We consider a version of the D.J.Price’s model for the growth of a bibliographic network, where in each iteration a constant number of citations is randomly allocated according to a weighted combination of accidental (uniformly distributed) and preferential (rich-get-richer) rules. Instead of relying on the typical master equation approach, we formulate and solve this problem in terms of the rank-size distribution. We show that, asymptotically, such a process leads to a Pareto-type 2 distribution with an appealingly interpretable parametrisation. We prove that the solution to the Price model expressed in terms of the rank-size distribution coincides with the expected values of order statistics in an independent Paretian sample. We study the bias and the mean squared error of three well-behaving estimators of the underlying model parameters. An empirical analysis of a large repository of academic papers yields a good fit not only in the tail of the distribution (as it is usually the case in the power law-like framework), but also across the whole domain. Interestingly, the estimated models indicate higher degree of preferentially attached citations and smaller share of randomness than previous studies.

I. INTRODUCTION

Citing papers is a common way to express appreciation of someone else’s research or to acknowledge the relevance thereof to our own results [1]. Regardless of the true motivation [2], the what-to-cite decisions of individual authors can be averaged over the whole bibliographic networks. This way, we may study the underlying mechanisms governing the emergence of particular citation distributions. Understanding these is of interest not only to network science, but also to the entire research community, especially bearing in mind that citation counts are widely used as a (controversial) proxy for articles’, authors’, and journals’ impact.

One such mechanism is known as rich-get-richer, success-breeds-success [3], the Matthew effect [4], or the preferential attachment rule [5]. It assumes that highly cited papers are most likely to receive even more citations (for the possible bibliometric applications of this rule see [6, 7]). Yet, recent research [8, 9] into the origins of success in science and beyond highlights the role of other factors — such as chance. Some [10, 11] claim that luck contributes more than the beneficiaries are eager to admit.

This leads to the question: how to measure the level of randomness in a citation network? It is noted in [12] that when citations follow the rich-get-richer rule, it is virtually impossible to distinguish between the merit- or non-merit-driven motivations. In this paper, however, we shall consider D.J.Price’s model [13] which explicitly combines the two said mechanisms. We shall develop a few methods for identifying the preferential-to-accidental ratio from real-world data. Let us also note that the literature knows numerous other approaches to modeling the structure of the citation networks, including those based on recursive searching and fitness factors [14–16].

The Price model was studied in, amongst others, [17, Chap. 14] and frequently appears in the literature under different names and modifications [18–23] and in different contexts: e.g., resistance to random failures and intentional attacks in complex networks [18] or computation of longest paths in random graphs [24].

The most typical approach (e.g., [21]) to deriving the citation distribution and thus the preferential-to-accidental ratio in the Price model is via master equations (see [25]). In this work, however, we shall apply a rank-size (order statistics) approach which was inspired by our earlier work [23], where we studied citation vectors of individual scientists, i.e., in small scale, but with similar accidental and preferential contributors. Here we shall modify the model’s boundary conditions so that we can focus on papers which have obtained a sufficiently large number of citations (e.g., 1, to avoid problems with computing and drawing on the log scale). Moreover, we shall consider citation networks in their entirety, i.e., study the model’s asymptotic behaviour. We shall show that this leads to the well-known Pareto-type 2 distribution, albeit with a new, appealing parametrisation. This way we make an interesting addition to the list of processes from which such a distribution emerges. Revealing the connection between the Price model and the Pareto
distribution will allow us to estimate the preferential-to-accidental ratio based on more statistically reliable methods than used previously in the literature.

This work is set out as follows. First we introduce the rank distribution approach to the preferential-accidental attachment process, which allows us to establish the relation between the Price model and the Pareto-type 2 distribution. Then we discuss three methods for estimating the model parameters from data and quantify their accuracy. Further, we apply them on the DBLP repository of computer science papers. Lastly, we discuss the implications of our findings and propose some directions for future research.

II. RANK DISTRIBUTION APPROACH TO PRICE’S MODEL

Let us consider a process where in every time step the citation network grows by one new paper with $\delta \geq 0$ initial citations. Then we distribute $m - \delta$ citations amongst the already published papers: $p = \rho/(m - \delta)$ citations according to the preferential attachment rule and $a = (1 - \rho)(m - \delta)$ completely at random, with $\rho \in (0, 1)$ representing the extent to which the rich-get-richer rule dominates over pure luck and $a + p = m - \delta$. Note that $m > \delta$ gives the total number of citations added into the system in every time step.

a. Exact solution. Let $X_k(t)$ denote the number of citations of the $k$-th most cited paper at time step $t$. Then we can write

$$X_k(t) = X_k(t-1) + \frac{a}{t} + \frac{p}{\delta} X_k(t-1) + \frac{1}{m} - \frac{p}{\delta}(t-1)m + a,$$

(1)

where $a$ and $p$ were defined above. We assume $X_k(k-1) = \delta$ for every $k$, i.e., the $k$-th publication enters the system with $\delta \in [0, m]$ citations. Solving the above (similarly as we did in [23] but with a less general boundary condition) leads to

$$X_k(t) = \left(\delta + m\frac{1-\rho}{\rho}\right) \frac{\Gamma(t+1)\Gamma(k+\phi)}{\Gamma(k)\Gamma(t+1+\phi)} - \frac{m-\rho}{m},$$

(2)

where $\Gamma$ is Euler gamma function and for brevity we denote $\phi = \rho(\delta/m - 1)$. Notice that the average number of citations in each iteration $t$ is preserved, i.e.,

$$\frac{1}{t} \sum_{k=1}^{t} X_k(t) = m.$$

b. Limiting case. Let us recall the Gautschi’s inequality (see Eq. (7) in [23]) which states that for any $k \geq 0$ and $\phi \in [-1, 0]$, it holds

$$k^\phi \leq \frac{\Gamma(k+\phi)}{\Gamma(k)} \leq (k-1)^\phi.$$

Eq. (2) can be rewritten as a function of $y = k/t$. Applying the above inequality this yields in the limit as $t \rightarrow \infty$

$$X(y) := \lim_{t \rightarrow \infty} X_y(t) = m\frac{1-\rho}{\rho} + \frac{m+\delta\rho - pm}{\rho} y^\phi$$

$$= \left(\delta + m\frac{1-\rho}{\rho}\right) y^{-\phi(1-\delta/m)} - \frac{m-\rho}{m}.$$

(3)

The inverse of $X$, denoted $S = X^{-1}$, is given by

$$S(x) = \left(\frac{x - m + m/\rho}{\delta - m + m/\rho}\right)^{1/\phi} = \left(1 + \frac{x - \delta}{m/\rho - m + \delta}\right)^{-1/\phi}$$

$$= \begin{cases} \frac{m-\rho(m-\delta)}{\phi x(t-1)(1-\rho)m} & \text{for } x \geq \delta \\ 1 & \text{otherwise} \end{cases},$$

(4)

It is a strictly decreasing continuous function onto $(0, 1]$ and hence we can treat it as the complementary cumulative distribution function (CCDF) of what we from now on shall call the $\delta$-truncated Price distribution with parameters $m > \delta \geq 0$ and $\rho$. Generally, if a random variable $Z$ has CCDF $S$, then $S(x) = \Pr(Z > x)$. It inverse $X = S^{-1}$ is referred to as the complementary quantile function.

c. Price meets Pareto. The CCDF given by Eq. (4) can be re-expressed for $x \geq \delta$ with

$$S(x) = \left(\frac{x - \delta}{\lambda}\right)^{-\alpha} = \left(\frac{\lambda}{x - \delta + \lambda}\right)^{\alpha}$$

(5)

where $\alpha = \frac{m}{\rho\delta} - 1$ and $\lambda = (m - \delta)(\alpha - 1) > 0$. This is nothing else than the standard parametrisation of the Pareto-type 2 distribution (e.g., [26]). Hence, the above derivations form an interesting addition to the catalogue of processes from which the Pareto distribution emerges, see [26] for an overview of other ones.

d. Back to the ranks. Despite our model’s stemming from a deterministic setting, its asymptotic expansion gives a description of a whole population of papers, from which we can then pick items at random. It might be interesting to see how the distribution we have just derived relates to Eq. (2).

First, however, let us note (e.g., [29] Eq. 21.9]) that the expected value of a random variable $X \geq \delta \geq 0$ with a continuous CCDF $S$ is given by

$$\mathbb{E}[X] = \int_0^\infty S(x) dx = \delta + \int_\delta^\infty S(x) dx = \int_0^\delta S^{-1}(y) dy.$$

In our case, where $S^{-1}$ is given by Eq. (3), it might be shown that it holds $\mathbb{E}[X] = m$, which is consistent with our model’s assumptions: a sample of randomly selected papers will have $m$ citations on average.

Further, given an independent identically distributed sample of random variables $X_1, \ldots, X_n$ following CCDF $S$, let $X_{r:n}$ denote the $r$-th order statistic, i.e., the $r$-th smallest value therein. Hence, $X_{1:n}$ is the minimum and $X_{n:n}$ is the maximum. Let us derive the formula for
\( \mathbb{E}[X_{r:n}] \). Of course, from Eq. (4.5.1) in \cite{30} we get that for large \( n \) it holds

\[
\mathbb{E}[X_{r:n}] = 1 - S^{-1}(r/(n+1)).
\]

However, it turns out that we can provide the exact formula. Namely, knowing that the \( r \)-th order statistic has CCDF (see \cite{30})

\[
S_{r:n}(x) = 1 - \sum_{j=r}^{n} \binom{n}{j} (1 - S(x))^j (S(x))^{n-j},
\]

we obtain

\[
\mathbb{E}[X_{r:n}] = \int_0^\infty S_{r:n}(x) \, dx = \delta + \int_\delta^\infty \left[ \sum_{j=0}^{n} \binom{n}{j} (1 - S(x))^j S^{-j}(x) + \right. \\
\left. \sum_{j=r}^{n} \binom{n}{j} (1 - S(x))^j S^{-j}(x) \right] \, dx
\]

Applying the binomial theorem to \((1 - S(x))^j\) and performing the elementary integration, we get

\[
\mathbb{E}[X_{r:n}] = \delta + (m - \delta) \sum_{j=0}^{n} \left( \sum_{\ell=0}^{j} \binom{j}{\ell} (-1)^{\ell} \delta^{-m+\ell} \right) \frac{\phi}{\Gamma(n-r+1)} (1+\phi) \frac{\Gamma(n+1)}{(n+r+1)}
\]

Let us also note that

\[
\frac{\Gamma(-(n+\phi))}{\Gamma(-(n-r+\phi))} = (-1)^r \frac{\Gamma(n-r+1+\phi)}{\Gamma(n+1+\phi)}. 
\]

Combining this with the above yields

\[
\mathbb{E}[X_{r:n}] = m \frac{\phi - 1}{\rho} + \frac{m \delta \rho - \rho m \phi}{\rho} \frac{\Gamma(n+1)}{\Gamma(n-r+1)} \frac{\Gamma(n-r+1+\phi)}{\Gamma(n+1+\phi)},
\]

which is equivalent to Eq. (2) with \( k = n - r + 1 \) and \( t = n \). This strengthens the rationale behind our asymptotic expansion even further: the rank-size approach gives the expected values of order statistics of any finite (including very small ones) sample therefrom.

C. Closer to real-world data: Discretisation. Typically, real-world empirical data are not continuous (for example, raw citation counts). In the sequel, we shall thus study a discretised version of our distribution. Let us assume that data come from the original distribution, \( X \), but what we observe is its truncated version, \( [X] \), being the greatest integer less than or equal to \( X \). The CCDF of such a random variable is given by

\[
\Pr([X] > x) = \Pr(X > [x] + 1) = S([x] + 1).
\]

On a side note, for \( \delta = 0 \), the master equation approach (see \cite{25}) applied to the same random process yields, in the limit as \( N \to \infty \), the probability mass function \( p(x) \) for any \( x \in \mathbb{N}_0 \) given by

\[
p(x) = \frac{1}{m+1} \frac{\Gamma(m/\rho - m + 1 + \frac{1}{\rho})}{\Gamma(m/\rho - m) \Gamma(x + m/\rho - m + 1 + \frac{1}{\rho})} \]

\[
= \frac{1}{m+1} \frac{\Gamma(m/\rho - m + 1 + \frac{1}{\rho})}{\Gamma(m/\rho - m) \Gamma(x + m/\rho - m + 1 + \frac{1}{\rho})},
\]

where \( (a)_x = \Gamma(a + x)/\Gamma(a) \) denotes the Pochhammer symbol.

It may be shown that for any integer \( k \):

\[
\sum_{\ell=0}^{k} p(\ell) = 1 - \frac{\Gamma(-m + m/\rho + 1)}{\Gamma(-m + m/\rho)} \frac{\Gamma(k - m + m/\rho + 1)}{\Gamma(k - m + m/\rho + 1 + \frac{1}{\rho})},
\]

hence the complementary cumulative distribution function of this distribution is defined for any \( x \geq 0 \) as:

\[
S(x) = \sum_{\ell=[x]+1}^{\infty} p(\ell) = \frac{\Gamma(m/\rho - m + 1 + \frac{1}{\rho})}{\Gamma(1 + \frac{m/\rho - m + 1 + \frac{1}{\rho}}{\rho})} \cdot
\]

By the aforementioned Gautschi’s inequality, we have that our discretisation leads to:

\[
S([x] + 1 | \delta = 0) = \frac{\Gamma(\rho - m + 1)}{\Gamma(\rho - m + 1 + \frac{1}{\rho})} \cdot S(x),
\]

which gives a nice correspondence between the master equation- and the rank distribution-approach.

III. ESTIMATING MODEL PARAMETERS

Fitting different heavy-tailed distributions to real-world data is quite commonly exercised in the complex systems practice; in particular, \cite{31} feature a comprehensive discussion related to the fitting of the power law distributions.

a. All parameters are interpretable. The model we have landed at is in fact a Pareto-type 2 distribution, which by itself has of course been studied extensively, see \cite{24}. However, in this paper we have arrived at a non-classic parametrisation, where all parameters are easily interpretable. Namely, \( \delta \) gives the cut-off, \( m \) is the expected number of citations or a randomly selected paper (and also the number of citations allocated per iteration), whereas \( \rho \) gives the ratio of preferentially to accidently attached citations.

We should thus be interested in studying the properties of various estimators thereof so that we can reliably fit our model to empirical data (the threshold \( \delta \geq 0 \) is usually known in advance, hence we consider it fixed).
and finally

form). From this we can get:

\[
\frac{m}{\rho - \text{continuous data}} = \begin{cases}
(0.27) & -0.00 \\
(0.25) & 0.01 \\
(0.29) & 0.01 \\
(0.18) & 0.00
\end{cases}
\]

\[
\frac{m}{\rho - \text{discretised data}} = \begin{cases}
(0.43) & 0.00 \\
(0.56) & 0.01 \\
(0.41) & 0.00 \\
(0.41) & 0.00
\end{cases}
\]

The first equation, involving \(\lambda\) only, can easily be solved numerically (trying to express the likelihood directly as a function of \(m\) and \(\rho\) does not lead to such a convenient form). From this we can get:

\[
m = \frac{\lambda}{\alpha - 1} + \delta, \quad \rho = \frac{m}{\alpha(m - \delta)}.
\]

We can thus consider the maximum likelihood estimator (MLE) of \(\alpha\) and \(\lambda\):

\[
\begin{align*}
0 &= \frac{1}{n} \sum_{i=1}^{n} \log(x_i + \lambda - \delta) + \frac{1}{\alpha} \sum_{i=1}^{n} \log \lambda - \frac{n}{\lambda} \sum_{i=1}^{n} (x_i + \lambda - \delta) - 1, \\
\alpha &= \frac{1}{\frac{1}{n} \sum_{i=1}^{n} \log(x_i + \lambda - \delta) - \log \lambda}
\end{align*}
\]

The first equation, involving \(\lambda\) only, can easily be solved numerically (trying to express the likelihood directly as a function of \(m\) and \(\rho\) does not lead to such a convenient form). From this we can get:

\[
m = \frac{\lambda}{\alpha - 1} + \delta, \quad \rho = \frac{m}{\alpha(m - \delta)}.
\]

We can thus consider the maximum likelihood estimator of \(\lambda\), which is determined by solving

\[
\frac{1}{n} \sum_{i=1}^{n} \log(x_i + \lambda - \delta) + 1 - \log \lambda - \frac{n}{\lambda} \sum_{i=1}^{n} (x_i + \lambda - \delta) - 1 = 0
\]

umerically (trying to express the likelihood directly as a function of \(m\) and \(\rho\) does not lead to such a convenient form). Then we get

\[
\alpha = 1 / \left[ \frac{1}{n} \sum_{i=1}^{n} \log(x_i + \lambda - \delta) - \log \lambda \right]
\]

and finally \(m = \frac{\lambda}{\alpha - 1} + \delta\) and \(\rho = \frac{m}{\alpha(m - \delta)}\).

For the sake of comparison, we will consider two other estimators based on the empirical CCDF, \(S(x|X_1, \ldots, X_n) = \left| \{i : X_i > x\} \right| / n\): one that minimises the maximal difference between the empirical and theoretical CCDF (from now on called SUP)

\[
\min_{m, \rho} \max_{x \in \{X_1, \ldots, X_n\}} \left| S(x|X_1, \ldots, X_n) - S(x|m, \rho) \right|
\]

and one that minimises the sum of squared errors (SSE) at all points of discontinuity of the empirical CCDF

\[
\min_{m, \rho} \sum_{x \in \{X_1, \ldots, X_n\}} \left( S(x|X_1, \ldots, X_n) - S(x|m, \rho) \right)^2
\]

We compute them using the Nelder–Mead method with a logarithmic barrier (via \texttt{constrOptim} in R) to enforce the constraints on \(m\) and \(\rho\) (the algorithm was started from, respectively, the 90% trimmed mean and 0.5).

Similarly, we will also consider the MLE, SSE, and SUP estimators for the discretised version of our distribution, \([X]\), with MLE being this time a (numerical) solution to

\[
\max_{m, \rho} \sum_{i=1}^{n} \log \rho(X_i|m, \rho),
\]

where the probability mass function \(\rho\) given by \(\rho(x) = S([x]) - S([x] + 1)\). Note that this time we are optimising the likelihood directly over \(m\) and \(\rho\).

\[c.\] Assessing estimator quality.

Thanks to the exact formula [1] for the inverse of the CCDF and the principle of inverse transform sampling, we can generate realisations of independents samples like \(X_1, \ldots, X_n\) following our distribution for specific \(\delta, m,\) and \(\rho\). To study the quality of the estimators, we have used \(M = 10,000\) Monte Carlo samples (which took a few hours to compute on a modern PC) each consisting of \(n = 100,000\) (i.e., we were interested in a large-sample behaviour) and also \(\delta = 1\) and \(m = 25\) (which resemble the setting from our empirical study in the next section) and a range of \(\rho\)s from the set \(\{0.5, 0.75, 0.9\}\).

The left-hand side of Table I gives the approximate bias and the root mean squared error of the various estimators of the \(m\) parameter. Recall that for a given estimator \(\hat{m}(X_1, \ldots, X_n)\) (which is a function of a sequence of random variables) of the true parameter \(m\) (which is a fixed value), its bias is defined as \(E[\hat{m}(X_1, \ldots, X_n) - m]\), whereas the root mean squared error is given by \(\sqrt{E[(\hat{m}(X_1, \ldots, X_n) - m)^2]}\). They can be estimated based on \(M\) Monte Carlo samples \(x_1^{(i)}, \ldots, x_n^{(i)}, \ i = 1, \ldots, M,\) using, respectively, the sample mean and
standard deviation of a vector \((e_1, \ldots, e_M)\) with \(e_i = \sqrt{\frac{1}{n} (x_1^{(i)}, \ldots, x_n^{(i)}) - m}\).

Interestingly, the one sample Student \(t\)-test with a significance level of 0.01 indicates that in many cases, the \(e_i\) values are, on average, not significantly different from 0. This indicates that the estimators might be unbiased, especially for smaller \(\rho\)s.

It might be tempting to use the methods of moments estimator for the \(m\) parameter, i.e., the arithmetic mean. Despite its being an unbiased estimator, it unfortunately tends to have quite high variance and hence it is a practically useless measure. We note that \(\alpha > 1\) guarantees the existence of the expected value (which always holds in our case), but the variance is only defined if \(\alpha > 2\), which for \(m = 25\) and \(\delta = 1\) holds whenever \(\rho < 25/48 \approx 0.521\) (but it is not the case in our empirical study in Sec. \(\text{III}\,0\,e\)).

Both in the continuous as well as in the discretised case, the MLE estimators work very well. The discretised case even seems more well-behaving. The MLE estimator should definitely be chosen if we suspect that the data really come from the distribution studied herein. Nevertheless, any deviations from the model such as data contamination might affect its performance. In such a case, the SSE estimator could also be noteworthy. SUP, on the other hand, is theoretically a consistent (i.e., with theoretical curve. It seems, however, that our (asymptotic) model expects the most highly cited papers in the (finite) sample to be cited more frequently than they really are.

It may also be interesting to study a different threshold \(\delta\). Figure \(\text{III}\,0\,e\) depicts the CCDFs in the case \(\delta = 50\) (which yields \(n = 320.007\) and the average number of citations of \(\bar{m} = 162.01\)). All estimators, this time, give similar results. The power law model agrees with them in the distribution tail, but is again less precise for small \(x\)s.

### IV. CONCLUSIONS

Our rank distribution approach marries the Price model and Pareto-type 2 distributions. We have shown how a combination of the rich-get-richer rule and sheer chance effects, in the long run, a well-known statistical distribution with an appealing, interpretable parametrisation. Reversely, the expected values of order statistics of any finite i.i.d. Paretoian sample are consistent with the baseline Price ranks.

The considered model fits the DBLP citation data quite well. Notably, the estimated \(\rho\) parameter \((\hat{\rho}_{\text{MLE}} = 0.87)\) corresponds to a very high fraction of preferentially-attached citations – other studies of different real-world databases (e.g., \(\text{[21]}\,\text{[33]}\)) suggested that success might be more accidental.

Future work will involve the study of the quality of the proposed estimators on smaller data samples as well as their application on data from other domains.

### ACKNOWLEDGMENTS

This research was supported by the Australian Research Council Discovery Project ARC DP210100227 (MG) and by the POB Research Centre Cybersecurity and Data Science of Warsaw University of Technology within the Excellence Initiative Program – Research University (ID-UB) (GS and PN). The authors would also like to thank Anna Cena, Barbara Żogała-Siudem, and Maciej J. Mrowiński for valuable remarks.
FIG. 1. Empirical (DBLP data, truncated at $\delta = 1$) and fitted complementary cumulative distribution functions; left: log scale on Ox, middle: log scale on both axes, right: error between the empirical and the theoretical CCDFs; the power-law model (straight line on the log-log scale) gives a particularly bad fit for smaller observations (which are the most prevalent); MLE, SSE, and SUP seem to give poorer fits at the distribution tail in the middle plot but note that the probabilities therein are on the log scale, which exaggerates how the error magnitudes are perceived.

FIG. 2. Histogram (DBLP data, truncated at $\delta = 1$) and the fitted probability mass functions; left: log scale on Ox, middle: log scale on both axes, right: error between the histogram and the theoretical PMFs; note again how the power-law model overestimates the density at smaller xs.

FIG. 3. Empirical (DBLP data, truncated at $\delta = 50$) and fitted complementary cumulative distribution functions, see Figure 1 for description. The power-law model (fitted on the log-log scale) gives a similar behaviour in the tails, but is inferior for small xs.
[1] M. Thelwall and P. Sud, Do new research issues attract more citations? A comparison between 25 Scopus subject categories, Journal of the Association for Information Science and Technology 72, 269 (2021).

[2] D. Lyu, X. Ruan, J. Xie, and Y. Cheng, The classification of citing motivations: A meta-synthesis, Scientometrics 126, 3243 (2021).

[3] D. J. de Solla Price, Little Science, Big Science (Columbia Univ. Press, New York, 1963).

[4] R. K. Merton, The Matthew effect in science, Science 159, 56 (1968).

[5] M. Perc, The Matthew effect in empirical data, Journal of The Royal Society Interface 11, 20140378 (2014).

[6] G. Ionescu and B. Chopard, An agent-based model for the bibliometric h-index, Eur. Phys. J. B 86, 426 (2013).

[7] B. Žoga la-Siudem, G. Siudem, A. Cena, and M. Gagolewski, Agent-based model for the h-index – Exact solution, European Physical Journal B, 21 (2016).

[8] R. Sinatra, D. Wang, P. Deville, C. Song, and A.-L. Barabási, Quantifying the evolution of individual scientific impact, Science 354 (2016).

[9] M. Janosov, F. Battiston, and R. Sinatra, Success and luck in creative careers, EPJ Data Science 9, 9 (2020).

[10] A. Pluchino, A. E. Biondo, and A. Rapisarda, Talent versus luck: The role of randomness in success and failure, Advances in Complex systems 21, 1850014 (2018).

[11] A. Pluchino, G. Burgio, A. Rapisarda, A. E. Biondo, A. Pulvirenti, A. Ferro, and T. Giorgino, Exploring the role of interdisciplinarity in physics: Success, talent and luck, PLOS ONE 14, e0218793 (2019).

[12] R. Heesen, Academic superstars: Competent or lucky?, Synthese 194, 4499 (2017).

[13] D. S. Price, A general theory of bibliometric and other cumulative advantage processes, Journal of the American Society for Information Science 27, 292 (1976).

[14] M. Golosovsky and S. Solomon, Growing complex network of citations of scientific papers: Modeling and measurements, Phys. Rev. E 95, 012324 (2017).

[15] M. Golosovsky, Mechanisms of complex network growth: Synthesis of the preferential attachment and fitness models, Phys. Rev. E 97, 062310 (2018).

[16] M. Golosovsky, Citation Analysis and Dynamics of Citation Networks (Springer, 2019).

[17] M. Newman, Networks (Oxford University Press, 2018).

[18] Z. Liu, Y.-C. Lai, N. Ye, and P. Dasgupta, Connectivity distribution and attack tolerance of general networks with both preferential and random attachments, Physics Letters A 303, 337 (2002).

[19] C. Bedogne and G. J. Rodgers, Complex growing networks with intrinsic vertex fitness, Physical Review E 74, 046115 (2006).

[20] Z.-G. Shao, X.-W. Zou, Z.-J. Tan, and Z.-Z. Jin, Growing networks with mixed attachment mechanisms, Journal of Physics A: Mathematical and General 39, 2035 (2006).

[21] G. J. Peterson, S. Pressé, and K. A. Dill, Nonuniversal power law scaling in the probability distribution of scientific citations, Proceedings of the National Academy of Sciences 107, 16023 (2010).

[22] Z. Néda, L. Varga, and T. S. Biró, Science and Facebook: The same popularity law!, PLOS ONE 12, 1 (2017).

[23] G. Siudem, B. Zogala-Siudem, A. Cena, and M. Gagolewski, Three dimensions of scientific impact, Proceedings of the National Academy of Sciences 117, 15896 (2020).

[24] T. S. Evans, L. Calmon, and V. Vasilakis-Kai, The longest path in the Price model, Scientific Reports 10, 1 (2020).

[25] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Structure of growing networks with preferential linking, Phys. Rev. Lett. 85, 4633 (2000).

[26] B. C. Arnold, Pareto Distributions (Chapman and Hall/CRC, New York, NY, USA, 2015).

[27] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, Arnetminer: Extraction and mining of academic social networks, in Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (2008) pp. 990–998.

[28] W. Gautschi, Some elementary inequalities relating to the gamma and incomplete gamma function, Journal of Mathematics and Physics 38, 77 (1959).

[29] P. Billingsley, Probability and Measure (Wiley, 1995).

[30] H. A. David and H. N. Nagaraja, Order statistics (Wiley, 2003).

[31] M. E. Newman, Power laws, Pareto distributions and Zipf’s law, Contemporary Physics 46, 323 (2005).

[32] A. Clauset, C. R. Shalizi, and M. E. Newman, Power-law distributions in empirical data, SIAM Review 51, 661 (2009).

[33] N. A. Arnold, R. J. Mondragon, and R. G. Clegg, Likelihood-based approach to discriminate mixtures of network models that vary in time, Scientific Reports 11, 1 (2021).