A Review of Full Counting and Fractional Counting Methods in Scientometric Research: Postprint

Authors: Chen Liyue, Yang Liying, Ding Jielan

Date: 2023-08-27T00:00:00+00:00

Abstract

[Purpose/Significance] This study defines relevant concepts of counting methods in scientometric research, constructs a classification system for counting methods, organizes and compares the characteristics and differences among counting methods, analyzes existing problems, and proposes directions for future improvement and recommendations for selecting counting methods. [Method/Process] First, it summarizes the components and usage processes of counting methods; proposes two elements for classifying counting methods from the perspective of credit allocation; divides counting methods into two major categories—full counting methods and fractional counting methods—and provides an overview of each method; taking the equal-weight algorithms of full counting and fractional counting—full counts and fractional counts—as examples, it compares the differences among counting methods from three perspectives: paper indicators, citation indicators, and network indicators. [Results/Conclusion] The article reflects on four aspects: the advantages and disadvantages of full counting and fractional counting methods, the consistency between counting units and counting objects, the rationality of credit allocation rules, and network influence measurement, and points out directions for further research in these four aspects in the future.

Full Text

Preamble

Vol. 62 No. 23 December 2018
ChinaXiv Collaborative Journal

A Review of Full Counting and Fractional Counting Methods in Scientometric Research

Chen Liyue[1],[2], **Yang Liying**[1], **Ding Jielan**[1],[2]
[1] National Science Library, Chinese Academy of Sciences, Beijing 100190
[2] University of Chinese Academy of Sciences, Beijing 100049

Abstract

[**Purpose/Significance**] This paper systematically defines the concepts related to counting methods in scientometric research, establishes a classification system for counting methods, compares the characteristics and differences among various counting methods, analyzes existing problems, and proposes future directions for improvement and recommendations for method selection. [**Method/Process**] First, we generalize the components and operational processes of counting methods, and propose two classification factors from the perspective of credit allocation, dividing counting methods into full counting and fractional counting categories, with an overview of each method. Taking the equal-weight algorithms of full and fractional counting—full counts and fractional counts—as examples, we compare differences across three perspectives: publication indicators, citation indicators, and network indicators. [**Result/Conclusion**] The paper reflects on four aspects: the advantages and disadvantages of full versus fractional counting, consistency between counting units and counting objects, the rationality of credit allocation rules, and network impact measurement. We identify future research directions in these four areas.

Keywords: counting methods; full counting method; fractional counting method

DOI: 10.13266/j.issn.0252-3116.2018.23.016

Research papers serve as the primary vehicle for scientific research outputs and play a crucial role in advancing and disseminating scientific knowledge. They represent both the intellectual property and knowledge rights of authors. In scientometric research, counting methods refer to computational approaches that allocate the attribution of research papers according to specific rules. These methods are used to calculate publication counts and citation frequencies for authors, institutions, and countries, and form the basis for other scientometric indicators. In the era of big data, the scientometric community has recognized that different counting methods possess distinct characteristics and advantages for addressing various problems. The growth of scientific collaboration has made the impact of counting methods on research conclusions increasingly significant. Existing studies lack comprehensive theoretical and practical analysis of counting methods. To fundamentally differentiate the design rationale, practical applications, and suitability of various counting methods, this paper takes the two foundational categories—full counting and fractional counting—as a starting point to define relevant concepts, construct a classification system, and compare the features and differences among common counting methods, aiming to provide references for method selection.

In research evaluation practice, quantitative data are often indispensable and support evaluation conclusions. Counting methods serve as the starting point and basis for allocating research output attribution, such as calculating publication counts, and directly affect the measurement of researcher contributions in evaluation processes. Therefore, research on counting method algorithms and their applicability holds important practical significance.

2 Overview of Counting Methods

Defining the connotation of various counting methods is a prerequisite for analyzing their principle differences. This section defines the basic concepts and components of counting methods, constructs a counting method system, and introduces various methods within this framework.

2.1 Conceptual Foundation of Counting Methods

To date, more than ten counting methods have been developed, most of which are related to author numbers and whose results are significantly affected by the degree of research collaboration. In the era of big science, collaboration has become the dominant form of experimental and theoretical research, with the average number of participants per study growing rapidly and the average number of authors per paper increasing continuously. Based on the computational process of counting methods, this paper defines concepts including counting unit, credit, credit allocation function, and counting object.

2.1.1 Counting Unit The counting unit is the basic unit of calculation in counting methods. This study defines the counting unit as the contributor (or supporter) of a paper. In this paper, paper contributors are not limited to authors but include participants or supporters at different levels, such as authors' institutions and countries. Therefore, for a given paper, counting units can be authors, institutions, countries, etc., with the author being the smallest counting unit.

2.1.2 Credit Counting units such as authors, their institutions, and affiliated countries are the contributors (or supporters) of academic achievements. While completing and supporting research work, they also share the prestige brought by academic achievements. Credit refers to the academic reputation that these counting units obtain by sharing and allocating the various rights of papers. In scientometric research, credit refers to the number of papers completed by counting units, citation counts, and other metrics calculated on this basis, such as authors' publication counts and received citations.

2.1.3 Credit Allocation Function The credit allocation function is the basis for counting units to obtain credit. For a paper involving multiple contributing entities (authors, institutions, countries), it is necessary to determine how each counting unit shares or allocates the rights (credit) of the paper. The design of credit allocation functions currently considers two factors:

Factor 1: The weight assigned to each counting unit (author, institution, country) during credit allocation. Since different participants in academic paper completion may have different contributions—for example, when counting units are authors, there are corresponding authors, first authors, and other authors—the credit allocation function assigns weights to each counting unit based on contribution levels.

Factor 2: Defining the credit value of a paper as 1 or defining the credit value of each counting unit as 1. When counting units are authors and a paper has N authors, under equal weight allocation, if a paper's credit value is defined as 1, each author's credit value is 1/N; if each author's credit value is defined as 1, a paper's credit value is N.

Since the credit allocation function determines credit distribution, it is the core of counting methods.

2.1.4 Counting Object Counting objects refer to the evaluation targets in research assessment, including authors, institutions, countries, disciplines, regions, journals, etc., with authors, institutions, and countries being the most common. In research evaluation practice, since counting objects are the evaluation targets, their credit value is the ultimate goal of counting methods, while the credit value of counting units forms the basis for calculating counting objects' credit. The relationship between counting objects and counting units' credit is: the credit value of a counting object equals the sum of credit values of counting units belonging to that object. For example, when the counting object is a country, its credit value can be the sum of credit values of counting units at different levels—country, institution, and author. The distinction between counting objects and counting units will be explained in detail with examples in Section 4.

2.1.5 Relationships Among Concepts Based on the above concepts, Figure 1 [Figure 1: see original paper] illustrates the relationships among counting unit, credit, credit allocation function, and counting object. As shown in Figure 1, after determining counting units, the credit allocation function first calculates each unit's credit; then, based on the affiliation relationship between counting units and counting objects (e.g., authors affiliated with institutions, institutions affiliated with countries), the credit values of counting objects are obtained by summing the credit values of their affiliated counting units.

2.2 Classification of Counting Methods

Researchers in scientometrics have proposed various counting methods from different perspectives. Organizing these methods into a classification system helps reveal the connections and differences among them from both common and individual perspectives. Some scholars have proposed different classification approaches. M. Gauffriau et al. [3] classified counting methods into five categories based on the correspondence between counting objects and counting units in various measurement problems: "complete," "complete-normalized," "whole," "whole-normalized," and "straight," where "whole" refers to full counts [4] and those with "normalized" refer to fractional counts [5], while "straight" represents first-author and corresponding-author counting methods [6]. L. Waltman et al. [7] subdivided counting methods by counting unit into seven types: "full counts," "country-level fractional counts," "organization-level fractional counts," "address-level fractional counts," "author-level fractional counts," "first-author counting," and "corresponding-author counting."

Through analysis, we found that existing classification systems mostly adopt perspectives of counting unit and counting object. Among counting method concepts, the credit allocation function is the basis for calculating credit, which is the ultimate goal of counting methods. Therefore, this study defines a classification system from the perspective of credit allocation, based on the two factors considered in credit allocation.

First, based on whether a paper's credit value is defined as 1 or each counting unit's credit value is defined as 1, counting methods are divided into two major categories: fractional counting and full counting. The conceptual difference between the two methods is that the former divides while the latter aggregates. Fractional counting divides a paper's credit value of 1 among counting units, while full counting aggregates counting units' credit values (each being 1) to constitute a paper's credit value.

On the basis of Factor 1, Factor 2 considers the weight of each counting unit (authors, institutions, countries) in credit allocation. In scientometric practice, the determination of weights for author counting units is related to author order, and weights can be either equal or unequal [8-12]. Factor 2 further subdivides the two categories into four types.

Equal weight among counting units means that all authors of a paper have equal contributions, and each counting unit has equal credit value. In full and fractional counting methods, the current equal-weight methods are full counts and fractional counts, respectively.

Unequal weight among counting units means that authors have different contributions, and author order determines the weight of credit allocation. In full counting, first-author and corresponding-author counting methods are most widely used. In fractional counting, there are seven types of unequal-weight methods. The classification system is shown in Figure 2 [Figure 2: see original paper].

2.3 Common Counting Methods

Based on counting method concepts and our classification system, this section introduces common counting methods using paper indicator calculation as an example. Considering that some methods' original English names are widely accepted, we retain them to avoid confusion.

(1) Full counts (also called normal counts, total counts, standard counts, whole counts, etc.): Full counts is the only equal-weight algorithm for full counting and the most widely used method, first proposed by D. Lindsey [4]. This method assigns each counting unit a credit score of 1.

(2) First-author counting and Corresponding-author counting: These two methods are unequal-weight algorithms of full counting, proposed by J.R. Cole et al. [6]. First-author counting assigns a credit value of 1 to the first author and 0 to others; corresponding-author counting assigns 1 to the corresponding author and 0 to others. Both methods emphasize the outstanding contributions of corresponding authors and first authors.

(3) Fractional counts (also called adjust counts): Proposed by D. Lindsey and D.D.S. Price [5], this is the only equal-weight algorithm for fractional counting. Its credit allocation function (CA) is:

$$CA_{u_j} = \frac{1}{n}$$

where u refers to the counting unit (author, institution, or country), u_j refers to author j, CA_{u_j} is the credit score of author j, n is the number of authors, and j is the author index, $j = (1, 2, \dots, n)$.

(4) Proportional counts: This is one of the unequal-weight algorithms for fractional counting, proposed by G.V. Hooydonk [8]. The credit allocation function is based on author order, decreasing sequentially. The contribution of the author in position j is represented by $n + 1 - j$, and the total contribution of all authors is $n(n + 1)/2$. The credit allocation function is:

$$CA_{u_j} = \frac{2(n + 1 - j)}{n(n + 1)}$$

where u refers to the counting unit (author), u_j refers to author j, CA_{u_j} is the credit score of author j, n is the number of authors, and j is the author index, $j = (1, 2, \dots, n)$.

(5) Geometric counts (pure geometric counts): This is another unequal-weight algorithm for fractional counting, proposed by L. Egghe and R. Rousseau et al. [9]. The credit allocation function decreases sequentially based on author order, using 2^{n-j} to represent the contribution of the author in position j, with total contribution $2^n - 1$. The credit allocation function is:

$$CA_{u_j} = \frac{2^{n-j}}{2^n - 1}$$

where u refers to the counting unit (author), u_j refers to author j, CA_{u_j} is the credit score of author j, n is the number of authors, and j is the author index, $j = (1, 2, \dots, n)$.

(6) Harmonic counts: This is another unequal-weight algorithm for fractional counting, proposed by N.T. Hagen [10]. The credit allocation function decreases based on author order, using $1/i$ to represent the contribution of the author in position j:

$$CA_{u_j} = \frac{1}{j}$$

where u refers to the counting unit (author), u_j refers to author j, CA_{u_j} is the credit score of author j, n is the number of authors, and j is the author index, $j = (1, 2, \dots, n)$.

(7) Correct credit distribution scores: This is another unequal-weight algorithm for fractional counting, proposed by I. Lukovits et al. [11]. This method incorporates author screening parameters, assuming authors receive credit only if their contribution exceeds a certain proportion:

$$CA_{u_1} = \frac{2F \cdot T}{n}$$

$$CA_{u_k} = \frac{2k \cdot F \cdot T}{n}$$

where u refers to the counting unit (author), u_1 refers to author 1, CA_{u_1} is the credit score of author 1, u_k refers to author k $(k = 2, 3, \dots, n)$, CA_{u_k} is the credit score of author k, n is the number of authors, and F is calculated as:

$$F = \frac{1 - n_T}{100}$$

where n_T = author contribution threshold (%, e.g., only contributions exceeding 5% or 10% of an article are counted).

(8) Exponential weighting: This is another unequal-weight algorithm for fractional counting, proposed by P. Vinkler [12]. This method was developed because research shows that in papers with five or more authors, the last author typically contributes more on average. Based on this principle, the last author is assigned relatively more credit:

$$CA_{u_j} = \frac{2Z_j}{\sum_{j=1}^{n} Z_j}$$

where u refers to the counting unit (author), u_j refers to author j, CA_{u_j} is the credit score of author j, n is the number of authors, j is the author index, $j = (1, 2, \dots, n)$, and Z_j represents the exponential weight of each author, calculated as:

$$Z_j = b^{j-1}$$

where $b = 0.8$.

(9) Modified weight/Refined weight methods: These are unequal-weight algorithms for fractional counting, proposed by F.J. Trueba et al. [13]. Both modified and refined weight methods assign more credit to the last author, as the last author is typically the research team leader. The modified weight method's credit allocation function is:

$$CA_{u_j} = \frac{2(2n - j + 2)}{3n(n + 1)}$$

The refined weight method's credit allocation function is:

$$CA_{u_j} = \frac{2(2n - j + 2)}{3n(n + 1)}(1 - f) + c_j \cdot f$$

where u refers to the counting unit (author), u_j refers to author j, CA_{u_j} is the credit score of author j, n is the number of authors, j is the author index, $j = (1, 2, \dots, n)$, f represents the share of credit allocated to priority authors (first, second, and last authors, with $0 < f < 1$), and c_j represents the number of priority authors with $\sum c_j = 1$.

2.4 Example of Counting Methods

The previous section introduced the design rationale and algorithms of various counting methods. This section demonstrates the principles of counting methods through examples while further explaining the relationship between counting objects and counting units.

This example uses a paper co-authored by six authors from four different institutions and three different countries, with author A1 as the first author. The goal is to calculate each country’s credit value, making the counting object the country. The relationships among authors, institutions, and countries are shown in Table 1 .

Table 1 Relationships among authors, institutions, and countries

This section selects one representative method from each of the four types in the classification system: full counts, first-author counting, fractional counts, and proportional counts for comparison. Table 2 shows the country credit values calculated based on these four methods when the counting unit is the country.

Table 2 Country credit values based on four counting methods

Note: In Table 2, each method uses country as the counting unit

3 Comparative Study of Full and Fractional Counting

Full counting and fractional counting are the two major categories of counting methods. Clarifying their essential characteristics helps fundamentally understand the applicability of counting methods. We compare the equal-weight algorithms of full and fractional counting—full counts and fractional counts—to examine differences between the two method types.

In Section 2.1, we defined concepts related to counting methods, among which counting unit and counting object are the foundation and ultimate goal of credit calculation. To further clarify their relationship, Table 3 lists the differences in credit values obtained by counting objects when counting units are country, institution, and author, showing country credit values calculated using full counts and fractional counts with different counting units.

Table 3 Country credit values when counting units are country, institution, and author

When the counting unit is country, under full counts, the credit values of the three countries are all 1; under fractional counts, each country receives 1/3 = 0.33 credit. When the counting unit is institution, under full counts, the four institutions each have a credit value of 1. Based on the affiliation relationships in Table 1, Country 1’s credit value is 1\$$\times 2 = 2, whileCountries2and3eachhaveacreditvalueof1.Underfractionalcounts, thefourinstitutionseachreceive1/ 0.25credit.Basedoninstitutionalaffiliations, Country1receives1/2 = 0.5credit, whileCountries2and3eachreceive1/4 = 0.25credit.Whenthecountingunitisauthor, underfullcounts countryaffiliationsinTable1, Country1'screditvalueis1\times 3 = 3, Country2'sis1\times 2 = 2, andCountry3'sis1.Underfractionalcounts, eachofthesixauthorsreceives1/6 = 0.17credit.Basedonauthor - countryaffiliations, Country1receives3\times 1/6 = 0.5credit, Country2receives2\times\$1/6 = 0.33$ credit, and Country 3 receives 1/6 = 0.17 credit. As shown in Table 3, as the granularity of counting units becomes finer—from country to institution to author—the differences in country credit values become more significant. The finer the granularity of counting units, the more accurately they reflect the actual contributions of counting objects, and the more precise the credit allocation.

Current research on the differences between full counts and fractional counts covers many perspectives. As mentioned earlier, counting methods can be used to calculate publication counts, citation counts, and other metrics for specified objects. Additionally, we find that counting methods also affect bibliometric network analysis. Therefore, this chapter compares full counts and fractional counts from three perspectives: publication indicators, citation indicators, and network indicators.

3.1 Comparison of Counting Methods in Publication Indicator Calculation

In scientometric research, publication counts and indicators based on publication volume are the most commonly used metrics. Current studies primarily rely on large-scale empirical analysis and objective facts to compare full counts and fractional counts, focusing on two aspects: the applicability of counting methods and the correlation of counting results.

3.1.1 Applicability of Counting Methods The applicability of counting methods refers to how different methods suit different counting objects and affect credit calculation conclusions. Leydesdorff [14], Braun [15], and Martin [16] discussed this issue in the context of calculating national publication world share indicators. Leydesdorff argued that since fractional counts assign more credit to single-author paper authors than to co-authors, increased international collaboration could paradoxically decrease a country's publication output, making fractional counts defective for this indicator. Braun et al. believed that full counts inflate overall publication numbers and are unsuitable, while fractional counts provide more accurate results. Other scholars have compared method applicability through national publication rankings. Gauffriau [17-18] used full counts and fractional counts to calculate national publication numbers and rankings, finding that full counts ranked some countries with weak research strength but extensive collaboration ahead of scientifically strong but less collaborative countries like Japan, concluding that full counts produce inaccurate results. Moed [19] demonstrated biases in full counts by comparing domestic publication counts and per-researcher publication rates, noting that while such biases are measurable at the country level, they are difficult to quantify at institutional and individual levels, thus recommending multiple counting methods.

3.1.2 Correlation of Counting Results Correlation studies compare credit calculation results from different methods like full counts and fractional counts, analyzing differences through correlation analysis. Pritychenko [20] constructed average publication trends for authors in nuclear physics based on both methods, finding negative correlations and that only fractional counts aligned with Plume et al.'s [21] described trends. Chudlarsky et al. [22] examined Czech universities and research institutions, finding that fractional counts produced similar results for evaluating both individual institutions and the total system, while full counts showed large disparities across different organizational levels.

3.2 Comparison of Counting Methods in Citation Indicator Calculation

Citation indicators in scientometrics measure impact based on citation data. Research on counting method selection for citation indicators can be viewed from single-discipline and cross-disciplinary perspectives.

3.2.1 Application in Single-Discipline Citation Measurement Narin [23] and Glänzel [24] noted that citation counts are more prone to errors and disputes than publication counts in comparative studies, as multi-author papers typically receive more citations than single-author papers, and internationally co-authored papers receive more citations than domestic ones. Rinia [25], Huang [26], and Lin [27] studied full counts, fractional counts, and first-author counting in physics. Huang et al. used universities as counting objects, while Lin et al. used countries, both examining publication counts and citations per paper. Their results showed that fractional counts and first-author counting were more scientifically accurate than full counts for evaluating impact at country and institution levels. Schreiber [28] and Egghe [29] used fractional counts for both papers and citations to calculate h-index and g-index, finding non-significant differences and concluding these indices are robust to counting methods, though cross-field comparisons were recommended.

3.2.2 Application in Cross-Disciplinary Citation Measurement Recent studies have found significant differences among counting methods in cross-disciplinary impact measurement. Aksnes et al. [30] used full counts and fractional counts to calculate the National Science Indicator (NSI), finding that fractional count-based indicators were generally lower than full count-based ones, and that differences correlated strongly with international collaboration proportions, suggesting countries with higher international collaboration appeared more competitive under full counts. Moya-Anegón et al. [31] calculated field-normalized indicators (MNCS) using full counts, fractional counts, and corresponding-author counting, proposing that differences between full counts and corresponding-author counting could reflect academic benefits from international collaboration—differences below 25% indicated scientifically advanced countries, while those above 40% indicated developing countries.

In 2012, CWTS first identified that full counts could not guarantee a baseline of 1 for the field-normalized indicator MNCS when releasing the Leiden Ranking report [33]. Subsequently, Waltman [7] et al. explored the applicability of full counts and fractional counts to normalized citation indicators like MNCS and PPtop10% [34], proposing that full counts produce a “full counting bonus” (FCB) and recommending fractional counts for field normalization, especially for countries or institutions. Perianes-Rodríguez et al. [35], building on Abbas et al. [36] and Foster et al. [37], proposed the additive decomposability of impact indicators like MNCS and proved full counts unsuitable. They introduced multiplicative counts, which assigns credit values of 1 to counting units but differs from full counts in calculating normalized indicators, and demonstrated it satisfies additive decomposability while maintaining a baseline of 1 for MNCS.

3.3 Counting Methods and Bibliometric Network Analysis

Bibliometric networks form the foundation of extensive analytical work. Credit calculation is fundamental to network construction, making counting method selection a prerequisite for network analysis. Specifically, the edge weights between nodes in a network—that is, the connection strength between nodes—depend on credit calculation. Using author collaboration networks as an example, assume a paper is co-authored by four authors. With network nodes representing the four authors, Figure 3 [Figure 3: see original paper] shows the collaboration strength (edge weights) calculated using full counts and fractional counts.

In the full counts collaboration network, all edges between authors have weight 1, with total edge weight sum of 6. In the fractional counts network, the total edge weight sum is 1. Clearly, when using fractional counts to calculate author collaboration weights, collaboration in different papers is equally weighted; under full counts, the total weight of edges in different collaboration scenarios is affected by the number of collaborators—the more collaborators, the greater the total weight.

Current research on counting methods and bibliometric networks focuses on two directions: optimizing network layout algorithms and developing fractional counts derivatives.

3.3.1 Optimization of Network Layout Algorithms Many scholars have proposed that weighted co-occurrence frequencies affect network layout and node clustering. Eck and Waltman [38] incorporated counting methods into the VOS network layout technique in VOSviewer [39] software, with the layout algorithm formula:

$$L(x_1, \dots, x_n) = \sum_{i<j} a_{ij} \|x_i - x_j\|^{\alpha} - \sum_{i<j} \|x_i - x_j\|^{\beta}$$

where a_{ij} represents the relatedness between nodes i and j, primarily calculated based on Association Strength (AS):

$$AS_{ij} = \frac{c_{ij}}{c_i c_j}$$

where c_{ij} is the co-occurrence frequency between nodes i and j, c_i is the total co-occurrence frequency of node i, and c_j is the total co-occurrence frequency of node j.

Different counting methods change the association strength between nodes, thereby altering the final network layout. Rodríguez et al. [40] used VOSviewer to construct collaboration networks of 750 universities from the 2015 Leiden Ranking based on full counts and fractional counts, finding that fractional counts networks more easily identified clusters than full counts networks, and that full counts networks were more influenced by papers with relatively large numbers of co-authors.

3.3.2 Fractional Counts and Its Derivatives In bibliometric network strength calculation, besides fractional counts shown in Figure 4 [Figure 4: see original paper], A.P. Rodríguez and W.P. Han [41] and Leydesdorff et al. [42] proposed several fractional counts derivatives based on different methodologies. Using author collaboration networks as an example, we summarize fractional counts and its derivatives.

Assume there are M papers including N authors, constructing an $N \times M$ author-paper co-occurrence matrix A. The number of authors in paper k is n_k:

$$n_k = \sum_{i=1}^{N} a_{ik}$$

where a_{ik} is an element in matrix A with value range $\{0, 1\}$, and k represents the kth paper.

Under full counts, the author collaboration matrix U is:

$$U = AA^T$$

with elements:

$$u_{ij} = \sum_{k=1}^{M} a_{ik} a_{jk}$$

Under fractional counts and its derivatives, elements u'_{ij} in author collaboration matrix U have several forms:

(1) Based on the number of author co-occurrence relationships in a single paper:

$$u'_{ij} = \sum_{k=1}^{M} a_{ik} a_{jk} \cdot \frac{2}{n_k(n_k - 1)}$$

where n_k is the number of authors in paper k, and a_{ik}, a_{jk} are elements in matrix A. This method assigns weight $2/n_k(n_k - 1)$ to each element, representing the number of co-occurrence relationships among all authors in paper k.

(2) Based on the number of co-occurrence relationships between any author and others in a single paper:

$$u'_{ij} = \sum_{k=1}^{M} a_{ik}a_{jk} \cdot \frac{1}{n_k - 1}$$

This method assigns weight $1/(n_k - 1)$, where $n_k - 1$ represents the number of co-occurrence relationships between any author and other authors in paper k.

(3) Based on fractionalizing the author-paper co-occurrence matrix by author count:

$$u'_{ij} = \sum_{k=1}^{M} \frac{a_{ik}}{n_k} \cdot \frac{a_{jk}}{n_k}$$

This method assigns weight $1/n_k$ to author-paper co-occurrence relationships when constructing matrix A.

(4) Based on single-paper author count:

$$u'_{ij} = \sum_{k=1}^{M} \frac{a_{ik}a_{jk}}{n_k}$$

where n_k is the number of authors in paper k, and a_{ik}, a_{jk} are elements in matrix A.

4 Discussion and Reflection

This paper introduces and proposes concepts including counting unit, credit, credit allocation function, and counting object to clarify the essential differences among counting methods. It designs and constructs a classification system to help researchers examine the commonalities and individualities of different methods from a holistic perspective. In reviewing application studies, we distinguish between counting units and counting objects and examine how different levels of counting units affect calculation results, aiming to draw the scientometric community's attention to this issue. Additionally, we introduce the rules and application effects of counting methods in citation calculation and network indicator calculation to provide references for quantitative evaluation.

During this research, we identified several issues worth attention or further study:

4.1 Advantages and Disadvantages of Full and Fractional Counting

Counting method research began in the 1980s and has become a prominent topic in scientometrics with the changing data environment. Debates about the merits of various methods frequently arise. In fact, different counting methods have distinct characteristics and applicability for different evaluation scenarios. Each method has its own premises and assumptions; no method is absolutely superior or inferior—only more or less suitable for specific analytical contexts. Moreover, the widespread use of full counts is partly limited by objective data environment constraints. When using counting methods, researchers should select appropriate methods based on analysis objectives and premises to avoid misjudging conditions and methods.

4.2 Consistency Between Counting Unit and Counting Object

This study defines counting unit and counting object and uses cases to reveal their differences and connections, comparing how different counting units affect credit calculation for the same counting object. This represents a gap not previously mentioned or noted in earlier research. Through case analysis, we observe that as counting unit granularity becomes finer—from country to institution to author—differences in counting objects' credit values become more significant. Theoretically, finer granularity more comprehensively reflects actual contributions and enables more precise credit allocation. Empirical research in this area awaits future validation with large-sample cases.

4.3 Rationality of Credit Allocation Rules

As the core of counting methods, credit allocation function design considers two factors: contribution division versus contribution aggregation, and equal versus unequal weighting of counting units. Both factors have rationality from different perspectives. For example, in some disciplines (e.g., high-energy physics, mathematics), author order is based entirely on alphabetical order, making contribution differences indistinguishable. While corresponding authors' contributions are undeniable, how can contribution shares be quantified? Without reasonable quantification basis, equal weighting may be the best choice by default.

4.4 Impact on Network Measurement Indicators

In research on counting methods and bibliometric networks, some scholars have found that fractional counts more easily identify clusters in collaboration networks and more accurately determine core nodes in coupling networks than full counts. However, the difference between fractional counts and full counts in this regard cannot yet be quantitatively compared. This issue resembles the comparison of cluster identification capabilities among layout algorithms in social network visualization tools and requires further exploration of quantitative methods to measure counting methods' impact on network analysis.

Acknowledgments

We thank members of the Scientometrics Center at the National Science Library of the Chinese Academy of Sciences—Yue Ting, Chen Fuyou, Zhai Yanqi, Tong Sichao, and Shi Shuangqing—and doctoral supervisor Liu Xiwen for valuable suggestions on the conception and revision of this paper. We also express sincere gratitude to the experts who reviewed this paper and provided valuable comments.

References

[1] WUCHTY S, JONES B F, UZZI B. The increasing dominance of teams in production of knowledge[J]. Science, 2007, 316(5827): 1036-1039.

[2] AAD G, ABBOTT B, ABDALLAH J, et al. Combined measurement of the Higgs Boson Mass in pp Collisions at sqrt[s]=7 and 8 TeV with the ATLAS and CMS Experiments[J]. Physical review letters, 2015, 114(19): 1362-1367.

[3] GAUFFRIAU M, LARSEN P O, MAYE I, et al. Publication, cooperation and productivity measures in scientific research[J]. Scientometrics, 2007, 73(2): 175-214.

[4] LINDSEY D. Production and citation measures in the sociology of science: the problem of multiple authorship[J]. Social studies of science, 1980, 10(2): 145-162.

[5] PRICE D D S. Multiple authorship[J]. Science, 1981, 212(4498): 986-986.

[6] COLE J R, COLE S, BEAVER D D. Social stratification in science[J]. American journal of sociology, 1977, 42(83): 923-945.

[7] WALTMAN L, ECK N J V. Field-normalized citation impact indicators and the choice of an appropriate counting method[J]. Journal of informetrics, 2015, 9(4): 872-894.

[8] HOOYDONK G V. Fractional counting of multiauthored publications[J]. Scientometrics, 2001, 51(1): 69-115.

[9] EGGHE L, ROUSSEAU R, HOOYDONK G V. Methods for accrediting publications to authors or countries: consequences for evaluation studies[J]. Journal of the American Society for Information Science and Technology, 2000, 51(2): 145-157.

[10] HAGEN N T. Harmonic allocation of authorship credit: source-level correction of bibliometric bias assures accurate publication and citation analysis[J]. PloS one, 2007, 3(12): 4021-4021.

[11] LUKOVITS I, VINKLER P. Correct credit distribution: a model for sharing credit among coauthors[J]. Social indicators research, 1995, 36(1): 91-98.

[12] VINKLER P. Research contribution, authorship and team cooperation[J]. Journal of the American Society for Information Science and Technology, 2010,

61(12): 2405-2416.

[13] TRUEBA F J, GUERRERO H. A robust formula to credit authors for their publications[J]. Scientometrics, 2004, 60(2): 181-193.

[14] LEYDESDORFF L. Problems with the ‘measurement’ of national scientific performance[J]. Science and public policy, 1988, 15(3): 149-152.

[15] BRAUN T, GLÄNZEL W, SCHUBERT A. Assessing assessments of British science. Some facts and figures to accept or decline[J]. Scientometrics, 1989, 15(3): 165-170.

[16] MARTIN B R. The bibliometric assessment of UK scientific performance: a reply to Braun, Glänzel and Schubert[J]. Scientometrics, 1991, 20(2): 333-357.

[17] GAUFFRIAU M, LARSEN P O. Different outcomes of different counting methods for publications and citations[C]//The international society for scientometrics and informetrics. Proceedings of ISSI 2005: International conference of the international society for scientometrics and informetrics. Sweden: ISSI, 2005: 242-246.

[18] GAUFFRIAU M, LARSEN P O, MAYE I, et al. Comparisons of results of publication counting using different methods[J]. Scientometrics, 2008, 77(1): 147-176.

[19] MOED H F. Citation analysis in research evaluation[J]. Information science & knowledge management, 2005, 57(1): 61-63.

[20] PRITYCHENKO B. Fractional authorship in nuclear physics[J]. Scientometrics, 2016, 106(1): 1-8.

[21] PLUME A, WEIJEN D. Publish or perish? The rise of the fractional author[M/OL]. [2016-10-16]. https://www.researchtrends.com/issue-38-september-2014/publish-or-perish-the-rise-of-the-fractional-author.htm.

[22] CHUDLARSKY T, DVORAK J, SOUCEK M. A comparison of research output counting methods using a national CRIS: effects at the institutional level[J]. Procedia computer science, 2014, 33(1): 147-152.

[23] NARIN F, STEVENS K, WHITLOW E S. Scientific co-operation in Europe and the citation of multinationally authored papers[J]. Scientometrics, 1991, 21(3): 320-323.

[24] GLÄNZEL W. National characteristics in international scientific co-authorship relations[J]. Scientometrics, 2001, 51(1): 69-115.

[25] RINIA E J, LANGE C D, MOED H F. Measuring national output in physics: delimitation problems[J]. Scientometrics, 1993, 28(1): 89-110.

[26] HUANG M H, LIN C S, CHEN D Z. Counting methods, country rank changes, and counting inflation in the assessment of national research productivity and impact[J]. Journal of the Association for Information Science and Technology, 2011, 62(12): 2427-2436.

[27] LIN C S, HUANG M H, CHEN D Z. The influences of counting methods on university rankings based on paper count and citation count[J]. Journal of informetrics, 2013, 7(7): 611-621.

[28] SCHREIBER M. The influence of self-citation corrections on Egghe's g-index[J]. Scientometrics, 2007, 76(1): 187-200.

[29] EGGHE L. Mathematical theory of the h- and g-index in case of fractional counting of authorship[J]. Journal of the American Society for Information Science & Technology, 2008, 59(10): 1608-1616.

[30] AKSNES D W, SCHNEIDER J W, GUNNARSSON M. Ranking national research systems by citation indicators: a comparative analysis using whole and fractionalized counting methods[J]. Journal of informetrics, 2012, 6(1): 36-43.

[31] MOYA-ANEGÓN F, GUERRERO-BOTE V P, BORRMANN L, et al. The research guarantors of scientific papers and the output counting: a promising new approach[J]. Scientometrics, 2013, 97(2): 421-434.

[32] WALTMAN L, ECK N J V, LEEUWEN T N V, et al. Towards a new crown indicator: some theoretical considerations[J]. Journal of informetrics, 2011, 5(1): 37-47.

[33] WALTMAN L, CALERO-MEDINA C, KOSTEN J, et al. The Leiden ranking 2011/2012: data collection, indicators, and interpretation[J]. Journal of the American Society for Information Science & Technology, 2012, 63(12): 2419-2432.

[34] WALTMAN L, SCHREIBER M. On the calculation of percentile-based bibliometric indicators[J]. Journal of the American Society for Information Science & Technology, 2014, 64(2): 372-379.

[35] PERIANES-RODRÍGUEZ A, RUIZ-CASTILLO J. Multiplicative versus fractional counting methods for co-authored publications: the case of the 500 universities in the Leiden Ranking[J]. Journal of informetrics, 2015, 9(4): 974-989.

[36] ABBAS A M. Weighted indices for evaluating the quality of research with multiple authorship[J]. Scientometrics, 2010, 88(1): 107-131.

[37] FOSTER J, THORBECKE E. A class of decomposable poverty measures[J]. Econometrica, 1984, 52(3): 761-766.

[38] ECK N J V, WALTMAN L, DEKKER R, et al. A comparison of two techniques for bibliometric mapping: multidimensional scaling and VOS[J]. Journal of the American Society for Information Science and Technology, 2010, 61(12): 2405-2416.

[39] ECK N J V, WALTMAN L. Software survey: VOSviewer, a computer program for bibliometric mapping[J]. Scientometrics, 2010, 84(2): 523-538.

[40] PERIANES-RODRÍGUEZ A, WALTMAN L, ECK N J V. Constructing bibliometric networks: a comparison between full and fractional counting[J]. Journal of informetrics, 2016, 10(4): 1178-1195.

[41] HAN W P, YOON J, LEYDESDORFF L. The normalization of co-authorship networks in the bibliometric evaluation: the government stimulation programs of China and Korea[J]. Scientometrics, 2016, 109(2): 1017-1036.

[42] LEYDESDORFF L, HAN W P. Full and fractional counting in bibliometric networks[J]. Journal of informetrics, 2016, 11(1): 117-120.

Author Contributions:

Chen Liyue: Conceived the research framework, conducted literature review and theoretical research, wrote the paper.

Yang Liying: Proposed research questions, revised research framework and paper.

Ding Jielan: Revised research framework and paper.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.