Journal of Informetrics 11 (2017) 117–120

Contents lists available at ScienceDirect

Journal of Informetrics

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

Correspondence

Full and fractional counting in bibliometric networks☆

In their study entitled "Constructing bibliometric networks: A comparison between full and fractional counting," Perianes-Rodriguez, Waltman, & van Eck (2016; henceforth abbreviated as PWvE) provide arguments for the use of fractional counting at the network level as different from the level of publications. Whereas fractional counting in the latter case divides the credit among co-authors (countries, institutions, etc.), fractional counting at the network level can normalize the relative weights of links and thereby clarify the structures in the network. PWvE, however, propose a counting scheme for fractional counting that is one among other possible ones. Alternative schemes proposed by Batagelj and Cerinšek (2013) and Park, Yoon, & Leydesdorff (2016; henceforth abbreviated as PYL) are discussed in an appendix. However, our approach is not correctly identified as identical to their Equation A3. Here below, we distinguish three approaches analytically; routines for applying these approaches to bibliometric data are also provided.

As is common in social-network analysis (SNA), the co-occurrence matrix is defined by PWvE as the multiplication of the occurrence matrix by its transposed (Eq. (3): $\mathbf{U} = \mathbf{A}\mathbf{A}^T$). Using an equation derived by Newman (2001), the authors posit that "(t)he number of fractional counting co-authorship links between researcher i and j, denoted by u^*_{ij}, is given by:"

$$u^*_{ij} = \sum_{k=1}^{N} \frac{a_{ik} a_{jk}}{n_k - 1} \tag{1}$$

Whereas the denominator is equal to one in the case of full counting, this denominator normalizes in Eq. (1). The normalization is at the paper level (k): n_k is the number of co-authorships of paper k; N is the number of publications in the set. The (n–1) rather than n in the denominator corrects for the self-link: each author has only (n–1) co-authors (Newman, 2001; p. 016132–5).

The argument is elaborated by the authors with both model and empirical examples. We focus here on the first example of a co-authorship matrix (Tables 2 and 3; their Table 3) which is based on the assumed authorship matrix in Table 1 (their Table 2, at p. 1182):

Table 1
Authorship matrix.

	P1	P2	P3	Total
R1	1	1	0	2
R2	1	0	1	2
R3	1	1	0	2
R4	0	0	1	1
Total	3	2	2	

☆ The peer review process of this paper was handled by Vincent Larivière, Associate Editor of Journal of Informetrics.

http://dx.doi.org/10.1016/j.joi.2016.11.007

Table 2
Co-authorship matrix on the basis of full counting.

	R1	R2	R3	R4	Total
R1		1	2	0	3
R2	1		1	1	2
R3	2	1		0	3
R4	0	1	0		1
Total	3	2	3	1	9

Table 3
Fractional counting PWvE.

	R1	R2	R3	R4	Total
R1		0.5	1.5	0.0	2.0
R2	0.5		0.5	1.0	2.0
R3	1.5	0.5		0.0	2.0
R4	0.0	1.0	0.0		1.0
Total	2.0	2.0	2.0	1.0	7

Table 4
Fractional counting of author credit.

	P1	P2	P3	Total
R1	0.33	0.50	0.00	0.83
R2	0.33	0.00	0.50	0.83
R3	0.33	0.50	0.00	0.83
R4	0.00	0.00	0.50	0.50
Total	1	1	1	3

1. Our alternative approach

The first document P1 is co-authored by three authors, each of whom would receive one-third point of the credit when counted fractionally (Table 4). In the fractionated co-authorship matrix, the cell value {R1, R2} is accordingly 1/3 * 1/3 = 1/9 or 0.11 (Table 5).

Eq. (2) formalizes this approach, as follows:

$$u_{ij}^{*} = \sum_{k=1}^{N} \frac{a_{ik}a_{jk}}{n_{ik}n_{jk}} = \sum_{k=1}^{N} \frac{a_{ik}a_{jk}}{{n_k}^2} \tag{2}$$

Note that our values are smaller than those of PWvE because the value of the denominator (n^2) is larger than (n–1). Whereas PWvE count a total of seven for three papers, we count three (after rounding). In other words, this method is consistent. Table 5 is also provided as Table A1 in the Appendix of PWvE (at p. 1194), but without the diagonal values so that this consistency is not noticed.

2. Directed versus undirected networks

In both methods, the numerator ($a_{ik}a_{jk}$) for paper k and authors i and j is based on the assumption that the relation of i with j is counted as one arrow in addition to the reverse relation of j with i. While relations are counted bi-directionally in SNA, from a bibliometrics perspective, co-authorship is conceptualized as a single edge instead of opposing arcs. When we accept the argument of PWvE to correct for the self-relation, only one of the two arcs is being corrected. The denominator would then be n * (n–1) instead of n^2. The resulting matrix (not shown here; but see Table 6 below) contains somewhat higher values than the ones in Table 5 and sums to 3.35.

Table 5
Fractionally counted co-authorship at the level of the network.

	R1	R2	R3	R4	Total
R1	0.36	0.11	0.36	0.00	0.83
R2	0.11	0.36	0.11	0.25	0.83
R3	0.36	0.11	0.36	0.00	0.83
R4	0.00	0.25	0.00	0.25	0.50
Total	0.83	0.83	0.83	0.50	2.98