Progressive nucleation mechanism for the growth behavior of items and its application to cumulative papers and citations of individual authors

Keshra Sangwal

Abstract A general expression based on the concepts of the progressive nucleation mechanism is proposed in the form $\alpha(t) = \frac{N(t)}{C} = \left[1 - \exp\left\{-\left(\frac{t}{\Theta}\right)^q\right\}\right]$ to describe the growth behavior of items in an individual system and a collective of systems. In the above relation, $\alpha(t)$ is the ratio of items $N(t)$ at time t to the maximum number C of possible items for the system, Θ is the corresponding time constant and q is the exponent. The above relation is then used to analyze: (1) the growth behavior of cumulative number $N(t)$ of papers published by individual authors and cumulative citations $L(t)$ of $N(t)$ papers of an author as a function of citation duration t, and (2) the relationship between cumulative citations $L(t)$ of papers and cumulative number $N(t)$ of papers. The proposed approach predicts that: (1) the fraction of items produced by successive systems is additive, (2) the cumulative fraction $\alpha_{sum}(t)$ of maximum number of sites is the sum of contributions of fractions of maximum number of items produced by different systems, and (3) the values of time constants Θ and exponent q increase with the addition of fraction of items produced by subsequent systems, but their values are the lowest for individual systems. The approach is applied to explain the growth behavior of cumulative $N(t)$ papers and $L(t)$ citations of four selected Polish professors.

Keywords Cumulative citations · Cumulative papers · Growth behavior of items · Individual authors · Progressive nucleation mechanism (PNM)

Introduction

It is well known that the growth of journals, articles and authors in different scientific fields occurs at a relatively slow rate initially, followed by an exponential increase, and, in some cases, finally the growth declines after a certain time, giving rise to a sigmoidal shaped (S-shaped) curve. In order to analyze the above type of growth behavior a variety of

K. Sangwal (✉)
Department of Applied Physics, Lublin University of Technology,
ul. Nadbystrzycka 38, 20-618 Lublin, Poland
e-mail: k.sangwal@pollub.pl

models have been developed and applied over years. Among the different equations of various models, power-law, exponential and logistic functions are commonly used (De Solla Price 1963; Egghe and Ravichandra Rao 1992; Gupta et al. 1995, 2002, Ravichandra Rao and Srivastava 2010, Wong and Goh 2010). Recently, Sangwal (2011a, b) applied a new equation, based on progressive nucleation mechanism (PNM) of a solid phase during its crystallization in a closed liquid system of fixed volume.

Using the new equation, the author analyzed the growth of the number of citations per year, denoted hereafter as $\Delta L(t)$, of individual authors (Sangwal 2011a), the cumulative growth of $N(t)$ articles, in three randomly selected databases in humanities, social sciences and science and technology (Sangwal 2011b) and the cumulative growth of $J(t)$ journals, $N(t)$ papers and $W(t)$ authors in malaria research (Sangwal 2011b). Analysis of the former data revealed that: (1) PNM describes the data better than the power-law relation, (2) the field of social sciences is saturated much earlier than science and technology but publication activity in humanities is saturated much later, and (3) that social sciences have the maximum growth, followed by lower growth in humanities and the lowest growth in science and technology. It was also observed that: (1) the data on $J(t)$ journals, $N(t)$ papers and $W(t)$ authors against publication year Y in malaria research can be described equally well by equations of the power law and PNM, and (2) the growth of journals $J(t)$ and articles $N(t)$ is intimately connected with the growth of authors $W(t)$.

The basic concepts of the PNM are well developed in the field of crystallization of a phase in crystallization medium (Kashchiev 2000). The basic idea of the PNM invovles the formation of crystallites progressively in the crystallizing volume. However, while analyzing different types of data using the PNM equation, the author (Sangwal 2011a, b) assumed ad hoc that the number $\Delta L(t)$ of citations per year of individual authors and cumulative growth of $J(t)$ journals, $N(t)$ papers and $W(t)$ authors in a scientific field are the analogs of crystallites. Despite the success of the equation based on PNM in describing the above data, the similarity between growth of crystallites during crystallization and different types of items in individual systems as well as collectives of systems remains unclear. For example, one can imagine a similarity between crystallites in a crystallization medium and $J(t)$ journals in a research area, but it remains obscure why the number $N(t)$ of papers published in different $J(t)$ journals of a research field should also follow the PNM equation.

The aim of the present study is twofold: (1) to develop the basic concepts of the progressive nucleation mechanism to describe the growth behavior of items in an individual system and a collective of systems, and (2) analyze the growth behavior of cumulative number $N(t)$ of papers $L(t)$ citations of individual authors.

The structure of the paper is as follows. In the following section the basic concepts and equations of the growth behavior of items, based on PNM described in Appendix, are presented. Then the citation data of the total publication output of four arbitrarily selected Polish professors are presented. These data are analyzed thereafter using the concepts and equations based on the PNM. Finally, the main findings of the study and the unclear aspects of the PNM as applied to the growth behavior of items are summarized.

Basic concepts and equations for growth behavior of items

Processes of the publication of articles by an author in journals, the citation of a particular article of a given author by other authors, the number of authors engaged in a research work in a given field or the number of cars produced by a factory since its inception are basically similar. Here articles published by an author, citations of a particular article,

researchers in a given field or cars produced by a factory are the analogs of crystallites forming during crystallization in a closed system (see Appendix). In the above cases, an author of the articles, an article fetching the citations, a research field and a car factory are the individual closed systems. However, individual closed systems producing similar or dissimilar items also constitute a closed group or collective of individual systems. For the collective of systems the total number of items produced during a given duration is the sum of items produced by all individual systems during that duration.

The process of growth of N items produced in a system is represented schematically in Fig. 1 as a function of time t. If N_1 and N_2 are the number of items at t_1 and t_2, respectively, from the change $(N_2 - N_1)$ in the number of items during the time interval $(t_2 - t_1)$ one may define the production rate v of item as

$$v = \frac{N_2 - N_1}{t_2 - t_1} = \frac{\Delta N}{\Delta t}, \tag{1}$$

where v (items/time) is constant during Δt. From Eq. 1 one may define instantaneous rate v_{ins} and average rate v_{av}:

$$v_{ins} = \lim_{\Delta t \to 0} \frac{\Delta N}{\Delta t} = \frac{dN}{dt}, \tag{2}$$

$$v_{av} = \sum_{i=1}^{n} \Delta N_i \bigg/ \sum_{i=1}^{n} \Delta t_i, \tag{3}$$
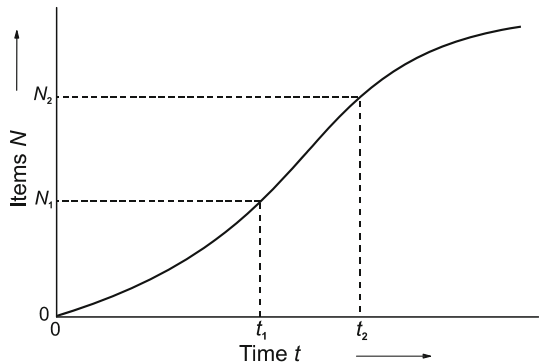
where ΔN_i is the number of items produced by an individual closed system i during Δt_i and n is the sum of intervals. In the above equations

$$N(t) = \sum_{i=1}^{n} \Delta N_i, \tag{4}$$

and when $\Delta t_i = 1$, $v_i = \Delta N_i$. It should be noted that the units of N and v are: item and item/time, respectively. With the above basic background well known in kinematics, we discuss the growth behavior of items produced in individual systems and in collectives of systems.

In the case of growth of items (such as articles, citations, authors and cars in the above examples) with time t in an individual system since the year Y_0, we assume that each real item has an imaginary volume such that the total volume of $N(t)$ items is $V_c(t)$ at time t and C is the maximum number of items possible in the volume V of the system. Denoting the

Fig. 1 Schematic presentation of growth of N items produced in a system is as a function of time t

number N_i of items produced by an individual closed system i at time t, the fraction $\alpha_i(t)$ of items $N_i(t)$ at time t may be given by (cf. Eq. 12)

$$\alpha_i(t) = \frac{N_i(t)}{C_i} = \left[1 - \exp\left\{-\left(\frac{t}{\Theta_i}\right)^{q_i}\right\}\right], \qquad (5)$$

where C_i is the maximum number of possible items for the system i, Θ_i the corresponding time constant and q_i is the exponent. The time constant Θ_i and the exponent q_i are given by Eqs. 22 and 14, respectively. As in the case of PNM for crystallization. Eq. 5 for individual systems predicts that $1 < q < 2.5$ and $1 < q < 4$ when the items grow to three-dimensional entities of visible size (i.e. when $d = 3$) by volume diffusion and mass transfer, respectively. When the items are already of visible size at the time of their nucleation, the exponent $q = 1$, which is the lowest limit for the validity of PNM.

In the case of a collective of n systems, the cumulative fraction α of items may be written as

$$\alpha(t) = \sum_{i=1}^{n} \alpha_i(t) = \sum_{i=1}^{n} \frac{N_i(t)}{C_i}, \qquad (6)$$

where $1 < i < n$. When a new individual system becomes active after a time lag Δ, the cumulative fraction $\alpha_{\text{sum}}(t)$ of items may be given by

$$\alpha_{\text{sum}}(t) = \sum_{i=1}^{n} \alpha_i(t) = \sum_{i=1}^{n} \frac{N_i(t)}{C_i} = \sum_{i=1}^{n} \left[1 - \exp\left\{-\left(\frac{t - (i-1)\Delta}{\Theta_i}\right)^{q_i}\right\}\right]. \qquad (7)$$

Equation 7 does not have a simple solution but can be solved numerically. This can be done by summing up all data of items generated by using Eq. 5 for systems characterized by different values of time constant Θ_i and exponent q_i, which become active successively with time lag Δ, to obtain cumulative fraction $\alpha_{\text{sum}}(t)$. Examples of the time dependence of the cumulative fraction $\alpha_{\text{sum}}(t)$ of items calculated in this way by adding contributions from systems are shown in Figs. 2, 3 and 4 for different values of parameters Δ, q_i and Θ_i, respectively. The points represent the $\alpha_{\text{sum}}(t)$ data of items produced as a function of time according to Eq. 7. The values of Δ and Θ_i are taken in arbitrary units abbreviated hereafter as a.u.

Figure 2 presents the cumulative fraction $\alpha_{\text{sum}}(t)$ of items produced by $n = 6$ systems of the same time constant $\Theta_i = 5$ a.u. as a function of time t. Among the six systems, the first five systems are considered to become active successively after equal time intervals $\Delta = 2$ a.u., whereas the last system becomes active after time interval $\Delta = 4$ a.u. Obviously, in

Fig. 2 Dependence of cumulative fraction $\alpha_{\text{sum}}(t)$ of items from different active systems as a function of time t for different values of parameter Δ, with $q_i = 2$ and $\Theta_i = 5$. See text for details
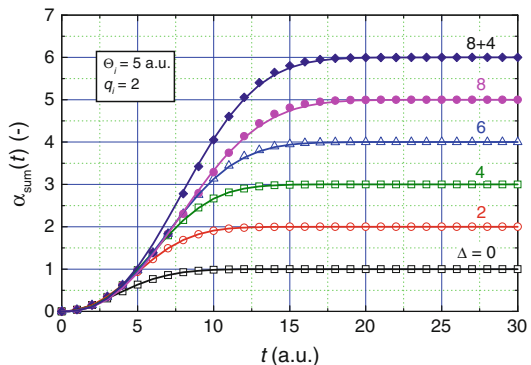
Fig. 3 Dependence of cumulative fraction $\alpha_{sum}(t)$ of items from different active systems as a function of time t for different values of parameters q_i with $\Delta = 0$ and $\Theta_i = 5$. See text for details
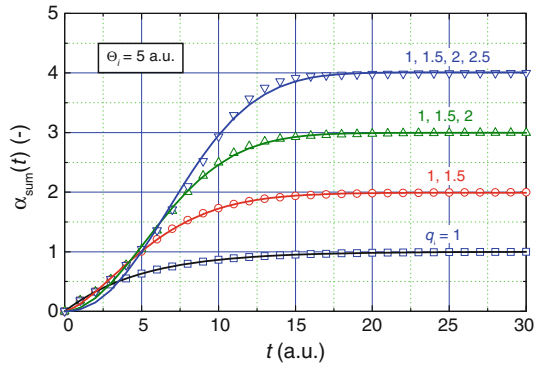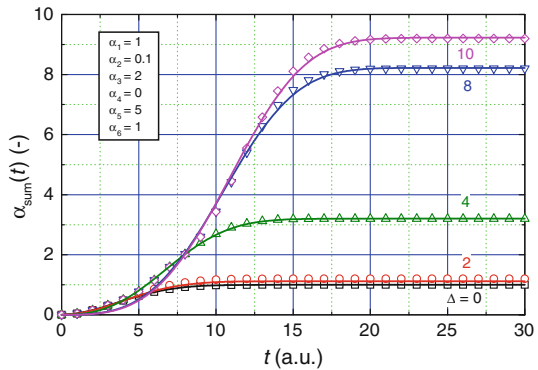

Fig. 4 Dependence of cumulative fraction $\alpha_{sum}(t)$ of items as a function of time t for unequal contributions of different systems and different values of parameters Δ, with $q_i = 2$ and $\Theta_i = 5$. See text for details

this case the time lags $(i - 1)\Delta$ when the systems 1, 2, 3, 4, 5 and 6 become active are 0, 2, 4, 6, 8 and 12 a.u. These values of $(i - 1)\Delta$ are indicated alongside the plots. Figure 3 presents the time dependence of cumulative fraction $\alpha_{sum}(t)$ of items from $n = 4$ systems of the same time constant $\Theta_i = 5$ a.u., which become active successively at time t equal to 0, 2, 4 and 6 a.u. and are characterized by the exponent q_i equal to 1, 1.5, 2 and 2.5, respectively. The values of q_i for the cumulative data are given alongside the plots. Figure 4 illustrates the cumulative fraction $\alpha_{sum}(t)$ of items from $n = 6$ systems characterized by the same time constant $\Theta_i = 5$ a.u. and $q_i = 2$ as a function of time t. The systems i become active at t equal to 0, 2, 4, 6, 8 and 10 a.u., but the contributions α_1, α_2, α_3, α_4, α_5 and α_6 of successive 6 systems to the cumulative fraction $\alpha_{sum}(t)$ of items are different as given in the inset whereas the values of Δ are given alongside the plots.

The data points generated above in Figs. 2, 3, 4 were fitted according to the relation

$$\alpha_{sum}(t) = \sum_{i=1}^{n} \alpha_i(t) = \alpha_0 \left[1 - \exp\left\{ -\left(\frac{t}{\Theta_0} \right)^{q_0} \right\} \right], \quad (8)$$

where Θ_0 and q_0 are new time constant and exponent describing the resultant growth behavior of the entire collective of systems, and the sum of all maximum fractions α_0 is defined as

$$\frac{\alpha_{sum}(t)}{\alpha_0} = \frac{N_{sum}(t)}{C_0}, \quad (9)$$

where the ratio $N_{sum}(t)$ is the sum of items produced by the collective of n systems at time t and C_0 is sum of the maximum numbers of items in the collective. The plots of the data of

the figures are drawn with the best-fit values of the constants of Eq. 8 for the data of the above figures are given Table 1. The data were analyzed using Origin software (version 4.1); see next section.

From Figs. 2, 3, 4 and Table 1 the following features may be noted:

(1) The fraction of items produced by successive systems is additive.
(2) Cumulative fraction $\alpha_{sum}(t)$ of maximum number of sites is the sum of contributions of fractions of maximum number of items produced by different systems.
(3) The fit of the generated data according to Eq. 8 is extremely good even in the case of highly different combinations of parameters Δt_i, q_i and Θ_i (see Table 1). This is due to the fact that a positive deviation of $\alpha_{sum}(t)$ at low values of t is somewhat compensated by a negative deviation of $\alpha_{sum}(t)$ at relatively high values of t. This feature of the mutual compensation of positive and negative deviations of $\alpha_{sum}(t)$ at low and high values of t may be noted from the plots for large values of $\alpha_{sum}(t)$ in Figs. 3 and 4.
(4) The values of constants Θ_0 and q_0 increase with the addition of fraction of items produced by subsequent systems.

From the above discussion it may be concluded that the growth of cumulative number N_i of items produced by an individual source i and the cumulative number N_{sum} of items produced by a collective of sources can be described by Eqs. 5 and 8, respectively. These equations are identical in form and are characterized by three parameters: the maximum number C_i or C_0 of items likely to be produced by the individual system i or collective of systems (cf. Eq. 9), and the corresponding time constant Θ_i or Θ_0 and the exponents q_i or q_0. This means that basically it is the same equation for the analysis of the growth behavior of cumulative papers N, cumulative citations L_i of an individual paper i and cumulative citations L_{sum} of all papers of an author. The former two cases are examples of individual systems, whereas the last one is an example of a collective of systems. This relation also explains the growth and decay of the number ΔL of citations per year of individual papers published by an author (Sangwal 2012).

Table 1 Best-fit values of constants of Eq. 8 for the data of Figs. 2, 3, 4

Figure	Constants	Variable	α_0	Θ_0	q_0
2	$\Theta_i = 5$, $q_i = 2$	$\Delta = 0$	1	5	2
		2	1.9992 ± 0.0008	6.1208 ± 0.0027	2.3074 ± 0.0072
		4	3.0001 ± 0.0018	7.2682 ± 0.0099	2.4510 ± 0.0106
		6	4.0048 ± 0.0049	8.4373 ± 0.0021	2.4801 ± 0.0194
		8	5.0153 ± 0.0106	9.6235 ± 0.0371	2.4497 ± 0.0288
		$8 + 4$	6.0109 ± 0.0117	9.5114 ± 0.0336	2.5554 ± 0.0288
3	$\Theta_i = 5$, $\Delta = 0$	$\Delta q_i = 1$	1	5	1
		1, 1.5	1.9909 ± 0.0037	6.2790 ± 0.0345	1.5064 ± 0.0168
		1, 1.5, 2	2.9958 ± 0.0107	7.5391 ± 0.0649	1.9254 ± 0.0415
		1, 1.5, 2, 2.5	4.0107 ± 0.0225	8.7813 ± 0.0100	2.2142 ± 0.0707
4	$\Theta_i = 5$, $q_i = 2$	$\Delta = 0$	1	5	2
		2	1.9999 ± 0.0001	5.3789 ± 0.0018	2.0589 ± 0.0019
		4	3.2012 ± 0.0060	8.0035 ± 0.0305	2.6488 ± 0.0343
		8	8.2253 ± 0.0393	11.7298 ± 0.0778	3.2738 ± 0.0855
–		10	9.2275 ± 0.0449	12.2145 ± 0.0788	3.3467 ± 0.0861

As described above, according to the PNM for crystallization in a given system defined by nucleation event alone, $q = 1$ (see Appendix). Thus, a natural consequence of the above observation that $q > 1$ for a group of systems characterized by different parameters producing items does not always mean that nuclei of items grow after their formation.

Citation data of selected authors

We used Thomson Reuters' ISI Web of Knowledge (Web of Science) to collect and analyze the citations of the publication output of four Polish professors: T. Dietl (affiliations: Institute of Physics, Polish Academy of Sciences, Warsaw, and University of Warsaw), J. Barnaś (affiliations: Adam Miśkiewicz University, Poznań, and Institute of Molecular Physics, Polish Academy of Sciences, Poznań), M. Kosmulski (Lublin University of Technology) and K. Sangwal (Lublin University of Technology). The basic bibliometric data involving the number of papers $\Delta N(t)$ published per year and the number $\Delta L(t)$ of citations with self-citations per year collected from the above database are given in Table 2 for the above authors. Here $\Delta N(t)$ and $\Delta L(t)$ are the increments in the values of the cumulative number N of papers and the cumulative number L of citations in the time interval between t and $(t - 1)$. The data were collected on 19–20 December 2010.

The total number N of papers ($N = \Sigma \Delta N(t)$) and their citations L ($L = \Sigma \Delta L(t)$) are: $N = 289, L = 10398$ for T. Dietl; $N = 257, L = 2509$ for Barnaś; $N = 139, L = 1789$ for M. Kosmulski; whereas $N = 149, L = 1505$ for K. Sangwal. The publications of these professors have spanned over a period t varying from 27 to 40 years. The publication period $t = Y - Y_0$, where Y_0 is the year of publication of the first paper whereas Y is the year of publication under consideration. The data were analyzed using Origin software (version 4.1). The procedure followed for the analysis of the data is described elsewhere (Sangwal 2011a). In cases when it was difficult to establish best fits, two different sets of the values of the constants were recorded. Examples of this type of sets of the constants are the $N(t)$ and $L(t)$ data of Barnaś given in Tables 3 and 4.

Results and discussion

Growth behavior of cumulative number $N(t)$ of papers

The number of papers published by an author is a typical case of an individual system in which the papers are the items produced in the system and C_i is the maximum number of possible sites in the system i. Then the traditional relation Eq. 5 following from the progressive nucleation mechanism may be applied. Figure 5 shows the data on the growth behavior of cumulative number $N(t)$ of papers published by the four authors whereas the curves are drawn with the best-fit values of the constants given in Table 3.

It may be seen from Fig. 5 that the data on the cumulative number of papers for all of the authors can be represented by Eq. 5 practically in the entire range. This smooth increase in the publication of papers by the authors is associated with a steady rate of publication of papers during their entire publication period.

According to PNM, the parameter q_i is a measure of the sluggish or fast growth of papers, time constant Θ_i is a measure of nucleation rate J_s (see Eq. 22), whereas the constant C_i defines the maximum possible number of papers for an author. Table 3 shows

Table 2 Papers N and citations ΔL of different authors

Year	T. Dietl		J. Barnaś		M. Kosmulski		K. Sangwal	
	N	ΔL	N	ΔL	N	ΔL	N	ΔL
2010	291	10,398	257	2,509	139	1,789	149	1,505
2009	278	9,314	240	2,130	130	1,570	145	1,384
2008	274	8,287	219	1,809	118	1,385	140	1,282
2007	260	6,978	192	1,415	117	1,195	139	1,189
2006	244	5,732	165	1,118	106	1,050	135	1,080
2005	227	4,670	133	931	101	930	134	1,000
2004	214	3,634	119	735	95	769	132	918
2003	197	2,907	109	626	86	658	126	814
2002	183	2,314	80	499	78	535	119	739
2001	170	1,825	68	391	68	458	115	661
2000	156	1,522	61	302	60	387	112	593
1999	147	1,399	55	240	57	326	106	525
1998	140	1,299	49	173	47	216	98	478
1997	126	1,176	42	123	46	179	94	435
1996	118	1,078	32	79	41	154	87	394
1995	108	954	24	43	37	122	81	355
1994	85	876	17	12	36	97	77	330
1993	76	810	10	1	33	74	74	315
1992	72	732	5	0	28	57	72	287
1991	66	650	3	0	24	47	70	273
1990	60	579	3	0	22	40	67	256
1989	50	506	3	0	16	29	63	212
1988	42	441	3	0	13	26	60	198
1987	36	325	3	0	12	24	57	176
1986	32	272	3	0	10	20	51	153
1985	24	194	3	0	8	12	46	135
1984	20	138	1	0	4	4	43	121
1983	17	108	–	–	2	2	37	96
1982	11	66	–	–	1	2	34	86
1981	10	48	–	–	1	2	24	54
1980	8	29	–	–	1	1	24	46
1979	8	21	–	–	1	0	21	25
1978	8	15	–	–	1	0	13	6
1977	4	5	–	–	–	–	7	3
1976	4	2	–	–	–	–	5	3
1975	1	0	–	–	–	–	4	2
1974	–	–	–	–	–	–	4	1
1973	–	–	–	–	–	–	2	1
1972	–	–	–	–	–	–	1	1
1971	–	–	–	–	–	–	1	0

Table 3 Best-fit values of the constants of Eq. 5 for $N_i(t)$ data

Author	Figures	Y range	Y_0 (year)	C_i (papers)	Θ_i (year)	q_i (–)
T. Dietl	5a	Entire	1975	824 ± 210	50.9 ± 8.1	2.155 ± 0.075
J. Barnaś	5b	Entire	1988	$1{,}134 \pm 1{,}437$	37.7 ± 23.5	2.509 ± 0.243
				832 ± 845	32.9 ± 17.1	2.545 ± 0.281
M. Kosmulski	5c	Entire	1977	531 ± 330	55.6 ± 18.9	2.306 ± 0.127
K. Sangwal	5d	Entire	1971	276 ± 51	45.7 ± 8.6	1.493 ± 0.076

Table 4 Best-fit values of the constants of Eq. 8 for $L(t)$ data

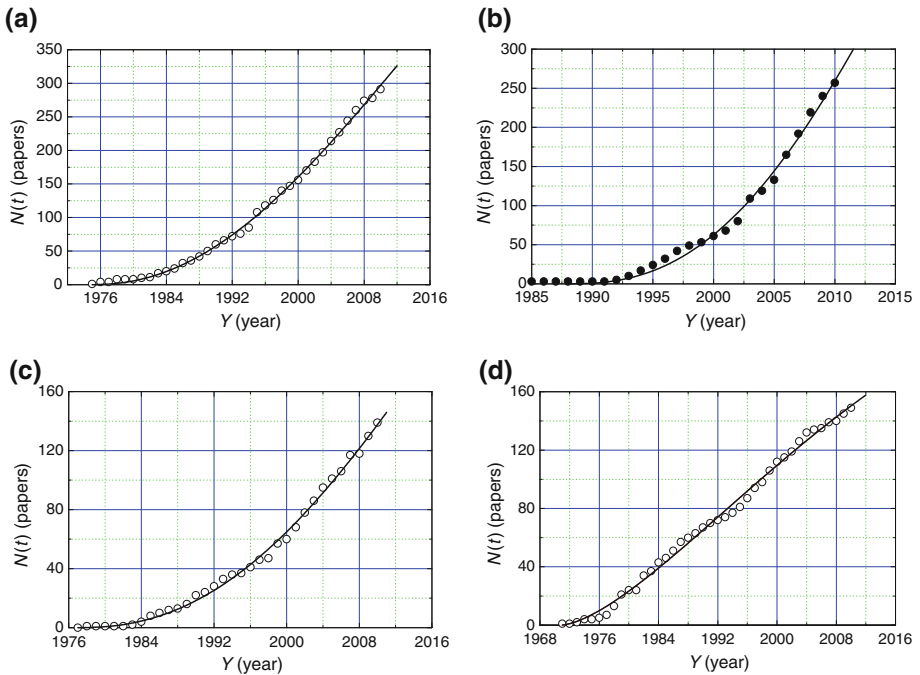| Author	Figures	Y range	Y_0 (year)	C_0 (papers)	Θ_0 (year)	q_0 (–)
T. Dietl	6a	<2,000	1975	$2{,}993 \pm 590$	29.1 ± 3.5	2.417 ± 0.109
		>2,000	1975	$13{,}940 \pm 966$	33.6 ± 0.5	7.851 ± 0.293
		Entire	1975	1.01×10^5	52.5 ± 39.3	5.422 ± 0.582
J. Barnaś	6b	Entire	1984	$6{,}118 \pm 3{,}192$	30.1 ± 4.4	4.905 ± 0.375
			1988	$5{,}836 \pm 2{,}391$	25.4 ± 3.5	4.220 ± 0.293
M. Kosmulski	6c	Entire	1977	6155 ± 1388	41.5 ± 2.6	4.676 ± 0.103
K. Sangwal	6d	Entire	1971	$4{,}523 \pm 3{,}187$	53.2 ± 16.0	3.077 ± 0.244

Fig. 5 Growth behavior of cumulative number $N(t)$ of papers published by four authors: **a** Dietl, **b** Barnaś, **c** Kosmulski and **d** Sangwal

that the highest growth of papers occurs for Barnaś, the slowest for Sangwal and intermediate for Dietl and Kosmulski. These trends are also reflected by the constant C_i for different authors.

The value of q_i lies between 1.5 and 2.55 for different authors. The value of q_i indicates that the growth of papers involves dissemination of information contained in them. The value of Θ_i lies between 33 and 56 years, indicating that the nucleation rate J_s for the papers of different authors differs among themselves by a factor of about 1.7. Using Eq. 22 one finds that J_s lies between $1.9 \cdot 10^{-10}$ and $3.3 \cdot 10^{-10}$ s^{-1} (assuming that the citation nuclei are spherical i.e. the shape factor $\kappa = 4\pi/3$, and the term $q^{1/q} = 1.42$; see Appendix).

Growth behavior of cumulative number $L(t)$ of citations

Figure 6 shows the growth behavior of cumulative number $L(t)$ of citations of papers published by the four authors while the best-fit values of the parameters of Eq. 8 for the data are listed in Table 4. As in the case of the steady increase in the cumulative number $N(t)$ of papers, the time dependence of the cumulative number $L(t)$ of papers published by Barnaś (Fig. 6b), Kosmulski (Fig. 6c) and Sangwal (Fig. 6d) as a function of publication time can be represented by Eq. 8 practically in the entire range. The values of the constants q_0 and Θ_0 of Eq. 8 for the data of these authors, shown in Fig. 6, are listed in Table 4. The value of the exponent q_0 lies between 2.3 and 4.9 whereas Θ_0 lies between 25 and 56 years for these authors.
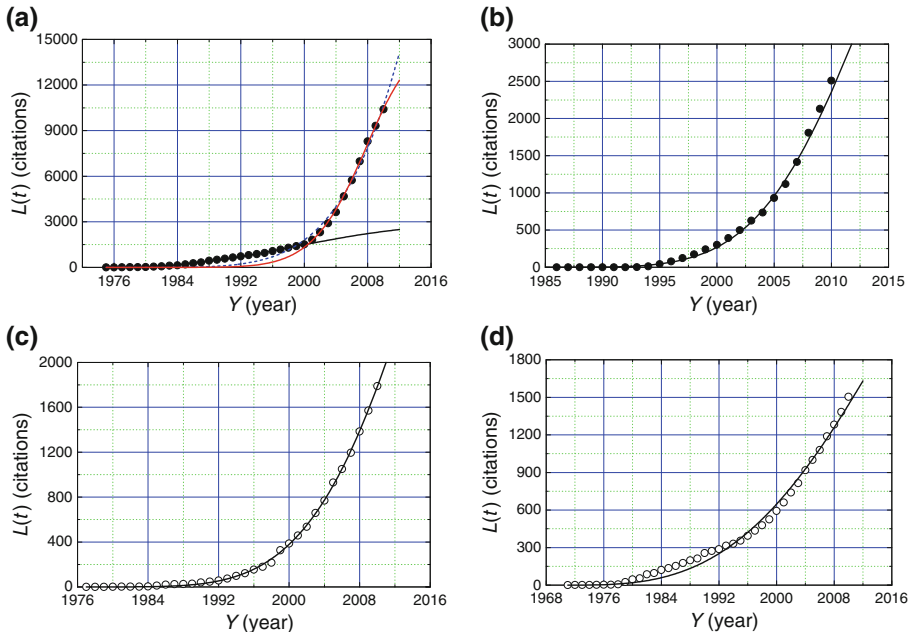

Fig. 6 Growth behavior of cumulative number $L(t)$ of citations of papers published by four authors: **a** Dietl, **b** Barnaś, **c** Kosmulski and **d** Sangwal. *Curves* are drawn with the best-fit values of parameters listed in Table 4. In **a** *solid curves* describe data before and after 2000 whereas *dashed curve* represent entire data. See text for details

In contrast to the data of the above three authors, it may be noted from that the data for Dietl can be interpreted in different ways (see Fig. 6a). For example, one can easily discern two different citation periods: before and after 2000, when the data are satisfactorily described by Eq. 8, as shown by the solid curves in Fig. 6a. The best-fit values of q_0 and Θ_0 are different in the two regions (Table 4). The value of q_0 increases from 2.42 for the data before 2000 to 7.85 after 2000, but the value of Θ_0 in the two citation regions remains essentially unchanged at about 30 years. However, one can equally attempt to describe the entire data by Eq. 8 with a single set of parameters C_0, q_0 and Θ_0 (dashed curve; Table 4). In this case, the data covering the period between about 1984 and 2005 are poorly represented by the best-fit curve. Now the exponent q_0 takes a value of 5.42, which is in between the values given above for the citation regions before and after 2000. The time constant increases to 52.5 years. This example of Dietl indicates that cumulative citations $L(t)$ of an author can have different well-defined citation periods.

It may be noted that the values of q_0 for the above collectives of systems are relatively high in comparison with those of q_i for individual systems (see Tables 3 and 4) These high values of q_0 are consistent with the prediction of the proposed PNM for the collectives of systems. In contrast to this, the values of Θ_0 are similar to those of Θ_i observed the case of cumulative number $N(t)$ of papers. The similar values of Θ_0 and Θ_i imply the nucleation rate J_s in the case of individual systems and collectives of systems are comparable. This also implies that the nucleation rate in the case of collectives of systems behaves as if it were an "effective" rate $J_{s(eff)}$ of nucleation and lies between 1.9×10^{-10} and 3.3×10^{-10} s^{-1}.

The smooth increase in the citations of papers published by Barnaś, Kosmulski and Sangwal is associated with a steady citation rate during their entire publication period. However, two different citation regions for the best fit of the data for Dietl are due to two different steady rates of citations before and after 2000. Obviously, the rate of citations of the papers published before 2000 is much lower than that of the papers published after 2000. The high rate of citations after 2000 is due to the high citability (i.e. more attractiveness) of some of the papers published by Dietl in the area of magnetic materials. Comprehensive review papers, as in the case of Sangwal, as well as more publications due to collaborations with other research groups can also lead to abrupt high rate of citations.

Relationship between cumulative number of papers and citations

As seen from Figs. 5 and 6, both the cumulative number $L(t)$ of citations of papers of an author and the cumulative number $N(t)$ of published papers increase with time t. Thus, one expects a relationship between cumulative citations $L(t)$ and cumulative papers $N(t)$ of an author. In this case, instead of time dependence of growth of cumulative citations or papers of an author as described by (5) of PNM, one may consider the dependence of citations $L(t)$ on the cumulative number $N(t)$ of papers, where $N(t)$ is an independent parameter, written in the form

$$L(t) = L_{max}\left[1 - \exp\left\{-\left(\frac{N(t)}{\Theta_N}\right)^{q*}\right\}\right], \tag{10}$$

where L_{max} is the maximum value of $L(N)$, Θ_N is a constant when L_{max} is attained, and the exponent $q*$ describes the steepness of the curve of $L(t)$ against $N(t)$. Equation 10 can be derived from Eq. 8 for the dependence of $L(t)$ on t and when the dependence of $N(t)$ on t in Eq. 5 follows the approximation

$$N(t) = N_{\max} \left[1 - \exp\left\{ -\left(\frac{t}{\Theta_N}\right)^{q*} \right\} \right] \approx N_{\max} \left(\frac{t}{\Theta_N}\right)^{q*}. \tag{11}$$

The above approximation holds for $(t/\Theta)^q \ll 1$ when $\exp[-(t/\Theta)^{q*}] = 1 - (t/\Theta)^{q*}$. As seen from the plots of Fig. 5, this approximation is indeed realistic. A similar behavior of $N(t)$ data was reported earlier (Sangwal 2011b).

Figures 7 and 8 show the cumulative number $L(t)$ of citations and cumulative number $N(t)$ of papers of different authors, whereas the curves are drawn with the best-fit values of the constants of Eq. 10 given in Table 5. As seen from Fig. 7, the data for Barnaś (Fig. 7a) and Kosmulski (Fig. 7b) can be described satisfactorily by Eq. 10 in the entire $N(t)$ range but those for Dietl (Fig. 8a) and Sangwal (Fig. 8b) follow two different dependences below and above 120 and 110 papers, respectively. The high rates of cumulative citations of Dietl and Sangwal after 120 and 110 papers, respectively, are due to comparatively very high citations of some of the papers published by them after 120 and 110 papers. These high rates are reflected by the high values of $q*$ in these regions of papers.

Summary and conclusions

For individual systems of items, the time dependence of the fraction $\alpha(t)$ of items, based on the concepts of PNM during crystallization, may be described by Eq. 5. The PNM as
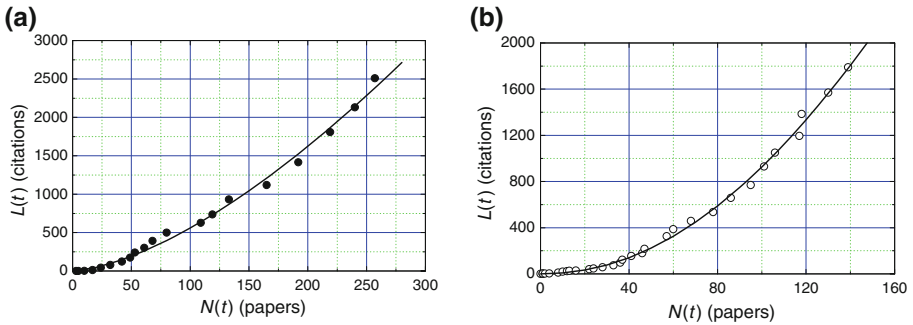

Fig. 7 Relationship between cumulative number $L(t)$ of citations and cumulative number $N(t)$ of papers of **a** Barnaś and **b** Kosmulski
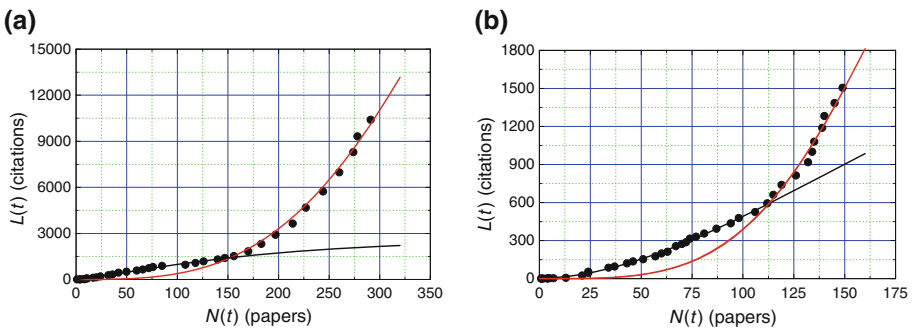

Fig. 8 Relationship between cumulative number $L(t)$ of citations and cumulative number $N(t)$ of papers of **a** Dietl and **b** Sangwal

Table 5 Best-fit values of the constants of Eq. 11 for $L(N)$ data

Author	Figures	N range	$L_{\max}$ (citations)	Θ_N (papers)	q^* (–)
T. Dietl	8a	<120	2,529 ± 671	178.1 ± 59.2	1.237 ± 0.096
		>120	51,123 ± 105,260	467.0 ± 362.4	3.204 ± 0.404
J. Barnaś	7a	Entire	11,222 ± 16,920	633.9 ± 718.6	1.615 ± 0.161
M. Kosmulski	7b	Entire	12,460 ± 16,724	338.9 ± 250.8	2.098 ± 0.120
K. Sangwal	8b	<110	2,195 ± 1,217	217.0 ± 82.7	1.834 ± 0.093
–		>110	4,374 ± 7,511	189.0 ± 118.3	3.749 ± 0.984

applied to describe the growth behavior of items in an individual system postulates that items are nucleated progressively during their formation (production) and, after their nucleation, they grow to visible size. In the case of items produced in a collective of systems, an equation (i.e.Eq. 8) similar in the form of Eq. 5 holds.

As in the case of PNM for crystallization, Eq. 5 for individual systems predicts that $1 < q < 2.5$ and $1 < q < 4$ when the items grow to three-dimensional entities of visible size (i.e. when $d = 3$) by volume diffusion and mass transfer, respectively. When the items are already of visible size at the time of nucleation, the exponent $q = 1$, which is the lowest limit for the validity of PNM.

An expression similar to that of Eq. 8 in the form of Eq. 10 is proposed to describe the relationship between the cumulative number $L(t)$ of citations and the cumulative number $N(t)$ of papers.

Equations 5 and 8, based on the concepts of PNM predict that: (1) the fraction of items produced by successive systems is additive, (2) the cumulative fraction $\alpha_{sum}(t)$ of maximum number of sites is the sum of contributions of fractions of maximum number of items produced by different systems, and (3) the values of time constants Θ_0 and exponent q_0 increase with the addition of fraction of items produced by subsequent systems, but their values are the lowest for individual systems.

Analysis of the growth behavior of cumulative number $N(t)$ of papers published by individual authors as a function of citation duration t revealed that the exponent q_i lies between 1.5 and 2.5 whereas that of the growth behavior of cumulative citations $L(t)$ of individual papers of an author shows that q_0 lies between 2.3 and 7.9. These values of q_i for individual system and q_0 for collectives of systems are consistent with the prediction of the proposed PNM. The values of the time constant Θ_i for papers and Θ_0 for citations lie between 30 and 60 years in the two cases. These values of the time constants give the nucleation rate J_s between $1.9 \cdot 10^{-10}$ and $3.3 \cdot 10^{-10}$ s^{-1}. However, two different regions of citations observed in some cases are associated with differences in the attractiveness of the papers in these regions.

Analysis of the dependence of $L(t)$ on $N(t)$ for the four authors reveals that the value of the exponent q^* of Eq. 10 lies between 1.2 and 3.8 whereas that of the constant Θ_N related to the saturation limit of citations lies between 178 and 634 papers. In these plots different regions of citations associated with the attractiveness of the published papers are clearly revealed.

Finally, it should be mentioned that the process of citations of an article published by an author is associated with its attractiveness or novelty, as recognized by its readers and future authors. The attractiveness of an article as recognized by its readers is the driving force for its citations. The greater the attractiveness of the article, the higher is the driving force for its citation. However, the probability of citations of an article is intimately connected with its

accessibility to its potential readers. The higher the accessibility of an article to its readers, the higher is the probability of its citations. In Eq. 17 the parameter Δc represents the driving force for citation of an article and the parameter B_1 describes the probability of its citation. The above concepts of citations of articles are in line with the results of a recent study by Stremersch et al. (2007), who found that the number of citations of an article in the marketing discipline depends on its quality and domain as well as on author visibility and personal promotion. However, more work is necessary in this direction.

Appendix: Progressive nucleation mechanism for overall crystallization

Overall crystallization of a melt or homogeneous solution involves simultaneous nucleation and growth of individual nuclei to crystallites of visible size. In the progressive nucleation mechanism, crystallites are nucleated progressively during crystallization. In the case when the nucleation rate J_s is time-independent (i.e. when nucleation is stationary), the fraction $\alpha(t)$ of the crystallized volume $V_c(t)$ at time t may be given by (Kashchiev 2000)

$$\alpha(t) = \frac{V_c(t)}{V} = \left[1 - \exp\left\{-\left(\frac{t}{\Theta}\right)^q\right\}\right], \tag{12}$$

where V is the initial volume of the crystallizing phase, the time constant

$$\Theta = \left(\frac{q}{\kappa G^{q-1} J_s}\right)^{1/q}, \tag{13}$$

and the exponent

$$q = 1 + vd. \tag{14}$$

In Eqs. 13 and 14, J_s is the rate of stationary nucleation, κ is the shape factor for the nuclei (for example, $\kappa = 4\pi/3$ for spherical nuclei) and the growth constant G is defined by

$$G = \frac{r^{1/v}}{t}, \tag{15}$$

where r is the radius of the growing nucleus and the constant $v > 0$. In Eq. 14 the parameter d denotes the dimensionality of growing nuclei. For nuclei growing in one-, two- and three-dimensions, $d = 1$, 2 and 3, respectively. Equation 15 describes the dependence of the radius r of the growth of *individual* nucleus on time t according to the traditional power-law relation

$$r(t) = At^v, \tag{16}$$

where the growth constant $G = A^{1/v}$ and the exponent v can have values between 0 and ∞.

From Eq. 14 it follows that the exponent $q = 1$ when $v = 0$, and $q \to \infty$ when $v \to \infty$. After their formation to a stable size, the nuclei do not grow at all in the former case, whereas they grow to infinite size immediately in the latter case. Thus, it is possible that $1 < q < \infty$ but, in reality, q can have finite values only because v is a number equal to 1 and 1/2 for growth controlled by mass transfer and volume diffusion, respectively (Kashchiev 2000). This means that $1 < q < 2.5$ and $1 < q < 4$ when crystal nuclei grow to

three-dimensional crystallites of visible size (i.e. when $d = 3$) by volume diffusion and mass transfer, respectively. When $d = 0$, $q = 1$, which represents particles (crystallites) as points (see Eq. 14),

In the above treatment the stationary nucleation rate J_s involves the formation of stable three-dimensional nuclei of critical size as a result of aggregation of growth units in the crystallization medium, and these stable nuclei subsequently grow into crystallites by the addition of more growth units (atoms/molecules). The nucleation rate may be given by (Kashchiev 2000; Mullin 2001)

$$J_s = A_1 \exp[-B_1/(\Delta c)^2], \tag{17}$$

where, at a given temperature, the constant A_1 is associated with the processes of aggregation of growth units, the constant

$$B_1 = \kappa_1 c_0^2 \left(\frac{\gamma \Omega^{2/3}}{k_B T} \right)^3, \tag{18}$$

and the excess solute concentration Δc (i.e. $\Delta c = c - c_0$, where c is the actual solute concentration) above equilibrium concentration c_0 is related to the chemical potential difference $\Delta\mu$ by

$$\Delta\mu = \mu - \mu_0 = k_B T (\Delta c/c_0). \tag{19}$$

In Eqs. 18 and 19, γ is the nucleus—medium interfacial tension, Ω is the molecular volume, k_B is the Boltzmann constant, T is the temperature in Kelvin, κ_1 is a factor related to the shape of the three-dimensional nuclei, and μ and μ_0 are the actual and equilibrium chemical potentials of the solute atoms/molecules, respectively. Equation 17 is used to describe the kinetics of homogeneous and heterogeneous three-dimensional nucleation. In heterogeneous nucleation, foreign particles in the medium or on the vessel surface catalyze the formation of stable nuclei. Nucleation described by Eq. 17 is frequently referred to as primary nucleation.

After the formation of three-dimensional stable nuclei, they grow larger by the addition of growth units on their surfaces. When the growth of stable nuclei to larger, visible sizes occurs by two-dimensional multiple nucleation, the growth rate constant G is given by (Mullin 2001)

$$G = \frac{r^{1/v}}{t} = A_2 \exp(B_2/\Delta c), \tag{20}$$

where A_2 is a new constant related to integration of growth units and B_2 is another constant related to the nucleus—medium interfacial tension γ, given by

$$B_2 = \kappa_2 c_0 \left(\frac{\gamma h \Omega^{1/2}}{k_B T} \right)^2. \tag{21}$$

In Eq. 21, h is the height of the two-dimensional nuclei, κ_2 is a factor related to their shape, whereas all other symbols have been defined above.

In Eqs. 17 and 21 the excess solute concentration Δc is the driving force for crystallization. The quantity Δc is related to the chemical potential difference $\Delta\mu$ as well as to the thermodynamic potential difference ΔF (Kashchiev 2000). Both of these quantities are measures of deviations of the crystallizing system from the equilibrium state.

Fig. 9 Dependence of $q^{1/q}$ on parameter q in Eq. 22. See text for details
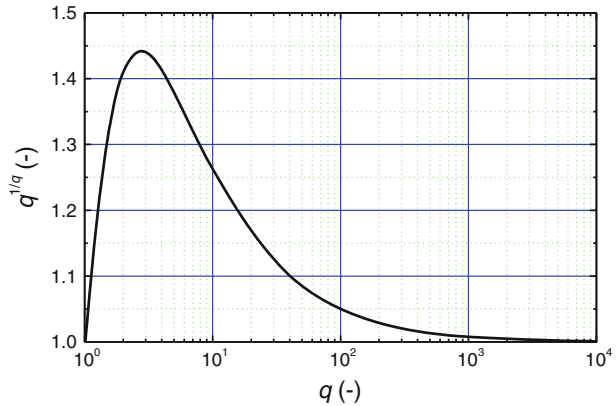
According to Eq. 13, when $q = 1$, the nucleation rate J_s alone determines the value of the time constant Θ. When $q > 1$, the dependence of Θ on nucleation and growth processes is relatively complex because it depends on three parameters: the nucleation rate J_s, the growth constant G and the exponent q. However, Eq. 13 of the time constant Θ may be written in the form

$$\Theta = \frac{q^{1/q}}{\kappa J_s},\tag{22}$$

when $\kappa J_s = G$. The condition for the validity of this equality may be obtained from Eqs. 17 and 20 by solving them for Δc, i.e.

$$\ln\left(\frac{\kappa A_1}{A_2}\right)(\Delta c)^2 - B_2 \Delta c + B_1 = 0.\tag{23}$$

Equation 23 has a single real root

$$\Delta c* = \frac{B_2}{2\ln(\kappa A_1/A_2)},\tag{24}$$

when

$$B_2^2 = 4B_1 \ln(\kappa A_1/A_2).\tag{25}$$

According to Eq. 22, the time constant Θ is inversely proportional to the nucleation rate J_s whereas it is a complex function of the exponent q. The factor $q^{1/q}$ initially increases steeply from 1 at $q = 1$–1.414 at $q = 2$ and then for $q > 4$ decreases slowly and approaches a value of unity at $q \approx 10^4$, as shown in Fig. 9. This implies that the factor $q^{1/q}$ is practically constant at 1.42 for the growing three-dimensional crystallites when $2 \geq q \geq 4$.

References

De Solla Price, D. J. (1963). *Little science, big science*. New York/London: Columbia University Press.
Egghe, L., & Ravichandra Rao, I. K. (1992). Classification of growth models based on growth rates and its applications. *Scientometrics, 25*(1), 5–46.

Gupta, B. M., Kumar, S., Sangam, S. L., & Karisiddappa, C. R. (2002). Modeling the growth of world social science literature. *Scientometrics, 53*(1), 161–164.

Gupta, B. M., Sharma, L., & Karisiddappa, C. R. (1995). Modeling the growth of papers in a scientific speciality. *Scientometrics, 33*(2), 187–201.

Kashchiev, D. (2000). *Nucleation: Basic theory with applications*. Oxford: Butterworth-Heinemann.

Mullin, J. W. (2001). *Crystallization* (4th ed.). Oxford: Butterworth-Heinemann.

Ravichandra Rao, I. K., & Srivastava, D. (2010). Growth of journals, articles and authors in malaria research. *Journal of Informetrics, 4*(1), 249–256.

Sangwal, K. (2011a). On the growth of citations of publication output of individual authors. *Journal of Informetrics, 5*(4), 554–564.

Sangwal, K. (2011b). Progressive nucleation mechanism and its application to the growth of journals, articles and authors in scientific fields. *Journal of Informetrics, 5*(4), 529–536.

Sangwal K. (2012). Application of progressive nucleation mechanism for the citation behavior of individual papers of different authors. *Scientometrics*. doi:10.1007/s11192-011-0564-x.

Stremersch, S., Verniers, I., & Verhoef, P. C. (2007). The quest for citations: Drivers of article impact. *Journal of Marketing, 71*(3), 171–193.

Wong, C.-Y., & Goh, K.-L. (2010). Growth behavior of publications and patents: a comparative study on selected Asian economies. *Journal of Informetrics, 4*(2), 460–474.