Full length article

Could bibliometrics reveal top science and technology achievements and researchers? The case for evaluatology-based science and technology evaluation

Guoxin Kang [b,a,c], Wanling Gao [b,a,c], Lei Wang [b,a,c], Chunjie Luo [b,a,c], Hainan Ye [a,b,c], Qian He [a], Shaopeng Dai [c], Jianfeng Zhan [a,b,c] [ID],*

[a] *The International Open Benchmark Council, China*
[b] *ICT, Chinese Academy of Sciences, Beijing, China*
[c] *University of Chinese Academy of Sciences, Beijing, China*

ARTICLE INFO

ABSTRACT

By utilizing statistical methods to analyze bibliographic data, bibliometrics faces inherent limitations in identifying the most significant science and technology achievements and researchers. To overcome this challenge, we present an evaluatology-based science and technology evaluation methodology. At the heart of this approach lies the concept of an extended evaluation condition (EC), encompassing nine crucial components derived from a field. We define four relationships that illustrate the connections among various achievements based on their mapped extended EC components, as well as their temporal and citation links. Within a relationship under an extended EC, evaluators can effectively compare these achievements by carefully addressing the influence of confounding variables. We establish a real-world evaluation system encompassing an entire collection of achievements, each of which is mapped to several components of an extended EC. Within a specific field like chip technology or open source, we construct a perfect evaluation model that can accurately trace the evolution and development of all achievements in terms of four relationships based on the real-world evaluation system. Building upon the foundation of the perfect evaluation model, we put forth four-round rules to eliminate non-significant achievements by utilizing four relationships. This process allows us to establish a pragmatic evaluation model that effectively captures the essential achievements, serving as a curated collection of the top N achievements within a specific field during a specific timeframe. We present a case study on the top 100 Chip achievements to demonstrate the effectiveness of our science and technology evaluatology. The case study highlights its practical application and efficacy in identifying significant achievements and researchers that otherwise cannot be identified by using bibliometrics.

1. Introduction

Science and technology (S&T) evaluation is a meticulous and comprehensive process. One of its paramount goals is to identify the most remarkable accomplishments in each field, duly recognize the individuals, institutions, or nations that have made significant contributions to these achievements, and delve deeper into the effective and efficient mechanisms and policies within the S&T ecosystems that profoundly shape the evolution of these achievements [1]. This article focuses on the first half of the task.

While bibliometrics methodologies have long relied on observable metrics such as publication numbers, citation counts, and the H-index to assess correlations and impact [2–5], as illustrated in Fig. 1. it is essential to recognize their inherent three limitations and the need for alternative approaches.

First, bibliometrics commonly employs publication numbers, citation counts, and related metrics to gauge scholarly works' quality, influence, and significance. However, various confounding variables can significantly impact citation counts. Moreover, citation counts are vulnerable to manipulation by malicious networks.

Second, bibliometrics often fails to consider critical non-bibliometric metrics, making them insufficient for evaluating significant technological achievements that may have limited publication outputs. For instance, the Linux operating system in computer science has made
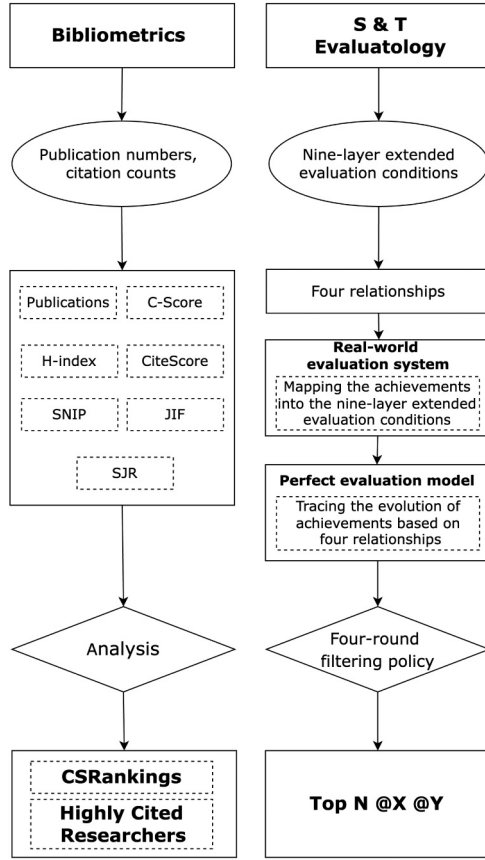
Fig. 1. Fundamental differences between bibliometrics and S&T evaluatology.

Fig. 2. Comparison of the number of scientists selected for the global top 2% in different disciplines.

a substantial impact despite having a modest publication record.

Third, many bibliometrics methodologies prioritize the quantity over the quality of publications, which can result in an incomplete assessment of the true value and impact of scholarly work.

To address these shortcomings, we introduce the S&T evaluatology, which exemplifies the application of evaluatology in evaluating S&T achievements. The S&T evaluatology is illustrated in Fig. 1 and presented in detail in [6,7]. The fundamental principle of evaluatology is to implement a well-defined evaluation condition (EC) on particular subjects to establish evaluation models or systems.

At the core of the S&T evaluatology is the notion of an extended EC, which comprises nine key components: (1) the field that can be broken down into several problem domains; (2) The set of problem domains, each of which can be broken down into various sub-problem domains; (3) the sub-problem domains, each of which can be decomposed into several problems; "(4) the set of a collective of equivalent problems, each of which can be broken down into multiple sub-problems; (5) the set of a collective of equivalent sub-problems; (6) the set of a collective of problems or sub-problem instances; (7) the algorithms or the algorithm-like mechanisms that tackles a problem or a sub-problem; (8) the implementations of algorithms or the algorithm-like mechanisms; (9) the support systems that provide necessary resources and environments [6,7]".

We define four relationships that illustrate the connections among various achievements based on their mapped extended EC components, as well as their temporal and citation links. We define two primary relationships: *pioneering and progressive* and two auxiliary relationships: *parallel* and *related but not connected*. Within a pioneering or progressive relationship under an extended evaluation condition, evaluators can effectively compare these achievements by carefully addressing the influence of confounding variables.

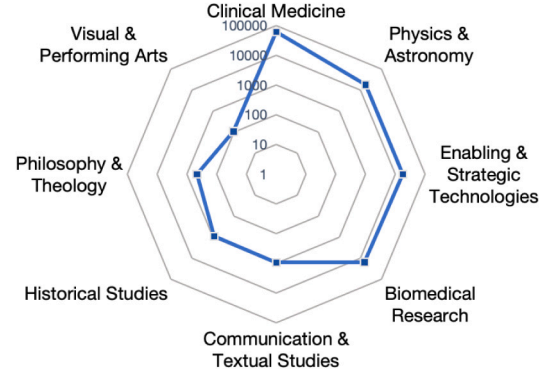We establish a real-world ES encompassing the complete collection of S&T achievements, each of which is mapped to several components of an extended EC. In line with the aim of identifying the top N S&T achievements, the proposed real-world S&T ES ignores the other components of the real-world S&T ecosystems, e.g., the mechanisms and policies that profoundly shape the evolution of these achievements [1].

Under the premise that all evaluated achievements belong to the same field, e.g., chip technology or open source, we construct "a perfect S&T EM" that can accurately trace the evolution and development of all achievements in terms of four relationships based on the real-world ES. We compare achievements that have a specific relationship under the extended EC they involve. Utilizing four relationships, we employ four rounds of rules to prune non-significant achievements to establish a pragmatic S&T EM that captures the fundamental S&T achievements. Essentially, the pragmatic S&T EM is a collection of top N achievements within a field during a timeframe.

The International Open Benchmark Council (BenchCouncil) utilized the S&T evaluatology principles and the instantiated Top N @X @Y methodology to systematically recognize the most 100 groundbreaking and influential achievements in chip technology (Chip100) [8]. The case study demonstrates the effectiveness of our proposed methodology compared to bibliometrics.

In the following sections, we will provide an in-depth examination of the S&T evaluatology. Section 2 enumerates the existing bibliometrics methodologies and analyzes their weakness. Section 3 presents the S&T evaluatology in detail. Section 4 provides an instantiated Top N @X @Y methodology. Section 5 introduces a case study on the Top 100 Chip Achievements. Section 6 concludes.

2. Motivation and related work

Bibliometrics is a field that applies statistical methods to analyze bibliographic data.

In this subsection, we first present the overall weakness of bibliometrics in Section 2.1. Then, we introduce the representative bibliometrics methodologies. Finally, we introduce the fundamental concept, theory, and methodology in evaluatology [6], based on which we will present the S&T evaluatology.

2.1. Motivation: The limitations of bibliometrics

Due to the nature of bibliometrics, there are several inherent drawbacks associated with its application.

First, bibliometrics commonly employs publication numbers, citation counts, and related metrics to gauge the quality, influence, and

significance of scholarly works. However, it is crucial to acknowledge that publication numbers and citation counts can be significantly impacted by various confounding variables. These may include the diverse disciplines involved, the reputation and networks linked to researchers and their institutional affiliations, as well as notable differences in researcher numbers and publication volumes across various fields. Moreover, citation counts may even be vulnerable to manipulation by malicious networks (Limitation One).

- (Limitation One-One). The same or similar works published in the same or different periods can be impacted by various confounding variables, such as the reputation and network of the researchers and their institutions, leading to significant variations in citation counts. Moreover, citation counts may be subject to manipulation by malicious networks.
- (Limitation One-Two). In fundamental disciplines like mathematics, once a problem has been effectively solved, there may be limited follow-up research on that specific topic. Consequently, the citation count for the original work in fundamental disciplines may not increase significantly. Hence, it cannot accurately reflect the impact or influence of the research in the field.
- (Limitation One-Three). Citation counts fail to account for the significant disparities in researcher numbers and publication volumes across different fields. In fields with fewer researchers and publications, citation counts are naturally lower, regardless of the quality of the research being conducted.
- (Limitation One-Four). Bibliometrics prioritize well-established disciplines, potentially overlooking emerging fields or unconventional research outputs that may have a significant impact but lower citation counts. The effectiveness of citation metrics is limited in representing contributions within emerging or specialized fields. Groundbreaking research in these domains might initially receive few citations due to the novelty of the subject matter or the field's limited scope. Consequently, as shown in Fig. 2, pivotal advancements in such areas risk being undervalued, as seen in the "top scientists" list created by Stanford University and the Elsevier data repository [3], which predominantly features scientists from well-established fields like Clinical Medicine and Physics & Astronomy. This bias is particularly harmful to innovators who spearhead new research directions, as their contributions may not be accurately captured by citation-based metrics.
- (Limitation One-Five). Self-citations occur when authors cite their previous work, potentially inflating the impact of their research. This practice skews the representation of a paper's or a researcher's genuine influence within the academic community. For instance, metrics like the H-index [4] are unable to circumvent the issue of self-citations, resulting in a biased assessment that may unfairly favor those who self-cite frequently.

Second, bibliometrics significantly ignores other fundamental non-bibliometric metrics and hence cannot be applied to significant technological achievements that have few or no publication outputs (Limitation Two). Bibliometrics primarily relies on analyzing published works. Limitation Two arises when considering groundbreaking technological advancements that may not be adequately represented in traditional scholarly publications.

In practical fields like computer science, substantial contributions frequently occur outside the conventional academic publishing framework. A prime example of this is the Linux operating system within the realm of computer science. As an open-source software, the Linux operating system boasts numerous contributors who may not publish extensively. Similarly, the computer mouse, one of the most universally adopted human–computer interaction technologies, demonstrates that significant impact does not necessarily stem from published research. Table 1 presents several significant technological achievements that are overlooked by bibliometrics. Therefore, bibliometrics alone may not

Table 1
Summary of significant achievements overlooked by bibliometrics.

Field	Achievements	Published paper	Citations
Chip	X86 ISA	No	N/A
	PCB	No	N/A
Open-sources systems	Linux Kernel	No	N/A
	Git	No	N/A
	MySQL	No	N/A
Benchmarks	Whetstone	No	N/A
	TPC-C	No	N/A
	TPC-H	No	N/A
	FIO	No	N/A

fully capture the impact and significance of these achievements. Consequently, non-academic metrics should be considered in the grading process to select the top-impact achievements.

Third, bibliometrics prioritizes the quantity over the quality of publications (Limitation Three). High citation counts of a researcher can result from either a large volume of modestly impactful publications or from many surveys on trending topics, such as timely topics on large language models. Although these works might garner significant attention, they do not necessarily represent substantial advancements within their disciplines. The emphasis on publication counts can lead to a skewed representation of research impact, as it fails to consider the significance, rigor, and originality of individual publications.

In summary, while bibliometrics provides a quantitative metric, like citation counts, for academic evaluation, they are beset with limitations that result in biased and incomplete assessments. Thus, the S&T evaluation urgently needs more nuanced and comprehensive evaluation metrics and methodologies that go beyond bibliometrics. Such metrics would ensure a fairer and more accurate depiction of scholarly impact, truly reflecting the multifaceted nature of academic contributions.

2.2. The representative bibliometrics methodologies

2.2.1. Csrankings in the computer science field

CSRankings is a specialized method for evaluating computer science achievements, favoring the conference publication over the journal. CSRankings adopts the metric of the number of publications at so-called top-tier conferences for gauging the academic influence of researchers or their affiliated institutions in computer science. Utilizing this metric, Emery Berger pioneered CSRankings [2], a tailored academic leaderboard specifically designed for the realm of computer science. CSRankings selects the Digital Bibliography & Library Project (DBLP) [9] as its data source, ensuring up-to-date and relevant rankings with quarterly updates.

However, this methodology has several serious flaws. First and foremost, it places a higher emphasis on publication quantity than quality, as outlined in Section 2.1 (Limitation Three), thereby having flaws in recognizing top researchers or groundbreaking achievements.

For example, David Patterson's influential works in chip technology, particularly with RISC, RISC-V, and RAID, have substantially shaped the field. Notwithstanding their extensive influence, Patterson is conspicuously absent from CSRankings, a glaring omission highlighting a significant shortcoming in the ranking system's ability to acknowledge key contributors even in leading institutions.

In addition, CSRankings cannot discern the varying impacts of different achievements. CSRankings quantifies the number of papers presented at top-tier conferences, but this approach fails to identify who pioneered a field. For instance, although the groundbreaking "Transformer" model [10] was presented at the 31st Conference on Neural Information Processing Systems (NeurIPS), it is erroneous to assume that all papers at this conference exert an influence comparable to that of the Transformer.

This situation underscores a fundamental flaw in the CSRankings system: overemphasizing top-tier conference publications can lead to

misleading representations, bypassing the real depth and enduring impact of substantial contributions.

Second, many influential works like the Linux operating system have never even sought publication in a so-called top-tier conference (Limitation Two). Table 1 provides other examples. The current metric focusing on publications fails to recognize significant achievements that are not encapsulated in conference papers. The Linux operating system's development and its widespread adoption stand as a prime example, achieving monumental impact without the endorsement of traditional academic publications.

Third, CSRankings overlooks the significant disparities in researcher numbers, publication frequency, and volumes across different fields within computer science (Limitation One-Three). This oversight has resulted in a skewed ranking from 1970–2022, where four out of the top seven institutions are led by faculty specializing in vision, a field known for its high paper acceptance volumes. For example, the field of computer vision, known for higher publication volumes, is overrepresented. As shown in Fig. 3, in 2022, The IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR), a leading conference in computer vision, accepted 2065 papers [11], whereas The IEEE/ACM International Symposium on Microarchitecture Conference (MICRO), a top conference in computer architecture, accepted just 83 [12].

The discrepancies in publication volume across different fields may lead to potential misleading outcomes when using CSRankings. These disparities raise concerns about the accuracy of CSRankings in providing equitable representation for all fields within computer science.

Fourth, as a consequence of being accepted by a so-called top-tier conference, this metric is impacted by various confounding variables, such as the reputation and network of the researchers and their institutions, and is subject to manipulation by malicious networks. Collusion among reviewers is not an isolated incident in numerous computer science conferences.

2.2.2. The standardized citation metrics (c-score)

The c-score, developed by John Ioannidis [3], assesses the influence of scientists. This standardized indicator amalgamates various elements, including citations, h-index, co-authorship-adjusted hm-index, and authorship-position-specific citations. Leveraging this metric, Ioannidis's team curated a global database for ranking scientists, categorized into career-long and single-year impacts based on the Scopus data. The former category spans citations from 1996 to now, while the latter focuses on the current calendar year alone. This innovative metric transcends traditional citation metrics, avoiding the evaluation biases introduced by self-citations. However, its primary focus on publications cannot completely encompass the wider spectrum of a scientist's influence, particularly in areas such as practical applications or cross-disciplinary collaborations. These critical dimensions, essential to the fabric of scientific progress, are often understated in conventional bibliometric measures (Limitation Two).

Despite its popularity in the scientific field, the standardized citation metric has limitations in acknowledging the impact of researchers in emerging disciplines (Limitation One-Four), leading to an underrepresentation of their contributions. The metric's proclivity to privilege well-established, voluminous fields is evidenced by the fact that over half of the top-ranked influential scientists in 2021 originated from fields like Clinical Medicine, Physics & Astronomy, Biomedical Research, and Enabling & Strategic Technologies. This trend reveals an inherent bias, favoring areas with more substantial publication frequencies and higher citation volumes (Limitation One-Three).

In addition, Limitation One remains a challenge that cannot be mitigated by the standardized citation metrics (c-score). Factors such as the reputation and network of researchers and their affiliated institutions can confound the evaluation process. For instance, even when two researchers from different institutions achieve similar achievements, the level of attention and recognition their work receives can vary significantly. In some cases, the earlier work may receive limited
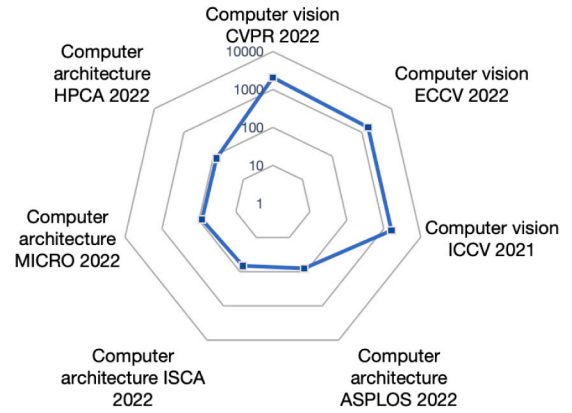

Fig. 3. Comparison of accepted papers by top conferences in the fields of computer vision and computer architecture.

attention, while subsequent work gains widespread acclaim. These disparities can be attributed to various factors, including the visibility and influence that researchers and their institutions hold within the academic community.

2.2.3. H-index

H-index [4] is a useful metric proposed by Jorge E. Hirsch to characterize a researcher's scientific output. The objective is to determine the highest value of h, where there are at least h papers with a citation number equal to or greater than h. The mathematical representation of H-index for a scientist is $h\_index(f) = \max\{i \in \mathbb{N} : f(i) \geq i\}$ [5]. Here, f is an array that contains the number of citations for the scientist's publications in decreasing order [5]. Instead of relying solely on single-number criteria like the total number of papers, H-index takes a more holistic approach by considering both productivity and academic impact. In addition to the limitations that we have discussed extensively in Section 2.1, in practice, vast self-citations can raise the H-index value easily.

2.2.4. CiteScore metrics

CiteScore Metrics [13], developed by Elsevier, extensively evaluates academic journals' citation impact and influence. These metrics are calculated yearly, considering a three-year citation window and considering the volume, quality, and field-normalized citation rates of articles published in a specific journal. Featuring indicators such as average citations per document, quartile ranking, and overall standing, CiteScore Metrics provides a transparent and comprehensive tool for researchers and institutions. While CiteScore metrics are designed to assess the quality and impact of scholarly journals rather than evaluate the quality of research within specific fields. In addition to the limitations that we have discussed in Section 2.1, it has another serious limitation. It is based on a three-year citation window. Consequently, achievements with a substantial long-term impact but relatively few citations in the short term may be undervalued.

2.2.5. Source Normalized Impact per Paper (SNIP)

The Source Normalized Impact per Paper [14] is a metric employed in assessing the influence of scholarly journals. It is determined by dividing an article's citation count within a journal by the anticipated citation rate within its particular field. SNIP considers the citation potential within the journal's discipline, enabling equitable comparisons across diverse areas of study. In essence, SNIP serves as a valuable gauge for evaluating the impact of a journal relative to its field. It provides researchers and institutions with a standardized measure to evaluate the influence of scholarly journals rather than the impact

Extended Evaluation Condition (extended EC)	
Component 1	the field
Component 2	the set of problem domains
Component 3	the set of sub-problem domains
Component 4	the set of a collective of equivalent problems
Component 5	the set of a collective of equivalent sub-problems
Component 6	the set of a collective of problems or sub-problem instances
Component 7	the algorithms or the algorithm-like mechanisms that tackles a problem or a sub-problem
Component 8	the implementations of algorithms or the algorithm-like mechanisms
Component 9	the support systems that provide necessary resources and environments

Fig. 4. The overview of an extended EC.

of the specific research achievement. Furthermore, SNIP compares a journal's citation count with the citation frequency in its field. However, it fails to consider the variations in citation practices across different subject areas. In addition, it has many inherent bibliometrics limitations we discussed in Section 2.1.

2.2.6. Journal impact factor (JIF)

The journal impact factor, devised by Eugene Garfield, is used by Clarivate's Web of Science to evaluate a journal's impact. The impact factor is calculated as $C / \sum_{i=0}^{n} P_i$, where C is the number of citations received in a given year for publications in a journal that were published in the n preceding years, and $\sum_{i=0}^{n} P_i$ is the total number of citable items published in that journal during the n preceding years.

2.2.7. SCImago Journal Rank (SJR)

SCImago Journal Rank (SJR) indicator, developed by the Scimago Lab, is a measure of the prestige of journals. SJR is calculated by using an algorithm similar to Google's PageRank, which assumes that important websites are linked to other important websites. Citations are used to link the journals. The algorithm begins by setting an identical amount of prestige to each journal, then using an iterative procedure to transfer each journal's achieved prestige to each other through citations until each journal's update reaches a minimum threshold. The limitations of SJR include the algorithm's complexity, the degree of transparency, and the reproducibility of the results.

Besides, Kevin W. Boyack [15] utilizes data mining and analysis techniques to map knowledge domains, specifically applying them to 20 years of PNAS publications. It combines various data sources to analyze the input–output ratio and diffusion between disciplines. However, its reliance on raw citation counts as the primary measure of impact, without adjusting for self-citations, potentially leads to a skewed and less meaningful assessment of true scholarly influence.

2.3. The basic concepts, theories, and methodologies in evaluatology

According to [6], an individual or system being evaluated is a subject. A stakeholder is defined as an entity that holds a stake of responsibility or interest in the subject matter. Evaluation is "the process of inferring the impact of subjects indirectly within evaluation conditions (EC) that cater to the requirements of stakeholders, relying on objective measurements and/or testing" [7].

The fundamental methodology for evaluating a single subject is outlined as follows. Zhan et al. [6] propose a universal methodology to define an EC, which consists of five basic components [6]: "(1) a set of equivalent definitions of problems; (2) the set of a collective of equivalent problem instances; (3) the algorithms or algorithm-like mechanisms; (4) the implementations of algorithms or algorithm-like mechanisms; (5) support systems that provide necessary resources and environments [7]".

Subsequently, it becomes crucial to implement a well-defined EC for a precisely defined subject, forming a well-defined evaluation model (EM) or system (ES).

In terms of complex scenarios, the evaluation methodology is to establish a series of EMs that ensure transitivity from a real-world ES to a perfect EM and a pragmatic EM [6].

Zhan et al. [6,7] characterize the real-world ES, perfect or pragmatic EMs. Because our S&T evaluation methodology is based on those concepts, we give a concise summary based on [6,7].

The real-world ES refers to "the entire population of real-world systems that are used to evaluate specific subjects". The real-world ES has several significant obstacles: "the presence of numerous confounding, prohibitive evaluation costs resulting from the huge state spaces".

A perfect EM replicates the real-world ES with utmost fidelity: "It eliminates irrelevant problems and has the capability to thoroughly explore and comprehend the entire spectrum of possibilities of an EC". However, it also has serious limitations: "possesses huge state space, entails a vast number of independent variables, and hence results in prohibitive evaluation costs".

Providing a means to estimate the parameters of the real-world ES or a perfect EM, a pragmatic EM simplifies the perfect EM in two ways: "reduce the number of independent variables that have negligible effect and sample from the extensive state space".

3. The science and technology evaluatology

This section presents the essence of S&T evaluatology.

3.1. The overview

Understanding the development of S&T is highly challenging. Sometimes, practice leads the way; at other times, theory does. Some individuals pose a significant problem and offer a preliminary solution, while others provide state-of-the-practice solutions without explicitly stating the problems. The relentless efforts of scientists and engineers make the landscape of S&T achievements intricate and dense, much like an interwoven forest, thereby making the objective evaluation of S&T contributions extremely challenging.

To tackle this challenge, we have adopted the evaluatology framework developed by Zhan et al. [6] as the theoretical foundation for our research. This framework serves as the basis for developing S&T evaluatology. The core principles and methodologies of S&T evaluatology are outlined as follows:

First, building upon the definition of an EC proposed in the referenced paper [1], we introduce the concept of an extended EC, as shown in Fig. 4.

With respect to the EC definition [7], an extended EC introduces several extra components to accommodate the new requirements of S&T evaluation, including the field that can be broken down into several problem domains, the set of problem domains, the set of sub-problem domains, and the set of a collective of equivalent sub-problems. The definition of the extended EC serves as the foundation for the proposed S&T evaluatology. It provides the framework upon which the evaluation of S&T achievements is based.

Second, in the realm of S&T evaluation, a subject refers to an accomplishment that can mapped onto the nine components of an extended EC.
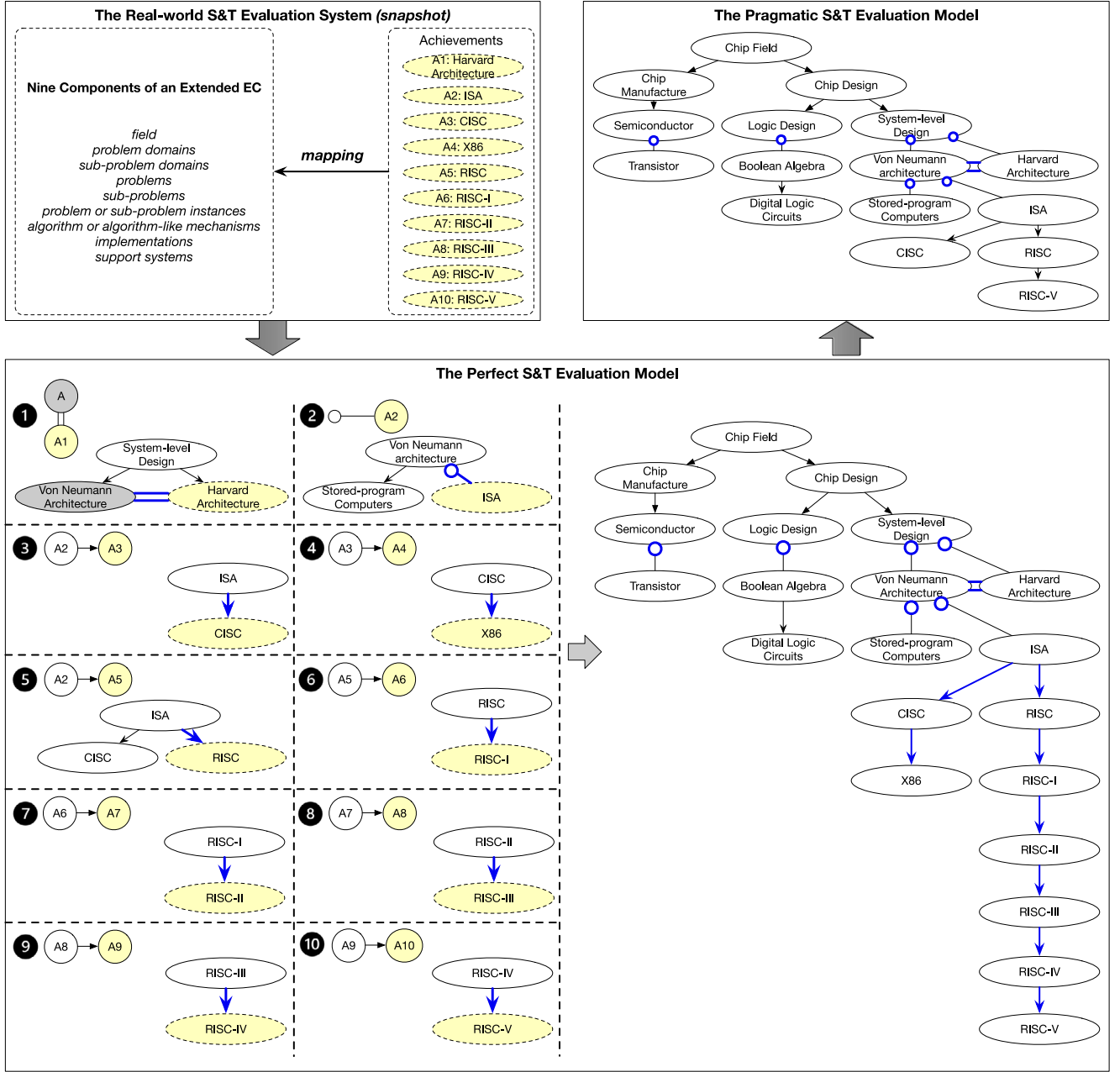
Fig. 5. Illustrating S&T Evaluatology with an example.

For instance, let us consider a scenario where a researcher proposes a new problem and provides a preliminary algorithm for solving that problem. In this case, the subject, a specific S&T achievement, comprises multiple components. These components include:

- problem: The specific problem being addressed or investigated.
- Algorithm: The preliminary algorithm proposed by the researcher to solve the given problem.

Third, based on their mapped extended EC components as well as their temporal and citation links, we establish two primary relationships: *pioneering and progressive* and two auxiliary relationships: *parallel* and *related but not connected* to illustrate the connections among different achievements. Section 3.3 will provide the details of four relationships.

Fourth, according to the theory of evaluatology, S&T evaluation involves applying a well-defined extended EC to the subject—a specific S&T achievement. This process allows for the creation of an EM or

ES. Within a relationship under an extended EC, evaluators can effectively compare different S&T achievements by carefully addressing the influence of confounding variables [6,7].

In the subsequent four steps, we will adhere to and implement the universal evaluation methodology proposed by Zhan et al. [6] to address the intricate S&T evaluation scenarios.

Fifth, we establish a real-world S&T ES, which encompasses the complete collection of S&T achievements. Moreover, each achievement will be decomposed into its respective components within an extended EC. In establishing a real-world S&T ES, it is crucial to characterize the real-world S&T ecosystems. In line with the aim of identifying the top N S&T achievements, the proposed real-world S&T ES in this article encompasses the entire collection of S&T achievements while ignoring the other components of the real-world S&T ecosystems.

Sixth, under the premise that all evaluated achievements belong to the same field, we assume the existence of a "perfect S&T EM" that can accurately trace the S&T evolution and development in terms of four
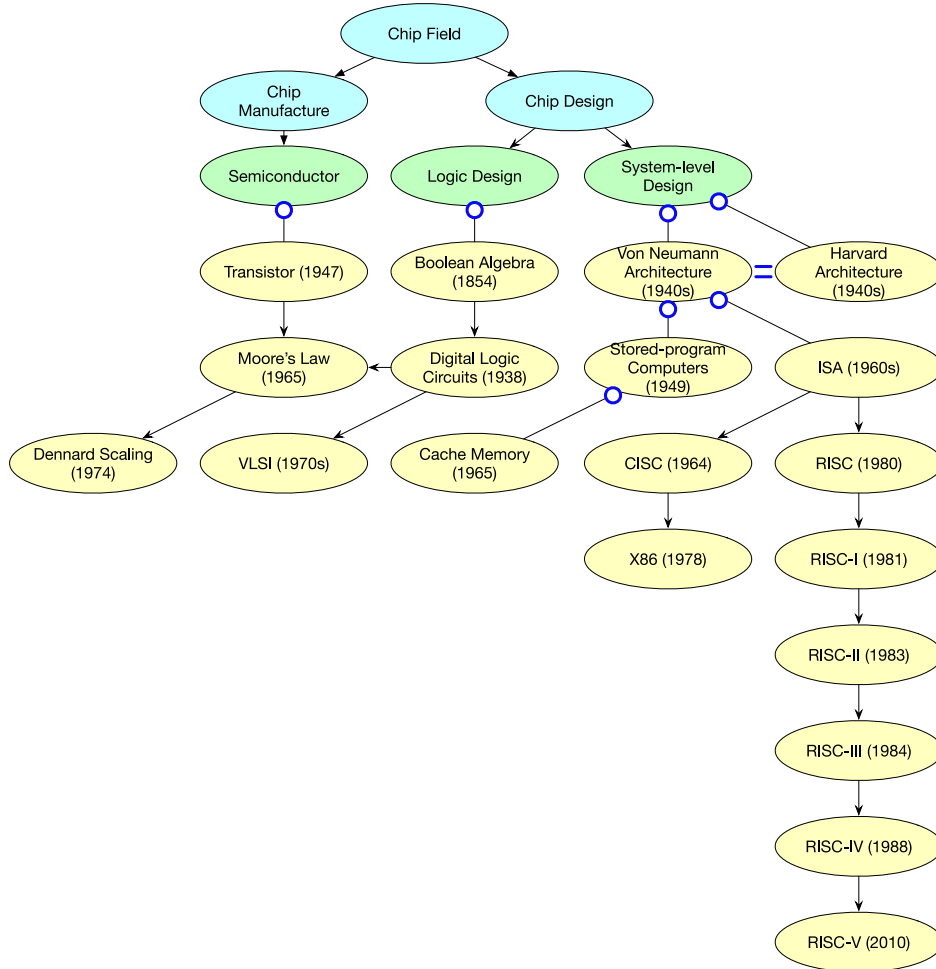
Fig. 6. The localized snapshot of a pragmatic EM in the field of chip technology.

relationships. That is to say, a "perfect S&T EM" can track the evolution of a real-world S&T ES from ES_i to ES_{i+1} in a rigorous manner. This model operates under the premise that only one change is made at a time. By implementing one change at a time, we ensure that only one achievement is added.

Seventh, as the perfect S&T EM contains huge states, we propose several simple rules to prune non-significant achievements to establish a pragmatic S&T EM that captures the fundamental S&T achievements. Essentially, the pragmatic S&T EM is a collection of top N achievements. The basic idea behind this process is that we compare achievements that have a pioneering or progressive relationship under the extended EC they involve. We will explain the simple rules in Section 3.6.

Fig. 5 illustrates S&T Evaluatology with an example, while Fig. 6 offers a localized snapshot of a pragmatic EM in the field of chip technology, showcasing individual achievements. Within the chip technology field are several critical problem domains like 'Chips Manufacture' and 'Chips Design'. This localized snapshot highlights the diversity and complexity within the chip technology field.

3.2. The definition of an extended EC

In [6], Zhan et al. emphasized that "understanding the composition of the problem domain is crucial in identifying the problem that best represents the whole. Across different disciplines, a field often exhibits a hierarchical structure, where a significant problem domain can be broken down into several problems", which provide the methodology to model an extended EC.

An extended EC consists of nine basic components [6,7], as shown in Fig. 4: (1) the field that can be broken down into several problem domains; (2) the set of problem domains, each of which can be broken down into various sub-problem domains; (3) the sub-problem domains, each of which can be decomposed into several problems; "(4) the set of a collective of equivalent problems, each of which can be broken down into multiple sub-problems; (5) the set of a collective of equivalent sub-problems; (6) the set of a collective of problems or sub-problem instances; (7) the algorithms or the algorithm-like mechanisms that tackle a problem or sub-problem; (8) the implementations of algorithms or the algorithm-like mechanisms; (9) the support systems that provide necessary resources and environments [6,7]".

As depicted in Fig. 7, the essential steps of the methodology can be summarized as follows. The first and second steps are to define the field and compose it into different problem domains. If necessary, the third step is to decompose each problem domain into several sub-problem domains. The fourth step is to break down problem domains or sub-problem domains into the problems. If necessary, the fifth step is to decompose each problem into several sub-problems. The sixth step proposes the problem instances or sub-problem instances. The seventh step is to figure out the algorithms or algorithm-like mechanisms to solve the problem or sub-problem. The eighth step encompasses the implementation of algorithms or algorithm-like mechanisms. The last step is to define the support system.

For example, chip design is a problem domain in the chip field. The system-level design is a typical sub-problem domain in chip design. The computer architecture design is one of the problems of the system-level design. The Von Neumann architecture was the pioneering

work that defined the computer architecture design problem and proposed algorithm-like mechanisms to address it. Any specific processor that aligns with the Von Neumann architecture can be viewed as an implementation of this mechanism.

3.3. The formal definition of four relationships

In this section, based on their mapped extended EC components as well as their temporal and citation links, we propose two primary relationships and two auxiliary relationships to connect achievements, as shown in Fig. 8.

3.3.1. Two primary relationships

Two fundamental relationships contain a pioneering relationship and a progressive relationship.

Relationship one: A pioneering relationship. *Definition:* A pioneering relationship pertains to an achievement that opens up a new research direction in the form of establishing a new field, problem domain, sub-problem domain, problem, sub-problem, algorithm or algorithm-like mechanism, implementation, or support system within an extended EC. The pioneering relationship recognizes the pioneering nature of such achievements, which lay the foundation for future advancements and innovations.

Formal expression: Let A represent an achievement. The pioneering relationship for A can be formally expressed as:

$$P(A) = \begin{cases} 1 & \text{if } A \text{ opens up a new research direction in the} \\ & \text{form of establishing a new field, problem do-} \\ & \text{main, sub-problem domain, problem, sub-} \\ & \text{problem, algorithm or algorithm-like} \\ & \text{mechanism, implementation, or support} \\ & \text{system within an extended EC,} \\ 0 & \text{otherwise} \end{cases}$$

This binary expression indicates whether A qualifies as a pioneering achievement (1) or not (0). It is based solely on the novelty and originality of the achievement A, without any preceding work.

Examples: Pioneering relationships manifest across various industries and disciplines, highlighting achievements that are the first to propose a novel field, problem domain, sub-problem domain, problem, sub-problem, solution, or support system within an extended EC.

- **Chip:** The Instruction Set Architecture (ISA) represents the pioneering work that defined the instruction set design sub-problem within the computer architecture design problem and proposed corresponding mechanisms to address it. The Reduced Instruction Set Computer (RISC) and Complex Instruction Set Computers (CISC) are subsequent developments following ISA.
- **AI:** The first computational model of a neuron, the McCulloh-Pitts neuron [16], is a pioneering algorithm-like mechanism in the field of neural networks.

Relationship two: A progressive relationship. *Definition*: For the achievements that involve the same component of an extended EC, e.g., a problem or sub-problem, a progressive relationship indicates subsequent achievements are inspired by preceding ones, and the latter publicly acknowledges this influence through citations.

Formal expression: A progressive relationship between two achievements A_i and A_j is defined as:

$$\begin{aligned} S(A_i, A_j) = 1 \iff & \left(Q(A_i) = Q(A_j) \right) \\ & \wedge \left(EC(A_i) \cap EC(A_j) \neq \emptyset \right) \\ \wedge \left((T(A_{i\_e}) < T(A_{j\_b})) \vee (T(A_{i\_b}) > T(A_{j\_e})) \right) & \\ & \wedge \left((A_i \in R(A_j)) \vee (A_j \in R(A_i)) \right) \end{aligned} \tag{1}$$
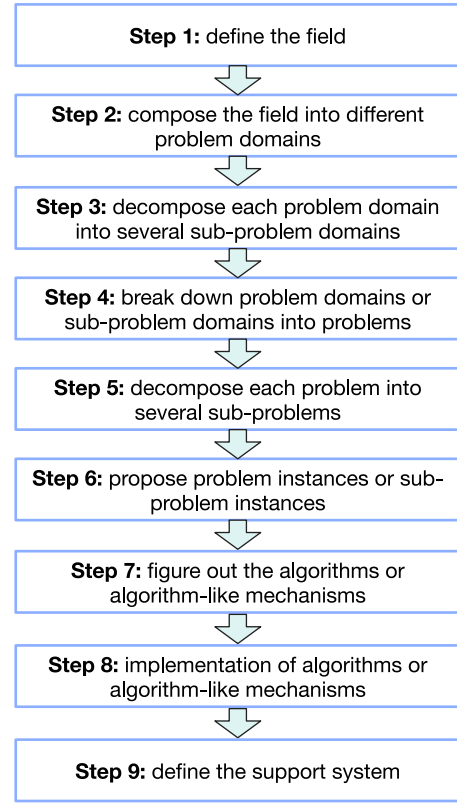
Fig. 7. Essential steps of S&T evaluatology.

where:

- $Q(a)$ is the key problem domain, sub-problem domain, problem, or sub-problem that achievement a addresses.
- $EC(a)$ denotes the EC involved in achievement a.
- $T(a\_b)$ and $T(a\_e)$ represent the begin time and end time of achievement a, respectively. Thus, $T(A_{i\_e}) < T(A_{j\_b})$ indicates that achievement A_i precedes achievement A_j in time. $T(A_{i\_b}) > T(A_{j\_e})$ indicates that achievement A_j precedes achievement A_i in time.
- $R(a)$ indicates the references of achievement a. Thus, $A_i \in R(A_j)$ indicates achievement A_j publicly acknowledge the influence of achievement A_i.

A many-to-one progressive relationship. *Definition*: A many-to-one progressive relationship is an instance of a progressive relationship, indicating multiple much preceding achievements inspire a single subsequent achievement.

Formal expression: A many-to-one progressive relationship between achievements $A_{i1}, A_{i2}, \ldots, A_{in}$ and A_j is defined as:

$$S(A_{i1}, A_j) \wedge S(A_{i2}, A_j) \wedge \cdots \wedge S(A_{in}, A_j) = 1 \tag{2}$$

where:

- $\{A_{i1}, A_{i2}, \ldots, A_{in}\}$ are multiple preceding achievements.
- A_j is a single subsequent achievement.

An one-to-many progressive relationship. *Definition*: A one-to-many progressive relationship is an instance of a progressive relationship, indicating a single preceding achievement inspires multiple subsequent achievements.

Formal expression: A one-to-many progressive relationship between achievement A_i and $A_{j1}, A_{j2}, \ldots, A_{jn}$ is defined as:

$$S(A_i, A_{j1}) \wedge S(A_i, A_{j2}) \wedge \cdots \wedge S(A_i, A_{jn}) = 1 \tag{3}$$
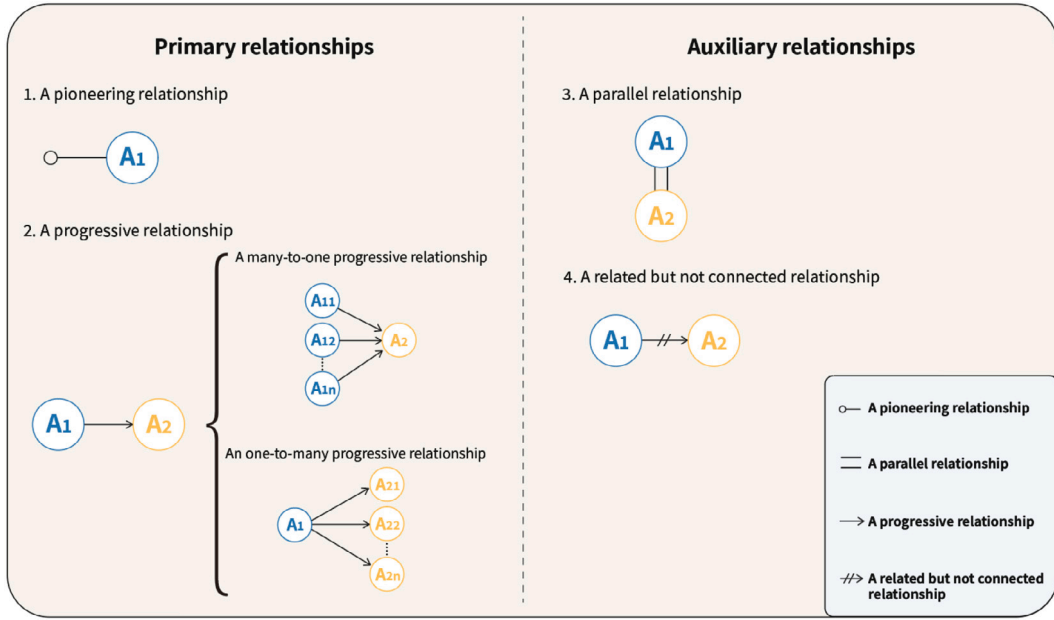
Fig. 8. Two fundamental relationships and two auxiliary relationships among the S&T achievements.

where:

- A_i is a single preceding achievement.
- $\{A_{j1}, A_{j2}, \ldots, A_{jn}\}$ are multiple subsequent achievements.

Examples: Progressive relationships demonstrate how knowledge and technology evolve over time, with each new development building on the previous ones.

- **Chip:** The RISC-V instruction set architecture has its origins in and was developed from the original RISC design.
- **AI:** LeNet [17,18] is a pioneering convolutional neural network that inspired AlexNet [19], a milestone in the field of deep learning.
- **Open-sources systems:** OpenBLAS [20] is a progressive achievement of GotoBLAS2 [21].
- **Benchmarks:** The CH-benCHmark [22] exemplifies a many-to-one progressive relationship as it integrates aspects from both the TPC-C [23] and TPC-H [24] benchmarks. This benchmark is designed to evaluate a hybrid workload by combining the transactional operations characteristic of TPC-C with the complex querying features of TPC-H.

3.3.2. Two auxiliary relationships

Two auxiliary relationships contain a parallel relationship and a connected but not related relationship.

Relationship three: A parallel relationship. *Definition:* A parallel relationship indicates that the achievements that involve the same component of an extended EC, e.g., problem or sub-problem, are proposed simultaneously within a brief and shared timeframe.

Formal expression: For a set of achievements A with each achievement $A_i \in A$, a parallel relationship between two achievements A_i and A_j is defined as:

$$
\begin{aligned}
P(A_i, A_j) = 1 \iff & \big(Q(A_i) = Q(A_j)\big) \\
& \wedge \big(EC(A_i) \cap EC(A_j) \neq \emptyset\big) \\
\wedge \big([T(A_i\_b), T(A_i\_e)] \cap [T(A_j\_b), T(A_j\_e)] \neq \emptyset\big)
\end{aligned}
\tag{4}
$$

where:

- $Q(a)$ is the key problem domain, sub-problem domain, problem, or sub-problem that achievement a addresses.
- $EC(a)$ denotes the EC involved in achievement a.
- $T(a\_b)$ and $T(a\_e)$ represent the begin time and end time of achievement a, respectively. The achievements A_i and A_j are considered to be in a parallel relationship if their time intervals overlap.

Examples: Parallel relationships occur across multiple fields where different approaches are employed simultaneously to address a common issue within a shared timeframe.

- **Chip:** The Von Neumann architecture and the Harvard architecture are two parallel works in computer system-level design in the 1940s.
- **AI:** BERT [25] and GPT [26] are two parallel works in the research of big models.
- **Open-sources systems:** Ubuntu [27] and CentOS [28] are two parallel works in open-source software.
- **Benchmarks:** BigDataBench [29] and BigBench [30] are two benchmarks specifically designed for evaluating big data systems, and they epitomize a parallel relationship as both were published within a year of each other, representing concurrent efforts in the problem domain of big data benchmarking.

Relationship four: A related but not connected relationship. *Definition:* For the achievements that involve the same component of an extended EC, e.g., a problem or sub-problem, a related but not connected relationship suggests that these achievements are not proposed simultaneously within a brief and shared timeframe. Instead, they are related in some way, but there is no explicit public acknowledgment cited by the later achievements indicating inspiration or influence from the earlier ones.

Formal expression: A related but not connected relationship characterizes that two achievements are not parallel and have similar components inheriting the same high-level component of an extended EC but lack a citation.

This relationship carries three implications. First, two achievements have similar components inheriting the same high-level component of an extended EC. Second, they are not parallel in nature, meaning they are not proposed simultaneously. Third, though the two achievements

---

Algorithm 1 Identify four fundamental relationships among numerous S&T achievements

---

1: **Input:** $IDs, TimeStamps\_b, TimeStamps\_e, References, EC, ProblemsQ$
2: **Output:** $PioneerRelationship, ParallelRelationship, ProgressiveRelationship, RelatedButNotConnectedRelationship$
3: Initialize $PioneerRelationship, ParallelRelationship, ProgressiveRelationship, RelatedButNotConnectedRelationship$ to empty sets
4: **for** each achievement i **do**
5:    **if** i opens up a new research direction in the form of establishing a new field, problem domain, sub-problem domain, problem, sub-problem, algorithm or algorithm-like mechanism, implementation, or support system within an extended EC. **then**
6:       Add i to $PioneerRelationship$
7:    **end if**
8: **end for**
9: **for** each pair of achievements (i, j) where $i \neq j$ **do**
10:    **if** $ProblemsQ[i] = ProblemsQ[j]$ **then**
11:       **if** TimeIntervalsExistOverlap($[TimeStamps\_b[i], TimeStamps\_e[i]], [TimeStamps\_b[j], TimeStamps\_e[j]]$) AND $EC[i] \cap EC[j] \neq \emptyset$ **then**
12:          Add (i, j) to $ParallelRelationship$
13:       **else if** ($TimeStamps\_e[i]$ precedes $TimeStamps\_b[j]$ OR $TimeStamps\_e[j]$ precedes $TimeStamps\_b[i]$) AND ($i \in References[j]$ OR $j \in References[i]$) AND $EC[i] \cap EC[j] \neq \emptyset$ **then**
14:          Add (i, j) to $ProgressiveRelationship$
15:       **end if**
16:    **end if**
17: **end for**
18: **for** each consecutive pair of achievements $(i, i+1)$, sorted by $TimeStamps\_e$ **do**
19:    **if** $ProblemsQ[i] = ProblemsQ[i+1]$ AND $A_i \notin References[A_{i+1}]$ AND $EC[i] \cap EC[i+1] \neq \emptyset$ AND TimeIntervalsNoOverlap($[TimeStamps\_b[i], TimeStamps\_e[i]], [TimeStamps\_b[i+1], TimeStamps\_e[i+1]]$) **then**
20:       Add $(i, i+1)$ to $RelatedButNotConnectedRelationship$
21:    **end if**
22: **end for**
23: **return** $PioneerRelationship, ParallelRelationship, ProgressiveRelationship, RelatedButNotConnectedRelationship$

---

have a chronological order, the later ones did not cite the earlier ones. While we cannot accurately disclose the underlying motivation, we emphasize the factual nature of these implications.

$$C(A_i, A_{i+1}) = 1 \iff (Q(A_i) = Q(A_{i+1}))$$
$$\wedge (EC(A_i) \cap EC(A_j) \neq \emptyset)$$
$$\wedge ([T(A_{i\_b}), T(A_{i\_e})] \cap [T(A_{i+1\_b}), T(A_{i+1\_e})] = \emptyset)$$
$$\wedge ((A_i \notin R(A_{i+1})))$$
(5)

where:

- $Q(a)$ is the key problem domain, sub-problem domain, problem, or sub-problem that achievement a addresses.
- $EC(a)$ denotes the EC involved in achievement a.
- $T(a\_b)$ and $T(a\_e)$ represent the begin time and end time of achievement a, respectively.
- R(a) indicates the references of achievement a. $A_i \notin R(A_{i+1})$ indicates achievement A_i does not in the reference list of achievement A_{i+1}.

Examples: related but not connected relationships trace the sequence of achievements that tackle similar issues across different timeframes. Although these developments may seem interconnected, they often evolve independently.

- **AI:** Condconv [31] and Dynamic Convolution [32] are two contemporary achievements for dynamical models with similar approaches.
- **Benchmarks:** TPC-C [23] and TPC-E [33], both developed to evaluate Online Transactional Processing (OLTP) databases, exemplify a related but not connected relationship. They sequentially advance the field of database benchmarking without direct influence from one another.

Fig. 8 illustrates the interplay among S&T achievements governed by four relationships: pioneering, progressive, parallel, and related but not connected. In S&T evaluatology, formalizing the four relationships is crucial for understanding and analyzing the interaction between various scientific achievements.

3.3.3. The algorithm to identify the four relationships

In this subsection, we present an algorithm designed to discern four significant types of relationships among a myriad of science and technology achievements: pioneering, progressive, parallel, and related but not connected relationships. The algorithm operates on a set of inputs comprising achievement IDs, timestamps, references (key references), evaluation conditions (EC), and the key problem domain, sub-problem domain, problem, or sub-problem Q addressed by each achievement. Subsequently, it outputs sets of achievement pairs categorized into pioneering, progressive, parallel, or related but not connected relationships.

Inputs:

- IDs: A list of achievement IDs or an optional list of pairs of achievement IDs for comparison.
- $TimeStamps\_b$: Timestamps indicating the beginning time of achievements.
- $TimeStamps\_e$: Timestamps indicating the end time of achievements.
- $References$: Key references or citations between achievements.
- EC: The involved EC components of each achievement.
- $ProblemsQ$: A compilation of key problem domain, sub-problem domain, problem, or sub-problem Q addressed by each achievement.

The algorithm proceeds as follows:

1. Identification of Pioneering Relationship:

- achievements that are the first to open up a new research direction by establishing a new field, problem domain, sub-problem domain, problem, sub-problem, algorithm or algorithm-like mechanism, implementation, or support system within an extended EC.

2. Identification of Parallel Relationship:

- achievements addressing the same problem domain, sub-problem domain, problem, or sub-problem are scrutinized.
- achievements occurring within overlapping time intervals are classified as having a parallel relationship.

3. Identification of Progressive Relationship:

- achievements sharing the same problem domain, sub-problem domain, problem, or sub-problem are paired.
- successive temporal order and mutual referencing between achievements, indicate a progressive relationship.

4. Identification of related but not connected Relationship:

- achievements within no-overlapping time intervals are evaluated.
- achievements addressing the same problem domain, sub-problem domain, problem, or sub-problem, without any mutual referencing, are considered to have a related but not connected relationship.

Outputs:

- *Pioneer Relationship*: A set of achievement pairs in a Pioneering relationship.
- *Parallel Relationship*: A set of achievement pairs in a Parallel relationship.
- *Progressive Relationship*: A set of achievement pairs in a Progressive relationship.
- *Related But Not Connected Relationship*: A set of achievement pairs in a related but not connected relationship.

This algorithm offers a systematic approach to unraveling the intricate interplay among S&T achievements, facilitating a deeper understanding of their underlying relationships.

3.4. Establishing the real-world S&T ES

This subsection presents how to model the real-world S&T ES (M_r), as depicted in Fig. 7. The proposed real-world S&T ES encompasses the entire collection of S&T achievements, each of which is mapped onto the several components of an extended EC. As the aim is to single out the top achievements, we ignore the other components of the S&T ecosystem, e.g., the mechanisms and policies within the S&T ecosystems that profoundly shape the evolution of these achievements.

Although this approach can identify all S&T achievements, the real-world S&T ES (M_r) is often susceptible to confounding factors. For instance, the communities tend to favor highly prestigious scientists, naturally drawing more attention to the research outcomes of well-known scientists. This bias stems from a real-world S&T ES (M_r) 's inability to track the developmental trajectory of S&T achievements and elucidate the relationships among these achievements.

To address these deficiencies in the real-world S&T ES (M_r), we will develop the perfect S&T EM (M_p) in Section 3.5, which systematically traces the evolution of S&T achievements and clarifies the interconnections among them.

3.5. Establishing the perfect S&T EM

The core objective of the perfect S&T EM is to track the evolution of S&T achievements. This model aims to capture these achievements' dynamic changes and progressions in terms of four relationships as they contribute to the S&T ecosystem. Doing so provides a full-picture understanding of the evolution of S&T within the real-world context.

Section 3.1 has offered a concise overview of the process for establishing a perfect S&T EM. This subsection will delve into the details,

comprehensively exploring the methodology.

The perfect S&T EM aims to track the evolution of the real-world S&T ES. A perfect S&T EM meticulously tracks the progression of a real-world S&T ES, from ES_i to ES_{i+1}, in a rigorous manner. This process ensures that only one achievement is added from ES_i to ES_{i+1}. This framework allows for an accurate description of the evolution of a field, starting from ES_0 and ultimately culminating in the development of a comprehensive real-world S&T ES.

In this framework, we also provide an auxiliary structure to depict the interconnected relationships among all the achievements. As we progress from ES_0 to ES_1, from ES_i, then to ES_{i+1}, and ultimately towards a real-world S&T ES, we adhere to the principle of adding only one achievement at a time. When a new achievement is introduced in ES_{i+1}, we compare it to its counterpart in ES_i and determine the relationship based on the rules defined in Section 3.3. This approach ensures a systematic and logical evaluation of the evolving achievements within the S&T evaluation framework.

Meanwhile, as discussed in [6,7], the perfect S&T EM also implies exploring and understanding the entire spectrum of possibilities within a research field.

By embracing the concept of a perfect S&T EM, researchers can push the boundaries of knowledge and innovation. It encourages them to explore new avenues, challenge existing assumptions, and uncover hidden potentials. Fig. 7 shows a sample of a perfect S&T EM. The perfect S&T EM has almost entirely replicated the real-world S&T ES. Not only can it establish an extended EC, but it can also organize a roadmap of achievements' evolution by identifying relationships among achievements.

3.6. Establishing the pragmatic EM

Building upon the perfect S&T EM, we can establish the pragmatic evaluation model after filtering out non-significant achievements. The process of filtering out non-significant achievements is essentially the reverse of the process outlined in Section 3.5, which explains how an achievement is added from ES_i to ES_{i+1}. In the filtering process, we employ four rounds of filtering rules.

In the first round, our focus is to identify and filter out non-significant achievements from those that demonstrate progressive relationships. For the achievements that have progressive relationships, as they involve one or several same components of an extended EC, e.g., a problem domain or a problem, we compare achievements under the shared components of the extended EC and filter out those that are not significant.

In the second round, we will identify the achievements that exhibit parallel relationships or related but not connected relationships to the achievements preserved in the first round. Once we have compiled these achievements, we will proceed with an additional filtering process to eliminate any non-significant ones.

We categorize an achievement that exhibits a pioneering relationship as a pioneering achievement. In the third round, we will compare the pioneering achievements under the shared components of the extended EC and filter out achievements that are deemed non-significant.

In the fourth round, we will identify the achievements that exhibit parallel relationships or related but not connected relationships to the pioneering achievements preserved in the third round. Once we have compiled these achievements, we will proceed with an additional filtering process to eliminate any non-significant ones.

According to Zhan [34], an achievement can exert a positive change force over a counterpart by significantly enhancing the simplicity, user experience, cost-effectiveness, efficiency, or other fundamental features by several orders of magnitude. On the other hand, significant deviation from existing technology ecosystems can generate a negative change force. Additionally, when different usage patterns require users to incur significant learning costs, it can also result in a negative change force. This empirical law helps to explain why a certain achievement

dominates over the other one.

In theory, it is possible to quantitatively measure two achievements under the same extended EC from different dimensions, and each dimension is defined as X_i. To summarize these dimensions, we propose a simple rule of thumb. We differentiate between positive and negative signs and sum up the positive or negative values of $lg(X_i)$ (metrics from different dimensions), and the formula is shown in Eq. (6). This approach allows for a holistic assessment of the achievements, taking into account their various dimensions and providing a comprehensive understanding of their overall impact. By considering both positive and negative values, we can gain insights into the strengths and weaknesses of each achievement, enabling a more nuanced evaluation and comparison.

$$V = \sum_i lg(X_i) \tag{6}$$

4. The top N @X @Y methodology

As a typical case study, this section presents how to apply S&T evaluatology.

We propose Top N @X @Y, aiming to recognize the top N achievements within a specific period X in a particular field Y. Here, N represents the number of top achievements, X represents a specific period, and Y represents a particular field.

To optimize the effectiveness of evaluating science and technology, a standardized procedure has been devised as outlined below.

First, during a particular timeframe X, we create a real-world S&T ES that encompasses all achievements. Each achievement is decomposed into various components within the specific extended EC.

Second, based on the real-world S&T ES during a timeframe X, we construct a perfect S&T EM that traces the evolution of S&T achievements in the field of Y according to the four relationships.

Third, considering the total number of achievements (N), we assign different percentages that add up to 100% to the achievements that have pioneering relationships and progressive relationships.

Finally, following the four-round filtering process defined in Section 3.6, we filter out non-significant achievements to establish a pragmatic S&T EM that comprises the top N achievements during a timeframe X in the field of Y. Please note that the final step is iterative.

Following the aforementioned procedures, the top N achievements are obtained and can be presented in a tree form, as depicted in Fig. 6. Subsequently, we can proceed to rank these achievements along with their corresponding contributors and institutions.

We propose a simple rule to score each achievement, with higher scores leading to higher rankings. Initially, each selected achievement is assigned a score of 1.0 points. However, we give an extra score to each pioneering achievement. With each groundbreaking achievement paving the way for new research directions, we aggregate the cumulative scores of progressive achievements by applying a weight, which we call a pioneering weight, to the original score of the pioneering achievement.

Once the scores for each achievement are determined, we proceed to assess the contribution shares of each author and their respective institutions. The specific criteria for assessing the main academic contributors are as follows:

1. If the number of authors is three or fewer, the score is evenly distributed among all authors involved.
2. If there are more than three authors, and their contributions are stated to be equal, the score is evenly divided among all authors.

3. When there are more than three authors and their contributions are not stated as equal, the first author is assigned a first-author ratio, i.e., 0.3. In cases where multiple individuals share the first authorship, the first-author ratio is equally divided among them.

The corresponding author (or the last author in the absence of a designated corresponding author) receives a corresponding author ratio, i.e., 0.3. Similarly, if multiple individuals share the corresponding author role, the corresponding author ratio is evenly distributed among them. The remaining ratio is equally divided among the other authors.

As per the aforementioned rule, the score assigned to each achievement is subsequently distributed among the respective contributors based on their designated ratios. For every contributor, the corresponding institutions (which may be one or multiple) can be determined at the time of their contribution. In cases where a contributor is associated with multiple institutions, the score will be evenly divided among all the affiliated institutions.

5. A case study on the top 100 chip achievements

The chip industry plays a crucial role in driving technological advancements across various sectors, encompassing a vast ecosystem involved in software, hardware, and application development to harness their capabilities. Utilizing S&T evaluatology principles, the International Open Benchmark Council (BenchCouncil) has developed a well-defined extended EC to assess various aspects of chips comprehensively. The first level is the chip field, while the second level encompasses three problem domains: chip design, chip manufacturing, and chip packaging. At the third level, chip design involves several sub-problem domains, including system-level design, logic design, physical design, timing design, verification, and simulation. Chip manufacturing covers semiconductors, materials, and optics. Then, using the Top N @X @Y methodology, BenchCouncil has launched an ambitious initiative to systematically recognize and honor the most 100 groundbreaking and influential achievements in chip technology (Chip100) [8].

The current version of Chip100 uses the Top N @X @Y methodology, where N stands for 100, X spans from the 1940s (the advent of the first computer) to 2023, and Y indicates the chip field and the percentages of pioneering achievements and progressive achievements are 40% and 60%, respectively. For the ranking in Chip100, the pioneering weight is set as 0.2, the first-author ratio is 0.3, and the corresponding author ratio is 0.3.

The major influential accomplishments in chips are encompassed within Chip100. For example, as depicted in Fig. 6, the Instruction Set Architecture (ISA) was first introduced by Frederick Brooks in the 1960s. It defines a crucial sub-problem of computer architecture design (problem) of the system-level design (sub-problem domain) within the chip design problem domain: the challenge of designing the instruction set and proposing effective mechanisms. This concept led to the development of Complex Instruction Set Computers (CISC) and Reduced Instruction Set Computer (RISC). Subsequently, Instruction Set Architectures such as X86 and RISC-V emerged, drawing from the principles of CISC and RISC. This examination provides valuable insights into the connections among these achievements. So, Chip100 identified and evaluated significant achievements and researchers in the chip field that could not be discerned through the application of bibliometrics.

We use the data of Chip100, CSRankings, and the Highly cited Researchers from Elsevier to find the top 100 achievements, contributors, and institutions in the chip field.

CSRankings uses the metric of the number of publications at the top-tier conferences for gauging the academic influence of researchers or their affiliated institutions in computer science. The CSRankings database utilized by us extends across a timeline from 1970 to 2023, representing the most extensive timeframe available for CSRankings. The most matched areas include Computer Architecture and Design Automation.

The Highly Cited Researchers list is the typical metric based on citations. The main criteria for inclusion are "the authorship of multiple Highly Cited Papers™ within the past decade and being ranked

Table 2
Comparing Chip100 against CSRankings and highly cited researchers from Elsevier.

Methods	Top 20 achievements	Top 20 contributors	Top 20 institutions
Chip100 [8] (By the end of 2023)	Von Neumann Architecture, ISA, Stored-program computers, Cache memory, Boolean Algebra, Floating Point Unit, Formal Verification, Out-of-Order Execution, Stream Architecture, Amdahl's Law, Verilog, FPGA, Branch Predictor, CC-NUMA, ECC, EDA, Electrostatic Discharge, Harvard Architecture, Multi-Core Processors, NOC, SIMD Architecture, Single-Chip Multiprocessor, SOC, The Principle of Locality, and Virtual address translation	John von Neumann, Maurice Wilkes, Frederick Brooks, David A. Patterson, Gene Amdahl, George Boole, Robert Tomasulo, William Kahan, Phil Moorby, John L. Hennessy, Aart de Geus, Claude Shannon, Jen-Hsun Huang, John Gustafson, Lisa Su, Mark Hill, Michael J. Flynn, Michel Mardiguian, Richard Hamming Ross H. Freeman, Wayne Wolf, and William M. Johnson	Princeton University, IBM, Univ. of California - Berkeley, University of Cambridge, Stanford University, AMD, Intel, Massachusetts Institute of Technology, NVIDIA, Xilinx, University of Michigan, Gateway Design Automation, ARM, Bell Labs, Georgia Institute of Technology, Google, Harvard University, Motorola, Sandia National Laboratories, Synopsys, University of Paris South, University of Pennsylvania, and University of Washington
CSRankings [2] (By the end of 2023)	Achievements are predicated on the number of publications in top-tier conferences.	David T. Blaauw, Andrew B. Kahng, Srini Devadas, Josep Torrellas, Diana Marculescu, Mark Horowitz, Alberto L. Sangiovanni Vincentelli, Mahmut T. Kandemir, Jason Cong, Yuan Xie, Moinuddin K. Qureshi, Giovanni De Micheli, Sheldon X.D. Tan, Onur Mutlu, David Z. Pan, Yiran Chen, ohsen Imani, Zhiru Zhang, Xiaoyao Liang, and Margaret Martonosi	University of Michigan, Univ. of California - San Diego, Massachusetts Institute of Technology, Univ. of Illinois at Urbana-Champaign, Carnegie Mellon University, Stanford University, Univ. of California - Berkeley, Pennsylvania State University, Univ. of California - Los Angeles, Univ. of California - Santa Barbara, Georgia Institute of Technology, EPFL, Univ. of California - Riverside, ETH Zurich, University of Texas at Austin, Univ. of California - Irvine, Duke University, Shanghai Jiao Tong University, Cornell University, and Princeton University
Highly cited researchers [35] (2023)	Achievements are highly cited papers	There are a total of 98 highly cited researchers in the field of computer science, listed in no particular order.[a]	Chinese Academy of Sciences, Harvard University, Stanford University, National Institutes of Health, Tsinghua University, Massachusetts Institute of Technology, University of California San Diego, University of Pennsylvania, University of Oxford, Max Planck Society, University of California San Francisco, University College London, University of Hong Kong, Washington University, University of California Berkeley, Johns Hopkins University, Memorial Sloan Kettering Cancer Center, University of Cambridge, Yale University, University of California Los Angeles, and University of Washington Seattle (based on the summary of highly cited researchers from all research fields)

[a] https://clarivate.com/highly-cited-researchers/.

in the top 1% based on citations in Web of Science™" [35]. Highly Cited Researchers™ represent a select group comprising only 0.1% of researchers in the world. The data of Highly Cited Researchers utilized by us was released in the year 2023, and hence, the timeframe is from 2013 to 2023, representing the most extensive timeframe available for this database. The matched area is Computer Science, as it cannot be narrowed down to focus solely on the chip field.

Table 2 outlines a compilation of the Top 20 outcomes from Chip100, CSRankings, and the Highly Cited Researchers list published by Elsevier. Throughout the remainder of this section, we will focus on analyzing the top five achievements, contributors, and institutions from various rankings to identify any notable distinctions.

First, we contrast the results of Chip100 with those from CSRankings. From Table 2, we can see that the results are totally different.

According to the analysis conducted by Chip100 (1940s–2023), The top five achievements include the Von Neumann Architecture, ISA, Stored-program computers, Cache memory, and Boolean Algebra. These achievements are crucial in driving the development of chips. Conversely, the achievements in CSRankings are solely based on the volume of publications in top-tier conferences.

Furthermore, the Top five institutions in the chip field encompass Princeton University (recognized for advancements like Von Neumann Architecture, The Principle of Locality, and Virtual address translation), IBM (recognized for advancements like ISA, CISC, Amdahl's Law, and Dennard Scaling Law), UC Berkeley (known for achievements

in Floating Point Unit design, RISC architecture, and RISC-V implementation), University of Cambridge (highlighted for innovations in Stored-program computers, Cache Memory, and Advanced RISC Machines), and Stanford University (acknowledged for progress in MIPS architecture, Superscalar processing, and Single-Chip Multiprocessor development).

In contrast, CSRankings only emphasizes the number of publications at top-tier computer science conferences. In the field of Computer Architecture and Design Automation, covering the period from 1970 to 2023, the top five research institutions include the University of Michigan, University of California-San Diego, Massachusetts Institute of Technology, University of Illinois at Urbana-Champaign, and Carnegie Mellon University.

Among the top five research institutions selected by CSRankings, only the University of Michigan (No. 11), Massachusetts Institute of Technology (No. 8), and Carnegie Mellon University (No. 24) are included within the Chip100 (1940s-2023), while the University of California-San Diego and the University of Illinois at Urbana-Champaign are not featured.

The Top five contributors in Chip100 are John von Neumann (recognized for Von Neumann Architecture), Maurice Wilkes (known for Stored-program computers and Cache Memory mechanism), Frederick Brooks (credited with ISA), David A. Patterson (recognized for the monograph "Computer Architecture: A Quantitative Approach", RISC, and RISC-V), and Gene Amdahl (recognized for CISC and Amdahl's

Law). Contrasting with this viewpoint, the top five chip research contributors according to CSRankings by the end of 2023 are Onur Mutlu (117 contribution papers), Yuan Xie (116 contribution papers), Jason Cong (112 contribution papers), Alberto L. Sangiovanni-Vincentelli (108 contribution papers), and David Z. Pan (104 contribution papers). The noticeable disparity between these rankings is apparent, with none of the top five researchers in CSRankings being featured in the Chip100 list spanning from the 1940s to 2023.

Another well-known ranking is the Highly Cited Researchers published by Elsevier. The achievements are constrained to highly cited papers as viewed through the lens of the Highly Cited Researchers. The top institutions listed in Table 2 are determined based on a roster of highly cited researchers from all research fields. In 2023, a total of 7125 researchers were recognized as Highly Cited Researchers, including 98 in the field of computer science. It is challenging to conduct precise searches for top institutions or researchers within a specific and focused field, such as Chip.

The criteria for this recognition clearly prioritize the impact of papers from a bibliometric perspective, as indicated by their citation counts. As a result, none of the top five contributors listed in the Chip100 have been encompassed in the Highly Cited Researchers list. On the other hand, none of the 98 Highly Cited Researchers in the field of computer science have been included in Chip100 as well.

6. Conclusion

This article systematically reveals three severe bibliometrics limitations in recognizing top science and technology achievements and researchers. To address these shortcomings, we introduce science and technology evaluatology, which exemplifies the application of evaluatology in evaluating science and technology achievements. At the heart of this approach lies the concept of an extended evaluation condition, encompassing nine crucial components. We define four relationships that illustrate the connections among various achievements based on their mapped extended EC components, as well as their temporal and citation links: pioneering, progressive, parallel, and related but not connected. Within a pioneering or progressive relationship under an extended evaluation condition, evaluators can effectively compare these achievements by carefully addressing the influence of confounding variables. The case studies show the effectiveness of the proposed methodology compared with bibliometrics.

CRediT authorship contribution statement

Guoxin Kang: Contributions to the summary of the related work for CSRankings and c-score in Section 2.2, the whole-session discussion, the mathematical formulations of four relationships, the algorithm for identifying the four fundamental relationships in Section 3.3, the benchmark examples in Section 3.3, and the presentations of Figure 1, 2, 3 and Table 1. **Wanling Gao:** Contributions to the summary of the related work for H-index in Section 2.2, the whole-session discussion, partial revision of mathematical formulations and algorithms in Section 3.3, and the presentations of Figure 4, 5, 6, 7. **Lei Wang:** Contributions to the summary of the related work for CiteScore and SNIP in Section 2.2, the whole-session discussion, the writing of Section 4 ,Section 5, the chip examples in 3.3, and the presentations of Figure 5, 6. **Chunjie Luo:** Contributions to the part of SJR in related work, the AI examples in Section 3.3, and the discussion. **Hainan Ye:** Contributions to the data for Chip100 rankings and the discussion. **Qian He:** Contributions to the presentations of Figure 8 and the discussion. **Shaopeng Dai:** Contributions to the discussion. **Jianfeng Zhan:** Contributions to the proposal for the evaluatology-based science and technology evaluation methodology, including the extended EC, four relationships, real-world ES, perfect S&T EM, pragmatic S&T EM, and Top N@X @Y methodology. Contributions to the presentation of most texts, excluding figures and tables. Also contributions to other aspects of the work unless otherwise explicitly stated.

Declaration of competing interest

Jiafeng Zhan is the editor in chef, Lei Wang, Wanling Gao, Chunjie Luo are the founding edtiors of BenchCouncil Transactions on Benchmarks, Standards and Evaluations and were not involved in the editorial review or the decision to publish this article. The other authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] BenchCouncil, BenchCouncil science and technology achievement evaluation, 2023, https://www.benchcouncil.org/evaluation/.

[2] Emery berger. CSRankings, 2023, https://csrankings.org/#/index?all&us.

[3] John P.A. Ioannidis, Jeroen Baas, Richard Klavans, Kevin W. Boyack, A standardized citation metrics author database annotated for scientific field, PLoS Biol. 17 (8) (2019) e3000384.

[4] Jorge E. Hirsch, An index to quantify an individual's scientific research output, Proc. Natl. Acad. Sci. 102 (46) (2005) 16569–16572.

[5] H-index, 2023, https://en.wikipedia.org/wiki/H-index.

[6] Jianfeng Zhan, Lei Wang, Wanling Gao, Hongxiao Li, Chenxi Wang, Yunyou Huang, Yatao Li, Zhengxin Yang, Guoxin Kang, Chunjie Luo, et al., Evaluatology: The science and engineering of evaluation, BenchCounc. Trans. Benchmark. Stand. Eval. 4 (1) (2024) 100162.

[7] Jianfeng Zhan, A short summary of evaluatology, BenchCounc. Trans. Benchmark. Stand. Eval. 4 (2) (2024).

[8] Top 100 chips achievements, 2023, https://www.benchcouncil.org/evaluation/.

[9] Digital Bibliography & Library Project, University of Trier, 2023, https://dblp.org/.

[10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, Illia Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).

[11] Accepted papers of CVPR 2022, 2022, https://cvpr2022.thecvf.com/accepted-papers.

[12] Accepted papers of MICRO 2022, 2022, https://microarch.org/micro55/index.php.

[13] Chris James, Lisa Colledge, Wim Meester, Norman Azoulay, Andrew Plume, Citescore Metr.: Creat. J. Metr. Scopus Cit. Index (2018) arXiv preprint arXiv: 1812.06871.

[14] Henk F. Moed, Measuring contextual citation impact of scientific journals, J. Informetr. 4 (3) (2010) 265–277.

[15] Kevin W. Boyack, Mapping knowledge domains: Characterizing PNAS, Proc. Natl. Acad. Sci. 101 (Suppl_1) (2004) 5192–5199.

[16] Warren S. McCulloch, Walter Pitts, A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biophys. 5 (1943) 115–133.

[17] Yann LeCun, Bernhard Boser, John S. Denker, Donnie Henderson, Richard E. Howard, Wayne Hubbard, Lawrence D. Jackel, Backpropagation applied to handwritten zip code recognition, Neural Comput. 1 (4) (1989) 541–551.

[18] Yann LeCun, Léon Bottou, Yoshua Bengio, Patrick Haffner, Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324.

[19] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst. 25 (2012).

[20] Xianyi Zhang, Qian Wang, Yunquan Zhang, OpenBLAS: a high performance blas library on loongson 3a cpu, J. Softw. 22 (Zk2) (2012) 208–216.

[21] Kazushige Goto, Robert A. van de Geijn, Anatomy of high-performance matrix multiplication, ACM Trans. Math. Softw. 34 (3) (2008) 1–25.

[22] Richard Cole, Florian Funke, Leo Giakoumakis, Wey Guy, Alfons Kemper, Stefan Krompass, Harumi Kuno, Raghunath Nambiar, Thomas Neumann, Meikel Poess, et al., The mixed workload CH-benchmark, in: Proceedings of the Fourth International Workshop on Testing Database Systems, 2011, pp. 1–6.

[23] TPC-c benchmark, 2010, http://www.tpc.org/tpcc/.

[24] TPC-H benchmark, 2010, http://www.tpc.org/tpch/.

[25] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018, arXiv preprint arXiv:1810.04805.

[26] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, Improving language understanding by generative pre-training, OpenAI (2018).

[27] Ubuntu. https://ubuntu.com/, Initial release in 2004.

[28] CentOS. https://www.centos.org/, Initial release in 2004.

[29] Lei Wang, Jianfeng Zhan, Chunjie Luo, Yuqing Zhu, Qiang Yang, Yongqiang He, Wanling Gao, Zhen Jia, Yingjie Shi, Shujie Zhang, et al., Bigdatabench: A big data benchmark suite from internet services, in: 2014 IEEE 20th International Symposium on High Performance Computer Architecture, HPCA, 2014, pp. 488–499.

[30] Ahmad Ghazal, Tilmann Rabl, Minqing Hu, Francois Raab, Meikel Poess, Alain Crolotte, Hans-Arno Jacobsen, Bigbench: Towards an industry standard benchmark for big data analytics, in: Proceedings of the 2013 ACM SIGMOD International Conference on Management of Data, 2013, pp. 1197–1208.

[31] Brandon Yang, Gabriel Bender, Quoc.V Le, Jiquan Ngiam, Condconv: Conditionally parameterized convolutions for efficient inference, Adv. Neural Inf. Process. Syst. 32 (2019).

[32] Yinpeng Chen, Xiyang Dai, Mengchen Liu, Dongdong Chen, Lu Yuan, Zicheng Liu, Dynamic convolution: Attention over convolution kernels, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11030–11039.

[33] Tpc Benchmark™ E. Citeseer, Transaction Processing Performance Council, 2010.

[34] Jianfeng Zhan, Three laws of technology rise or fall, BenchCounc. Trans. Benchmark. Stand. Eval. 2 (1) (2022) 100034.

[35] Highly cited researchers, 2023, https://clarivate.com/highly-cited-researchers/.