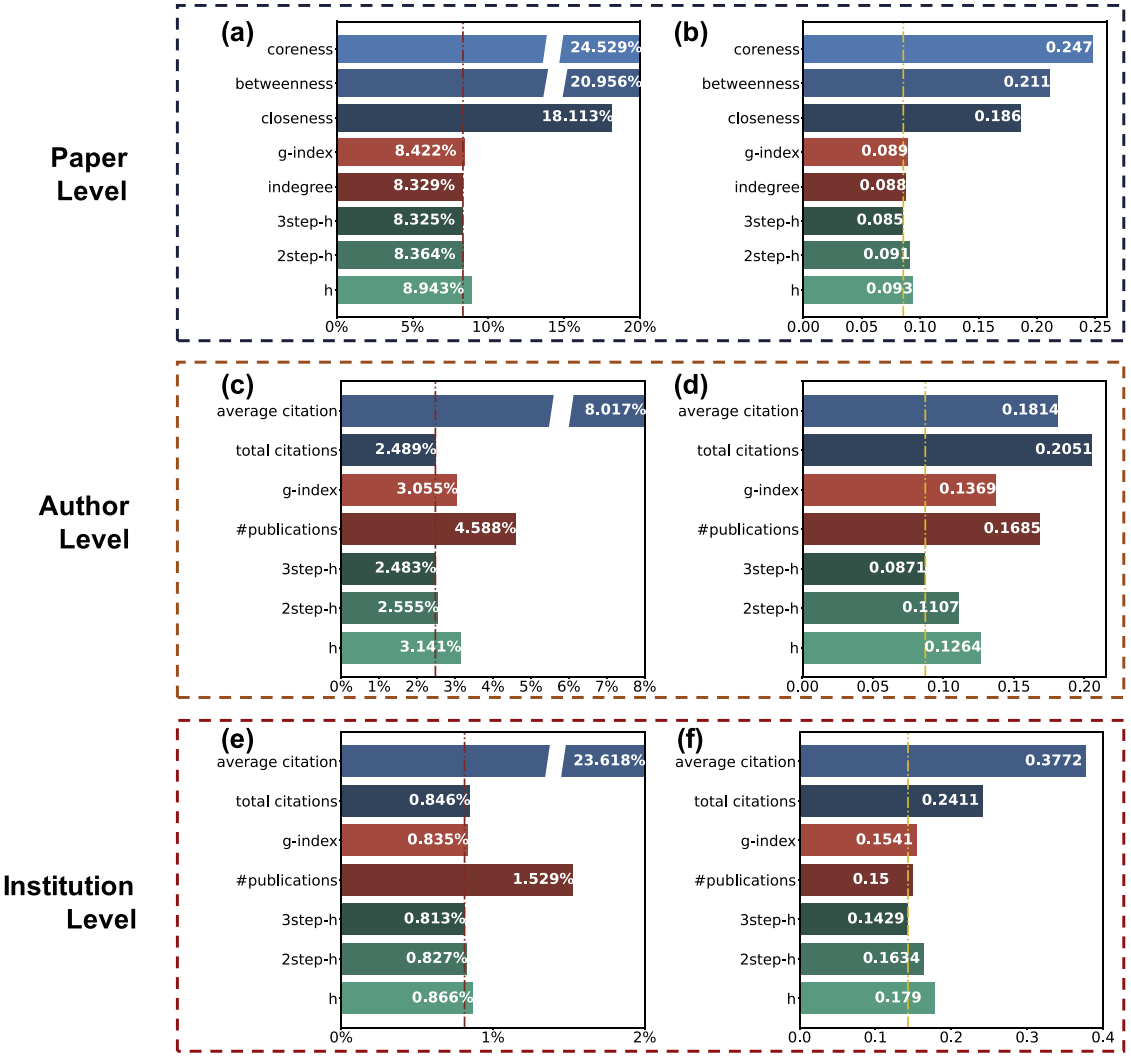


Fig. 6. The average ranking (AR) and $\widetilde{AR}$ of expert-selected items under each metric at the paper, author and institution levels, respectively. (a, b) The AR and $\widetilde{AR}$ of the k-step *h-indices* and benchmark measures at the paper level. (c, d) The AR and $\widetilde{AR}$ of the k-step *h-indices* and benchmark measures at the author level. (e, f) The AR and $\widetilde{AR}$ of the k-step *h-indices* and benchmark measures at the author level. Note that lower AR and $\widetilde{AR}$ means higher convergent validity.

distinguished Physics awards laureates (such as Nobel Prize, Wolf Prize, and Dirac Medal recipients) at the author level, and top-ranking institutions identified by the Shanghai Ranking at the institution level, we can validate the convergent validity of the k-step *h-indices*. These expert-selected items were matched within our dataset (refer to S.M. Note 1 for detailed information).

To assess convergent validity, we calculated the average ranking (AR) of the metrics in ranking critical items. The underlying concept of this task involves determining the average ranking of all critical nodes in networks using a given metric m. We further introduced the concept of $\widetilde{AR}$ to measure the performance of each metric, defined as $\widetilde{AR}(m,v_i) = \frac{R(m,v_i)}{\max(R(m,v))-\min(R(m,v))}$, where v_i represents critical nodes (i.e., publications, authors, and institutions), m denotes the metrics, and the average value $\widetilde{AR}(m,v_i)$ represents $\widetilde{AR}$. In this manner, we eliminate the influence of non-seminal items by subtracting the ranking of the lowest-ranking critical item (with a larger value) from the ranking of the highest-ranking critical item. The critical item refers to the expert-selected item (e.g., milestone papers, authors, and institutions) identified by APS. Utilizing $\widetilde{AR}$ helps penalize biased ranking results.

The results of the average ranking (AR) and $\widetilde{AR}$ of the k-step *h-indices* and conventional metrics at the paper, author, and institution levels are presented in Fig. 6. At each level, the average ranking of the k-step *h-indices* outperforms other benchmarks. Furthermore, the average ranking of higher-step *h-indices* surpasses that of lower-step ones. Specifically, the 3-step *h-index* achieves an average ranking within the top 8.325 % for milestone papers at the paper level (Fig. 6a), within the top 2.438 % for esteemed authors at the author level (Fig. 6c), and within the top 0.813 % for prominent institutions at the institution level (Fig. 6e). These results demonstrate superior performance compared to lower k-step *h-indices*, the original *h-index*, g-index, and other benchmarks.

It is important to note that we chose a selection of commonly used benchmarks for comparison and did not include more sophisticated measures such as the PageRank (Brin & Page, 1998), disruptive indegree, or SDC (Yang, Hu, et al., 2023) as benchmarks. However, we did compare the k-step *h-index* with these state-of-the-art benchmarks and found that the k-step *h-index* may not be the optimal method for ranking expert-selected items (refer to S.M. Note 4 and Fig. S3 for detailed information).

6.4. RQ4: Predicting power of expert-selected Items

In addition to the average ranking task, previous studies (Dunaiski et al., 2018; Mariani et al., 2016; Xu et al., 2020) have also explored the concept of identification rate, which measures the metrics' ability to accurately identify prominent scientific entities. This concept aligns with the notion of recall in recommender tasks (Lü et al., 2012). The identification rate (Xu et al., 2020) serves as a score for this task. In our study, we take a step further and introduce a prediction task by setting a specific threshold to predict expert-selected items. We then calculate various performance metrics, including accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve (AUC), to assess the predictive power of the metrics. The threshold, denoted as z, is set to 1 % in our main results (refer to Table 4), and we also consider z values of 0.1 %, 5 %, 10 %, and 20 % for further analysis (refer to S.M. Table S2-S5). It is important to note that the number of expert-selected items is relatively small compared to the overall sample size, which may result in low precision and F1 scores. However, we observe remarkably high accuracy for all metrics, and the recall value holds particular significance based on existing literature (Dunaiski et al., 2018; Mariani et al., 2016; Xu et al., 2020).

The results demonstrate that, at each level, the average ranking of the k-step *h-indices* outperforms other benchmarks. Furthermore,

Table 4
The predictive Performance of Expert-Selected Items for k-Step *h-indices* (z = 1 %).

	Metrics	Accuracy	Precision	Recall	F1	AUC
Paper level	**h**	99.00 %	2.31 %	51.20 %	4.42 %	75.11 %
	2step-h	99.00 %	2.46 %	54.55 %	4.71 %	76.78 %
	3step-h	**99.01** %	**2.50** %	**55.50** %	**4.79** %	**77.26** %
	indegree	98.99 %	1.96 %	43.54 %	3.76 %	71.28 %
	g-index	99.00 %	2.31 %	51.20 %	4.42 %	75.11 %
	closeness	98.98 %	1.32 %	29.19 %	2.52 %	64.10 %
	betweenness	98.98 %	1.04 %	22.97 %	1.98 %	60.99 %
	coreness	98.96 %	0.22 %	4.78 %	0.41 %	51.89 %
Author level	**h**	99.00 %	2.95 %	53.08 %	5.59 %	76.05 %
	2step-h	99.01 %	3.33 %	60.00 %	6.32 %	79.52 %
	3step-h	**99.01** %	**3.46** %	**62.31** %	**6.56** %	**80.67** %
	#publications	98.98 %	1.62 %	29.23 %	3.08 %	64.12 %
	g-index	99.00 %	2.91 %	52.31 %	5.51 %	75.67 %
	total citations	99.01 %	3.12 %	56.15 %	5.91 %	77.59 %
	average citation	98.97 %	1.41 %	25.38 %	2.67 %	62.20 %
Institution level	**h**	99.53 %	88.64 %	70.91 %	78.79 %	85.40 %
	2step-h	99.57 %	90.91 %	72.73 %	80.81 %	86.32 %
	3step-h	**99.75** %	**100.00** %	**80.00** %	**88.89** %	**90.00** %
	#publications	99.08 %	65.91 %	52.73 %	58.59 %	76.19 %
	g-index	99.53 %	88.64 %	70.91 %	78.79 %	85.40 %
	total citations	99.53 %	88.64 %	70.91 %	78.79 %	85.40 %
	average citation	97.77 %	0.00 %	0.00 %	0.00 %	49.50 %

Note: Utilizing the top 1 % of items based on each metric to make predictions for expert-selected items. The items that performed the best are highlighted in bold, while the second best-performing ones are underlined.

the average ranking of higher-step *h-indices* surpasses that of lower-step ones. Specifically, the 3-step *h-index* achieves a recall rate of 55.50 % at the paper level, 62.31 % at the author level, and 80.00 % at the institution level. These results highlight the superior performance of higher k-step *h-indices* compared to lower ones, as well as the original *h-index*, g-index, and other benchmarks.

To mitigate the influence of confounding variables and mitigate potential biases associated with fields and time, we employ an OLS regression analysis to assess the robustness of the convergent validity of the metrics at the paper level (refer to S.M. Table S1). The results reinforce our previous findings, confirming the superior performance of higher k-step *h-indices* compared to lower ones, as well as the other benchmarks. This further strengthens the validity and reliability of our conclusions, indicating the consistent and reliable nature of the k-step *h-indices* in assessing scholarly impact.

However, it is important to acknowledge that k-step *h-indices* are not the state-of-the-art measures for identifying milestone papers or laureates. In fact, they perform worse than certain network-based metrics, such as the PageRank (Brin & Page, 1998) and disruptive indegree (Yang, Hu, et al., 2023) at the paper level and SDC (Yang, Hu, et al., 2023) at the author level (refer to S.M. Note 4 for detailed information).

7. Discussion

7.1. Theoretical and practical implications

The implications of this paper extend to several key areas within the field of scientometrics and science of science. First, the introduction of the k-step *h-index* expands our conceptualization of scientific impact beyond traditional citation-based metrics (Waltman, 2016). While the conventional *h-index* measures the impact of a paper based on direct citations, the k-step *h-index* considers the influence propagated through citation cascades across multiple generations. This conceptual shift acknowledges the complex nature of scholarly impact, highlighting the importance of indirect effects and long-term influence (C. Min et al., 2021). By incorporating higher generations of citation information, the k-step *h-index* offers a more comprehensive and nuanced assessment of scientific impact, capturing latent patterns and potential disruptive entities that may be overlooked by traditional metrics. This theoretical advancement contributes to a more holistic understanding of scholarly influence.

Second, the analysis of k-step citations and the distribution properties of the k-step *h-indices* provides insights into the dynamics of citation networks (Holme & Saramäki, 2012; Iniguez et al., 2022). The observed exponential growth in k-step citations indicates the rich interconnectedness of scholarly works and the propagation of influence across multiple generations. This finding highlights the complex nature of citation cascades and the cumulative effects of citations over time. Additionally, the power law distribution characteristics of the k-step *h-indices* reveal the coexistence of highly influential entities and less influential ones, emphasizing the presence of preferential attachment mechanisms and the Matthew effect within citation networks. These findings deepen our theoretical understanding of the structure and dynamics of scholarly communication, shedding light on the mechanisms that drive the formation of citation patterns and the concentration of influence within scientific communities.

Third, the examination of the k-step *h-index* in relation to conventional metrics and innovation measures contributes to the theoretical underpinnings of impact assessment (Fontana et al., 2020; Xu et al., 2022). The strong correlations between the k-step *h-indices* and conventional metrics, as well as their alignment with measures of novelty and disruption (Funk & Owen-Smith, 2017; Uzzi et al., 2013), provide evidence for the convergent validity of the k-step *h-index*. This suggests that the k-step *h-index* captures important dimensions of scientific impact that are captured by existing metrics but also extends our understanding by incorporating higher generations of citation information. Furthermore, the superior performance of higher k-step *h-indices* in accurately ranking expert-selected items and predicting prominent scientific entities indicates their potential to provide a more comprehensive and accurate assessment of scholarly impact. These findings enhance our theoretical knowledge of impact evaluation methodologies and highlight the value of considering higher generations of citation information in impact assessment frameworks.

Last, the findings of this research extend beyond academia and have practical implications for research evaluation and policy making. By providing a more comprehensive and nuanced understanding of scientific impact, the k-step *h-index* offers valuable insights for decision-making processes related to research funding, career advancement, and institutional assessments (Yin et al., 2022; Yin et al., 2021). The incorporation of higher generations of citation information can help identify and recognize influential papers, exceptional authors, and prestigious institutions that may have been overlooked by traditional metrics. This has implications for promoting fairness and inclusivity in research evaluation processes, as the k-step *h-index* considers the long-term impact and the potential for disruptive contributions that may not be immediately apparent. These theoretical implications can inform the development of more robust and informed research evaluation frameworks and policies.

7.2. Limitations

This study acknowledges several limitations that should be taken into consideration when interpreting the findings. First, this study is based on analysis using specific citation datasets and focused on the field of Physics. Consequently, caution should be exercised when generalizing the results to other disciplines and databases. The characteristics of citation networks, citation patterns, and impact dynamics may differ across scientific fields, and therefore, the applicability and performance of the k-step *h-index* in those domains may vary.

Second, like any metric based on citations, the k-step *h-index* is susceptible to biases inherent in citation practices. Biases such as self-citations, disciplinary citation norms, and the impact of highly influential but under-cited works can influence the calculation and interpretation of the k-step *h-index*. While the k-step *h-index* attempts to mitigate some limitations of traditional metrics by

incorporating higher generations of citation information, it may still be influenced by these biases.

Third, although expert-selected items were employed in this study to assess the performance of the k-step *h-index*, further external validation using measures of impact and expert judgments is necessary. Relying solely on expert-selected items from the same dataset used for analysis may introduce circularity and limit the robustness of the findings. External validation through expert evaluations or comparison with alternative impact indicators would strengthen the credibility and validity of the k-step *h-index*. Additionally, it is essential to acknowledge that the selection of expert-selected items, such as the Shanghai Ranking for identifying distinguished institutions, may be subject to certain biases, which can affect the accuracy and representativeness of the assessment.

Addressing these limitations in future research will contribute to refining and advancing the understanding and application of the k-step *h-index*. By exploring diverse datasets, accounting for biases, and validating the findings through external measures and expert evaluations, researchers can enhance the reliability, generalizability, and credibility of the k-step *h-index* as a tool for evaluating scientific impact.

7.3. Outlook

In light of the findings and limitations identified in this study, several future avenues for research can be explored to further enhance the understanding and application of the k-step *h-index*. The following areas offer promising opportunities for future investigation:

First, future research should focus on refining the methodology for calculating and interpreting the k-step *h-index*. This could involve exploring the optimal value of k for different levels of analysis, such as papers, authors, and institutions, taking into account the specific characteristics and dynamics of each level. Additionally, alternative weighting schemes could be investigated to assign different weights to citations from different generations, considering the evolving importance of past and recent citations. Incorporating additional information, such as co-citation networks or text-based analysis (Wang et al., 2020), can also enrich the assessment of scholarly impact and contribute to a more comprehensive understanding of the k-step *h-index*.

Second, while this study primarily focused on the field of Physics, there is a need to examine the applicability and effectiveness of the k-step *h-index* across diverse disciplines. Different fields exhibit variations in citation patterns, publication practices, and collaboration dynamics (Wang & Barabási, 2021), which may influence the performance and relevance of the k-step *h-index*. Comparative studies across disciplines can provide insights into the disciplinary-specific nature of scholarly impact and shed light on the utility and adaptability of the k-step *h-index* in different research domains.

Third, to provide a more comprehensive evaluation of scientific impact, future research could integrate Altmetrics (Amiri et al., 2023; Xia et al., 2022) into the assessment framework. Altmetrics capture online attention, social media engagement, and other non-traditional indicators of impact, offering complementary perspectives to citation-based metrics. By incorporating Altmetrics, the assessment of scholarly impact can encompass broader societal impact, interdisciplinary collaborations, and the transformative nature of research. Additionally, qualitative assessments by experts, peer review, and expert panels can be valuable in capturing nuanced dimensions of impact that may not be captured by quantitative metrics alone.

Last, conducting dynamic analyses can uncover temporal dynamics in citation patterns and impact trajectories (Holme & Saramäki, 2012; Wang et al., 2013). Studying the evolution of the k-step *h-index* over time can provide insights into the changing nature of scientific impact, the emergence of influential entities, and the impact of scientific breakthroughs. Longitudinal analyses can identify trends and patterns in scholarly impact, enabling researchers to track and understand the dynamics of impact over extended periods. This longitudinal perspective can inform policy decisions, strategic planning, and the development of predictive models for future impact assessment.

By pursuing these future avenues of research, scholars can deepen their understanding of the k-step *h-index* and its theoretical implications, refine the methodology for impact assessment, and expand the applicability of the k-step *h-index* across disciplines and research contexts. This will contribute to a more comprehensive and nuanced evaluation of scientific impact, promoting fair and accurate assessments of scholarly contributions.

Data availability

The APS citation dataset is available at https://journals.aps.org/datasets, the MAG dataset can be accessed through the Get Microsoft Academic Graph on Azure storage - Microsoft Academic Services | Microsoft Learn. The version number of MAG data used in our study is the final (latest) version, published in December 20, 2021.

The Nobel-winning Papers and Laureates data used in this study are open-access at Harvard Dataverse https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/6NJ5RN. For milestone papers identified by the American Physical Society, please visit https://journals.aps.org/125years, https://journals.aps.org/prl/50years/milestones, and https://journals.aps.org/pre/collections/pre-milestones. The disambiguated author data for APS with a column indicating whether the author is a laureate or non-laureate can be accessed at https://github.com/AlexJieYang/Disambiguated-authors-for-APS-datasets. The disambiguated institution data can be accessed at https://zenodo.org/record/8150239.

Other data used in this study can be obtained by making reasonable requests.

CRediT authorship contribution statement

Alex Jie Yang: Writing – original draft, Conceptualization, Methodology, Software, Visualization, Validation, Data curation.

Linwei Wu: Writing – review & editing. **Qi Zhang:** Conceptualization, Writing – review & editing. **Hao Wang:** Supervision, Validation. **Sanhong Deng:** Conceptualization, Writing – review & editing, Supervision.

Declaration of Competing Interest

The authors declare no competing interest relevant to this article.

Acknowledgments

This research was supported by the National Social Science Fund of China (No.19BTQ062), the Jiangsu Key Laboratory Fund, the International Joint Informatics Laboratory Fund, and the Fundamental Research Funds for the Central Universities.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.joi.2023.101456

References

Acemoglu, D., Akcigit, U., & Kerr, W. R. (2016). Innovation network. *Proceedings of the National Academy of Sciences, 113*(41), 11483–11488. https://doi.org/10.1073/pnas.1613559113

Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. (2009). h-Index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics, 3*(4), 273–289. https://doi.org/10.1016/j.joi.2009.04.001

Alstott, J., Bullmore, E., & Plenz, D. (2014). powerlaw: A python package for analysis of heavy-tailed distributions. *PLoS One, 9*(1), e85777. https://doi.org/10.1371/journal.pone.0085777

Amiri, M. R., Saberi, M. K., Ouchi, A., Mokhtari, H., & Barkhan, S. (2023). Publication performance and trends in altmetrics: A bibliometric analysis and visualization. *International Journal of Information Science and Management, 21*(1), 95–115. https://doi.org/10.22034/ijism.2022.1977686.0

An, X., Sun, X., & Xu, S. (2022). Important citations identification with semi-supervised classification model. *Scientometrics, 127*(11), 6533–6555. https://doi.org/10.1007/s11192-021-04212-6

Arencibia-Jorge, R. (2009). New indicators of institutional scientific performance based on citation analysis: the successive H indices. *Revista Espanola De Documentacion Cientifica, 32*(3), 101–106. https://doi.org/10.3989/redc.2009.3.692

Barabási, A. L. (2016). *Network science.* Cambridge University Press.

Barabasi, A. L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512. https://doi.org/10.1126/science.286.5439.509

Bi, H. H. (2023). Four problems of the h-index for assessing the research productivity and impact of individual authors. *Scientometrics, 128*(5), 2677–2691. https://doi.org/10.1007/s11192-022-04323-8

Blumm, N., Ghoshal, G., Forró, Z., Schich, M., Bianconi, G., Bouchaud, J. P., & Barabási, A. L. (2012). Dynamics of ranking processes in complex systems. *Physical Review Letters, 109*(12), Article 128701. https://doi.org/10.1103/PhysRevLett.109.128701

Bornmann, L., & Leydesdorff, L. (2018). Count highly-cited papers instead of papers with h citations: use normalized citation counts and compare "like with like"! *Scientometrics, 115*(2), 1119–1123. https://doi.org/10.1007/s11192-018-2682-1

Bornmann, L., & Tekles, A. (2021). Convergent validity of several indicators measuring disruptiveness with milestone assignments to physics papers by experts. *Journal of Informetrics, 15*(3), Article 101159. https://doi.org/10.1016/j.joi.2021.101159. Article.

Braun, T., Glanzel, W., & Schubert, A. (2006). A Hirsch-type index for journals [Article]. *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*(1-7), 107–117. https://doi.org/10.1016/s0169-7552(98)00110-x

Brito, R., & Navarro, A. R. (2021). The inconsistency of h -index: A mathematical analysis. *Journal of Informetrics, 15*(1), Article 101106. https://doi.org/10.1016/j.joi.2020.101106. Article.

Ding, Y., & Cronin, B. (2011). Popular and/or prestigious? Measures of scholarly esteem. *Information Processing & Management, 47*(1), 80–96. https://doi.org/10.1016/j.ipm.2010.01.002

Dunaiski, M., Geldenhuys, J., & Visser, W. (2018). How to evaluate rankings of academic entities using test data. *Journal of Informetrics, 12*(3), 631–655. https://doi.org/10.1016/j.joi.2018.06.002

Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics, 69*(1), 131–152. https://doi.org/10.1007/s11192-006-0144-7

Egghe, L. (2008). Modelling successive h-indices. *Scientometrics, 77*(3), 377–387. https://doi.org/10.1007/s11192-007-1968-5

Egghe, L. (2010). On the relation between Schubert's h-index of a single paper and its total number of received citations. *Scientometrics, 84*(1), 115–117. https://doi.org/10.1007/s11192-009-0062-6

Fassin, Y. (2023). The ha-index: The average citation h-index. *Quantitative Science Studies*, 1–22. https://doi.org/10.1162/qss_a_00259

Fontana, M., Iori, M., Montobbio, F., & Sinatra, R. (2020). New and atypical combinations: An assessment of novelty and interdisciplinarity. *Research Policy, 49*(7), Article 104063. https://doi.org/10.1016/j.respol.2020.104063

Fortunato, S., Bergstrom, C. T., Boerner, K., Evans, J. A., Helbing, D., Milojevic, S., Petersen, A. M., Radicchi, F., Sinatra, R., Uzzi, B., Vespignani, A., Waltman, L., Wang, D., & Barabasi, A. L. (2018). Science of science. *Science, 359*(6379), Article eaao0185. https://doi.org/10.1126/science.aao0185. Article.

Funk, R. J., & Owen-Smith, J. (2017). A dynamic network measure of technological change. *Management Science, 63*(3), 791–817. https://doi.org/10.1287/mnsc.2015.2366

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature, 520*(7548), 429–431. https://doi.org/10.1038/520429a

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

Hirsch, J. E. (2019). hα: An index to quantify an individual's scientific leadership. *Scientometrics, 118*(2), 673–686. https://doi.org/10.1007/s11192-018-2994-1

Holme, P., & Saramäki, J. (2012). Temporal networks. *Physics Reports, 519*(3), 97–125. https://doi.org/10.1016/j.physrep.2012.03.001

Hu, X. J., & Rousseau, R. (2017). Nobel Prize winners 2016: Igniting or sparking foundational publications? *Scientometrics, 110*(2), 1053–1063. https://doi.org/10.1007/s11192-016-2205-x

Hutchins, B. I., Yuan, X., Anderson, J. M., & Santangelo, G. M. (2016). Relative citation ratio (RCR): A new metric that uses citation rates to measure influence at the article level. *PLOS Biology, 14*(9), Article e1002541. https://doi.org/10.1371/journal.pbio.1002541

Iniguez, G., Pineda, C., Gershenson, C., & Barabasi, A. L. (2022). Dynamics of ranking. *Nature Communications, 13*(1). https://doi.org/10.1038/s41467-022-29256-x. Article 1646.
Jo, W. S., Liu, L., & Wang, D. (2022). See further upon the giants: Quantifying intellectual lineage in science. *Quantitative Science Studies, 3*(2), 319–330. https://doi.org/10.1162/qss_a_00186
Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika, 30*(1/2), 81–93.
Koltun, V., & Hafner, D. (2021). The h-index is no longer an effective correlate of scientific reputation. *PLoS One, 16*(6), Article e0253397. https://doi.org/10.1371/journal.pone.0253397
Liu, L., Jones, B. F., Uzzi, B., & Wang, D. (2023). Data, measurement and empirical methods in the science of science. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-023-01562-4
Lü, L., Medo, M., Yeung, C. H., Zhang, Y. C., Zhang, Z. K., & Zhou, T. (2012). Recommender systems. *Physics Reports, 519*(1), 1–49. https://doi.org/10.1016/j.physrep.2012.02.006
Lü, L., Zhou, T., Zhang, Q. M., & Stanley, H. E. (2016). The H-index of a network node and its relation to degree and coreness. *Nature Communications, 7*(1), 10168. https://doi.org/10.1038/ncomms10168
Maji, G. (2020). Influential spreaders identification in complex networks with potential edge weight based k-shell degree neighborhood method. *Journal of Computational Science, 39*, Article 101055. https://doi.org/10.1016/j.jocs.2019.101055
Margolis, J. (1967). Citation indexing and evaluation of scientific papers. *Science, 155*(3767), 1213–1219. https://doi.org/10.1126/science.155.3767.1213
Mariani, M. S., Medo, M., & Lafond, F. (2019). Early identification of important patents: Design and validation of citation network metrics. *Technological Forecasting and Social Change, 146*, 644–654. https://doi.org/10.1016/j.techfore.2018.01.036
Mariani, M. S., Medo, M., & Zhang, Y. C. (2016). Identification of milestone papers through time-balanced network centrality. *Journal of Informetrics, 10*(4), 1207–1223. https://doi.org/10.1016/j.joi.2016.10.005
Merton, R. K. (1968). The Matthew effect in science. *Science, 159*(3810), 56–63. https://doi.org/10.1126/science.159.3810.56
Min, C., Bu, Y., Wu, D., Ding, Y., & Zhang, Y. (2021). Identifying citation patterns of scientific breakthroughs: A perspective of dynamic citation process. *Information Processing & Management, 58*(1), Article 102428. https://doi.org/10.1016/j.ipm.2020.102428
Mohammed, S., Morgan, A., & Nyantakyi, E. (2020). On the influence of uncited publications on a researcher's h-index. *Scientometrics, 122*(3), 1791–1799. https://doi.org/10.1007/s11192-020-03356-1
Park, M., Leahey, E., & Funk, R. J. (2023). Papers and patents are becoming less disruptive over time. *Nature, 613*(7942), 138–144. https://doi.org/10.1038/s41586-022-05543-x
Peterson, G. J., Pressé, S., & Dill, K. A. (2010). Nonuniversal power law scaling in the probability distribution of scientific citations. *Proceedings of the National Academy of Sciences, 107*(37), 16023–16027. https://doi.org/10.1073/pnas.1010757107
Poirrier, M., Moreno, S., & Huerta-Canepa, G. (2021). Robust h-index. *Scientometrics, 126*(3), 1969–1981. https://doi.org/10.1007/s11192-020-03857-z
Prathap, G. (2019). Letter to the editor: Revisiting the h-index and the p-index. *Scientometrics, 121*(3), 1829–1833. https://doi.org/10.1007/s11192-019-03240-7
Radicchi, F., & Castellano, C. (2011). Rescaling citations of publications in physics. *Physical Review E, 83*(4), Article 046116. https://doi.org/10.1103/PhysRevE.83.046116
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
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*(1), 4630. https://doi.org/10.1038/ncomms5630
Rousseau, R. (1987). The Gozinto theorem: Using citations to determine influences on a scientific publication. *Scientometrics, 11*(3), 217–229. https://doi.org/10.1007/BF02016593
Rowlands, I. (2018). What are we measuring? Refocusing on some fundamentals in the age of desktop bibliometrics. *FEMS Microbiology Letters, 365*(8). https://doi.org/10.1093/femsle/fny059. fny059.
Ruan, X., Lyu, D., Gong, K., Cheng, Y., & Li, J. (2021). Rethinking the disruption index as a measure of scientific and technological advances. *Technological Forecasting and Social Change, 172*, Article 121071. https://doi.org/10.1016/j.techfore.2021.121071
Schubert, A. (2007). Successive h-indices. *Scientometrics, 70*(1), 201–205. https://doi.org/10.1007/s11192-007-0112-x
Schubert, A. (2009). Using the h-index for assessing single publications. *Scientometrics, 78*(3), 559–565. https://doi.org/10.1007/s11192-008-2208-3
Shen, H. W., & Barabási, A. L. (2014). Collective credit allocation in science. *Proceedings of the National Academy of Sciences, 111*(34), 12325–12330. https://doi.org/10.1073/pnas.1401992111
Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A. L. (2016). Quantifying the evolution of individual scientific impact. *Science, 354*(6312). https://doi.org/10.1126/science.aaf5239. Article aaf5239.
Ullah, A., Wang, B., Sheng, J., Long, J., Khan, N., & Sun, Z. (2021). Identifying vital nodes from local and global perspectives in complex networks. *Expert Systems with Applications, 186*, Article 115778. https://doi.org/10.1016/j.eswa.2021.115778
Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science, 342*(6157), 468–472. https://doi.org/10.1126/science.1240474
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. (2012). The Inconsistency of the h-index. *Journal of the American Society for Information Science and Technology, 63*(2), 406–415. https://doi.org/10.1002/asi.21678
Wang, D., & Barabási, A. L. (2021). *The science of science*. Cambridge University Press.
Wang, D., Song, C., & Barabási, A. L. (2013). Quantifying long-term scientific impact. *Science, 342*(6154), 127–132. https://doi.org/10.1126/science.1237825
Wang, D., & Uzzi, B. (2022). Weak ties, failed tries, and success. *Science, 377*(6612), 1256–1258. https://doi.org/10.1126/science.add0692
Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics, 94*(3), 851–872. https://doi.org/10.1007/s11192-012-0775-9
Wang, J., Xu, S., Mariani, M. S., & Lü, L. (2021). The local structure of citation networks uncovers expert-selected milestone papers. *Journal of Informetrics, 15*(4), Article 101220. https://doi.org/10.1016/j.joi.2021.101220
Wang, M. Y., Zhang, J. Q., Jiao, S. J., Zhang, X. R., Zhu, N., & Chen, G. S. (2020). Important citation identification by exploiting the syntactic and contextual information of citations. *Scientometrics, 125*(3), 2109–2129. https://doi.org/10.1007/s11192-020-03677-1
Wolcott, H. N., Fouch, M. J., Hsu, E. R., DiJoseph, L. G., Bernaciak, C. A., Corrigan, J. G., & Williams, D. E. (2016). Modeling time-dependent and -independent indicators to facilitate identification of breakthrough research papers. *Scientometrics, 107*(2), 807–817. https://doi.org/10.1007/s11192-016-1861-1
Wu, L., Kittur, A., Youn, H., Milojević, S., Leahey, E., Fiore, S. M., & Ahn, Y. Y. (2022). Metrics and mechanisms: Measuring the unmeasurable in the science of science. *Journal of Informetrics, 16*(2), Article 101290. https://doi.org/10.1016/j.joi.2022.101290
Wu, X., Wei, W., Tang, L., Lu, J., & J, L (2019). Coreness and h -index for weighted networks. *IEEE Transactions on Circuits and Systems I: Regular Papers, 66*(8), 3113–3122. https://doi.org/10.1109/TCSI.2019.2907751
Xia, W., Li, T., & Li, C. (2022). A review of scientific impact prediction: tasks, features and methods. *Scientometrics*. https://doi.org/10.1007/s11192-022-04547-8
Xu, F., Wu, L., & Evans, J. (2022). Flat teams drive scientific innovation. *Proceedings of the National Academy of Sciences, 119*(23), Article e2200927119. https://doi.org/10.1073/pnas.2200927119
Xu, S. Q., Mariani, M. S., Lu, L. Y., & Medo, M. (2020). Unbiased evaluation of ranking metrics reveals consistent performance in science and technology citation data. *Journal of Informetrics, 14*(1). https://doi.org/10.1016/j.joi.2019.101005
Yan, Z. B., Wu, Q., & Li, X. C. (2016). Do Hirsch-type indices behave the same in assessing single publications? An empirical study of 29 bibliometric indicators. *Scientometrics, 109*(3), 1815–1833. https://doi.org/10.1007/s11192-016-2147-3
Yang, A. J., Deng, S., Wang, H., Zhang, Y., & Yang, W. (2023). Disruptive coefficient and 2-step disruptive coefficient: Novel measures for identifying vital nodes in complex networks. *Journal of Informetrics, 17*(3), Article 101411. https://doi.org/10.1016/j.joi.2023.101411

Yang, A. J., Hu, H., Zhao, Y., Wang, H., & Deng, S. (2023). From consolidation to disruption: A novel way to measure the impact of scientists and identify laureates. *Information Processing & Management, 60*(5), Article 103420. https://doi.org/10.1016/j.ipm.2023.103420

Ye, F. Y., & Rousseau, R. (2010). Probing the h-core: an investigation of the tail-core ratio for rank distributions [Article]. *Scientometrics, 84*(2), 431–439. https://doi.org/10.1007/s11192-009-0099-6

Yin, Y., Dong, Y., Wang, K., Wang, D., & Jones, B. F. (2022). Public use and public funding of science. *Nature Human Behaviour, 6*(10), 1344–1350. https://doi.org/10.1038/s41562-022-01397-5

Yin, Y., Gao, J., Jones, B. F., & Wang, D. (2021). Coevolution of policy and science during the pandemic. *Science, 371*(6525), 128–130. https://doi.org/10.1126/science.abe3084

Yin, Y., Wang, Y., Evans, J. A., & Wang, D. (2019). Quantifying the dynamics of failure across science, startups and security. *Nature, 575*(7781), 190–194. https://doi.org/10.1038/s41586-019-1725-y

Zeng, A., Shen, Z. S., Zhou, J. L., Wu, J. S., Fan, Y., Wang, Y. G., & Stanley, H. E. (2017). The science of science: From the perspective of complex systems. *Physics Reports, 714*, 1–73. https://doi.org/10.1016/j.physrep.2017.10.001

Zhai, L., Yan, X., & Zhang, G. (2018). Bi-directional h-index: A new measure of node centrality in weighted and directed networks. *Journal of Informetrics, 12*(1), 299–314. https://doi.org/10.1016/j.joi.2018.01.004

Zhai, L., Yan, X., & Zhang, G. (2019). The bi-directional h-index and B-core decomposition in directed networks. *Physica A: Statistical Mechanics and its Applications, 531*, Article 121715. https://doi.org/10.1016/j.physa.2019.121715

Zhang, F., & Wu, S. L. (2021). Measuring academic entities' impact by content-based citation analysis in a heterogeneous academic network. *Scientometrics, 126*(8), 7197–7222. https://doi.org/10.1007/s11192-021-04063-1

Zhao, S. X., Rousseau, R., & Ye, F. Y. (2011). h-Degree as a basic measure in weighted networks. *Journal of Informetrics, 5*(4), 668–677. https://doi.org/10.1016/j.joi.2011.06.005