Journal of Informetrics 7 (2013) 784–791

Contents lists available at ScienceDirect

Journal of Informetrics

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

Harmonic coauthor credit: A parsimonious quantification of the byline hierarchy☆

Nils T. Hagen*

Faculty of Biosciences and Aquaculture, University of Nordland, N-8049 Bodø, Norway

ARTICLE INFO

Article history:
Received 27 March 2013
Received in revised form 28 May 2013
Accepted 26 June 2013

Keywords:
Byline hierarchy
Coauthor credit
Harmonic formula
Bibliometric bias
Parsimony

ABSTRACT

In this paper the accuracy of five current approaches to quantifying the byline hierarchy of a scientific paper is assessed by measuring the ability of each to explain the variation in a composite empirical dataset. Harmonic credit explained 97% of the variation by including information about the number of coauthors and their position in the byline. In contrast, fractional credit, which ignored the byline hierarchy by allocating equal credit to all coauthors, explained less than 40% of the variation in the empirical dataset. The nearly 60% discrepancy in explanatory power between fractional and harmonic credit was accounted for by equalizing bias associated with the omission of relevant information about differential coauthor contribution. Including an additional parameter to describe a continuum of intermediate formulas between fractional and harmonic provided a negligible or negative gain in predictive accuracy. By comparison, two parametric models from the bibliometric literature both had an explanatory capacity of approximately 80%. In conclusion, the results indicate that the harmonic formula provides a parsimonious solution to the problem of quantifying the byline hierarchy. Harmonic credit allocation also accommodates specific indications of departures from the basic byline hierarchy, such as footnoted information stating that some or all coauthors have contributed equally or indicating the presence of a senior author.

1. Introduction

Most scientific papers are multi-authored (Wuchty, Jones, & Uzzi, 2007), by coauthors who have not contributed equally (Waltman, 2012). Nevertheless, it is still routine practice to quantify authorship credit by dividing one credit equally among all coauthors of a paper irrespective of their actual contribution, thereby underestimating the credit of primary authors and overestimating the credit of secondary authors (Hagen, 2008). A more tenable approach is to estimate each coauthor's share of credit as accurately as possible by including all relevant information about the relative size of each contribution when calculating coauthor credit (Hagen, 2010a, 2010b).

Accurate quantification of coauthor credit requires a formulaic interpretation of the byline hierarchy which by convention lists coauthors in order of decreasing contribution (Lake, 2010; Zuckerman, 1968). It is also important that the quantification procedure accommodates all specific indications of departures from the basic byline hierarchy, for example footnoted information stating that some or all coauthors have contributed equally (Akhabue & Lautenbach, 2010; Frandsen & Nicolaisen,

* Tel.: +47 755 17 398.
E-mail address: Nils.Hagen@uin.no

1751-1577/$ – see front matter © 2013 The Author. Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2013.06.005

2010; Hu, 2009) or indicating the presence of a senior author (Buehring, Buehring, & Gerard, 2007; Mattsson, Sundberg, & Laget, 2011).

The harmonic formula, first proposed by Hodge and Greenberg (1981), fits these requirements by providing a straightforward quantification of the byline hierarchy that is easily modified to include specific information about the seniority or equality of some coauthors (Hagen, 2008), while demonstrating a robust fit when validated against empirical data from medicine, psychology, and chemistry (Hagen, 2010b). Recently, Liu and Fang (2012a, 2012b) proposed adding an additional parameter to the harmonic formula, in order to define a continuum of intermediate formulas ranging from fractional equal credit at one extreme to harmonic credit at the other, and suggested that a parameter value corresponding to two thirds of the distance between fractional and harmonic would be the preferred alternative.

In this paper I use the harmonic and fractional formulas as reference points when comparing the performance of Liu and Fangs's (2012a, 2012b) model with earlier parametric models from the bibliometric literature proposed by Lukovits and Vinkler (1995), and by Trueba and Guerrero (2004). I assess these formulas by quantifying their ability to explain the variation in a composite set of empirical data on perceived notions of coauthor contribution in chemistry (Vinkler, 2000), medicine (Wren et al., 2007), and psychology (Maciejovsky, Budescu, & Ariely, 2009), adapted for bibliometric analysis by Hagen (2010b). Finally, I conclude by discussing the results with reference to the problem of overfitting and the principle of parsimony.

2. Material and methods

The empirical data consists of three independent samples from the scientific subfields of chemistry (Vinkler, 2000), medicine (Wren et al., 2007), and psychology (Maciejovsky et al., 2009), that were extracted from the original publications as described by Hagen (2010b).

The consolidated data set describes perceived notions of coauthor credit for papers with up to 6 coauthors (Table 1). The data for medicine (Wren et al., 2007), imply the presence of a senior last author and support the assumption that the senior and first authors have contributed equally (Hagen, 2010b).

3. Theory/calculation

Harmonic authorship credit for the ith author of a publication with N coauthors is calculated according to the following formula (Hagen, 2008; Hodge & Greenberg, 1981):

$$\text{Harmonic}i\text{th author credit} = \frac{1/i}{[1 + 1/2 + \cdots + 1/N]} \tag{1}$$

Evidence of senior authorship in the scientific subfield of medicine (Table 1), was included in the calculation as described by Hagen (2008), by assuming that the senior author and the first author had contributed equally (Hagen, 2010b).

Fractional credit is calculated as follows:

$$\text{Fractional}i\text{th author credit} = \frac{1}{N} \tag{2}$$

Liu and Fang's (2012a, 2012b) model for coauthor credit is identical to fractional credit when the tuning parameter $q = 0$, identical to harmonic credit when $q = 1$, and provides a continuous range of potential formulas for $0 < q < 1$.

$$\text{Liu and Fang}'\text{s (2012a, 2012b)}\, i\text{th author credit} = \frac{i^{-q}}{\sum_{j=1}^{N} j^{-q}} \tag{3}$$

Evidence of senior authorship in the scientific subfield of medicine (Table 1), was included in the calculation of Liu and Fang's (2012a, 2012b) model as described above for harmonic credit.

According to Lukovits and Vinkler's (1995) model, coauthor credit is calculated as follows:

$$\text{1st author credit} = \frac{N+1}{2NF}, \text{ and} \tag{4}$$

$$i\text{th author credit} = \frac{i+T}{2iFT}, \text{ for} i = 2, \ldots, N, \text{ where} \tag{5}$$

$$F = \frac{1}{2}\left[\frac{1}{N} + \frac{N+1}{T} + \sum\nolimits_{j=1}^{N} \frac{1}{j}\right], \text{ and} T = \frac{100}{H}.$$

The tuning parameter H is the percentage value of their contribution threshold. To facilitate comparison among the different tuning parameters, H is expressed as a fraction in Fig. 3.

According to Trueba and Guerrero's (2004) model, coauthor credit is calculated as follows:

$$\text{1st author credit} = \frac{2N+1}{N \cdot (N+1)} \cdot \frac{2}{3} \cdot (1 - f) + c_1 \cdot f \tag{6}$$