Three dimensions of scientific impact

Grzegorz Siudem[a,1], **Barbara Żogała-Siudem**[b], **Anna Cena**[c], **and Marek Gagolewski**[b,c,d]

[a]Faculty of Physics, Warsaw University of Technology, 00-662 Warsaw, Poland; [b]Systems Research Institute, Polish Academy of Sciences, 01-447 Warsaw, Poland; [c]Faculty of Mathematics and Information Science, Warsaw University of Technology, 00-662 Warsaw, Poland; and [d]School of Information Technology, Deakin University, Geelong, VIC 3220, Australia

Edited by Anthony Van Raan, Leiden University, Leiden, The Netherlands, and accepted by Editorial Board Member Adrian E. Raftery May 11, 2020 (received for review January 18, 2020)

The growing popularity of bibliometric indexes (whose most famous example is the *h* index by J. E. Hirsch [J. E. Hirsch, *Proc. Natl. Acad. Sci. U.S.A.* 102, 16569–16572 (2005)]) is opposed by those claiming that one's scientific impact cannot be reduced to a single number. Some even believe that our complex reality fails to submit to any quantitative description. We argue that neither of the two controversial extremes is true. By assuming that some citations are distributed according to the rich get richer rule (success breeds success, preferential attachment) while some others are assigned totally at random (all in all, a paper needs a bibliography), we have crafted a model that accurately summarizes citation records with merely three easily interpretable parameters: productivity, total impact, and how lucky an author has been so far.

science of science | scientometrics | bibliometric indexes | rich get richer

Ever since Garfield's (1) impact factor for journals and Hirsch's (2) h index for individual researchers, the popularity of bibliometric impact measures has been growing rapidly. The fact that they summarize one's scientific performance with just a single number is appealing to many. However, some argue (3) that the nature of scientific activities is too multidimensional for such a simple description to be possible and a few quantitative metrics will never be sufficient to capture this complex reality in its entirety.

In this paper we address this issue from the perspective of the increasingly popular science of science (Sci-Sci) (4, 5) approach, which can be dated back to the classical book by de Solla Price, *Little Science, Big Science* (6). The modern Sci-Sci utilizes complex systems methodology and can be considered a fusion of agent-based modeling and big data analysis.

We have developed a model of an author's research activity that is based on two simple assumptions: 1) In each time step one new paper is added into the simulation. 2) Each newly added paper cites the existing publications according to a combination of a) the preferential attachment rule—highly cited papers are more likely to attract even more citations [compare the rich get richer mechanism (7), the success breeds success phenomenon (8), and the effect of a scientist's reputation (9)]—and b) sheer chance—papers might be discovered by the citing authors by accident or be included in the bibliography completely at random.

While the importance of the rich get richer rule (7) in bibliometrics is unquestionable [first part of Merton's (10) Matthew effect, referred to as the cumulative advantage process by de Solla Price (8) or success-breeds-success phenomenon (6, 11), confirmed experimentally (12)], we argue here that a purely preferential model is incapable of explaining our reality well enough and the accidental component is necessary (13, 14).

Furthermore, in our case we adopt different levels of analysis [as known from social sciences (15)] (Fig. 1) for generated bibliometric data. Agent-based models are formulated at the microlevel—from the perspective of an individual paper. The Sci-Sci perspective usually investigates the structure of the citation network in its entirety, for instance to describe general citation patterns across the whole scientific discipline (macrolevel). Here we are mainly focusing on the rarely considered mesolevel (Table 1), which is the perspective of a single scientist, i.e., a small-sample one. As such, the above publication–citation process can be thought of as an extension of the iterative procedure known as the Ionescu–Chopard model (16, 17) (*Materials and Methods*, *Model Description*).

Model Derivation

Assume $X_1, X_2, \ldots, X_N$ is a descending sequence of citation counts for each of the N papers of an author. In other words, X_1 denotes the number of bibliographic references to the author's most cited paper, X_2 is the second most cited, …, and X_N is the least cited one. Famous approaches (18) to the problem of approximating observed citation records $X_1, \ldots, X_N$ with simple mathematical models $\hat{X}_1(\cdots), \ldots, \hat{X}_N(\cdots)$ that depend on a small number of parameters were mostly based on the power law (19) or other functions (20). Unfortunately, they do not provide a good fit at the mesolevel—they are usually applied for describing papers sampled from the whole citation network (21, 22).

Our model, on the other hand, not only has a clear interpretation (recall the two simple assumptions above), but also provides high-accuracy approximations of citation records of individuals. Due to this, we are able to describe this complex reality with merely three self-explanatory parameters: the number of papers N; the total number of citations $C = X_1 + X_2 + \cdots + X_N$; and the ratio of citations distributed according to the preferential attachment rule ρ, where $\rho \simeq 0$ means that all papers receive citations completely at random and $\rho \simeq 1$ that all of them follow the rich get richer rule.

Significance

What are the mechanisms behind one's research success as measured by one's papers' citability? By acknowledging the perceived esteem might be a consequence not only of how valuable one's works are but also of pure luck, we arrived at a model that can accurately recreate a citation record based on just three parameters: the number of publications, the total number of citations, and the degree of randomness in the citation patterns. As a by-product, we show that a single index will never be able to embrace the complex reality of the scientific impact. However, three of them can already provide us with a reliable summary.

Author contributions: G.S., B.Ż.-S., A.C., and M.G. designed research; G.S., B.Ż.-S., A.C., and M.G. performed research; B.Ż.-S., A.C., and M.G. analyzed data; and G.S., B.Ż.-S., A.C., and M.G. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. A.V.R. is a guest editor invited by the Editorial Board.

Data deposition: The raw citation sequences, estimated parameters, and source code used to perform the data analysis can be accessed at the GitHub repository: https://github.com/gagolews/three_dimensions_of_scientific_impact.

[1]To whom correspondence may be addressed. Email: grzegorz.siudem@pw.edu.pl.

First published June 8, 2020.

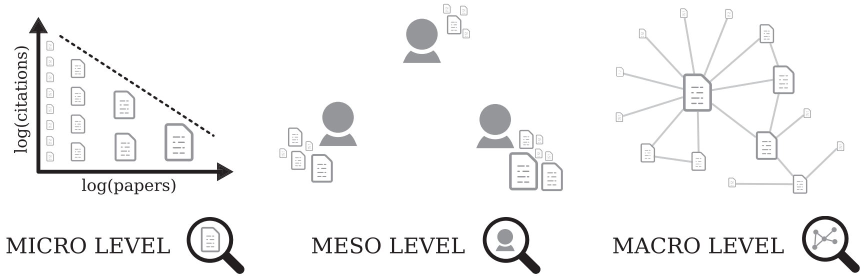

Fig. 1. Different levels of analysis of bibliometric datasets. On the microlevel we describe the distribution of the number of citations of individual papers, irrespective of who authored them as well as which articles actually referenced them. The rarely studied mesolevel, which is the perspective of this contribution, accounts for the author-specific differences. The structure of the citation network in its entirety is studied at the macrolevel.

For the derivation of the model please refer to *Materials and Methods*, *Model Description*. The citation process proposed above, after all of the N papers have been published and all of the citations have been distributed, yields the following analytic formula for the estimated number of citations of the kth most cited paper (*Materials and Methods*, *Exact Solution of the Model*):

$$\hat{X}_k(N, C, \rho) = \frac{1-\rho}{\rho}\frac{C}{N}\left(\prod_{i=k}^{N}\frac{i}{i-\rho} - 1\right) \quad [1]$$
$$= \frac{1-\rho}{\rho}\frac{C}{N}\left(\frac{k}{k-\rho}\cdot\frac{k+1}{k+1-\rho}\cdots\frac{N}{N-\rho} - 1\right).$$

Dataset Description

To demonstrate the usefulness of the model, we study the DBLP Computer Science Bibliography (47) dataset of computer science papers; see *Materials and Methods*, *Data Availability* for description. We consider citation records of all 123,621 scholars whose h index is at least 5. To determine the three model parameters characterizing each author, we omit the papers with no citations (as overfitting to a tail composed of zeros cannot lead to a good overall description). Then we compute the author's N (number of papers that were cited at least once) and C (the total number of citations) and then estimate ρ using the least-squares fit with respect to the Cauchy loss $\sum_{k=1}^{N}\log(1+(\hat{X}_k(N,C,\rho)-X_k)^2)$ to weaken the influence of any potential outliers.

Once we obtain an author's N, C, and ρ, we can reproduce the author's citation record quite accurately (Fig. 2). The high variance of ρ for each fixed N and C (Fig. 3) indicates that this parameter is necessary for a precise description of data. This suggests that indeed the modeled reality might be three-dimensional (3D), which roughly agrees with the estimates in ref. 48.

Results and Discussion

It turns out that ca. 30% of the authors have their corresponding $\rho \approx 0$, which means that, under our model, their citations appear to be distributed in an almost purely accidental manner. These authors publish on average half as many papers as those with $\rho > 0$, which might indicate that they are at the beginning of their careers or their best papers are still yet to come. We observe a positive correlation between ρ and N as well as C (Fig. 3). In other words, more productive and/or influential authors tend to have more papers distributed according to the rich get richer rule. This observation is consistent with the well-known fact (5) that one's highest-impact paper can occur at any time during the course of one's career; thus, authors with more papers are more likely to have published their best work already. However, as there is a considerable variability in ρ at all levels, even some outstanding careers might still be a result of more luck than reason (13, 49).

By indicating that the citation record space is 3D, we have proved that any single citation measure, including the h index and the author's ranking it generates, necessarily yields an oversimplified projection of a more complex space (3). In other words, whenever one chooses a single citation index, some information must inherently be lost; we will never be able to see the whole picture through the lenses of any single measure.

The proposed model emphasizes the use of multiple indexes in the evaluation of scientific work. We have indicated that merely three parameters are sufficient to provide an accurate description of our reality. In the near future, we plan to perform a broad study of bibliometric indexes to come up with an intuitive and insightful classification for which of the three dimensions each index focuses on the most. This will allow policy makers to make better-informed decisions when choosing particular evaluation tools. The questions of how to best combine N, C, and ρ to cause the least information loss and how well popular citation indexes perform with regard to the quality of data approximation will also be explored.

Materials and Methods

Model Description. Let us introduce the proposed model in a formal manner. For the description of the citation dynamics we use the following parameters: the total number of papers N, the total number citations C that will be distributed among all papers, and ratio of the number of preferential citations to the total number of citations $\rho \in (0, 1)$.

Due to the assumed boundary conditions in Eq. **3**, we disallow both $\rho = 0$ and $\rho = 1$.

Table 1. Overview of the related literature on the modeling of the distribution of citations

	Microlevel	Mesolevel	Macrolevel
Purely preferential	Distribution of the number of citations (21–30)	Lotkaian informetrics (19), Ionescu–Chopard model (16, 17)	Barábasi–Albert model and its modifications (31)
Preferential and/or accidental	Microscopic model (14) implies Tsalis–Pareto distribution (32)	This paper	Empirical data (33, 34), models studied in refs. 35–46

By assuming that citations might be assigned completely at random as well as follow the rich get richer rule, we revealed the underlying dimensionality of the mesolevel, leading to an accurate description of the output of an individual author.

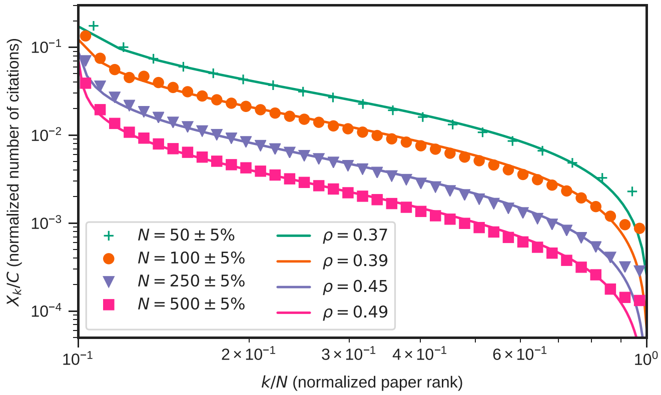

Fig. 2. Normalized average number of citations X_k/C as a function of the normalized paper rank k/N on a double-logarithmic scale. Each plotting character corresponds to citation sequences of different lengths: "+," all of the 2,624 authors with 48 to 52 papers in total; "•," 1,113 authors with 95 to 105 papers; "▼," 131 authors with 238 to 262 papers; and "■," 18 authors with 475 to 525 papers. The curves represent the corresponding predictions $\hat{X}_k/C$ as generated by our model with ρ equal to the averages over the individual authors' fitted rich get richer ratios. A particularly good fit is observed in the case of highly and moderately influential papers.

The stages of the model's simulation are strictly connected to the scientific activity of the considered author. Each of the N steps corresponds to the publication of one of the author's papers. At the tth step, the t articles already in existence are to receive $n_a + n_p = \frac{C}{N}$ citations in total, where $n_p = \rho\frac{C}{N}$ citations are distributed according to the preferential attachment rule, and $n_a = (1-\rho)\frac{C}{N}$ citations are uniformly distributed between the t papers

Note that both n_p and n_a do not need to be integers—we consider them as averages.

The rate equation for the number of citations of the kth mostly cited paper at the tth stage of the simulation, $X_k^{(t)}$, takes the form

$$X_k^{(t)} = \underbrace{X_k^{(t-1)}}_{\text{previous value}} + \underbrace{\frac{n_a}{t}}_{\text{accidental income}} + \underbrace{n_p \frac{X_k^{(t-1)} + \frac{n_a}{t}}{n_a + \sum_{l=1}^{t-1} X_l^{(t-1)}}}_{\text{preferential income}}, \qquad [2]$$

for $k = 1, \ldots, t$. As each paper has initially no citations, we introduce the following boundary conditions:

$$X_k^{(k-1)} = 0, \text{ for } k = 1, 2, \ldots \qquad [3]$$

Note that in the rightmost term in Eq. **2**, i.e., the preferential part, we assume that accidental citations are distributed first to avoid singularities with the very natural boundary conditions of the form given by Eq. **3**. This explains the occurrence of n_a there. The structure of the preferential part is the expected value of the Bernoulli distribution with the number of trials n_p and the probability resulting from the assumed rich get richer mechanism—the number of citations thus obtained is proportional to the actual number of citations (i.e., $X_k^{(t-1)} + n_a/t$).

Exact Solution of the Model. Below we derive the exact formula for $X_k^{(t)}$. Note that Eq. **2** can be simplified as

$$X_k^{(t)} = \left[X_k^{(t-1)} + \frac{n_a}{t}\right]\left[1 + \frac{n_p}{n_a + \sum_{l=1}^{t-1} X_l^{(t-1)}}\right].$$

Moreover, the second term can be further simplified due the fact that in each of the $(t-1)$ steps, the papers receive $n_a + n_p$ citations; i.e.,

$$\sum_{l=1}^{t-1} X_l^{(t-1)} = (n_a + n_p)(t-1).$$

Therefore,

$$1 + \frac{n_p}{n_a + \sum_{l=1}^{t-1} X_l^{(t-1)}} = 1 + \frac{n_p}{n_a + (n_a + n_p)(t-1)}$$
$$= \frac{(n_a + n_p)t}{(n_a + n_p)t - n_p} = \frac{t}{t - \frac{n_p}{n_a + n_p}}.$$

Furthermore, since $\rho = n_p/(n_a + n_p)$, the following holds:

$$X_k^{(t)} = X_k^{(t-1)} \frac{t}{t-\rho} + \frac{n_a}{t-\rho}. \qquad [4]$$

Moreover,

$$\begin{aligned} X_k^{(t)} &= \left[X_k^{(t-2)} \frac{t-1}{t-1-\rho} + \frac{n_a}{t-1-\rho}\right] \frac{t}{t-\rho} + \frac{n_a}{t-\rho} \\ &= \left[X_k^{(t-3)} \frac{t-2}{t-2-\rho} + \frac{n_a}{t-2-\rho}\right] \frac{t(t-1)}{(t-\rho)(t-1-\rho)} \\ &\quad + \frac{n_a t}{(t-\rho)(t-1-\rho)} + \frac{n_a}{t-\rho} \\ &= X_k^{(t-3)} \frac{t(t-1)(t-2)}{(t-\rho)(t-1-\rho)(t-2-\rho)} \\ &\quad + \frac{n_a t(t-1)}{(t-\rho)(t-1-\rho)(t-2-\rho)} \\ &\quad + \frac{n_a t}{(t-\rho)(t-1-\rho)} + \frac{n_a}{t-\rho}. \end{aligned} \qquad [5]$$

Keeping in mind that the Euler gamma function Γ (e.g., ref. 50, chap. 5), defined as

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x}\, dx,$$

satisfies the factorial-like relation (equation 5.5.1 in ref. 50)

$$\Gamma(z+1) = z\Gamma(z), \qquad [6]$$

for every number z, we can transform Eq. **5** as

$$\begin{aligned} X_k^{(t)} = X_k^{(t-3)} &\frac{\Gamma(t-2-\rho)\Gamma(t+1)}{\Gamma(t+1-\rho)\Gamma(t-2)} + n_a \frac{\Gamma(t-2-\rho)\Gamma(t+1)}{\Gamma(t+1-\rho)\Gamma(t-1)} \\ &+ n_a \frac{\Gamma(t-1-\rho)\Gamma(t+1)}{\Gamma(t+1-\rho)\Gamma(t)} + n_a \frac{\Gamma(t-\rho)\Gamma(t+1)}{\Gamma(t+1-\rho)\Gamma(t+1)}. \end{aligned} \qquad [7]$$

By continuing evaluation of Eq. **4** of the form given by Eq. **7**, we obtain

$$\begin{aligned} X_k^{(t)} = \underbrace{X_k^{(k-1)}}_{=0} &\frac{\Gamma(t+1)}{\Gamma(t+1-\rho)} \frac{\Gamma(k-\rho)}{\Gamma(k)} \\ &+ n_a \frac{\Gamma(t+1)}{\Gamma(t+1-\rho)} \sum_{r=0}^{t-k} \frac{\Gamma(t-r-\rho)}{\Gamma(t-r+1)}. \end{aligned} \qquad [8]$$

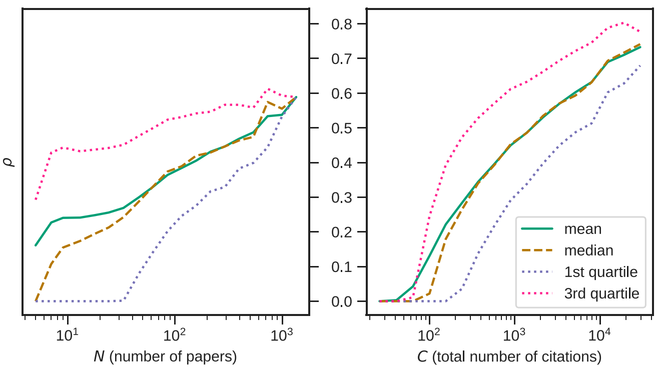

Fig. 3. The more productive and/or influential an author is, the more likely the author's papers are cited according to the rich get richer rule.

In Eq. **8** we can stop the nesting procedure by using the boundary conditions given by Eq. **3**. The final formula for $X_k^{(t)}$ with the change of the summation variable $\ell = t - r$ takes the form

$$X_k^{(t)} = n_a \frac{\Gamma(t+1)}{\Gamma(t+1-\rho)} \sum_{\ell=k}^{t} \frac{\Gamma(\ell-\rho)}{\Gamma(\ell+1)}.$$

This can be simplified further, because the sum of the ratios of gamma functions satisfies the identity

$$\sum_{\ell=k}^{t} \frac{\Gamma(\ell-\rho)}{\Gamma(\ell+1)} = \frac{1}{\rho}\left[\frac{\Gamma(k-\rho)}{\Gamma(k)} - \frac{\Gamma(t+1-\rho)}{\Gamma(t+1)}\right], \qquad [9]$$

which leads to

$$X_k^{(t)} = \frac{n_a}{\rho}\left[\frac{\Gamma(k-\rho)}{\Gamma(k)}\frac{\Gamma(t+1)}{\Gamma(t+1-\rho)} - 1\right]. \qquad [10]$$

Finally, we put $t = N$, which leads to the situation where each paper has been published and every citation has been distributed. This yields $\hat{X}_k := X_k^{(N)}$ such that

$$\hat{X}_k(N, C, \rho) = \frac{1-\rho}{\rho}\frac{C}{N}\left[\frac{\Gamma(k-\rho)}{\Gamma(k)}\frac{\Gamma(N+1)}{\Gamma(N+1-\rho)} - 1\right]. \qquad [11]$$

Gamma functions, although very elegant, are not computationally well behaving. This is the reason why we should be interested in deriving the following equivalent of Eq. **11**. Due to Eq. **6**, we can substitute the gamma functions with the following product:

$$\hat{X}_k(N, C, \rho) = \frac{1-\rho}{\rho}\frac{C}{N}\left(\prod_{\ell=k}^{N}\frac{\ell}{\ell-\rho} - 1\right). \qquad [12]$$

The Pochhammer symbol (section 5.2 in ref. 50) is defined as

$$(k)_m = \frac{\Gamma(k+m)}{\Gamma(k)} = k(k+1)\ldots(k+m-1). \qquad [13]$$

Employing it in Eq. **11** yields

$$\hat{X}_k(N, C, \rho) = \frac{1-\rho}{\rho}\frac{C}{N}\frac{(k)_{N-k+1} - (k-\rho)_{N-k+1}}{(k-\rho)_{N-k+1}}. \qquad [14]$$

Note that the Pochhammer symbol is implemented in many numerical software packages, thus enabling fast and accurate computations.

Data Availability. Empirical data analysis conveyed in this paper is based on the DBLP V10 bibliography database (47) (https://aminer.org/citation), consisting of 3,079,007 papers and 25,16,994 citation relationships. DBLP includes most of the journals related to computer science. It also tracks numerous conference proceedings papers from the field.

We have extracted citation records of 1,762,044 authors. Most of them have published a small number of papers or have received very few citations. Therefore, we restricted the analysis to the subset of researchers characterized by the *h* index not less than 5. This gave 123,621 citation records. Moreover, papers with 0 citations have been omitted from the analysis, as they are problematic when performing computations on the log scale. Note that most impact indexes, including the *h* index, ignore zeros anyway.

The raw citation sequences, estimated parameters, and source code used to perform the data analysis can be accessed at the GitHub repository: https://github.com/gagolews/three_dimensions_of_scientific_impact (51).

ACKNOWLEDGMENTS. We thank Maciej J. Mrowiński, Tessa Koumoundouros, and the reviewers for valuable feedback and constructive remarks.

1. E. Garfield, Citation indexes for science: A new dimension in documentation through association of ideas. *Science* **122**, 108–111 (1955).
2. J. E. Hirsch, An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16569–16572 (2005).
3. M. Gagolewski, Scientific impact assessment cannot be fair. *J. Informetrics* **7**, 792–802 (2013).
4. A. Clauset, D. B. Larremore, R. Sinatra, Data-driven predictions in the science of science. *Science* **355**, 477–480 (2017).
5. S. Fortunato *et al.*, Science of science. *Science* **359**, eaao0185 (2018).
6. D. J. de Solla Price, *Little Science, Big Science* (Columbia University Press, New York, NY, 1963).
7. M. Perc, The Matthew effect in empirical data. *J. R. Soc. Interface* **11**, 20140378 (2014).
8. D. J. de Solla Price, A general theory of bibliometric and other cumulative advantage processes. *J. Am. Soc. Inf. Sci.* **27**, 292–306 (1976).
9. A. M. Petersen *et al.*, Reputation and impact in academic careers. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15316–15321 (2014).
10. R. K. Merton, The Matthew effect in science. *Science* **159**, 56–63 (1968).
11. J. Tague, The success-breeds-success phenomenon and bibliometric processes. *J. Am. Soc. Inf. Sci.* **32**, 280–286 (1981).
12. A. van de Rijt, S. M. Kang, M. Restivo, A. Patil, Field experiments of success-breeds-success dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 6934–6939 (2014).
13. A. L. Barabási, Luck or reason. *Nature* **489**, 507–508 (2012).
14. Z. Néda, L. Varga, T. S. Biró, Science and Facebook: The same popularity law!. *PLoS ONE* **12**, 1–11 (2017).
15. H. M. Blalock, *Social Statistics* (McGraw-Hill, New York, NY, ed. 2, 1972).
16. G. Ionescu, B. Chopard, An agent-based model for the bibliometric h-index. *Eur. Phys. J. B* **86**, 426 (2013).
17. B. Żogała-Siudem, G. Siudem, A. Cena, M. Gagolewski, Agent-based model for the h-index – Exact solution. *Eur. Phys. J. B* **89**, 21 (2016).
18. A. M. Petersen, H. E. Stanley, S. Succi, Statistical regularities in the rank-citation profile of scientists. *Sci. Rep.* **1**, 181 (2011).
19. L. Egghe, Lotkaian informetrics and applications to social networks. *Bull. Belg. Math. Soc. Simon Stevin* **16**, 689–703 (2009).
20. K. Sangwal, Comparison of different mathematical functions for the analysis of citation distribution of papers of individual authors. *J. Informetrics* **7**, 36–49 (2013).
21. M. Thelwall, Are the discretised lognormal and hooked power law distributions plausible for citation data?. *J. Informetrics* **10**, 454–470 (2016).
22. F. Radicchi, S. Fortunato, C. Castellano, Universality of citation distributions: Toward an objective measure of scientific impact. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 17268–17272 (2008).
23. S. Redner, How popular is your paper? An empirical study of the citation distribution. *Eur. Phys. J. B Condens. Matter Complex Syst.* **4**, 131–134 (1998).
24. M. L. Wallace, V. Larivière, Y. Gingras, Modeling a century of citation distributions. *J. Informetrics* **3**, 296–303 (2009).
25. M. Brzezinski, Power laws in citation distributions: Evidence from Scopus. *Scientometrics* **103**, 213–228 (2015).
26. T. Fenner, M. Levene, G. Loizou, A model for collaboration networks giving rise to a power-law distribution with an exponential cutoff. *Soc. Network.* **29**, 70–80 (2007).
27. M. Thelwall, Are there too many uncited articles? Zero inflated variants of the discretised lognormal and hooked power law distributions. *J. Informetrics* **10**, 622–633 (2016).
28. J. A. G. Moreira, X. H. T. Zeng, L. A. N. Amaral, The distribution of the asymptotic number of citations to sets of publications by a researcher or from an academic department are consistent with a discrete lognormal model. *PLoS ONE* **10**, e0143108 (2015).
29. M. Thelwall, P. Wilson, Distributions for cited articles from individual subjects and years. *J. Informetrics* **8**, 824–839 (2014).
30. M. Thelwall, The discretised lognormal and hooked power law distributions for complete citation data: Best options for modelling and regression. *J. Informetrics* **10**, 336–346 (2016).
31. A. L. Barabási, Scale-free networks: A decade and beyond. *Science* **325**, 412–413 (2009).
32. S. Thurner, F. Kyriakopoulos, C. Tsallis, Unified model for network dynamics exhibiting nonextensive statistics. *Phys. Rev. E* **76**, 036111 (2007).
33. E. A. Leicht, G. Clarkson, K. Shedden, M. E. Newman, Large-scale structure of time evolving citation networks. *Eur. Phys. J. B* **59**, 75–83 (2007).
34. A. Barabási *et al.*, Evolution of the social network of scientific collaborations. *Phys. Stat. Mech. Appl.* **311**, 590–614 (2002).
35. A. L. Barabási, R. Albert, H. Jeong, Mean-field theory for scale-free random networks. *Phys. Stat. Mech. Appl.* **272**, 173–187 (1999).
36. F. Papadopoulos, M. Kitsak, M. A. Serrano, M. Boguñá, D. Krioukov, Popularity versus similarity in growing networks. *Nature* **489**, 537–540 (2012).
37. Z. G. Shao, X. W. Zou, Z. J. Tan, Z. Z. Jin, Growing networks with mixed attachment mechanisms. *J. Phys. Math. Gen.* **39**, 2035–2042 (2006).
38. Z. G. Shao, T. Chen, B.-q. Ai, Growing networks with temporal effect and mixed attachment mechanisms. *Phys. Stat. Mech. Appl.* **413**, 147–152 (2014).
39. M. L. Goldstein, S. A. Morris, G. G. Yen, Group-based yule model for bipartite author-paper networks. *Phys. Rev. E* **71**, 026108 (2005).
40. Z. X. Wu, P. Holme, Modeling scientific-citation patterns and other triangle-rich acyclic networks. *Phys. Rev. E* **80**, 037101 (2009).

41. Z. Xie, Z. Ouyang, P. Zhang, D. Yi, D. Kong, Modeling the citation network by network cosmology. *PLoS ONE* **10**, e0120687 (2015).
42. L. Zalányi *et al.*, Properties of a random attachment growing network. *Phys. Rev. E* **68**, 066104 (2003).
43. S. R. Goldberg, H. Anthony, T. S. Evans, Modelling citation networks. *Scientometrics* **105**, 1577–1604 (2015).
44. M. V. Simkin, V. P. Roychowdhury, A mathematical theory of citing. *J. Am. Soc. Inf. Sci. Technol.* **58**, 1661–1673 (2007).
45. M. Golosovsky, S. Solomon, Growing complex network of citations of scientific papers: Modeling and measurements. *Phys. Rev. E* **95**, 012324 (2017).
46. Y. H. Eom, S. Fortunato, Characterizing and modeling citation dynamics. *PLoS ONE* **6**, e24926 (2011).
47. J. Tang *et al.*, "ArnetMiner: Extraction and mining of academic social networks" in *Proceedings of the Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD'2008)* (Association for Computing Machinery, New York, NY, 2008), pp. 990–998.
48. J. R. Clough, T. S. Evans, What is the dimension of citation space? *Phys. Stat. Mech. Appl.* **448**, 235–247 (2016).
49. R. Heesen, Academic superstars: Competent or lucky? *Synthese* **194**, 4499–4518 (2017).
50. F. W. J. Olver, *et al.*, Eds., NIST digital library of mathematical functions, Version 1.0.24. http://dlmf.nist.gov/. Accessed 1 January 2020.
51. G. Siudem, B. Żogała-Siudem, A. Cena, M. Gagolewski, Three dimensions of scientific impact: Supplementary files and data, estimated_parameters_aminer_dblp_v10.csv.gz. https://github.com/gagolews/three_dimensions_of_scientific_impact. Deposited 26 April 2020.