Journal of Informetrics 8 (2014) 972–984

Contents lists available at ScienceDirect

Journal of Informetrics

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

Distributions of citations of papers of individual authors publishing in different scientific disciplines: Application of Langmuir-type function

Keshra Sangwal*

Department of Applied Physics, Lublin University of Technology, ul. Nadbystrzycka 38, 20-618 Lublin, Poland

ARTICLE INFO

Article history:
Received 25 August 2014
Received in revised form
23 September 2014
Accepted 30 September 2014
Available online 25 October 2014

Keywords:
Citation analysis
Citation distribution
Langmuir-type function

ABSTRACT

The distribution of cumulative citations L and contributed citations L_f to individual multi-authored papers published by selected authors working in different scientific disciplines is analyzed and discussed using Langmuir-type function: $y_n = y_0[1 - \alpha Kn/(1 + Kn)]$, where y_n denotes the total number of normalized cumulative citations $l_n{}^*$ and normalized contributed citations $l_{nf}{}^*$ received by individual papers of rank n, y_0 is the maximum value of y_n when $n = 0$, $\alpha \geq 1$ is an effectiveness parameter, and K is the Langmuir constant related to the dimensionless differential energy $Q = \ln(KN_c)$, with N_c as the number of papers receiving citations. Relationships between the values of the Langmuir constant K of the distribution function, the number N_c of papers of an individual author receiving citations and the effectiveness parameter α of this function, obtained from analysis of the data of rank-size distributions of the authors, are investigated. It was found that: (1) the quantity KN_c obtained from the real citation distribution of papers of various authors working in different disciplines is inversely proportional to $(\alpha - 1)$ with a proportional constant $(KN_c)_0 < 1$, (2) the relation $KN_c = (KN_c)_0/(\alpha - 1)$ also holds for the citation distribution of journals published in countries of two different groups, investigated earlier (Sangwal, K. (2013). Journal of Informetrics, 7, 487–504), and (3) deviations of the real citation distribution from curves predicted by the Langmuir-type function are associated with changing activity of sources of generation of items (citations).

1. Introduction

Distributions of the number of items such as citations, papers, authors, journals and journal impact factors by their rank and size is an important research area in informetrics (Bornmann & Daniel, 2009; Egghe & Waltman, 2011; Egghe, 2009, 2011, 2013; Guerrero-Bote, Zapico-Alonso, Espinosa-Calvo, Gomez-Crisostomo, & Moya-Anegon, 2007; Kretschmer & Rousseau, 2001; Lancho-Barrantes, Guerrero-Bote, & Moya-Anegon, 2010; Leherrere & Sornette, 1998; Perc, 2010; Radicchi, Fortunado, & Castellano, 2008; Redner, 1998, 2005; Tsallis & de Albuquerque, 2000; Vieira & Gomes, 2010; Wallace, Lariviere, & Gingras, 2009). In the literature various laws (e.g. Lotka's and Zipf's laws) and functions have been proposed to describe these informetric distributions and to explain the mechanism underlying their occurrence. The approaches used in the investigations of rank and size distributions of citations may be classified in the following categories:

* Tel.: +48 81 5384 504; fax: +48 81 5384 731.
E-mail address: k.sangwal@pollub.pl

http://dx.doi.org/10.1016/j.joi.2014.09.009
1751-1577/

(1) Theoretical studies of modeling of citation behavior carried out using preselected mathematical functions to generate citations (Burrell, 2001, 2002, 2013; Egghe, 2009, 2013; Kretschmer & Rousseau, 2001; Nadarajah & Kotz, 2007). However, the main limitation of these functions is that they contain adjustable empirical parameters.
(2) Empirical studies devoted to the analysis of a dataset of citation distributions, constructed over a selected time window or a long period of time for a single discipline, speciality or journal, carried out using known mathematical functions (Bornmann & Daniel, 2009; Clauset, Shalizi, & Newman, 2009; Companario, 2010; Perc, 2010; Radicchi et al., 2008; Redner, 1998, 2005; Vieira & Gomes, 2010; Wallace et al., 2009). The main limitation of these functions is that they result in poor fitting at very high or very low rank (Mansilla, Köppen, Cocho & Miramontes, 2007; Naumis & Cocho, 2007; Sangwal, 2013a).
(3) Phenomenological approach used to describe citation data in terms of theoretical equations based on specific microscopic models (Barabasi & Albert, 1999; Guerrero-Bote et al., 2007; Gupta, Campanha, & Schinaider, 2008; Mansilla et al., 2007; Naumis & Cocho, 2007; Price, 1965, 1976; Sangwal, 2013a, 2013b; Simkin & Roychowdhury, 2007; Tsallis & de Albuquerque, 2000; Vieira & Gomes, 2010; Wallace et al., 2009). The advantage of this approach is that the mathematical functions contain parameters attributed to some physical processes.

Sangwal (2013a) analyzed citation distributions of papers of different selected authors using five mathematical functions. The main conclusion drawn from the above study is that Zipf-type power law and logarithmic function previously proposed by Guerrero-Bote et al. (2007) for their iceberg hypothesis are inadequate to describe the citation distribution of individual papers of the authors, and that the new stretched exponential, Langmuir-type and empirical binomial mathematical functions can be employed to analyze citation distributions. In a later paper, Sangwal (2013b) analyzed distributions of citations L, two- (IF2) and five-year impact factors (IF5), and citation half-lives λ of journals published in different selected countries using Langmuir-type relation. It was found that the general features of the rank-order distributions of L, IF2 or IF5 of the journals published in different individual countries are similar to those of the citation distribution of papers of individual authors, and that the product of the Langmuir constant K and the number N of journals for the processes of citations and two- and five-year impact factors of journals published in different countries is constant for a process.

Size- and rank-order distributions may be defined in terms of information production processes (Egghe & Rousseau, 1990). An information production process consists of sources which produce or have items. A country in which N published journals receive L citations, a journal or a discipline in which N published papers receive L citations and an author publishing N papers which receive L citations since their inception are typical examples of information production processes. An information production process may be considered as a system or set of sources generating items. Sangwal (2013a, 2013b) derived his equations of the rank-order distribution of items generated by individual sources of the same activity using the concepts of adsorption processes involved during crystal growth.

It is well known that the average number of citations per paper differs among various scientific disciplines due to their citation behavior (for example, see: Abramo, Cicero & D'Angelo, 2012; Alonso, Cabrerizo, Herrera-Viedma & Herrera, 2009; Hirsch, 2005; Iglesias & Pecharroman, 2007; Lundberg, 2007; Podlubny, 2005; Radicchi et al., 2008). Therefore, the total number of citations or any index such as the Hirsch index h based on citation distribution cannot be used to compare the research performance of researchers working in different scientific disciplines. Using the total number of citations (Iglesias & Pecharroman, 2007; Lundberg, 2007) or distributions of citations to the papers published in different fields (Abramo et al., 2012; Alonso et al., 2009; Podlubny, 2005; Radicchi et al., 2008) different scaling parameters have been proposed to compare the research output of researchers in various fields. Several studies have reported the average number of citations in different fields as an effective scaling parameter (Abramo et al., 2012; Alonso et al., 2009; Bornmann & Daniel, 2009; Hirsch, 2005; Iglesias & Pecharroman, 2007; Podlubny, 2005; Radicchi et al., 2008) but median or geometric mean of citations has also been proposed (Lundberg, 2007).

All citation-related measures for the publication output of an author assume that the papers receiving citations are written by him/her alone. However, since various papers published by an author in most disciplines are multiple-authored, it is illogical to award full credit to each author. The above measures in reality penalize authors who publish alone in comparison with others publishing with a large number of coauthors, juniors or seniors. In view of frequently unknown contributions of different authors in multiauthored papers devising of a fair method of counting of contributions of individual authors in multiauthored papers has drawn considerable attention for over three decades (for example, see: Assimakis & Adam, 2010; Batista, Campiteli, Kinouchi & Martinez, 2006; Hodge & Greenberg, 1981; Pereira de Araújo, 2008; Price, 1981; Tol, 2011; Vinkler, 1993).

A survey of the published literature shows that analysis of the distribution of citations received by multiauthored papers published by individual authors using physical models has drawn relatively poor attention so far. Moreover, until now no study has been devoted to the investigation of the distribution of citations contributed by individual authors to their different multiauthored papers and comparison of distributions of cumulative citations and contributed citations to every multiauthored papers published by individual authors. The present paper is addressed to this topic. The aim of the paper is two-fold: (1) to analyze the distribution of cumulative citations L and contributed citations L_f to every multiauthored papers published by individual authors working in different scientific disciplines using the newly proposed Langmuir-type function, and (2) to investigate the relationship between the Langmuir constant K of the distribution function, the number N_c of papers of an individual author receiving citations and the effectiveness parameter α of this function.