Journal of Informetrics 8 (2014) 478–485

Contents lists available at ScienceDirect

Journal of Informetrics

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

Empirical study of *L*-Sequence: The basic *h*-index sequence for cumulative publications with consideration of the yearly citation performance

Yu Liu [a,*], Yongliang Yang [b]

[a] School of Software, Dalian University of Technology, Dalian 116024, PR China
[b] Center for Molecular Medicine, School of Life Science and Biotechnology, Dalian University of Technology, Dalian 116024, PR China

ARTICLE INFO

Article history:
Received 12 August 2013
Received in revised form 14 February 2014
Accepted 19 March 2014

Keywords:
Bibliometrics
Citations
Research evaluation
H-index sequence

ABSTRACT

Most current *h*-type indicators use only a single number to measure a scientist's productivity and impact of his/her published works. Although a single number is simple to calculate, it fails to outline his/her academic performance varying with time. We empirically study the basic *h*-index sequence for cumulative publications with consideration of the yearly citation performance (for convenience, referred as *L*-Sequence). *L*-Sequence consists of a series of *L* factors. Based on the citations received in the corresponding individual year, every factor along a scientist's career span is calculated by using the h index formula. Thus *L*-Sequence shows the scientist's dynamic research trajectory and provides insight into his/her scientific performance at different periods. Furthermore, L_{α}, summing up all factors of *L*-Sequence, is for the evaluation of the whole research career as alternative to other *h*-index variants. Importantly, the partial factors of the *L*-Sequence can be adapted for different evaluation tasks. Moreover, *L*-Sequence could be used to highlight outstanding scientists in a specific period whose research interests can be used to study the history and trends of a specific discipline.

1. Introduction

Bibliometrics has played an increasingly important role in evaluating individual researchers. It could be used as a quantitative analysis and assessment tool in tasks such as faculty promotion, funding allocation and awarding scientific prizes, etc (King, 1987). Bibliometrics provides a reliable and cost–effective way to evaluate scientific publications and their citations compared to the resource-expensive peer review (Abramo & DAngelo, 2011).

In 2005, a simple indicator for the assessment of the academic performance was suggested by Hirsch (2005), with consideration of both productivity and impact. The *h*-index has received a lot of attention from the scientific community in the last few years owing to its excellent properties (Ball, 2005). Many variants of the *h*-index were proposed to improve the original *h*-index (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2009; Egghe, 2010; Zhang, Thijs, & Glänzel, 2011). However, most of these *h*-type indicators use only a single number to measure scientists' life-long performance. One dimensional indicator lacks the ability to reveal the evolution details of a scientist's career at different periods.

* Corresponding author. Tel.: +86 15524582688.
E-mail address: yuliu@dlut.edu.cn (Y. Liu).

http://dx.doi.org/10.1016/j.joi.2014.03.002
1751-1577/

In contrast to single number, a series of index can describe a scientist's academic performance along with his/her career period. Liang first proposed an *h*-index sequence that was calculated in the reverse direction (Liang, 2006). However, Egghe pointed out that the calculation in the forward direction is more practical and easy to understand (Egghe, 2009b). Another interesting work is by Liu and Rousseau (2008) in which they defined 10 types of *h*-index sequences. Unfortunately, empirical study of practical examples based on large datasets are still lacking. Further, Egghe investigated four important sequences of which three sequences were defined by Liu & Rousseau before. These sequences were well explained and some practical examples were also discussed. Egghe's study has stimulated more research in this direction: for example, Fred Y. Ye and Ronald Rousseau studied the relationship between the power law model and total career *h*-index sequences (Ye & Rousseau, 2008), denoted as h_4 in Egghe's sequences; Wu, Lozano and Helbing performed the empirical study of the real career *h*-index sequence (Wu, Lozano, & Helbing, 2011), denoted as h_3 in Egghe's work. Based on h_3, Lin Zhang and Wolfgang Glänzel proposed the age dynamics of its *h*-core. Moreover, the other two time series, evolution of co-authorship and the age pyramids, were also presented in order to capture various facets of individual academic careers (Zhang & Glänzel, 2012).

In this paper, we performed the empirical study of another important *h*-index sequence with consideration of yearly citation performance for cumulative publications, denoted as h_2 in Egghe's work. Here we present the sequence as *L*-Sequence for convenience. As stated by Egghe, the sequence is rather necessary and challenging to study. The fact of slightly increasing in the sequence should be further studied and interpreted (Egghe, 2009a). We use a large bibliographic data set for computer scientists to evaluate the performance of *L*-Sequence. Our experimental results demonstrate that *L*-Sequence could effectively reflect the dynamic properties of a scientist's productivity and citation impact. Particularly, L_∞, summing up all factors of *L*-Sequence, can be used for the evaluation of the whole research career of a scientist as an alternative to other *h*-index variants. Importantly, the partial factors of the *L*-Sequence can be adapted to various evaluation contexts. In addition, the scientific impact of researchers is normalized to the year coordination. Hence, it is easy to compare researchers' performance in specified years, which becomes feasible to study the history and trends of a research field or discipline.

2. *L*-Sequence

Instead of using a single number, *L*-Sequence uses a sequence of measuring factors along with a scientist's career period to reflect his/her research performance. A series of process variables over the time can provide an accurate and sufficient description of a scientist's dynamic research trajectory, consistent with underlying natural mechanism of scientific research.

2.1. Definition of L-Sequence

Suppose a scientist has published n papers, $P_1, P_2, \ldots, P_n$, along his/her research career. The year of the first publication is T_1 and the current year is T_2. *L*-factor of each year (L_t) is calculated by the h index formula based on the citations received in year t for all papers. Thus, a series of *L*-factor for each year $L_{T_1}, L_{T_1+1}, \ldots, L_{T_2}$ constitute the *L*-Sequence.

2.2. Graphical illustration of L-Sequence

To explicitly understand the *L*-Sequence, Fig. 1 demonstrates the calculation process based on real bibliographic data of Judea Pearl, 2011 Tuning Award Winner. The citation counts for each of his papers in each individual year are recorded and then plotted in a single color as in Fig. 1. The citation counts received in one year are defined as the citation slice for this year, which is subsequently used to calculate a factor of *L*-Sequence. For example, the citation slice in 2000 is employed to calculate L_{2000}. In the top left corner of Fig. 1, citation slice in 2000 is displayed in the form of paper-citation distribution and the *h*-index formula is used to calculate L_{2000} based on the citation counts received in 2000. In the same way, a series of factors which consists of *L*-Sequence, are achieved one year by one year during a scientist's career. Thus Judea Pearl' *L*-Sequence is shown in Fig. 2.

Notably, in the course of calculating *L*-Sequence, a highly-cited paper for a long period could be involved in the calculation of L_t factors for several years, thereby making substantial contribution to the entire *L*-Sequence. We provided herein an example of a highly-cited paper from 1990 to 2010 as displayed as blue-color dash line (asterisk) in Fig. 1. Conversely, lowly-cited paper will be ignored in the calculation of L_t factors and outdated paper will barely make contribution to L_t factors when few citation counts were received. Again, we provided another example of an outdated paper from 1978 to 1988 as displayed as purple-color dash line (asterisk) in Fig. 1. This paper received its highest citation in 1981 and contributed to the L_t factors in the following two years. However, since 1984, this paper barely contributes to the calculation of L_t factors because it has received very few citations. This is consistent with the evolution process of research activities.

2.3. Application of L-Sequence

L-Sequence is a flexible evaluation tool and provides more details for researcher's academic performance varying with time. More importantly, users could select continuous or discrete parts of the factors in *L*-Sequence for different evaluation purposes. For instance, for awarding purpose, the performance of individual's entire research career should be taken into account. Therefore, all factors of *L*-Sequence could be summed up to indicate the life-long performance; For promotion and project support, the performance of recent m years is more valuable. Consequently, the recent m year factors could be