Methodological frameworks for counting methods in bibliometrics

Gauffriau, Marianne

23rd Nordic Workshop on Bibliometrics and Research Policy, 8th – 9th November, 2018
University of Borås, Sweden

Counting methods in bibliometrics

Marianne Gauffriau (mgau@kb.dk)
Coordinator of the Bibliometric Service
Copenhagen University Library, The Danish Royal Library, Denmark

Introduction: In many bibliometric studies is it difficult to see how publication and citations are counted, and very few studies justify the choice of counting method. In this study, I give an overview of counting methods. The goal is to facilitate that bibliometric studies report the used counting method and importantly, motivate the choice.

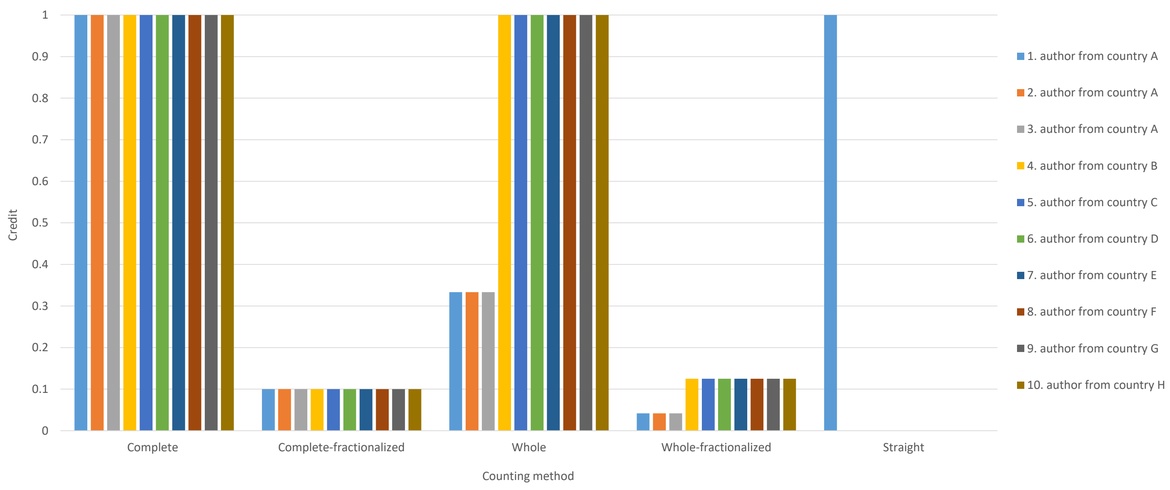
Method: The overview of counting methods is based on two methodological frameworks: 1) The five mathematical properties in Table 1 used to describe counting methods (Gauffriau *et al*., 2007), and 2) four groups of arguments for choosing a counting method: Group 1: The indicator measures participation / production / contribution / output / volume of an object of study, e.g. a country. Group 2: Additivity of counting method. Group 3: Pragmatic reasons (availability of data, prevalence of counting method, simplification of indicator, or the conclusion of the study is insensitive to a change of counting method). Group 4: Influence on / from the research community (Gauffriau, 2017).

Results: In Gauffriau *et al*., (2007) five counting methods often used in bibliometric studies are described according to five mathematical properties (Table 1). In Figure 1, I show how the countries of a publication with ten authors from eight countries share the credit according to each of the five counting methods. There is no agreement in the bibliometric literature on when to apply one or the other counting method. More of the four groups of arguments are being used for the same counting method (Larsen, 2008; Gauffriau, 2017).

Table 1: Description of traditional counting methods based on five mathematical properties (Gauffriau *et al*., 2007)

Defined for all objects	Based on a fixed crediting scheme	Additive	Rank-independent	Fractionalized	Counting methods
Yes	Yes		Yes	No	Complete
				Yes	Complete-fractionalized
			No	No	
				Yes	Straight (first author counting)
	No	Yes	Not applicable	No	
				Yes	
		No		No	Whole
				Yes	
No	No	No	Not applicable	No	
				Yes	Whole-fractionalized

Figure 1: How countries of a publication with ten authors from eight countries share the credit. Traditional counting methods.



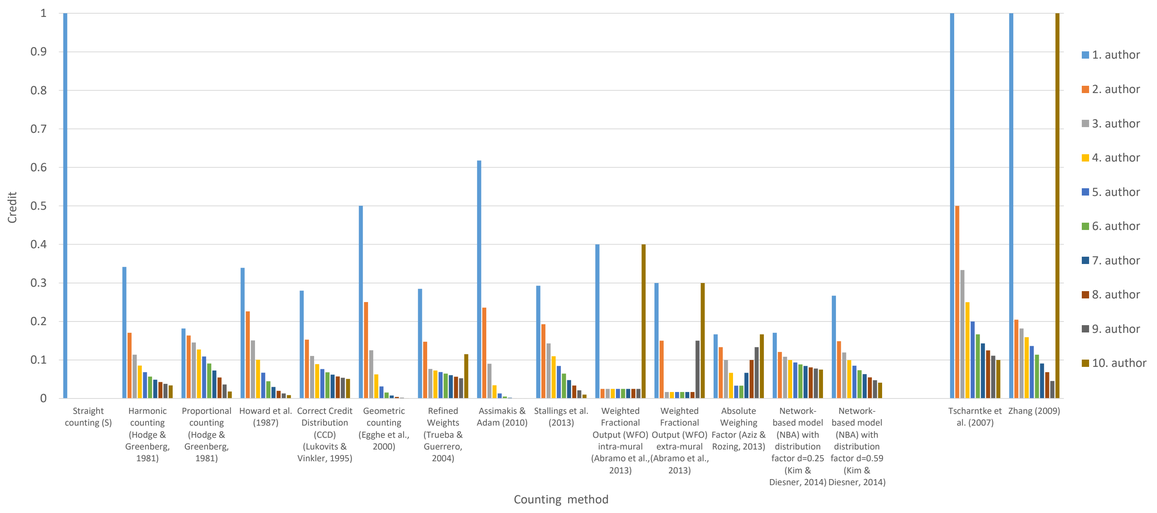
There are, however, many other counting methods in the bibliometric literature. Below, I have grouped these counting methods according to their mathematical properties, and the arguments for their introduction into the literature.

The largest group is like straight counting defined for all objects, based on a fixed crediting scheme, additive, rank-dependent, and fractionalized. Another group of counting methods has the same mathematical properties except that it is non-fractionalized as the sum of credits for a publication can exceed 1. The arguments for introducing these two groups of counting methods are to measure the contribution or production of an object of study (Group 1).

In addition to the counting methods in Figure 2, there are four fractionalized counting methods where the distribution of credit is not depending on ranks but on other characteristics:

- Modified straight count: The credit of 1 for a publication is divided between the authors with weights depending on the productivity (overall number of publications) of each author (Pravdic *et al*., 1986).
- Pareto weights: The credit of 1 for a publication is divided between the authors. An author "receives the greater credit for the joint publication if the number of actual citations is more in line with her citation record" (Tol, 2011).
- The credit of 1 for a publication is divided between the authors with weights depending on the author's share of authorships in the co-citation network of the publication and the number of co-citations. The more publications and citations an author has in the research field the more credit will be assigned to her / him (Shen & Barabasi, 2014).
- Relative intellectual contribution (IC^r): In publications where the authors state their contribution guided by a taxonomy (e.g. CRediT), the types of contributions can be weighted and these weights credited to the contributing authors. All author contributions of a publication sum up to 1 (Rahman *et al*., 2017).

Figure 2: How authors of a publication with ten authors share the credit. Rank-dependent counting methods, fractionalized (left) and non-fractionalized (right).

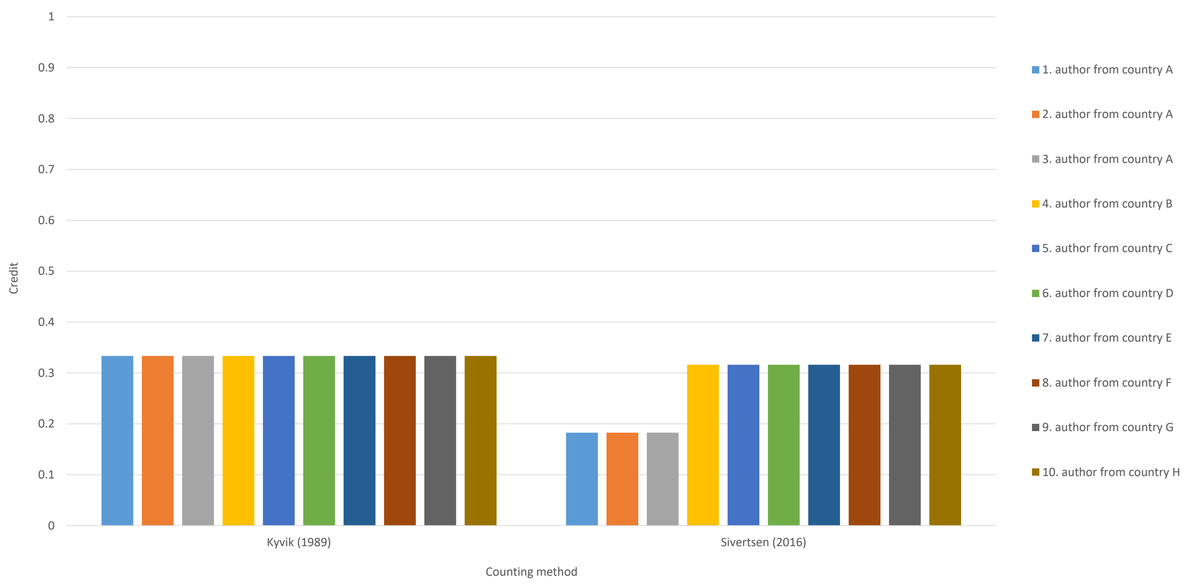


The last two groups of counting methods contain one counting method each. Both aim to give a balanced representation of productivity across research disciplines. This type of argument is not included in four groups of arguments for choosing a counting method. Despite the overlapping argumentation, the two counting methods do not share all mathematical properties.

The first counting method (Kyvik, 1989) is defined for all objects, based on a fixed crediting scheme, additive, rank-independent and non-fractionalized. This is close to complete counting where each author of a publication gets a full credit of 1. Kyvik assigns 1/2 credit to each author of publications with two or three authors and 1/3 credit to each author of publications with four or more authors.

The other counting method is used in the Norwegian Publication Indicator for institutions (Sivertsen, 2016). The credit is equal to the square root of an institution's complete-fractionalized credit for a publication. It could be seen as defined for all objects, not based on a fixed crediting scheme, non-additive, rank-independent and non-fractionalized. However, applying the square root as done in Norwegian Publication Indicator does not comply with measure theory, which is the theoretical basis for the properties in Table 1 (von Ins 2018, personal communication, 30th Oct.).

Figure 3: How countries of a publication with ten authors from eight countries share the credit. Counting methods aiming to give a balanced representation of productivity across research disciplines.



Discussion and conclusion: The two methodological frameworks facilitate the description of counting methods, and thus have potential to support improved transparency in bibliometric studies. Further research could investigate if counting methods can help us to learn more about what it is we are actually measuring when we count publications.

References
1. Gauffriau, M., Larsen, P. O., Maye, I., Roulin-Perriard, A. & von Ins, M. Publication, cooperation and productivity measures in scientific research. *Scientometrics* **73,** 175–214 (2007).
2. Gauffriau, M. A categorization of arguments for counting methods for publication and citation indicators. *Journal of Informetrics* **11,** 672–684 (2017).
3. Larsen, P. O. The state of the art in publication counting. *Scientometrics* **77,** 235–251 (2008).
4. Hodge, S. E. & Greenberg, D. A. Publication credit. *Science* **213,** 950 (1981).
5. Howard, G. S., Cole, D. A. & Maxwell, S. E. Research Productivity in Psychology Based on Publication in the Journals of the American Psychological Association. *American Psychologist* 12 (1987).
6. Lukovits, I. & Vinkler, P. Correct Credit Distribution: A Model for Sharing Credit among Coauthors. *Social Indicators Research* **36,** 91–98 (1995).
7. Egghe, L., Rousseau, R. & Van Hooydonk, G. Methods for accrediting publications to authors or countries: Consequences for evaluation studies. *Journal of the American Society for Information Science* **51,** 145–157 (2000).
8. Trueba, F. J. & Guerrero, H. A robust formula to credit authors for their publications. *Scientometrics* **60,** 181–204 (2004).
9. Assimakis, N. & Adam, M. A new author's productivity index: p-index. *Scientometrics* **85,** 415–427 (2010).
10. Stallings, J. *et al.* Determining scientific impact using a collaboration index. *Proceedings of the National Academy of Sciences* **110,** 9680–9685 (2013).
11. Abramo, G., D'Angelo, C. A. & Rosati, F. The importance of accounting for the number of co-authors and their order when assessing research performance at the individual level in the life sciences. *Journal of Informetrics* **7,** 198–208 (2013).
12. Aziz, N. A. & Rozing, M. P. Profit (p)-Index: The Degree to Which Authors Profit from Co-Authors. *PLoS ONE* **8,** e59814 (2013).
13. Kim, J. & Diesner, J. A network-based approach to coauthorship credit allocation. *Scientometrics* **101,** 587–602 (2014).
14. Tscharntke, T., Hochberg, M. E., Rand, T. A., Resh, V. H. & Krauss, J. Author Sequence and Credit for Contributions in Multiauthored Publications. *PLoS Biology* **5,** e18 (2007).
15. Zhang, C.-T. A proposal for calculating weighted citations based on author rank. *EMBO reports* **10,** 416–417 (2009).
16. Pravdić, N. & Oluić-Vuković, V. Dual approach to multiple authorship in the study of collaboration/scientific output relationship. *Scientometrics* **10,** 259–280 (1986).
17. Tol, R. S. J. Credit where credit's due: accounting for co-authorship in citation counts. *Scientometrics* **89,** 291–299 (2011).
18. Shen, H.-W. & Barabasi, A.-L. Collective credit allocation in science. *Proceedings of the National Academy of Sciences* **111,** 12325–12330 (2014).
19. Rahman, M. T., Regenstein, J. M., Kassim, N. L. A. & Haque, N. The need to quantify authors' relative intellectual contributions in a multi-author paper. *Journal of Informetrics* **11,** 275–281 (2017).
20. Kyvik, S. Productivity differences fields of learning, and Lotka's law. *Scientometrics* **15,** 205–214 (1989).
21. Sivertsen, G. A bibliometric indicator with a balanced representation of all fields. In *Proceedings of the 21ST International Conference on Science and Technology Indicators* 910–914 (2016).