

RESEARCH ARTICLE

Science and Facebook: The same popularity law!

Zoltán Néda^{1*}, Levente Varga¹, Tamás S. Biró²

1 Babeş-Bolyai University, Department of Physics, Cluj-Napoca, Romania, **2** HIRG, HAS Wigner Research Centre for Physics, Budapest, Hungary

* zneda@phys.ubbcluj.ro



Abstract

The distribution of scientific citations for publications selected with different rules (author, topic, institution, country, journal, etc. . .) collapse on a single curve if one plots the citations relative to their mean value. We find that the distribution of “shares” for the Facebook posts rescale in the same manner to the very same curve with scientific citations. This finding suggests that citations are subjected to the same growth mechanism with Facebook popularity measures, being influenced by a statistically similar social environment and selection mechanism. In a simple master-equation approach the exponential growth of the number of publications and a preferential selection mechanism leads to a Tsallis-Pareto distribution offering an excellent description for the observed statistics. Based on our model and on the data derived from PubMed we predict that according to the present trend the average citations per scientific publications exponentially relaxes to about 4.

OPEN ACCESS

Citation: Néda Z, Varga L, Biró TS (2017) Science and Facebook: The same popularity law! PLoS ONE 12(7): e0179656. <https://doi.org/10.1371/journal.pone.0179656>

Editor: Pablo Dorta-González, Universidad de las Palmas de Gran Canaria, SPAIN

Received: February 9, 2017

Accepted: June 1, 2017

Published: July 5, 2017

Copyright: © 2017 Néda et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The data is available on the Internet by following the links in the Reference section and using the Methodology described for processing them. All data are public data and no privacy issues were violated. We have uploaded to Figshare (public repository: https://figshare.com/articles/Science_and_Facebook_the_same_popularity_law_/4986551; DOI:10.6084/m9.figshare.4986551.v2) all data used for plotting the Figures.

Funding: Nemzeti Kutatási, Fejlesztési és Innovációs Hivatal, <http://nkfih.gov.hu/>, NKFIH/OTKA project Nr. 104260, funded the research of

Introduction

The number of citations for a publication is basically a social popularity measure for it, while it is considered to reflect the quality and impact of the research.

Citations are thus in our focus when evaluating researchers, groups and institutes [1–3]. The statistics and dynamics of citations are studied in several works [4–9] and lately we assisted to many serious debates on their use for quantifying objectively the quality and impact of a given research [1–3, 10–13]. In view of this, further scientific arguments or novel information regarding the citation statistics and its similarity to other social selection mechanisms is of enhanced importance.

It has been reported [4–6] that citations for scientific papers, selected according to an arbitrary collection rule, like author, topic, publication year, institution, journal, etc. . . , rescale on a common curve if considering their value relative to the average. More specifically, if one computes for the selected set the probability density $f(x)$ for one paper to have x citations, and represent graphically the $\langle x \rangle \cdot f(x)$ value as a function of $x/\langle x \rangle$, the data obtained for different sets will collapse on the same curve (see the figures in [4–6] and Fig 1). We denoted here with $\langle x \rangle = \int_0^\infty x f(x) dx$, the mean value of x , or the first moment of the probability distribution function (PDF). For high citation numbers a clear power-law trend is visible, especially if one

TSB. The authors also acknowledge Universitatea Babeş-Bolyai, <http://www.ubbcluj.ro/en/>, STAR UBB fellowship for allowing the collaboration between ZN and TSB.

Competing interests: The authors have declared that no competing interests exist.

considers datasets where the $x/\langle x \rangle > 10$ domain is visible. There is, however, a very active debate on fitting this rescaled curve [4–6, 9, 14–19]. Researchers have suggested lognormal, negative binomial, Wakeby and power-law tailed distributions to fit the entire curve. Recent results [16, 18, 19] favour a Tsallis-Pareto (TP) [20, 21] type hooked distribution, albeit the log-normal distribution is still in use [6]. The obvious scale-free nature of the tail and accordingly the observed invariance relative to mixing or selecting just a part of the ensemble is however a major argument in favour of the TP distribution. It worth noting here that recently it has been shown [22] that the frequency distribution of scientific memes follows also a universal distribution with a power-law like tail. In such aspect the TP distribution could be a good candidate for fitting also the results presented in [22].

Biology, physics and socio-economic phenomena offer many intriguing examples of scale-free distributions in complex systems [23–25]. The celebrated Zipf law [26], or many other power-law tailed distributions are widely known and well-studied [27]. The pure power-law, however, is not a distribution in the strict mathematical sense since it cannot be normalized for the whole interval between zero and infinity. Quite frequently we do not even have a large enough scaling interval to prove or disprove the presence of pure power-law distributions

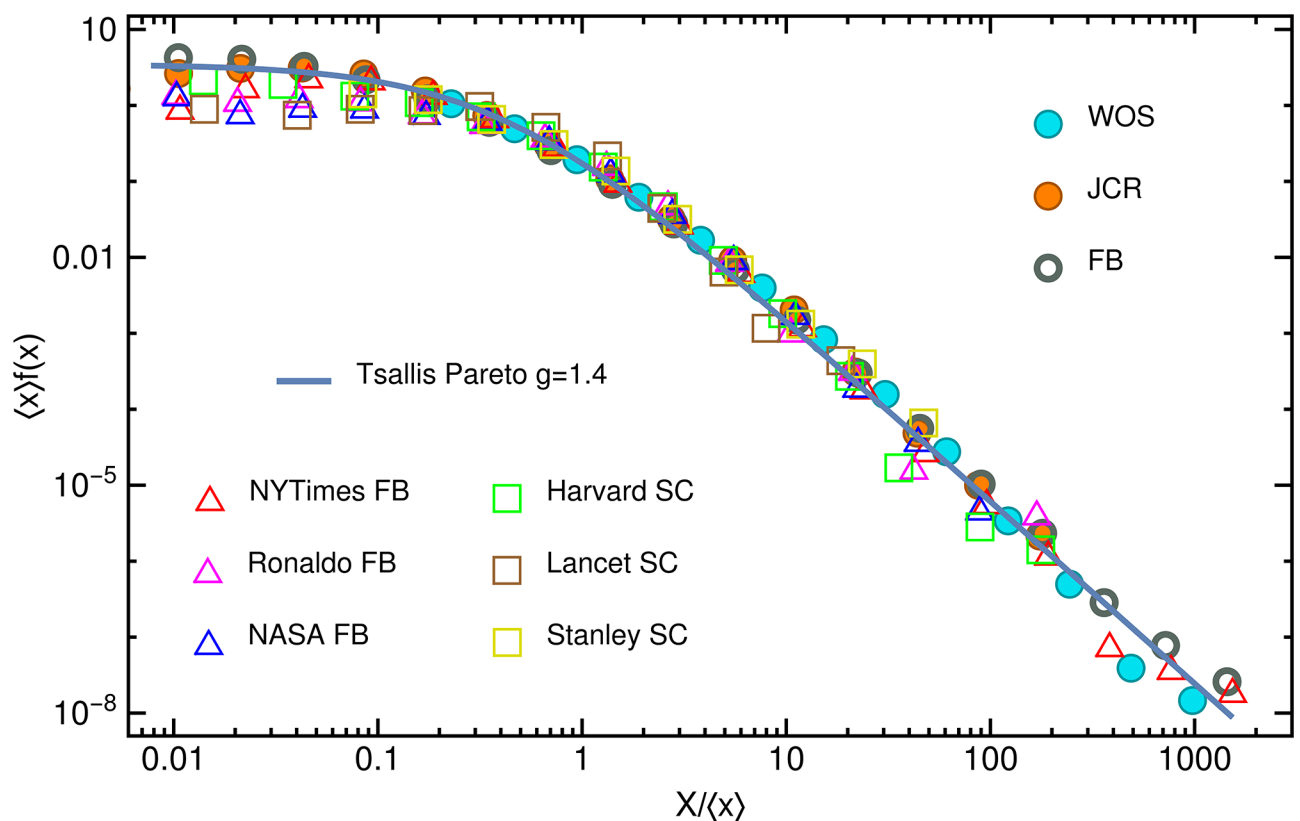


Fig 1. Rescaled distribution of the citation (share) numbers. $f(x)$ is the probability density (PDF) for one paper (post) to have x citations/shares. We present the $\langle x \rangle \cdot f(x)$ value as a function of $x/\langle x \rangle$ ($\langle x \rangle$ the mean value, or first moment of the PDF). For high citation number a clear power-law trend is visible. Different symbols are for different datasets as illustrated in the legend. The considered datasets are described in the Methods section. For high $x/\langle x \rangle$ a clear power-law trend is visible. The entire curve can be well-fitted with a TP distribution Eq (1) with $g \approx 1.4$ and $\langle x \rangle = 1$.

<https://doi.org/10.1371/journal.pone.0179656.g001>

[28]. On the other hand the Tsallis-Pareto distribution [20, 21]

$$f(x) = \frac{g}{(g-1)\langle x \rangle} \left(1 + \frac{x}{(g-1)\langle x \rangle} \right)^{-1-g} \quad (1)$$

is a proper probability density function (PDF) with a power-law like tail. It has been found that many heavy-tailed distributions are well fitted by using the above PDF [24]. Although this is not strictly a scale-free distribution, one can numerically check that for $g > 1$ exponents and for large enough $x/\langle x \rangle$ the scale free properties and invariance under mixing or splitting of the dataset are well satisfied.

Results and discussions

A simple exercise on citation data collected from more than 600 000 ISI Web of Science (WOS) publications (mapping a part of the WOS citation network by using an Internet robot, please see the [Methods](#) section), draws the shape of the universal form for the studied distribution. If one makes a simple data processing exercise from the total number of citations received in ten years for all ISI indexed journals (InCites, Journal Citation Reports [29]), the data (JCR) scale on the very same curve. If we select now data for the publications authored by one researcher, for the publications published in a given journal in one given year or by authors associated to a given institute, the data rescale again. For $x/\langle x \rangle \geq 0.1$ the collapsed data can be nicely fitted with a one-parameter TP type PDF, using $g \approx 1.4$ and $\langle x \rangle = 1$. (see [Fig 1](#)). As already emphasised, this type of fit has the advantage that the scale-free property for $x/\langle x \rangle \geq 0.1$ is evident and also explains the invariance of the distribution when combining several data sets.

A similar study can be performed on different Facebook pages for their posts (for details please consult again the [Methods](#) section). Instead of citations the popularity proxy for a post is the number of “shares” it receives. “Share” is a stronger selection rule than the simple “like”, and it’s role is similar to citations in Science. Interestingly the PDF for “shares” collected from 16 different Facebook users (in total more than 150 000 posts) scale on the very same curve with the Scientific Citations ([Fig 1](#)). The universal TP type distribution with $g \approx 1.4$ suggests a common growth mechanism for the Facebook Shares and Scientific Citations. Reducing now the Facebook data on users, the rescaled PDF behaves in a similar manner (for details on the used data see again the [Methods](#) section). Due to the larger scatter for the data points resulting from the reduced data size (both for scientific citations and Facebook shares) we cannot conclude however the same Pareto exponent, just note the similar trend. The invariance of the distributions relative to the splitting of the data is in agreement with the scale-free properties of this distribution.

The appropriateness of the TP type fit can be proved by computing the generally used R^2 *coefficient of determination*. We have computed this for the datasets that contained a fair number of elements (e.g. the statistics is reliable): WOS, JCR and FB. The $R^2 > 0.9$ values presented in [Table 1](#) indicates that the TP type fit is justified. For the smaller datasets such statistics is not relevant, and one can just note that the data points are following the trend of the TP fit with $g = 1.4$.

Many models have been already considered for explaining the dynamics of citations [30–32] and the observed universality in the rescaled PDF [33, 34]. A simple explanation for this intriguing universality can be given by considering a coarse-grained master equation for the growth process and assuming an exponential growth of publications (post) number as a function of time together with a linear preferential growth rate in the flow (for details please see the technical subsection: The Master equation approach).

Table 1. Coefficient of determination for the TP fit in Fig 1.

	WOS	JCR	FB
<i>n</i>	629 575	12 026	160 889
<i>R</i> ²	0.982	0.967	0.940

Goodness of fit for the large datasets (WOS, JCR and FB) in Fig 1. We denote by *n* the size of the dataset and by *R*² the coefficient of determination. The fit is given by the TP distribution, Eq (1), with *g* = 1.4 and $\langle x \rangle = 1$. For more details on the experimental data please consult the Methods section.

<https://doi.org/10.1371/journal.pone.0179656.t001>

The approach considered here is the simplest mean-field type approximation where only the stochastic nature of the growth process is taken into account and the specificity of the posts quality are coarse-grained. The exponential growth of the number of publications which are the carriers of the citations is known (see for example [35, 36]). In a recent statement from Mark Zuckerberg we also learn that the information sharing activity on Facebook is also growing exponentially (see for example [37]).

On the other hand the linear preferential growth rate hypothesis or the commonly known Matthew effect (“For to all those who have, more will be given”) has been highlighted in various social systems [38–40]. The presence of the Matthew effect in citation and science was also discussed in many previous publications [41, 42]. In such manner the two main assumptions of our simple model are all reasonable, and can be applied both to Facebook posts and scientific articles. The Markov-like process constructed on these bases can be analytically solved also in the continuous limit where it leads to a TP Eq (1) probability distribution (see the technical subsection: The Master equation approach). From the model we learn that the parameter *g* in the TP distribution, governing the power-law tail, is just the ratio of the exponential growth rate γ to the proportionality constant σ for the linear preferential growth: $g = \gamma/\sigma$. The fact that the obtained *g* value is independent from the way we construct the studied ensemble and it is the same for Facebook posts and Scientific Publications is intriguing. It can be understood by taking into account that both phenomena are taking place on a social network with similar topological properties, where the released information amount is increasing exponentially and the selection rules for its transmission are adapted to the increase rate.

The Master equation approach- technical details

We consider a classical master equation approach for the growth phenomenon. This approach is the simplest possible mean-field like description where the properties of different elements (posts, publications) are coarse-grained and only the stochastic character of the process is kept. In this framework, the stochastic growth process is quantified by a mean growth rate μ_n describing the transition rate from state with *n* quanta (citations, shares, likes. . .) to a state with *n* + 1 quanta. Since there is no reverse process inside the chain, just a continuous growth a detailed balance condition cannot be fulfilled. We illustrate this process in the left panel of Fig 2, where *N_n(t)* denotes the number of elements having *n* quanta at time moment *t*. A master equation for this process writes as:

$$\frac{dN_n(t)}{dt} = \mu_{n-1}N_{n-1}(t) - \mu_n N_n(t) \tag{2}$$

In order to achieve a non trivial steady-state distribution, parallel with this continuous growth a continuous dilution should be present in the system. This can be achieved by assuming that

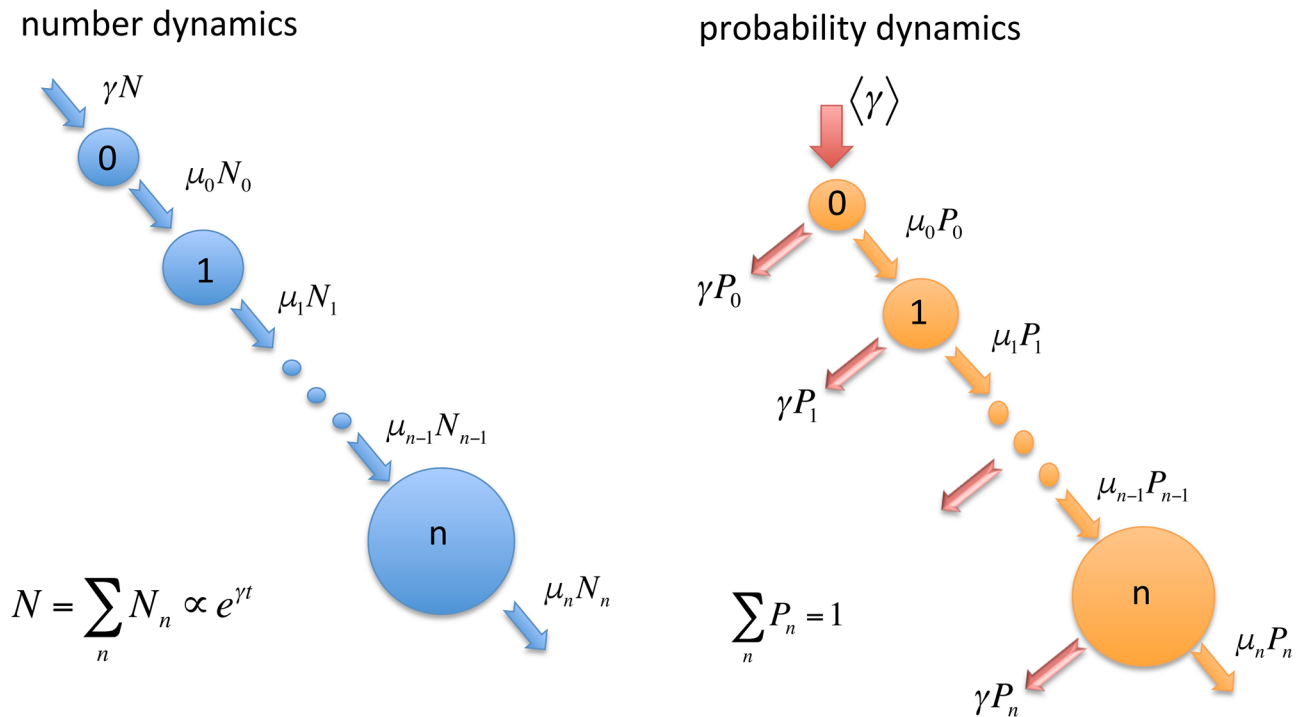


Fig 2. Schematic representation of the coarse-grained random growth model. The panel on the left side indicates the growth process in the number of elements with n quanta: N_n . Due to the fact that the total number of elements is exponentially increasing, the probability P_n that an element will have n quanta, experiences the dynamics sketched on the right panel of the figure.

<https://doi.org/10.1371/journal.pone.0179656.g002>

the number of elements are continuously increasing in time. This means that

$$N(t) = \sum_n N_n(t) \tag{3}$$

is increasing in time. Considering now the probability $P_n(t)$ that an element has n quanta at time moment t

$$P_n(t) = \frac{N_n(t)}{N} \tag{4}$$

we rewrite the master equation using instead of $N_n(t)$ the $P_n(t)$ distribution:

$$\frac{d}{dt}(NP_n) = N \frac{dP_n}{dt} + P_n \frac{dN_n}{dt} = \mu_{n-1}NP_{n-1} - \mu_n NP_n \tag{5}$$

The number of elements in the considered systems is exponentially increasing. Assuming thus an exponential growth in $N(t)$ with a rate γ , characteristic for each ensemble in part (scientific papers, Facebook posts, etc. . .):

$$N(t) = N(0)e^{\gamma t} \rightarrow \frac{dN(t)}{dt} = \gamma N(t), \tag{6}$$

From Eq (5) we arrive now to a master equation in $P_n(t)$:

$$\frac{dP_n}{dt} = \mu_{n-1}P_{n-1} - (\mu_n + \gamma)P_n \tag{7}$$

The flow diagram for this process is illustrated in the right panel of Fig 2. The corresponding equation for the $n = 0$ term can be obtained from the normalization condition $\sum_n P_n(t) = 1$:

$$\frac{dP_0}{dt} = \gamma - (\mu_0 + \gamma)P_0 \tag{8}$$

We consider now the continuous limit of Eq (7) (see for example [4]), where the discrete states n are replaced by continuous x states:

$$\frac{\partial P(x, t)}{\partial t} = -\frac{\partial}{\partial x} (\mu(x)P(x, t)) - \gamma P(x, t) + \gamma \delta(x) \tag{9}$$

This equation describes a flow with a general velocity field $\mu(x)$, a loss rate γ and a feeding at $x = 0$. (We denoted by $\delta(x)$ the Dirac functional). The $P_s(x)$ stationary probability density can be derived from the condition: $\frac{\partial P_s(x,t)}{\partial t} = 0$, and according to Eq (9) it satisfies

$$\frac{d}{dx} (\mu(x)P_s(x)) = -\gamma P_s(x) . \tag{10}$$

The solution of this equation writes as

$$P_s(x) = \frac{K}{\mu(x)} e^{-\gamma \int \frac{1}{\mu(x)} dx} \tag{11}$$

In order to write up the solution one has to specify a kernel for the $\mu(x)$ growth rate. From several social-economic phenomena we learn that the growth is usually governed by a preferential selection, in the simplest case by a linear preferential growth rate (the well-known “rich gets richer” phenomenon or the Matthew effect [38, 39]). According to this

$$\mu(x) = \sigma(x + b) , \tag{12}$$

where the σ and b values are characteristic to the considered ensemble (scientist, Facebook users). Accepting this kernel, Eq (11) leads to the Tsallis-Pareto distribution [20, 21]:

$$P_s(x) = \frac{\gamma}{b\sigma} \left(1 + \frac{x}{b}\right)^{-1-\gamma/\sigma} \tag{13}$$

Denoting $g = \gamma/\sigma$ and using $b = \langle x \rangle (g - 1)$, where $\langle x \rangle$ is the first moment of the distribution, we get:

$$P_s(x) = \frac{g}{(g - 1)\langle x \rangle} \left(1 + \frac{x}{(g - 1)\langle x \rangle}\right)^{-1-g} \tag{14}$$

This is the scaling Tsallis-Pareto distribution, which for $g = 1.4$ and $\langle x \rangle = 1$ offers a good fit for the collapsed data on Fig 1. The prediction of our simple model is in agreement with the more technical approach considered in [32].

From this simple mean-field type model we learn that the popularity measures both for scientific publications and Facebook are the results of an exponential growth and a preferential retransmission of the received information. The collapse for the Facebook popularity measures and scientific citations indicate that for their coarse-grained dynamics the ratio $g = \gamma/\sigma$ should be similar. Seemingly this ratio is also independent on the precise manner in how we construct

the ensembles (institutes, journals, individuals, etc. . .). This is an exciting finding which inspires further studies.

Trend for the average number of citations per paper

From the promising fit indicated in Fig 1, using $g \approx 1.4$, we gain confidence in the statistical prediction capability of our simple mean-field type approximation. We elaborate thus further on our model and make some statistical predictions on the expected evolution of the average number of citations (shares) per publication (post). The total number of citations at time t can be written as: $C(t) = \sum_n nN_n(t)$. According to our hypothesis the increase in the total number of citations in unit time is given as:

$$\frac{dC(t)}{dt} = \sum_n \frac{dN_n^+}{dt} = \sum_n \sigma(n + b)N_n = \sigma C(t) + \sigma bN(t) \tag{15}$$

Combining this with the exponential growth of $N(t)$: $\frac{dN(t)}{dt} = \gamma N(t)$ leads to a simple differential equation for the $m(t) = C(t)/N(t)$ average number of citations per work:

$\frac{dm(t)}{dt} = (\sigma - \gamma)m(t) + \sigma b$. The solution is an exponential relaxation

$$m(t) = \frac{K}{\gamma - \sigma} e^{-(\gamma - \sigma)t} + \frac{b}{g - 1} , \tag{16}$$

where K is an integration constant. Due to the fact that $g = \gamma/\sigma \approx 1.4$ we get $\gamma > \sigma$ and therefore $c(t)$ has an exponentially relaxing trend. From the previous section we learned that $b/(g - 1) = \langle m \rangle_{ech}$, the equilibrium value for the average citation per paper in the considered ensemble.

We can now determine the time-evolution of the total citations number per year. Let us assume now that we measure the time in years, and introduce the yearly published article number $n(t) = \frac{dN(t)}{dt} = n_0 e^{-\gamma(t-t_0)}$, and the new citations that appear in one year: $c(t) = \frac{dC(t)}{dt}$. If we assume that at time t_0 we have $c(t_0) = c_0$ and $n(t_0) = n_0$ we get that

$$K = \left(\frac{c_0}{n_0} (g - 1) - b \right) \frac{\gamma}{g} , \tag{17}$$

and

$$c(t) = c_0 e^{\sigma(t-t_0)} + \frac{bn_0}{(g - 1)} (e^{\gamma(t-t_0)} - e^{\sigma(t-t_0)}) . \tag{18}$$

For the case of scientific articles indexed in MEDLINE/PubMed (see the Methods section) we can determine the $\gamma \approx 0.06$ value (Fig 3(a)), which leads to $\sigma \approx 0.043$. A simple fitting exercise on the $c(t)$ curve using the data for PubMed, leads us to $b \approx 1.6$ (Fig 3(a)). According to these results we predict that the average number of citations per article (for the case of PubMed indexed articles) will relax to $b/(g - 1) \approx 4$ (Fig 3(b)).

Conclusion

Our **conclusions** are pretty clear: Science and Facebook show the same popularity pattern which can be simply understood by a coarse-grained master equation approach where we admit the exponentially increasing amount of information together with a “rich gets richer” preferential information filtering mechanism. Our model predicts that the average number of citations per publication (or shares per Facebook posts) exponentially relaxes to a constant value. This suggests that our society acts in responsible and selective manner in retransmitting

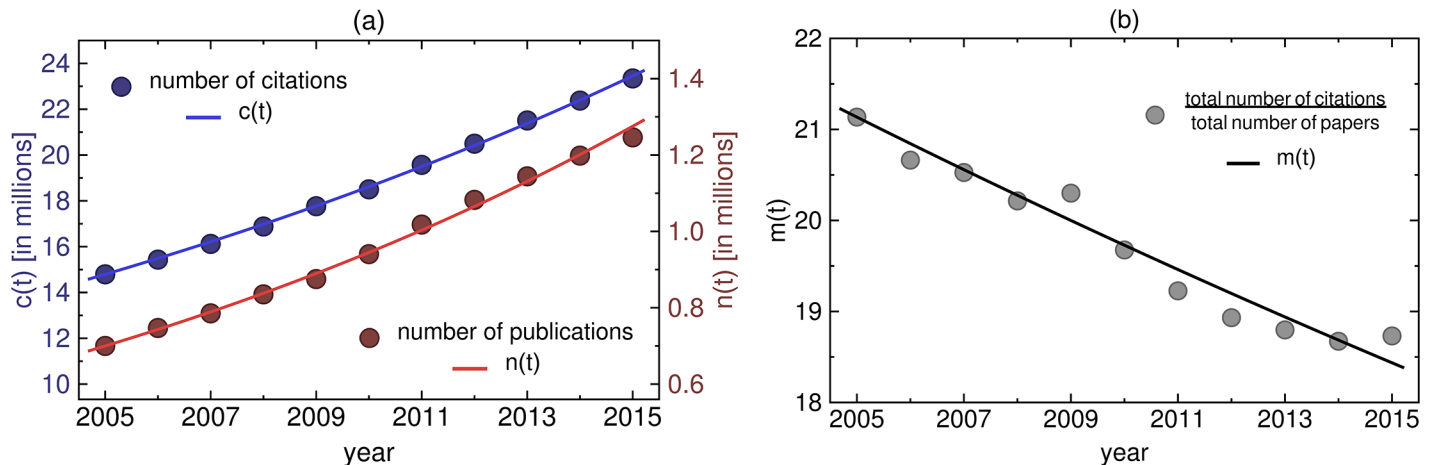


Fig 3. Results for the MEDLINE/PubMed database. Fig 3(a) illustrates the time evolution of the yearly indexed papers, $n(t)$, and the total number of citations, $c(t)$, introduced by them for each year in the 2005-2015 time interval. The trend $n(t)$ can be nicely fitted (red curve) with an exponential curve with $\gamma = 0.06$ using $t_0 = 2005$ and $n_0 = 699\,915$. For $t_0 = 2005$, $n_0 = 699\,915$, $c_0 = 14\,792\,864$, $g = 1.4$ and $\gamma = 0.06$ ($\sigma = \gamma/g \approx 0.043$) the trend given by Eq (18) can be approximated with $b \approx 1.6$. Fig 3(b) illustrates the time evolution for the yearly incoming citations divided by the total number of new papers, $m(t)$. Using the parameters from $n(t)$ and $c(t)$, the $m(t)$ trend given by Eq (16) is plotted by the black curve.

<https://doi.org/10.1371/journal.pone.0179656.g003>

the informations. For scientific articles we predict that their average number of citations converges to a value of approximately 4.

Methods

The data plotted in Fig 1 was collected as follows:

For the WOS dataset (Scientific Citations from ISI Web of Science) we used an Internet robot that started from a given article and reached all the papers that were cited by this. We extracted only the article's identification code and have done this for a depth of four levels recording the total number of citation for all ISI indexed articles that were accessed with this procedure. In total more than 600 000 articles were reached. The use of such robots for accessing an incomplete part (only accession numbers) of the database is not prohibited by the terms of use for the Web of Science [43].

For the JCR dataset we have downloaded the table from InCites, Journal Citation Reports [29], and recorded "total number of citations" for each of the (more than 12 000) indexed journals. For the reduced datasets we followed the methodology described in [6] selecting by random some Institutes, Journals and researchers. We extracted from ISI Web of Science the citations up to the present date for articles published in 1990 with authors from Harvard University. In the same manner for journals we selected papers published in The Lancet (Elsevier) in 1990 and recorded their citations up to the present date. Since our results were in agreement with the one published in [6], we concluded that the results for other Institutes and Journals rescale on the very same curve as it is illustrated in [6]. To complete the study on citation distribution with an even more challenging dataset we have selected a single author from physics (Prof. H. E. Stanley from the Boston University, USA) with an impressive number of publications (965 ISI papers) and ISI citations (62 996) and constructed the citation distribution for all his papers up to the present date independently of the publication year.

In collecting the statistics for MEDLINE/PubMed articles we have used the trend for the total number of publications from [44], and the yearly total number of citations statistics from [45].

For Facebook we used only public pages and informations that are publicly visible. In order to do this in an automatic manner we registered as a developer, and used a publicly available page scraper [46]. Since all collected data are already public on Facebook no privacy issues were violated, for more information on this page scraper please see the relevant information available at: [46]. We have selected 16 Facebook pages that have a relatively high number of shares for their posts in comparison to other users in the same field of activity. For news we selected: New York Times, CNN, BBC news; for science we selected: NASA, and National Geography; for sport celebrity we selected Cristiano Ronaldo; for festivals we have chosen Burning Man, for nightclubs: Sugar Factory; for administration: USA gov., European Council, European Parliament; for movies and TV celebrities: IMDB and for politics: Democratic Party of USA and the Republican Party of USA. From their metadata we have extracted the number of “shares” for all posts independently of their publication date. We have also combined all “share” numbers for the posts of all 16 users and considered as the combined FB database.

All data used to plot the figures are available for download [47].

Probability distribution functions were constructed using a logarithmic binning method, considering bins of sizes 2^n . In order not to overload Fig 1 we have plotted the results only for some selected datasets (see the legend). The other collected data, follows the same general trend. All the rescaled data can be nicely fitted with a TP distribution with $g = 1.4$ and $\langle x \rangle = 1$.

Privacy statement for the collected data

All data collected on Facebook are publicly available, and no other personal data was collected. No privacy issues were violated.

Acknowledgments

The work of Tamás Bíró was supported by the NKFIH/OTKA project Nr. 104260 and a STAR-UBB fellowship.

Author Contributions

Conceptualization: Zoltán Néda.

Data curation: Levente Varga.

Formal analysis: Zoltán Néda, Levente Varga, Tamás S. Biró.

Funding acquisition: Zoltán Néda, Tamás S. Biró.

Investigation: Zoltán Néda, Levente Varga.

Methodology: Zoltán Néda, Tamás S. Biró.

Software: Levente Varga.

Visualization: Zoltán Néda, Levente Varga.

Writing – original draft: Zoltán Néda.

Writing – review & editing: Zoltán Néda, Tamás S. Biró.

References

1. Petersen AM, Wang F, Stanley HE. Methods for measuring the citations and productivity of scientists across time and discipline. *Phys. Rev. E*. 2010; 81:036114 <https://doi.org/10.1103/PhysRevE.81.036114>
2. Azoulay P. Research efficiency: Turn the scientific method on ourselves. *Nature*. 2012; 484: 31–32 <https://doi.org/10.1038/484031a> PMID: 22481340
3. Sorzano COS, Vargas J, Caffarena-Fernandez G, Iriarte A. Comparing scientific performance among equals. *Scientometrics*. 2014; 101: 1731–1745 <https://doi.org/10.1007/s11192-014-1368-6>
4. Radicchi F, Fortunato S, Castellano C. Universality of citation distributions: Toward an objective measure of scientific impact. *PNAS*. 2008; 105: 17268–17272 <https://doi.org/10.1073/pnas.0806977105> PMID: 18978030
5. Radicchi F, Castellano C. Rescaling scientific publications in physics. *Phys. Rev. E*. 2011; 83: 046116 <https://doi.org/10.1103/PhysRevE.83.046116>
6. Chatterjee A, Ghosh A, Chakrabarti BK. Universality of Citation Distributions for Academic Institutions and Journals. *Plos One*. 2016; 11: 0146763 <https://doi.org/10.1371/journal.pone.0146762>
7. Hsu Jw, Huang Dw. Dynamics of citation distribution. *Computer Physics Communication*. 2011; 182: 185–187 <https://doi.org/10.1016/j.cpc.2010.07.031>
8. Petersen AM, Stanley HE, Succi S. Statistical regularities in the rank-citation profile of scientists. *Scientific Reports*. 2011; 1: 181 <https://doi.org/10.1038/srep00181> PMID: 22355696
9. Brzezinski M. Power law in citation distributions: evidence from Scopus. *Scientometrics*. 2015; 103: 213–228 <https://doi.org/10.1007/s11192-014-1524-z> PMID: 25821280
10. Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I. Bibliometrics: The Leiden Manifesto for research metrics. *Nature*. 2015; 520: 429–431 <https://doi.org/10.1038/520429a> PMID: 25903611
11. Lehmann S, Jackson AD, Lautrup BE. Measures of measures. *Nature*. 2006; 444: 1003–1004 <https://doi.org/10.1038/4441003a> PMID: 17183295
12. Waltman L, van Eck NJ. Field-normalized citation impact indicators and the choice of an appropriate counting method. *J. of Informetrics*. 2015; 9: 872–894 <https://doi.org/10.1016/j.joi.2015.08.001>
13. Lehmann S, Jackson AD, Lautrup BE. A quantitative analysis of indicators of scientific performance. *Scientometrics*. 2008; 76: 369–390 <https://doi.org/10.1007/s11192-007-1868-8>
14. Sangwal K. Comparison of different mathematical functions for the analysis of citation distribution of papers of individual authors. *J. of Informetrics*. 2013; 7: 36–49 <https://doi.org/10.1016/j.joi.2012.09.002>
15. Thelwall M, Wilson P. Regression for citation data: An evaluation of different methods. *J. of Informetrics*. 2014; 8: 963–971 <https://doi.org/10.1016/j.joi.2014.09.011>
16. Thelwall M. The discretised lognormal and hooked power law distributions for complete citation data: Best options for modelling and regression. *J. of Informetrics*. 2016; 10: 336–346 <https://doi.org/10.1016/j.joi.2015.12.007>
17. Katchanov YL, Markova YV. On a heuristic point of view concerning the citation distribution: introducing the wakeby distribution. *Springer Plus*. 2015; 4: 94 <https://doi.org/10.1186/s40064-015-0821-1> PMID: 25763305
18. Thelwall M. Are the discretised lognormal and hooked power law distributions plausible for citation data? *J. of Informetrics*. 2016; 10: 454–470 <https://doi.org/10.1016/j.joi.2016.03.001>
19. Golosovsky M, Solomon S. Runaway events dominate the heavy tail of citations distributions. *Eur. Phys. J. Special topics*. 2012; 205: 303–311 <https://doi.org/10.1140/epjst/e2012-01576-4>
20. Pareto V. La legge della domanda. *Giornale degli Economisti*. 1895; 10: 59
21. Thurner S, Kyriakopoulos F, Tsallis C. Unified model for network dynamics exhibiting nonextensive statistics. *Phys. Rev. E*. 2007; 76: 036111 <https://doi.org/10.1103/PhysRevE.76.036111>
22. Kuhn T, Perc M and Helbing D. Inheritance Patterns in Citation Networks Reveal Scientific Memes. *Phys. Rev. X*. 2014; 4: 041036
23. Newman MEJ. Complex systems: A survey. *Am. J Phys*. 2010; 79: 800–810 <https://doi.org/10.1119/1.3590372>
24. Clauset A, Shalizi CR, Newman MEJ. Power-law distributions in empirical data. *SIAM Rev*. 2009; 51: 661–703 <https://doi.org/10.1137/070710111>
25. Gabaix X. Power Laws in Economics and Finance. *Annu. Rev. Econ*. 2009; 1: 255–293 <https://doi.org/10.1146/annurev.economics.050708.142940>
26. Zipf GK. *Human Behavior and Principle of Least Effort* Cambridge, MA: Addison-Wesley; 1949

27. Newman MEJ. Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*. 2007; 46: 323–351 <https://doi.org/10.1080/00107510500052444>
28. Stumpf MPH, Porter MA. Critical truths about power laws. *Science*. 2012; 335: 665–666 <https://doi.org/10.1126/science.1216142> PMID: 22323807
29. InCites, Journal citation report, Thomson Reuter Available: <https://jcr.incites.thomsonreuters.com/JCRJournalHomeAction.action>
30. Eom YH, Fortunato S. Characterizing and Modeling Citation Dynamics. *Plos One*. 2011; 6: e24926 <https://doi.org/10.1371/journal.pone.0024926> PMID: 21966387
31. Wang D, Song C, Barabási AL. Quantifying long-term scientific impact. *Science*; 2013; 342: 127–132 <https://doi.org/10.1126/science.1237825> PMID: 24092745
32. Sinatra R, Wang D, Deville P, Song C, Barabási AL. Quantifying the evolution of individual scientific impact. *Science*. 2016; 354: aaf5239 <https://doi.org/10.1126/science.aaf5239> PMID: 27811240
33. Kathanov YL. Towards a simple mathematical theory of citation distributions. *SpringerPlus*. 2015; 4: 677 <https://doi.org/10.1186/s40064-015-1467-8>
34. Medo M, Cimini G. Model-based evaluation of scientific impact indicators. *Phys. Rev. E*. 2016; 94: 032312 <https://doi.org/10.1103/PhysRevE.94.032312> PMID: 27739778
35. Sinatra R, Deville P, Szell M, Wang D, Barabási AL. A century of physics. *Nature Physics*. 2015; 11: 791–796 <https://doi.org/10.1038/nphys3494>
36. Martin-Martin A, Orduna-Malea E, Ayllon JM, Lopez-Cozar ED. Reviving the past: the growth of citations to old documents. *arXiv*. 2015;1501.02084
37. Zuckerberg M. Online Sharing Is Growing At An Exponential Rate. Accessed at: <https://www.youtube.com/watch?v=HNy9uxcRedU>
38. Jeong H, Néda Z, Barabási AL. Measuring preferential attachment in evolving networks. *Europhys. Lett*. 2003; 61: 567–572 <https://doi.org/10.1209/epl/i2003-00166-9>
39. Perc M. The Matthew effect in empirical data. *J. R. Soc. Interface*. 2014; 11: 20140378 <https://doi.org/10.1098/rsif.2014.0378> PMID: 24990288
40. Perc M. Self-organization of progress across the century of physics. *Sci. Rep*. 2013; 3: 1720 <https://doi.org/10.1038/srep01720>
41. Goldstone JA. A deductive explanation of the Matthew effect in Science. *Social Studies of Science*. 1979; 9: 385–391 <https://doi.org/10.1177/030631277900900306>
42. Wang J. Unpacking the Matthew effect in citations. *J. of Informetrics*. 2014; 8: 329–339 <https://doi.org/10.1016/j.joi.2014.01.006>
43. Web of Science, Web of Knowledge Terms of Use. Accessed at 2014-09-01, <http://wokinfo.com/terms/>
44. Corlan AD. Medline trend: automated yearly statistics of PubMed results for any query. Accessed at 2016-11-10, <http://dan.corlan.net/medline-trend.html>
45. MEDLINE/PubMed, Baseline repository—background Accessed at 2016-11-10, <https://mbr.nlm.nih.gov/Background.shtml>
46. Facebook Page Post Scraper, Please note for this the disclosers concerning the Privacy issues: “This scraper can only scrape public Facebook data which is available to anyone, even those who are not logged into Facebook. No personally-identifiable data is collected in the Page variant; the Group variant does collect the name of the author of the post, but that data is also public to non-logged-in users. Additionally, the script only uses officially-documented Facebook API endpoints without circumventing any rate-limits.” Accessed at 2016-11-1, <https://github.com/minimaxir/facebook-page-post-scraper>
47. Rough data for the figures (free download), Database: figshare [Internet] Accessed at 2017-11-5, <https://figshare.com/articles/s/4986551>