Journal of Informetrics 7 (2013) 36–49

Contents lists available at SciVerse ScienceDirect

Journal of Informetrics

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

Comparison of different mathematical functions for the analysis of citation distribution of papers of individual authors

Keshra Sangwal*

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

ARTICLE INFO

Article history:
Received 30 April 2012
Received in revised form
17 September 2012
Accepted 18 September 2012

Keywords:
Adsorption isotherms
Citation analysis
Citation rank-order distribution
Rank-frequency functions

ABSTRACT

The citation distribution of papers of selected individual authors was analyzed using five mathematical functions: power-law, stretched exponential, logarithmic, binomial and Langmuir-type. The former two functions have previously been proposed in the literature whereas the remaining three are novel and are derived following the concepts of growth kinetics of crystals in the presence of additives which act as inhibitors of growth. Analysis of the data of citation distribution of papers of the authors revealed that the value of the goodness-of-the-fit parameter R^2 was the highest for the empirical binomial relation, it was high and comparable for stretched exponential and Langmuir-type functions, relatively low for power law but it was the lowest for the logarithmic function. In the Langmuir-type function a parameter K, defined as Langmuir constant, characterizing the citation behavior of the authors has been identified. Based on the Langmuir-type function an expression for cumulative citations L relating the extrapolated value of citations l_0 corresponding to rank $n = 0$ for an author and his/her constant K and the number N of paper receiving citation $l \geq 1$ is also proposed.

1. Introduction

Investigation of distribution of authors, citations and publications is an active research area in informetrics (Egghe & Waltman, 2011; Egghe, 2009, 2011, 2012; 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). Various laws (e.g. Lotka's and Zipf's laws) and functions have been proposed in the literature to describe these informetric distributions and to explain the mechanism underlying their occurrence. Citation distributions, for example, have been studied using the following approaches: (1) theoretical studies involving modeling of citation behavior using a preselected mathematical function to generate citations (Burrell, 2001, 2002; Egghe, 2009, 2012; Kretschmer & Rousseau, 2001; Nadarajah & Kotz, 2007), (2) empirical studies devoted to the analysis of a dataset, constructed over a selected time window or a long period of time for a single discipline, speciality or journal, 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), and (3) phenomenological approach based on describing citation data using specific microscopic models (Barabasi & Albert, 1999; Gupta, Campanha, & Schinaider, 2008; Naumis & Cocho, 2007; Price, 1965, 1976; Simkin & Roychowdhury., 2007; Tsallis & de Albuquerque, 2000; Wallace et al., 2009).

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

1751-1577/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2012.09.002

A power-law type of behavior of citation distribution was suggested in the first empirical studies in the area (Naranan, 1971; Price, 1965, 1976; Seglen, 1992), but it is now recognized that a single function such as power law is unable to describe citatation distributions over the whole range of citations (Perc, 2010; Radicchi et al., 2008; Redner, 1998, 2005; van Raan, 2001; Wallace et al., 2009). Redner (1998, 2005) examined the citation distributions of large data sets of articles published in Physical Review and found that a power-law behavior dominates at high number of citations whereas a stretched exponential function provides a better fit at a small numer of citations. Based on the general, nonextensive thermostatistical formalism, Tsallis and de Albuquerque (2000) proposed a new function, now known as Tsallis distribution function, to describe the citation distribution in the entire citation range. Radicchi et al. (2008) used the lognormal distribution function, derived from reorganization of the stretched exponential function, to fit data on 14 among more than 200 subject categories. Wallace et al. (2009) examined the citation distribution of papers published between 1900 and 2006 in natural sciences and engineering, medicine and social sciences and found that stretched-exponential and Tsallis's distribution functions fit the entire citation data satisfactorily. Vieira and Gomes (2010) analyzed the distribution of five-year citations of papers published in 2004 in chemistry, biology and biochemistry, mathematics and physics using one- and two-parameter (double) exponential-Poisson distributions and found that the double exponential-Poisson distribution describes the data well. Perc (2010) examined the distributions of citations of individual papers published during 1970 and 2009 by researchers in Slovenia using Zipf's plots, power law and lognormal distributions. It was found that the data follows power law at low and high values of citations.

Guerrero-Bote et al. (2007) and Lancho-Barrantes et al. (2010) studied Journal Impact Factor (JIF) rank-order distribution and found that the distributions of JIFs were fairy close to exponential, which could be fitted to a logarithmic function. However, these authors also encountered subject areas having shapes of their JIF rank-order distributions with more sharply defined peaks and relatively long tails, something like icebergs. These authors suggested that icebergs (i.e. scientific areas) are exporters of ideas because the knowledge generated within them is visible from other areas which then import it (iceberg hypothesis).

A general feature of many informetric distributions is that the shape of the size-frequency distribution f and the shape of the rank-frequency distribution g are interrelated (Egghe & Rousseau, 2006, 2012). When the size-frequency distribution is a monotonically decreasing function, the corresponding rank-frequency distribution is convex in the entire range. In contrast to this, when the size-frequency distribution increases first and then steadily decreases, thus passing through a maximum, the rank-frequency distribution is convex initially and concave at the end. The curvature of the concave part of the rank-frequency distribution is related to the position of the maximum in the size-frequency distribution. These size- and rank-frequency distributions are explained by empirical functions other than those following from Lotka's and Zipf's laws (Companario, 2010; Egghe & Waltman, 2011; Mansilla, Köppen, Cocho, & Miramontes, 2007). However, this type of distribution behavior can also be explained by Tsallis distribution function based on nonextensive thermostatistical formalism (Gupta et al., 2008; Tsallis & de Albuquerque, 2000).

The above literature shows that distribution of authors, citations and publications in single discipline, speciality or journals has been investigated until now using known mathematical functions. However, no study has been devoted so far to the analysis of the distribution of citations of papers as a function of paper rank at the level of individual authors. The aim of the present paper is to analyze the citation distribution of papers of different selected authors using five mathematical functions. Two of these, the power law and the extended exponential function, are well known in the citation literature, whereas the remaining three are novel mathematical functions. Among the new functions, the logarithmic function proposed for the analysis is similar to that used by Guerrero-Bote et al. (2007) and Lancho-Barrantes et al. (2010) for their iceberg hypothesis. The new mathematical functions proposed in this work are derived following the concepts of growth kinetics of crystals in the presence of additives which act as inhibitors of growth (Appendices A and B). An additional aim of the study is to propose a possible mechanism of the citation rank-order distribution in terms of physical processes at the elementary level.

2. Mathematical functions

In this section the mathematical functions used in this study for the analysis of the citation distribution of the papers of different authors are briefly described.

If an author publishes N papers and l_n denotes the number of citations of the nth paper such that n is ranked in the order of decreasing citations l_n, the relation between l_n and n is given by the power-law distribution

$$l_n = \frac{l_0}{n^{\delta}}, \quad \text{power law,} \tag{1}$$

where $l_0 > 0$, $\delta > 0$ and $1 < n < N$. Here l_0 is the extrapolated value of l_n when $n \to 0$. The value of the exponent δ has been reported to lie between 2.4 and 3.1 (Perc, 2010; Redner, 1998). Power law distribution may be derived from preferential attachment models and are considered as representative of complex networks (Albert & Barabasi, 2002; Leherrere & Sornette, 1998). Zipf's law is a typical case of power law but the exact mechanism behind it remains unclear. It is found that Zipf's law describes the citation distribution reasonably well at relatively high values of n (Perc, 2010; Redner, 1998).

For sufficiently large values of n the relation between l_n and n is described by the stretched exponential function

$$l_n = l_0 \exp\left[-\left(\frac{n}{n_0}\right)^{\beta}\right], \quad \text{stretched exponential function,} \tag{2}$$

where l_0 denotes citations of the maximally cited paper, and β and n_0 are empirical constants. The constant $\beta \le 1$. For real citation distributions analyzed in the literature, β is found to lie between 0.39 and 0.57 (Redner, 1998; Wallace et al., 2009). In the deterministic model of Hirsch (2005) the parameter $n_0 = ha$, where h the Hirsch index and a is related to the cumulative citations L and is about 0.5. The parameters β and n_0 may be obtained empirically by performing a least-squares fit over all values of n using Eq. (2).

Stretched exponential distribution can easily be distinguished from power law distribution by plotting $l_n(n)$ data on logarithmic and semi-logarithmic scales (Leherrere & Sornette, 1998; Perc, 2010; Redner, 1998; Wallace et al., 2009). The case $\beta = 1$ corresponds to the usual exponential distribution which gives a linear dependence of $\ln l_n$ on n with slope $1/n_0$. However, the linear part decreases when $\beta < 1$.

In this paper the following mathematical relations are proposed:

$$l_n = l_0 b'(1 - Z_1 \ln n), \quad \text{logarithmic function,} \tag{3}$$

$$l_n = l_0(1 - k_1 n^p + k_2 n^{2p}), \quad \text{binomial relation,} \tag{4}$$

$$l_n = l_0 \left[1 - \alpha \left(\frac{Kn}{1 + Kn}\right)\right], \quad \text{Langmuir-type function.} \tag{5}$$

In the above equations l_0 is the extrapolated value of l when $n = 0$, and K, α, Z_1, b', k_1, k_2 and p are positive constants, which may be considered as fitting parameters for the analysis of the citation distribution data. In Eq. (4) when $k_1 n^p \gg k_2 n^{2p}$ and $k_1 = \alpha_1$, it reduces to the form

$$l_n = l_0(1 - \alpha_1 n^p), \quad \text{decreasing power law.} \tag{6}$$

where $k_1 = \alpha_1$. Note that relation (6) is completely different from power law (1).

As discussed in Appendix B, logarithmic function (3), Langmuir-type function (5) and decreasing power-law relation (6) are derived following the concepts of adsorption processes involved during crystal growth. However, it is suggested that binomial function (4) is an extended form of decreasing power law (6) when coverage θ of possible adsorption sites is described by Eq. (C.1).

In Eq. (5) the parameter K, defined as Langmuir constant (see Appendix A), characterizes the citation behavior of a given author. However, the units of K are inverse of paper rank (i.e. paper-rank^{-1}) and are determined by the way the paper rank n is expressed. In analogy with dimensionless pressure P_A/P_A^* defined in Appendix A, in Eq. (5) one can define a dimensionless paper rank n/N, and a new dimensionless Langmuir constant $K' = KN$, where N is the number of papers which receive citations. Then the dimensionless Langmuir constant K' for citations of different authors is related to their corresponding dimensionless differential energy Q by (see Appendix B)

$$K' = KN = \exp Q. \tag{7}$$

The fitting parameters α of Eq. (5), b' and Z_1 of Eq. (3) and α_1 and p of (6) also have clear physical interpretation (Appendix B). As shown in Appendix B, the parameter $b' < 1$ in Eq. (3) and $p = m < 1$ in Eq. (6). Eq. (3) is exactly the same as that reported by Guerrero-Bote et al. (2007) for their iceberg hypothesis and applies for $n \ge 1$.

3. Citation data of selected authors

We used Thomson Reuters' ISI Web of Knowledge (Web of Science) to collect and analyze the citations of nine arbitrarily selected scientists from different research disciplines. J. Barnaś (JB), T. Ditl (TD), S. Krukowski (SK), K. Sangwal (KS), and Z.R. Żytkiewicz (ZRZ), are physicists, M. Kosmulski (MK) is a chemist, K.J. Kurzydłowski (KJK) is a materials scientist, whereas Q.L. Burrell (QLB) and L. Egghe (LE) are informetricians. The first seven of these scientists are from Poland whereas the last two are from United Kingdom and Belgium, respectively. The Polish scientists are from different institutions and were selected from a consideration of their known academic activities. TD and JB are among the most cited physicists in Poland, MK and KS are the most cited scientists in the Lublin University of Technology, SK and ZRZ are among the most active crystal growers in Poland, whereas KJK is an eminent materials scientist and administrator at the national level. Their data covering the period up to 2010 were collected in December 2010. The data for LE and QLB cover the period up to 2011 and were collected in April 2012. Both of them are well-known informetricians. The basic bibliometric data involving the number of citations of individual papers collected from the above database are given in Tables 1 and 2, whereas the number of papers N with citations equal to and exceeding one, the total number of papers $N_{\max}$ and the cumulative citations L with self-citations are included in Table 3.

It should be mentioned that the publication output of the Polish scientists considered in this work has previously been analyzed (Sangwal, 2011, 2012a, 2012b).

Table 1
Citation data for different authors.

M. Kosmulski		K. Sangwal		S. Krukowski		K.J. Kurzydłowski		Z.R. Żytkiewicz	
n	l_n	n	l_n	n	l_n	n	l_n	n	l_n
1	104	1	96	1	119	1	46	1	70
2	93	2	67	2	82	2	43	2	36
3	86	3	60	3	42	3	32	3	35
4	85	4	45	4	36	4	28	4	24
5	74	5	36	5	33	5–6	25	5	21
6	64	6	35	6–8	30	7	24	6	19
7	62	7	31	9	26	8	22	7	18
8	56	8	28	10	23	9–10	21	8	16
9	53	9	27	11–12	20	11	19	9	14
10	49	10–13	25	13–15	19	12	18	10–11	13
11–13	37	14	24	16–17	18	13	17	12–13	12
14–15	36	15	23	18	17	14	15	14	11
16	34	16	22	19–20	16	15–16	14	15–18	10
17	33	17–19	21	21–22	15	17–18	13	19–20	9
18	32	20	20	23–24	14	19	12	21–23	7
19	31	21	20	25–26	11	20–23	11	24–25	6
20	30	22	18	27–28	10	24	10	26–31	5
21	29	23–26	17	29–30	9	25–28	9	32–37	4
22	27	27–30	16	31	8	29–37	8	38–45	3
23	24	31	15	32–35	7	38–42	7	46–53	2
24	23	32–39	14	36–37	6	43–53	6	54–64	1
25	21	40–43	13	38–45	5	54–61	5	65–89	0
26	20	44–45	12	46–52	4	62–79	4		
27–29	19	46–49	11	53–59	3	80–99	3		
30–31	17	50–55	10	60–66	2	100–121	2		
32	16	56–62	9	67–80	1	122–169	1		
33–34	15	63–68	8	81–103	0	170–268	0		
35	14	69–77	7						
36–37	13	78–82	6						
38–40	12	83–90	5						
41–44	11	91–102	4						
45–49	10	103–113	3						
50–53	9	114–121	2						
54–61	8	122–131	1						
62–64	7	132–155	0						
65–68	6								
69–75	5								
76–83	4								
84–90	3								
91–105	2								
106–114	1								
115–140	0								

4. Results and discussion

4.1. Analysis of citation distribution data

The real $l_n(n)$ data for different authors were confronted with Eqs. (1)–(5) mentioned above using two approaches. In the first case, the applicability of power law (1), stretched exponential function (2) and logarithmic function (3) was checked by plotting the $l_n(n)$ data in the form of dependences of (a) $\ln l_n$ on $\ln n$, (b) $\ln l_n$ on n and (c) l_n on $\ln n$, as shown in Figs. 1–3, respectively. These forms of the dependences follow from Eqs. (1)–(3), and are usually used to distinguish between different functions and the range of their applicability (Redner, 1998).

Fig. 1 shows that there is an observable curvature in the log–log plots of the power-law distribution over the entire n rage, but there are narrow regions of n in some of the plots. This indicates that some other distribution alone can represent the data in a wide range of n. However, when the $l_n(n)$ data for a given scientist are presented as a plot of $\ln l_n$ against n expected from the stretched exponential function, there are large ranges of n in which the dependence is linear. For example, in the case of KS, the linear dependence covers n between about 15 and 110 (Fig. 2) but there are no large regions of constant slope in the log–log plot of Fig. 1. One also observes no specific trends of a linear dependence in the plots of Fig. 3. From these figures it may be concluded that stretched exponential distribution respresents the $l_n(n)$ data most satisfactorily.

In the second approach, the data of the above authors were analyzed using Eqs. (1)–(5) directly. Nonlinear least-squares fitting, involving chi-square residual, was carried out with Macrocal^{TM} "Origin 4.1" package, which yields values of the fitting parameters, their standard deviations and the corresponding "dependency" for each parameter of an equation. The values of the fitting "dependency" for different parameters are different for a dataset. For example, in the case of fitting the data

Table 2
Citation data for different authors.

J. Barnaś		T. Dietl				L. Egghe		Q.L. Burrell	
n	l_n	n	l_n	n	l_n	n	l_n	n	l_n
1	197	1	3566	58–62	37	1	224	1	39
2	146	2	391	63	25	2	122	2	38
3	94	3	294	64–67	24	3	107	3	34
4	90	4	291	68–69	23	4–5	60	4–5	35
5	89	5	285	70	22	6	48	6	34
6	73	6	256	71–75	21	7	47	7	33
7	68	7	210	76–77	20	8	41	8	32
8	60	8	172	78–82	19	9	40	9	24
9–10	57	9	140	83–86	18	10	34	10	23
11–13	53	10	112	87	17	11	30	11–13	22
14–16	49	11–12	109	88	16	12–13	29	14	19
17–18	48	13	107	89–92	15	14	26	15	18
19	43	14	103	93–99	14	15	24	16	15
20	39	15	95	100–105	13	16	23	17–18	14
21	37	16	90	106–110	12	17–20	22	19	13
22	36	17	88	111–113	11	21	21	20–21	11
23	35	18–19	85	114–120	10	22–23	20	22	10
24	32	20	78	121–127	9	24–25	19	23–26	9
25–27	31	21	76	128–134	8	26–28	18	27	7
28–30	30	22	71	135–144	7	29–30	17	28–30	6
31–32	28	23	69	145–155	6	31–34	15	31	5
33	27	24	65	156–164	5	35–38	14	32–37	4
34	25	25	62	165–177	4	39–40	13	38–41	3
35–36	24	26	59	178–187	3	41–44	12	42–47	2
37–38	23	27	58	188–207	2	45–47	11	48–50	1
39–41	22	28–29	54	208–232	1	48–53	10	51–61	0
42–44	21	30	53	233–292	0	54–55	9		
45–47	19	31	49			56–68	8		
48	18	32	48			69–73	7		
49–53	17	33	47			74–78	6		
54	15	34	46			79–90	5		
55–56	13	35–37	45			91–98	4		
57–60	12	38	44			99–114	3		
61–65	11	39–40	42			115–132	2		
66–71	10	41	41			133–162	1		
72–79	9	42	40			163–198	0		
80–86	8	43	37						
87–96	7	44	36						
97–103	6	45–46	34						
104–114	5	47	33						
115–127	4	48–49	32						
128–144	3	50–52	31						
145–168	2	53–54	30						
169–206	1	55	29						
207–290	0	56–57	28						

Table 3
Total numbers of papers and citations of different scientists and estimated parameters of Eq. (1) for real $l_n(n)$ data[a].

Scientist	N (N_{max})	L (citations)	l_0 (citations)	δ
M. Kosmulski	114 (140)	1790	142.7 ± 6.2	0.630 ± 0.020
K. Sangwal	131 (152)	1505	104.1 ± 2.4	0.623 ± 0.010
J. Barnaś	206 (290)	3017	219.3 ± 4.8	0.669 ± 0.010
T. Dietl	232 (292)	10,405	798.3 ± 20.1	0.812 ± 0.012[b]
			3534 ± 48	2.40 ± 0.08
S. Krukowski	80 (103)	916	124.0 ± 2.9	0.803 ± 0.002
K.J. Kurzydłowski	169 (268)	962	59.3 ± 1.7	0.611 ± 0.012
Z.R. Żytkiewicz	64 (89)	510	71.5 ± 1.6	0.799 ± 0.016
Q.L. Burrell	50 (61)	633	56.9 ± 5.0	0.556 ± 0.043
L. Egghe	162 (198)	1889	226.3 ± 2.3	0.832 ± 0.007

[a] The R^2 coefficient was between 0.6 and 0.77 for all authors other than SK and TD.
[b] Point corresponding to $n = 1$ was omitted during analysis.

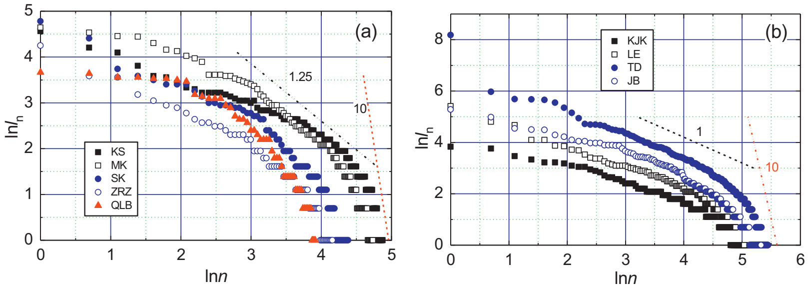

Fig. 1. Plots of $\ln l_n$ against $\ln n$ for papers published by different authors: (a) KS, MK, SK, ZRZ and QLB and (b) KJK, LE, TD and JB. Lines with different slopes are shown for visual reference. With increasing value of $\ln n$ the slope δ increases from zero to values up to 10.

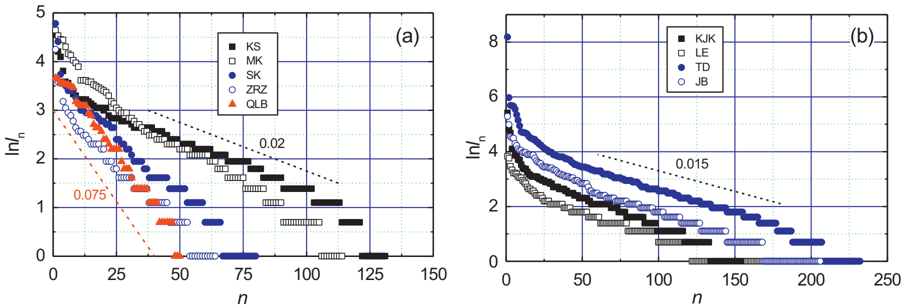

Fig. 2. Plots of $\ln l_n$ against n of papers published by different authors: (a) KS, MK, SK, ZRZ and QLB and (b) KJK, LE, TD and JB. Lines with different slopes are shown for visual reference. Large regions of linear dependence may be seen in these plots.

with a quadratic equation this fitting program gives three "dependencies" with different values. The closer the value of the dependency to unity for a parameter, the better is the fit. We observed empirically that for a simple two-parameter equation this dependency is related to the goodness-of-the-fit parameter R^2, i.e. dependency$^{1/2} \approx R^2$. However, in view of more than one value of the "dependencies" corresponding to the best-fit parameters of a mathematical equations used for the analysis of a given dataset, we considered the lowest value of the "estimated" goodness-of-the-fit parameter R^2. The best-fit values of the parameters of different equations obtained in this way are given in Tables 3–7, where the corresponding standard deviations, denoted by "±" sign, are also included. Since our analysis program does not give the values of the goodness-of-the-fit parameter R^2 directly, the R^2 values "estimated" from the "dependency" of the best-fit parameters are given below the tables merely for reference purposes. The best-fit plots of the $l_n(n)$ data for all authors according to Eq. (5), as an example, are presented in Fig. 4.

It was found that the value of the goodness-of-the-fit parameter R^2 is the highest for the empirical binomial relation (4), it is high and comparable for relations (2) and (5), relatively low for power law (1) and the lowest for the logarithmic function (3). In the case of Eq. (4) it was also observed that the value of l_0 is very sensitive to the value of p and, in general, it was difficult to obtain the best fit with simultaneous variation of all variables for all authors because of the dominance of the third term related to the exponent $2p$. Therefore, during the analysis when the nature of the fitting curve representing the entire data for an author did not change significantly, two or three sets of the l_0, k_1, k_2 and p parameters are given (see Table 6). The exponent p is relatively low and usually lies between 0.03 and 0.22. Moreover, it was found that omission of

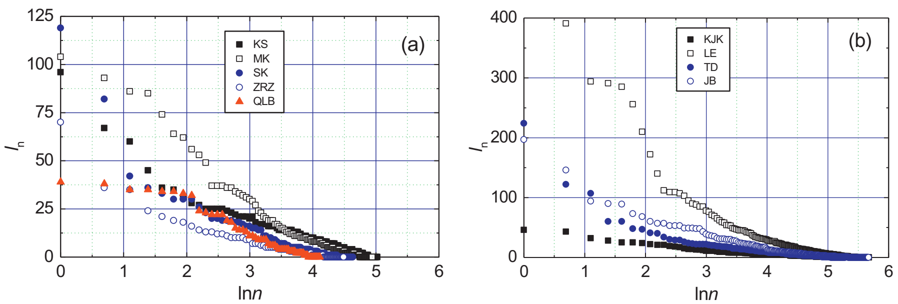

Fig. 3. Plots of l_n against $\ln n$ for papers published by different authors: (a) KS, MK, SK, ZRZ and QLB and (b) KJK, LE, TD and JB.

Table 4
Estimated parameters of stretched exponential function (2) for real $l_n(n)$ data of different scientists.[a]

Scientist	l_0 (cites)	n_0	β
M. Kosmulski	140 ± 4	9.12 ± 0.45	0.606 ± 0.015
K. Sangwal	574 ± 164	0.0555 ± 0.0553	0.211 ± 0.021
J. Barnaś	849 ± 141	0.0188 ± 0.098	0.248 ± 0.016
T. Dietl	2766 ± 659	0.1658 ± 0.1078	0.264 ± 0.216[b]
	$(6.3 \pm 59.0) \times 10^5$	0.0244 ± 0.2524	0.443 ± 0.764
S. Krukowski	550 ± 5913	0.0002 ± 0.0007	0.158 ± 0.037
K.J. Kurzydłowski	93 ± 5	2.8235 ± 0.3875	0.365 ± 0.012
Z.R. Żytkiewicz	1233 ± 781	0.0041 ± 0.0083	0.195 ± 0.033
Q.L. Burrell	40.9 ± 0.8	17.34 ± 0.41	1.280 ± 0.048
L. Egghe	11,676 ± 7827	0.00013 ± 0.00029	0.155 ± 0.022

[a] The R^2 coefficient was between 0.8 and 0.99 for all authors other than TD.
[b] Point corresponding to $n = 1$ was omitted during analysis.

Table 5
Estimated parameters of logarithmic function (3) for real $l_n(n)$ data of different scientists.[a]

Scientist	$l_0 b'$ (cites)	Z_1
M. Kosmulski	93.0 ± 1.9	0.2176 ± 0.0012
K. Sangwal	61.4 ± 1.4	0.2073 ± 0.0014
J. Barnaś	123.1 ± 3.1	0.2094 ± 0.0015
T. Dietl	285 ± 9.7	0.2109 ± 0.0019[b]
	703 ± 89	0.2247 ± 0.0070
S. Krukowski	65.7 ± 2.9	0.2358 ± 0.0030
K.J. Kurzydłowski	36.6 ± 0.7	0.2056 ± 0.0012
Z.R. Żytkiewicz	40.2 ± 1.7	0.2433 ± 0.0030
Q.L. Burrell	50.5 ± 1.3	0.2512 ± 0.0023
L. Egghe	96.3 ± 4.7	0.2151 ± 0.0028

[a] The R^2 coefficient was between 0.30 and 0.46 for all authors other than TD.
[b] Point corresponding to $n = 1$ was omitted during analysis.

Table 6
Estimated parameters of binomial function (4) for real $l_n(n)$ data of different scientists.[a]

Scientist	l_0 (citations)	k_1	k_2	p
M. Kosmulski	267.9 ± 14.2	0.704 ± 0.052	0.124 ± 0.018	0.216 ± 0.017
K. Sangwal	167.0 ± 40.7	0.696 ± 0.142	0.122 ± 0.049	0.206 ± 0.045
	111.4 ± 19.9	0.476 ± 0.109	0.057 ± 0.026	0.285 ± 0.051
J. Barnaś	338.3 ± 44.2	0.677 ± 0.074	0.115 ± 0.025	0.218 ± 0.023
	1084 ± 312	1.219 ± 0.108	0.372 ± 0.066	0.096 ± 0.018
T. Dietl	3994 ± 1783	1.325 ± 0.143	0.440 ± 0.094	0.084 ± 0.022
	21,623 ± 25,556	1.701 ± 0.174	0.723 ± 0.148	0.033 ± 0.021
S. Krukowski	431.5 ± 213.1	1.076 ± 0.216	0.290 ± 0.116	0.144 ± 0.049
	500.9 ± 274.8	1.138 ± 0.226	0.324 ± 0.129	0.131 ± 0.048
K.J. Kurzydłowski	151.0 ± 22.2	0.905 ± 0.075	0.207 ± 0.034	0.155 ± 0.018
	365.4 ± 118.5	1.281 ± 0.113	0.411 ± 0.072	0.083 ± 0.018
	791.4 ± 448.4	1.508 ± 0.150	0.568 ± 0.112	0.051 ± 0.019
Z.R. Żytkiewicz	257.0 ± 119.9	1.077 ± 0.205	0.291 ± 0.110	0.147 ± 0.047
Q.L. Burrell	45.9 ± 1.6	0.091 ± 0.014	0.0021 ± 0.0007	0.778 ± 0.044
L. Egghe	409.1 ± 121.1	0.847 ± 0.150	0.179 ± 0.064	0.183 ± 0.039
	3259 ± 3129	1.555 ± 0.209	0.605 ± 0.163	0.054 ± 0.029

[a] The R^2 coefficient was better than 0.999 for all authors.

the $k_2 n^{2p}$ term leads to deterioration of the fit. This suggests that the empirical binomial equation alone gives the best fit of the $l_n(n)$ data.

Eq. (5) was found to represent the $l_n(n)$ data satisfactorily for all scientists except for Dietl, where the first paper with the exceptionally high number of citations was omitted. The best-fit values of the constants are listed in Table 7. It may be noted from Table 7 that the effectiveness parameter α is essentially unity. An exception here is QLB probably because of small data size. The estimated value of $\alpha \approx 1$ implies that there are only uninhibited citations l_0 and empty sites in the stacking of citations of a paper in the citation column (see Appendix B and Fig. B1). The value of the constant K is enormously different for various authors and its value increases with the initial increasing steepness of the $l_n(n)$ plots (see Fig. 4). Therefore, with $\alpha = 1$ the $l_n(n)$ data were fitted again for these authors and the values of l_0 and K are listed in the table.

It should be mentioned that, except in the case of binomial function, the $l_n(n)$ data of TD gave the best fit only when the first-ranked paper with the exceptionally high number of citations was omitted during the analysis by the functions used in this work. Inclusion of this first-ranked paper with the exceptionally high number of citations always yielded very high l_0

Table 7
Estimated parameters of Langmuir-type function (5) for real $l_n(n)$ data of different scientists.[a]

Scientist	l_0 (cites)	α	K (paper^{-1})	Q
M. Kosmulski	123.6 ± 1.4	1.0576 ± 0.0025	0.138 ± 0.004	2.96
	167.0 ± 5.0	1	0.237 ± 0.049	3.50
K. Sangwal	108.9 ± 4.6	0.9953 ± 0.0035	0.276 ± 0.023	3.74
	105.3 ± 3.8	1	0.252 ± 0.015	3.65
J. Barnaś	224.3 ± 7.3	1.0033 ± 0.0029	0.277 ± 0.018	4.39
	229.4 ± 6.8	1	0.294 ± 0.014	4.45
T. Dietl	683.5 ± 25.8	1.0083 ± 0.0017	3.812 ± 2.522	11.35[b]
	767.8 ± 31.2	1	0.450 ± 0.026	4.88[b]
	$(1.2 \pm 0.7) \times 10^4$	1.0034 ± 0.0029	3.999 ± 2.881	7.06
	1.21×10^7	1	4718 ± 25211	14.14
S. Krukowski	243.8 ± 22.9	0.9992 ± 0.0023	1.093 ± 0.159	4.72
	239.0 ± 18.3	1	1.053 ± 0.115	4.69
K.J. Kurzydłowski	51.8 ± 0.9	1.0127 ± 0.0025	0.161 ± 0.006	3.76
	54.2 ± 0.9	1	0.187 ± 0.005	3.91
Z.R. Żytkiewicz	117.9 ± 7.4	1.0076 ± 0.0026	0.791 ± 0.083	4.25
	135.8 ± 9.3	1	1.031 ± 0.101	4.52
Q.L. Burrell	46.4 ± 1.3	1.320 ± 0.035	0.060 ± 0.006	1.30
	57.3 ± 4.9	1	0.197 ± 0.030	2.49
L. Egghe	607.9 ± 38.8	0.9978 ± 0.0005	1.781 ± 0.154	5.87
	533.5 ± 26.8	1	1.456 ± 0.098	5.66

[a] The R^2 coefficient was between 0.90 and 0.99 for all authors.
[b] Point corresponding to $n = 1$ was omitted during analysis.

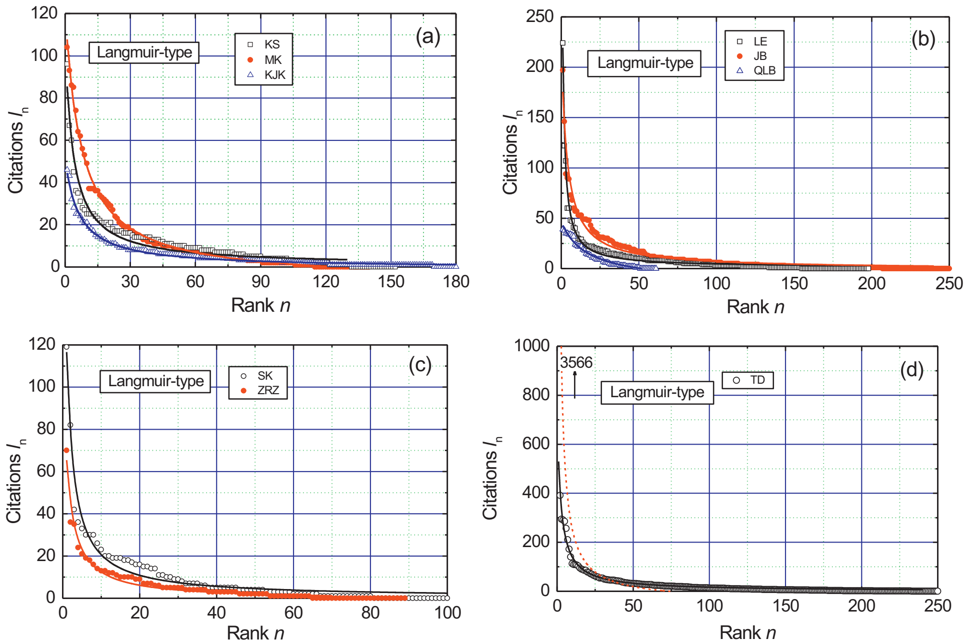

Fig. 4. Best-fit plots of the $l_n(n)$ data for different authors according to Langmuir-type relation (5). Constants used for fitting are listed in Table 7: (a) MK, KS and KJK, (b) JB, LE and QLB, (c) SK and ZRZ, and (d) TD. In (d) continuous curve represents fit for data without first-ranked paper whereas dashed curve represents data for all papers. Somewhat poor fit of the data for KS, SK and TD may be noted in this case. See text for details.

and high or low values of the constant(s) of the equations. For example, when analyzing the data of TD with Langmuir-like function (5), inclusion of the first-ranked paper increases l_0 and K by more than three orders in comparison with those without it (see Table 7).

Regarding the values of l_0 obtained by fitting different functions, one finds that power law (1) and Langmuir-type function (5) give comparable values of l_0 for individual authors, but, as expected since $b' < 1$ (cf. Eq. (B.6), logarithmic relation (3) gives $l_0 b'$ lower than l_0 obtained by analyzing the data using the above relations (cf. Tables 3, 5 and 7). However, although stretched exponential function (2) and empirical binomial relation (4) describe the data satisfactorily, the values of l_0 obtained by them are usually much higher than those obtained by Eqs. (1) and (3) and wide ranges of the values of the constants of Eqs. (2) and (4) give comparable best fit of the $l_n(n)$ data for different authors (see Tables 4 and 6).

From Table 4 it may be observed that, except in the case of QLB, the constant $\beta < 1$ in exponential function (2), as is expected from stretched exponential distribution, but the constant n_0 lies between 0.0001 and 17 for different authors. A comparison of the values of n_0 for two authors, such as SK and KJK or MK and LE, having comparable cumulative citations L shows that its value differs by a factor of 3–5 orders of magnitude. Since $n_0 \approx 0.5\,h$, it is difficult to explain this enormous difference in the value of n_0 for these pairs of authors. The values of n_0 obtained from the $l_n(n)$ data are probably reasonable only in the case of MK and QLB.

The observation that power law (1) and logarithmic function (3) describe the $l_n(n)$ data of individual papers of the authors considered in this study for different authors extremely poorly suggests that preferential attachment (Barabasi & Albert, 1999) or cumulative advantage models (Price, 1965, 1976) and iceberg hypothesis (Guerrero-Bote et al., 2007; Lancho-Barrantes et al., 2010) advanced for the power law and logarithmic distributions, respectively, can be discarded. Although stretched exponential function (2) respresents the $l_n(n)$ data satisfactorily, there is a problem with the interpretation of the values of the parameter n_0 since it is expected that $n_0 \approx 0.5\,h$ (see Section 2). Therefore, stretched exponential function (2) also remains essentially an empirical relation. Binomial function (4), on the other hand, represents the data most satisfactorily with the exponent $p < 1$. However, as discussed in Appendix B, this equation does not follow directly from an adsorption isotherm. This suggests that binomial function (4) is also empirical. As pointed out in Appendix C, binomial relation (4) implies that the number of citations of an author increases with the number of papers published by him/her and is associated with the increasing visibility of the author in the scientific field.

4.2. Comments on Eq. (5) and significance of parameter K

It is interesting to note the equivalence of Eq. (5) with the power law (e.g. Zipf's law) and Tsallis distribution. As observed in the present work, the parameter α in Eq. (5) is approximately 1. Therefore, with $\alpha = 1$, Eq. (5) may be rewritten in the form

$$l_n = \frac{l_0}{1 + Kn}. \tag{8}$$

When $Kn \gg 1$, this reduces to the form of power law (1) with the exponent $\delta = 1$. The form of Eq. (8) is also a special case of Tsallis's distribution (Tsallis & de Albuquerque, 2000)

$$l_n = \frac{l_0}{[1 + (q-1)\lambda n]^{q/q-1}}, \tag{9}$$

where q and λ are empirical parameters. Eq. (9) reduces to (8) for high values of q when $q/(q-1) \to 1$ and $(q-1)\lambda = K$.

The area under an $l_n(n)$ curve defines the total number of citations L, which can be obtained by integrating the equations given in Section 3. Integration of Eq. (5) gives

$$L = l_0 \int_0^N \left[1 - \alpha\left(\frac{Kn}{1+Kn}\right)\right] dn = l_0\left[(1-\alpha)N + \frac{\alpha}{K}\ln(1+KN)\right] \approx \frac{l_0}{K}\ln(1+KN) \tag{10}$$

where we have used the fact that $\alpha \approx 1$ (see Table 7) and N is the number of papers which received citations. Using the values of N given in Table 3 and the values of l_0 and K in Table 7, it can be verified that Eq. (10) predicts cumulative citations L which are in good agreement with those given in Table 3 for different authors.

In Eqs. (8) and (10) the parameter K, defined as Langmuir constant (see Appendix A), characterizes the citation behavior of a given author. The values of K obtained from the analysis of the $l_n(n)$ data for different authors (Table 7) and the number N of papers receiving citations (Table 3) enable to calculate the values of the corresponding dimensionless differential energy Q using Eq. (7). The values of Q calculated in this way are included in Table 7.

In the derivation of Eqs. (8) and (10), it is a priori assumed that all of the $N_{\max}$ papers of an author have the same number of sites available for inhibition and all of them are initially inhibited completely. However, later they either remain inhibited (i.e. $\alpha\theta = 1$) or are uninhibited partially to different extent (i.e. $0 < \alpha\theta < 1$). The situation $\alpha\theta = 1$ implies that a paper has no citations whereas the situation $0 < \alpha\theta < 1$ means that, depending on the value of $\alpha\theta$, the n-ranked paper receives citations $0 < l_n < l_0$. In other words, when relations like Eqs. (3) and (5) based on adsorption isotherms describing the $l_n(n)$ data of an author apply, the number of readers likely to cite a paper remains constant and the so-called linear coverage θ in the citation process is a measure of the citability of a paper. The higher the value of l_0 for an author, the higher is the citability of his/her papers. Similarly, the lower is the value of θ for the nth paper of an author, the higher is its citability.

The process of receiving citations by $N_{\max}$ papers of an author, described above, may be presented graphically, in the order of decreasing citations, as columns containing l_n filled squares followed by $(l_0 - l_n)$ empty squares arranged side-by-side along the citations axis. In other words, corresponding to a given paper rank n, the filled squares occupy lower levels whereas the open squares are placed above the filled ones (see Fig. B1). The resulting diagram is a rectangle of sides l_0 and $N_{\max}$. This type of presentation is similar to that of Ferrers' graphs (Andrews, 1998) but each filled and where the empty squares are of an n-ranked paper represents contribution (citation) 1 and 0, respectively.

Eqs. (3)–(5) are derived on the ideas of inhibition of adsorption sites available on linear steps on the flat surface of a crystal. Eq. (5) based on Langmuir adsorption isotherm assumes that all adsorption sites have the same activity and the dimensionless differential energy $Q = Q_{\text{diff}}/R_G T$ involved during adsorption does not depend on the linear coverage θ (see

Appendix A; Eq. (A.3)). Eq. (3) based on Temkin adsorption isotherm assumes that Q depends on θ (see Eq. (A.4)), whereas Eq. (6) based on Freundlich isotherm assumes that different energies are involved in adsorption at different sites (see Eq. (A.5)). During the adsorption of gas molecules on linear steps of solids Temkin isotherm occurs when the adsorption energy depends on linear coverage θ whereas Freundlich isotherm occurs when adsorption sites in the linear steps have different activity. The fact that, in contrast to Eqs. (3) and (6), Eq. (5) describes the citation rank-order distribution for different authors considered in this work satisfactorily indicates that all active sites are similar. However, as discussed in Appendix C, the form of binomial relation (4), which describes the citation data most satisfactorily, suggests that linear coverage θ given by Eq. (B.1) is more realistic. This form of the linear coverage θ indicates that the adsorption sites are of different activity and their activity decreases with coverage θ.

5. Summary and conclusions

The citation rank-order distribution of papers of different selected authors was analyzed in this work using five functions: power law (1), stretched exponential function (2), logarithmic function (3), binomial function (4) and Langmuir-type function (5). The former two functions have previously been proposed in the literature whereas the remaining three are novel and are derived following the concepts of growth kinetics of crystals in the presence of additives which act as inhibitors of growth. Among these new functions, the logarithmic function (3) proposed for the analysis is similar to that used by Guerrero-Bote et al. (2007) and Lancho-Barrantes et al. (2010) for their iceberg hypothesis, whereas the binomial empirical function (4) is an extension of the theoretical relation (6) with the addition of the empirical k_2n^{2p} term.

The derivation of functions (3) and (5) and function (4) without the empirical k_2n^{2p} term (Eq. (6)) is based on the assumption that all of the $N_{\max}$ papers of an author have the same number of sites available for inhibition and all of them are initially inhibited completely. However, later they either remain inhibited (i.e. $\alpha\theta = 1$) or are uninhibited partially to different extent (i.e. $0 < \alpha\theta < 1$). The situation $\alpha\theta = 1$ implies that a paper has no citations whereas the situation $0 < \alpha\theta < 1$ means that, depending on the value of $\alpha\theta$, the n-ranked paper receives citations $0 < l_n < l_0$. In other words, when relations like Eqs. (3) and (5) based on adsorption isotherms describing the $l_n(n)$ data of an author apply, the number of readers likely to cite a paper remains constant and the so-called linear coverage θ in the citation process is a measure of the citability of a paper. The higher the value of l_0 for an author, the higher is the citability of his/her papers. Similarly, the lower is the value of θ for the nth paper of an author, the higher is its citability.

Analysis of the data of citation rank-order distribution of papers of different authors revealed that the value of the goodness-of-the-fit parameter R^2 is the highest for the empirical binomial relation (4), it is high and comparable for stretched exponential (2) and Langmuir-type functions (5), relatively low for power law (1) but it is the lowest for the logarithmic function (3). The observation that power law (1) and logarithmic function (3) describe the $l_n(n)$ data of individual papers of the authors considered in this study for different authors extremely poorly suggests that preferential attachment (Barabasi & Albert, 1999) or cumulative advantage models (Price, 1965, 1976) and iceberg hypothesis (Guerrero-Bote et al., 2007; Lancho-Barrantes et al., 2010) advanced for the power law and logarithmic distributions, respectively, can be discarded. Although stretched exponential function (2) represents the data satisfactorily, there is a problem with the interpretation of the values of the parameter n_0. Binomial function (4) represents the data most satisfactorily with the exponent $p < 1$. In fact, the value of $p < 1$ is predicted by Eq. (B.5) based on Freundlich adsorption isotherm when the k_2-term is omitted. Addition of the k_2-term to empirical binomial relation (4) is attributed to the increasing visibility of an author in his/her scientific field.

Langmuir-type function (5) not only describes the $l_n(n)$ data for different authors satisfactorily but physical significance can also be assigned to its parameters l_0 and K. It was found that the value of Langmuir constant K is enormously different for various authors and its value increases with the initial increasing steepness of the $l_n(n)$ plots. In fact, in this function the parameter K characterizes the citation behavior of an author.

Based on the Langmuir-type function (5) expression (10) for cumulative citations L relating the extrapolated value of citations l_0 corresponding to rank $n = 0$ for an author and his/her constant K and the number N of paper receiving citation $l \geq 1$ is proposed.

Acknowledgements

The author expresses his gratitude to the anonymous referees for their advice and suggestions. He is also grateful to Dr. K. Wójcik for preparing Figs. A1 and B1.

Appendix A. Adsorption processes and crystal growth

Relations similar to Eqs. (5) and (6) are frequently used to explain the effect of concentration c of additives on the kinetics of growth rates R of faces of crystals from solutions or the vapor phase and displacement rates v of steps on flat faces (Chernov, 1984, chap. 4; Sangwal, 2007, chap. 4). In the case of crystal growth, these relations are derived on the premise that: (1) there are $n_{\max}$ possible sites on a growing face or on a step spreading on the growing faces, (2) a part of the possible sites, say n_{ad}, is occupied by additive particles whereas the remaining $(n_{\max} - n_{\text{ad}})$ sites remain unoccupied, and (3) additive particles occupy the sites following traditional adsorption isotherms. The basic concepts of adsorption isotherms and displacement rates of steps in the presence of inhibitors are given below.

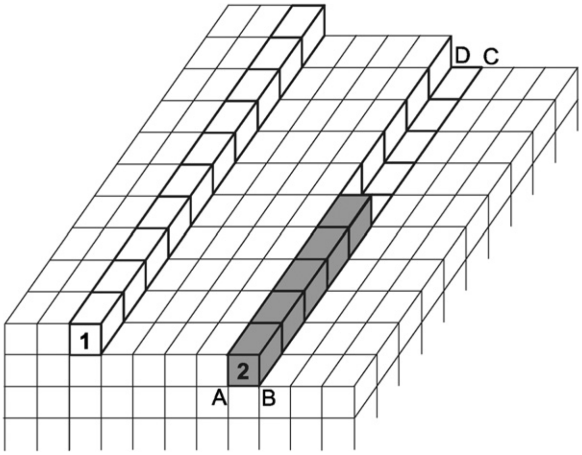

Fig. A1. Schematic illustration of two different types of steps, completed step 1 and developing step 2, on the surface of a crystal composed of square-shaped building blocks. Attachment of new growth units in step 2, shown by dark blocks, growing away from the reader, occurs at its terminal by forming three bonds with growth units: to the left, at the bottom and at the previously attached growth unit before it.

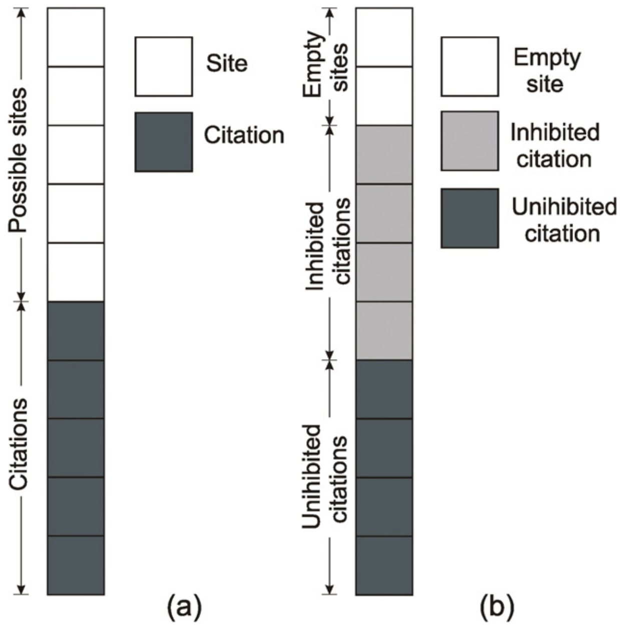

Fig. B1. Development of citations of a particular paper (a) without and (b) with inhibition in citation. In (a) already received citations and empty sites for new citations are shown in analogy with step 2, as seen from above in the form of area ABCD, in Fig. A1. Column of (b) is composed of additional gray squares representing inhibited citations.

For the reaction between additive particle A and site s, given by (Eggers, Gregory, Halsey, & Rabinovitch, 1964, chap. 18)

$$\mathrm{A} + s = \mathrm{A}s, \tag{A.1}$$

the equilibrium constant, usually called as Langmuir constant K, is given by

$$K = \frac{\theta n_{\max}}{(1-\theta) n_{\max} P_{\mathrm{A}}} = \exp\left(\frac{Q_{\mathrm{diff}}}{R_{\mathrm{G}} T}\right), \tag{A.2}$$

where Q_{diff} is the differential heat of adsorption corresponding to the coverage θ, R_{G} is the gas constant, T is the temperature in Kelvin and P_{A} is the pressure of adsorbing particles. From Eq. (A.2) one obtains the surface coverage θ in the form

$$\theta = \frac{KP_{\mathrm{A}}}{1 + KP_{\mathrm{A}}}. \tag{A.3}$$

The constant K has dimensions of pressure^{-1}. However, to calculate the differential heat Q_{diff} of adsorption from Eq. (A.4) one requires dimensionless $K' = KP_A^*$ which can be obtained by defining a dimensionless pressure P_A/P_A^* where P_A^* is the value of pressure when the coverage θ approaches a maximum value θ_{max}.

The surface coverage θ represented by Eq. (A.3) is the Langmuir adsorption isotherm. It is based on the postulate that the adsorption sites have the same activity and the differential heat of adsorption Q_{diff} does not depend on surface coverage θ. When Q_{diff} depends on surface coverage θ, the dependence of θ on P_A is given by Temkin adsorption isotherm (Eggers et al., 1964), i.e.

$$\theta = Z \ \ln C_0 + Z \ln P_A, \tag{A.4}$$

where $Z = 1/b \ln C_0$, with $b < 1$. When the active adsorption sites are of different adsorption energy, the surface coverage θ is given by Freundlich adsorption isotherm (Eggers et al., 1964), i.e.

$$\theta = m\left(\frac{P_A}{P_A^*}\right)^m, \tag{A.5}$$

where m and P_A^* are constants. P_A^* is the value of pressure when θ approaches unity whereas the parameter $m < 1$ and is related to the distribution of energies of adsorption sites.

The above equations are for adsorption on surfaces but they can equally be employed to describe adsorption of additive molecules/atoms on active sites in linear steps occurring on a crystal face during growth. Then the parameter θ represents linear coverage of the linear step by additive molecules/atoms and the adsorption isotherms refer to linear adsorption (Sangwal, 2007).

It should be noted that development of a linear step in a particular direction can take place only by the successive attachment of its building units (i.e. molecules/atoms) at active sites formed on the step. An active site for the attachment of a building unit is produced repeatedly by another building unit attached before. This process of generation of an active site, called kink, at the end of a developing step is shown in Fig. A1. This figure illustrates schematically two different types of steps, completed step 1 and developing step 2, composed of square-shaped building units. In the case of molecules/atoms other than the building units (i.e. additive molecules/atoms), they inhibit the attachment of the building units to the active sites because both additive molecules/atoms and step building units compete for the same active sites. Therefore, in the absence of additive molecules/atoms available for adsorption (i.e. $\theta = 0$), one expects uninhibited development of the linear step. However, for the highest adsorption when $\theta = 1$, maximum inhibition occurs in the development of the step.

Appendix B. Derivation of Eqs. (3)–(5) based on adsorption processes

Following an analogy with the development of a linear step during growth in the absence and presence of additive molecules/atoms described above, the development of citations of a particular paper is illustrated schematically in columns in Fig. B1. The column of Fig. B1a, without inhibition process, is composed of empty and dark squares representing empty sites and uninhibited citations, respectively, whereas the column of Fig. B1b, involving inhibition process, is composed of additional squares representing inhibited sites, respectively. Empty, dark and gray squares representing empty sites, uninhibited citations and inhibited citations are analogs of possible active sites, building units and additive molecules/atoms, respectively, during the development of a linear step during growth.

It should be mentioned that there is a fundamental difference between the stacking of citations in the column of Fig. B1 and the building units in a growth step on the underlying surface of a crystal. In the case of growth steps, the adjoining building units are held together at equal distances by attractive interactions and the terminal building unit in the step serves as a sink for the attachment of a new building unit at the following empty site. In contrast to this, there is no attractive interaction between adjoining citations in the stacking of citations in columns such as those shown in Fig. B1. The stacking of citations and empty sites in columns is merely our presentation of citations in which empty sites are replaced by citations formed outside the previously existing citations.

In order to describe the dependence of citations l_n of papers published by an author on their rank n in the citation-rank sequence of N papers, we arrange the columns of citations of the N papers side-by-side such that the arrangement of citation columns of the papers is similar to Ferrers' graphs (Andrews, 1998). Then we make the following assumptions:

(1) All N papers published by an author receive the same number of l_0 citations.
(2) The l_0 citations are brought by n_{max} possible active sites.
(3) Of the n_{max} sites n_{ad} sites are inhibited and remain only partly accessible for citation whereas the remaining $(n_{max} - n_{ad})$ sites remain completely uninhibited (i.e. they are accessible for citation).
(4) Inhibition process during the process of citation of papers may be described by the adsorption isotherms given above.
(5) The rank n of a paper as a measure of inhibition of the citation probability of the paper such that the pressure P_A of the adsorbing particles in the adsorption isotherms may be replaced by the paper rank n.

Assumption (5) means that the citations l_n of a paper of rank n decreases with increasing n whereas l_0 citations are received by the paper with rank $n = 0$.

Denoting the number of citations of a paper in Fig. B1b without and with inhibition by l_0 and l_i, respectively, the resulting number of citations l_n of paper with rank n in the presence of inhibition with linear coverage $\theta = n_\mathrm{ad}/n_\mathrm{max}$ of inhibited sites may be given by (cf. Chernov, 1984; Sangwal, 2007)

$$l_n = l_0(1-\theta) + l_i\theta = l_0(1-\alpha\theta), \tag{B.1}$$

where the effectiveness parameter

$$\alpha = \left(1 - \frac{l_i}{l_0}\right), \tag{B.2}$$

which lies between 0 and 1, depending on the value of l_i/l_0. When $l_i = 0$, $\alpha = 1$. The situation $\alpha = 1$ implies that there are no inhibited citations, and a column is entirely composed of uninhibited citations and empty sites, as in Fig. B1a.

On substituting for θ from Eqs. (A.3)–(A.5) in (B.1), one obtains

$$l_n = l_0\left[1 - \alpha\left(\frac{Kn}{1+Kn}\right)\right], \tag{B.3}$$

$$l_n = l_0 b'(1 - Z_1 \ln n), \tag{B.4}$$

$$l_n = l_0(1 - \alpha_1 n^m), \tag{B.5}$$

where the new parameters are given by

$$b' = (1 - \alpha Z \ln C_0) < 1, \quad Z_1 = \frac{\alpha Z}{1 - \alpha Z \ln C_0}, \tag{B.6}$$

$$\alpha_1 = \frac{\alpha m}{n^{*m}}. \tag{B.7}$$

In Eq. (B.3) the constant K has dimensions of paper-rank^{-1} and n^* is the maximum number of papers when θ approaches unity.

Using an argument similar to that advanced above to define dimensionless Langmuir constant $K' = KP_\mathrm{A}^*$ for gas adsorption, one can define a dimensionless Langmuir constant $K' = KN$, where N is the number of papers receiving citations when the coverage $\theta = 1$. However, calculation of the differential heat Q_diff of adsorption from Eq. (A.2) does not carry sense because this equation contains the $R_\mathrm{G}T$ term which has no significance in the citation process. Therefore, one may define a new dimensionless differential heat Q of adsorption for citation such that $Q = Q_\mathrm{diff}/R_\mathrm{G}T$.

Appendix C. Derivation of binomial relation (4)

Freundlich adsorption isotherm described by Eq. (A.5) of Appendix A is based on the postulate that surface coverage θ by adsorbing gas molecules involves active adsorption sites of different adsorption energy. Now we assume that the total surface coverage θ follows the empirical relation

$$\theta = \theta_0(1 - z\theta_0), \tag{C.1}$$

where θ_0 denotes the coverage given by the traditional Freundlich isotherm (Eq. (A.5)) and z is an empirical parameter ($0 \le z \le 1$). The parameter z essentially determines the contribution of the correction term $(1 - z\theta_0)$ to the total coverage θ_0.

Substitution of the value of surface coverage θ_0 given by Freundlich isotherm (A.5) in Eq. (C.1) gives the total surface coverage θ in the form

$$\theta = m\left(\frac{P_\mathrm{A}}{P_\mathrm{A}^*}\right)^m \left[1 - zm\left(\frac{P_\mathrm{A}}{P_\mathrm{A}^*}\right)^m\right]. \tag{C.2}$$

As mentioned in Appendix A, Eq. (C.2) also holds for adsorption on a linear step. In the case of citations, we replace the adsorbing gas pressure P_A in Eq. (C.2) by paper rank n. Then substitution of the value of θ from Eq. (C.2) in (B.1) gives

$$l_n = l_0(1 - k_1 n^p + k_2 n^{2p}), \tag{C.3}$$

where $k_1 = \alpha_1$ and is given by Eq. (B.7), $k_2 = z\alpha_1{}^2/\alpha$ and $p = m$. Note that this relation holds when the values of the parameters k_1 and k_2 remain constant for a citation distribution. Physically, relation (C.3) implies that the number of citations of an author increases with the number of papers published by him/her and is associated with the increasing "visibility" of the author in his/her scientific field.

References

Albert, R., & Barabasi, A. L. (2002). Statistical mechanics of complex networks. *Rev. Mod. Phys.*, *74*(1), 47–97.

Andrews, G. E. (1998). *The theory of partitions*. Cambridge: Cambridge University Press.

Barabasi, A. L., & Albert, R. (1999). Emerging of scaling in random networks. *Science*, *286*(5439), 509–512.

Bornmann, L., & Daniel, H.-D. (2009). Universality of citation distribution – A validation of Radicchi et al.'s relative indicator $c_\mathrm{f} = c/c_0$ at the micro level using data from chemistry. *Journal of the American Society for Information Science and Technology*, *60*(8), 1664–1670.

Burrell, Q. L. (2001). Stochastic modeling of the first-citation distribution. *Scientometrics*, *52*(1), 3–12.
Burrell, Q. L. (2002). The *n*th-citation distribution and oblescence. *Scientometrics*, *53*(3), 309–323.
Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, *51*(4), 661–703.
Chernov, A. A. (1984). *Modern crystallography III: Crystal growth*. Berlin: Springer.
Companario, J. M. (2010). Distribution of ranks of articles and citations in journals. *Journal of the American Society for Information Science and Technology*, *61*(2), 419–423.
Eggers, D. F., Gregory, N. W., Halsey, G. D., & Rabinovitch, B. S. (1964). *Physical chemistry*. New York: Wiley.
Egghe, L. (2009). A rationale for the Hirsch-index rank-order distribution and a comparison with the impact factor rank-order distribution. *Journal of the American Society for Information Science and Technology*, *60*(10), 2142–2144.
Egghe, L. (2011). The impact factor rank-order distribution revisited. *Scientometrics*, *87*(3), 683–685.
Egghe, L. (2012). Study of rank- and size-frequency functions in the case of power law growth of sources and items and proof of Heaps' law. *Information Processing and Management*, 1–2. http://dx.doi.org/10.1016/ipm.2012.02.004
Egghe, L., & Rousseau, R. (2006). An informetric model for the Hirsch-index. *Scientometrics*, *69*(1), 121–129.
Egghe, L., & Rousseau, R. (2012). The Hirsch index of a shifted Lotka function and its relation with the impact factor. *Journal of the American Society for Information Science and Technology*, *63*(5), 1048–1053. http://dx.doi.org/10.1002/asi.22617
Egghe, L., & Waltman, L. (2011). Relations between the shape of a size-frequency distribution and the shape of a rank-frequency distribution. *Information Processing and Management*, *47*(2), 238–245.
Guerrero-Bote, V. P., Zapico-Alonso, F., Espinosa-Calvo, M. E., Gomez-Crisostomo, R., & Moya-Anegon, F. (2007). Import–export of knowledge between scientific subject categories: The iceberg hypothesis. *Scientometrics*, *71*(3), 423–441.
Gupta, H. M., Campanha, J. R., & Schinaider, S. J. (2008). Size limiting in Tsallis statistics. *Physica A*, *387*(27), 6745–6751.
Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572.
Kretschmer, H., & Rousseau, R. (2001). Author inflation leads to a breakdown of Lotka's law. *Journal of the American Society for Information Science and Technology*, *52*(8), 610–614.
Lancho-Barrantes, B. S., Guerrero-Bote, V. P., & Moya-Anegon, F. (2010). The iceberg hypothesis revisited. *Scientometrics*, *85*(2), 443–461.
Leherrere, J., & Sornette, D. (1998). Stretched exponential distributions in nature and economy "fat tails" with characteristic scales. *European Physical Journal B*, *8*(2), 535–539.
Mansilla, R., Köppen, E., Cocho, G., & Miramontes, P. (2007). On the behavior of journal impact factor rank-order distribution. *Journal of Informetrics*, *1*(2), 155–160.
Nadarajah, S., & Kotz, S. (2007). Models for citation behavior. *Scientometrics*, *72*(2), 291–305.
Naranan, S. (1971). Power law relations in science bibliography: A self-consistent interpretation. *Journal of Documentation*, *27*(2), 83–97.
Naumis, G. G., & Cocho, G. (2007). The tails of rank-size distributions due to multiplicative processes, from power laws to stretched exponentials and beta-like functions. *New Journal of Physics*, *9*(00), 286–293.
Perc, M. (2010). Zipf's law and log-normal distributions in measures of scientific output across fields and institutions: 40 years of Slovenia's research as an example. *Journal of Informetrics*, *4*(2), 358–364.
Price, D. D. S. (1965). Networks of scientific papers. *Science*, *149*(3683), 510–515.
Price, D. D. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science and Technology*, *27*(5), 292–306.
Radicchi, F., Fortunado, S., & Castellano, C. (2008). Universality of citation distribution: Towards an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(45), 17268–17272.
Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B*, *4*(1), 131–134.
Redner, S. (2005). Citation statistics from 110 years of physical review. *Physics Today*, *58*(1), 49–54.
Sangwal, K. (2007). *Additives and crystallization processes: From fundamentals to applications*. Chichester: Wiley.
Sangwal, K. (2011). On the growth of citations of publication output of individual authors. *Journal of Informetrics*, *5*(4), 554–564.
Sangwal, K. (2012a). Progressive nucleation mechanism for the growth behavior of items and its application to cumulative papers and citations of individual authors. *Scientometrics*, *92*(3), 575–591. http://dx.doi.org/10.1007/s11192-011-0610-8
Sangwal, K. (2012b). Application of progressive nucleation mechanism for the citation behavior of individual papers of different authors. *Scientometrics*, *92*(3), 643–655. http://dx.doi.org/10.1007/s11192-011-0564-x
Seglen, P. O. (1992). The skewness of science. *Journal of the American Society for Information Science and Technology*, *43*(9), 628–638.
Simkin, M. V., & Roychowdhury, V. P. (2007). A mathematical theory of citing. *Journal of the American Society for Information Science and Technology*, *58*(11), 1661–1673.
Tsallis, C., & de Albuquerque, M. P. (2000). Are citations of scientific papers a case of nonextensivity? *European Physical Journal B*, *13*, 778–780.
van Raan, A. F. J. (2001). Two-step competition process leads to quasi power law income distribution. Application to scientific publication and citation distribution. *Physica A*, *298*(3–4), 530–536.
Vieira, E. S., & Gomes, J. A. N. F. (2010). Citations of scientific articles: Its distribution and dependence on the article features. *Journal of Informetrics*, *4*(1), 1–13.
Wallace, M. L., Lariviere, V., & Gingras, Y. (2009). Modeling a century of citation distributions. *Journal of Informetrics*, *3*(1), 296–303.