Please cite this paper as: Geras A., Siudem G., Gagolewski M., Should we introduce a dislike button for academic papers?, Journal of the Association for Information Science and Technology 71(2), 221–229, 2020, DOI:10.1002/ASI.24231

Should we introduce a dislike button for academic papers?

Agnieszka Geras[1]

Grzegorz Siudem[2]

Marek Gagolewski[1,3]

[1] Faculty of Mathematics and Information Science Warsaw University of Technology,
ul. Koszykowa 75, 00-662 Warsaw, Poland

[2]Faculty of Physics,Warsaw University of Technology, ul. Koszykowa 75, 00-662
Warsaw, Poland

[3]Systems Research Institute, Polish Academy of Sciences, ul. Newelska 6, 01-447
Warsaw, Poland

Correspondence and proofs should be addressed to Agnieszka Geras. E-mail:
A.Geras@mini.pw.edu.pl

Should we introduce a dislike button for academic papers?

Abstract

On the grounds of the revealed, mutual resemblance between the behaviour of users of the *Stack Exchange* and the dynamics of the citations accumulation process in the scientific community, we tackled an outwardly intractable problem of assessing the impact of introducing "negative" citations.

Although the most frequent reason to cite a paper is to highlight the connection between the two publications, researchers sometimes mention an earlier work to cast a negative light. While computing citation-based scores, for instance the h-index, information about the reason why a paper was mentioned is neglected. Therefore it can be questioned whether these indices describe scientific achievements accurately. In this contribution we shed insight into the problem of "negative" citations, analysing data from Stack Exchange and, to draw more universal conclusions, we derive an approximation of citations scores. Here we show that the quantified influence of introducing negative citations is of lesser importance and that they could be used as an indicator of *where* attention of scientific community is allocated.

Introduction

Various citation-based scores, such as the number of citations, the h-index, the g-index etc. are amongst quantitative indices based on which scientific achievements are assessed and the funding for members of the scientific community is divided (Dorogovtsev & Mendes, 2015; Hirsch, 2005). A citation occurs when an author gives a reference to an earlier paper. Typically, it is assumed that the more citations an author has, the greater their scientific impact (Deville et al., 2014). Many reasons why papers are cited can be differentiated, see (L. Bornmann & Daniel, 2008; Lutz Bornmann, de Moya Anegón & Leydesdorff, 2010). The most obvious and frequent reason is to highlight the connection between the two papers: demonstrating in what manner the previous work has influenced the subsequent, how it was applied, continued, or developed. By doing so, one author expresses approbation for another researcher's results and shows a **positive** attitude towards it. However, researchers sometimes cite an earlier paper in completely different circumstances, to cast a **negative** light, thus in order to point out flaws, mistakes or even to criticise it entirely. When we calculate basic bibliometric indices such as the aforementioned h-index, we do not take into consideration whether the paper was cited positively or negatively – this information is lost, hence reliability of those indices can be called into question.

In this contribution we investigate what would happen if we distinguished two types of citations:

- Positive: expressing commendation or simply stating "information re-use",
- Negative: expressing criticisms.

In particular, we address the following SciSci (the science of science, see (Clauset, Larremore & Sinatra, 2017; Fortunato et al., 2018)) questions:

1. Is criticism a frequent phenomenon and does it vanish when one achieves a certain level of accomplishment?
2. Does distinguishing negative citations influence an author's h-index?
3. Does the "rich get richer" rule apply for negative citations as well?
4. Which mathematical models may appropriately describe the aforementioned bipolar citation accumulation process?

Real-world bibliometric data obviously lack information about the character of citations. A possible approach towards the problem of negative citation in science is to analyse the sentiment of citations. Examples of this can be found, in e.g., (Catalini, Lacetera & Oettl, 2015; Kumar, 2016). In our view, such an approach is imperfect due to the fact that text-mining methods can be imprecise and there are no reference benchmark data to precisely validate the quality of results they generate. While looking for a structure akin to the scientific community, where a parallel phenomenon to negative citations can be identified, our attention was drawn to the *Stack Exchange*

Network, a cluster of Q&A websites, where evaluating positively and negatively human performance is an inherent feature. *Stack Exchange'* users pose questions on a great variety of topics– from quite trivial ones to strictly scientific. Questions and answers posted on *Stack Exchange* can be voted for (UpVote) and against (DownVote). Based on those positive and negative votes, a user's reputation index is calculated, which enables us to quantify their level of expertise. *Stack Exchange* sites can serve as a model for the scientific community, because of their mutual resemblance (however, such an analysis has not been previously performed, see (Vasilescu et al., 2018)). An apparent advantage of this solution is that we analyse data describing users that are intrinsically aware of the possibility of receiving a negative vote, which would lower their reputation. Thus, they might hesitate before publishing low-quality content. We presume that an analogous behaviour of members of the scientific communities would arise in the event of introducing negative citations, a pattern which cannot be captured by post-factual text-mining analyses (Catalini et al., 2015; Kumar, 2016). To some extent, Stack Exchange suffers from the problem of inequalities among disciplines, since, for instance, some programming languages are more popular than others. Despite many analogies, evidently, there exist a few discrepancies between the accumulation of votes on Stack Exchange and on real-world citation networks. First of all, Stack Exchange's votes do not suffer from the problem of multiple authorship, which also entails a clear division into well defined, separated fields, differently than in the scientific community, where stating research disciplines very rigidly might be a difficult or irrelevant task. What is more, the Stack Exchange community allows one to remain anonymous and do not reveal their real name.

In this contribution we focus on describing the behaviour of 1,922,770 users of *StackOverflow*, by far the largest of the *Stack Exchange* sites that focuses on computer programming. They wrote 24,492,237 posts in total, which received 70,493,274 votes. A vast number of answers, up to 37.6%, remain without a single vote. This may indicate that they were uninteresting, poorly written, of a lower importance, or too recent to be given attention. Only 2.9% of all the votes accounted for negative ones, hence we see that expressing criticism is rare – users are much more eager to express approbation. The frequency of negative scores is relatively low, although a noticeable number of posts received at least one negative score (6%). What is more, as far as real bibliometric data is concerned, (Catalini et al., 2015) reported that 2.4% of the citations are negative, while (Kumar, 2016) – stated that 8.6% were negative. This indicates the similarity between the scientific and the *StackOverflow* users' communities.

The relationship between positive and negative attentions

Let us consider a user and assume she answered N questions. Her scores can be described by two vectors, each of length N: a vector of positive scores $U = [U_1, U_2, ..., U_N]$ and a vector of negative scores $D = [D_1, D_2, ..., D_N]$. Therefore, U_i denotes the number of UpVotes received by the i-th answer. Likewise, D_i denotes the number of its DownVotes. Furthermore, $P = \sum_{i=1}^{N} U_i$ denotes the total number of UpVotes, while $M = \sum_{i=1}^{N} D_i$ denotes the total number of DownVotes.

One might wonder if it is plausible that a high number of positive votes intensify the upward trend in receiving negative votes. Are we allowed to draw a conclusion that even a truly excellent piece of work cannot avoid criticism? In other words, does being controversial help to gain more attention or simply one cannot "please" everyone and no matter how thorough a work is, critique cannot be avoided?

Similarly as in (Abisheva, Garcıa & Schweitzer, 2016) we explore the relationship between vectors U and D, but in an aggregated manner – we consider the relationship between values of P and M for all authors. We observe an appearance of two trends: at first slower but then more rapid growth of the number of DonwVotes for higher values of UpVotes. Figure 1 presents the results of fits between the logarithms of UpVotes and DownVotes using standard linear regression against a non-linear Multivariate adaptive regression splines (MARSplines) fits (see Methods). The R squared coefficient suggests that dual model ($R^2 = 0.4531$) suits the data slightly better than the simple linear model ($R^2 = 0.44$).

While taking a closer look at Figure 1 we observe that the highest values of positive scores are obtained only when the highest values of negative scores are present as well, which indicates that indeed excellent work also draws a considerable amount of criticism.

Afterwards, we computed the Kendall's τ correlation coefficient (see Methods) between U and D vectors. On the face of it, one would expect that a high number of negative citations entail a low number of positive citations, but in a vast number of cases (see the histogram in Figure 2), there exist a small yet statistically significant *positive* correlation between U and D.

Figure 1 and 2 suggest that the *StackOverflow* community filters out low quality work, focusing and discussing selected, compelling topics. Users express their disapproval on popular matters, which accumulated many votes "for". Therefore, negative citations may be used as an indicator of *where* the attention of the society is actually located.

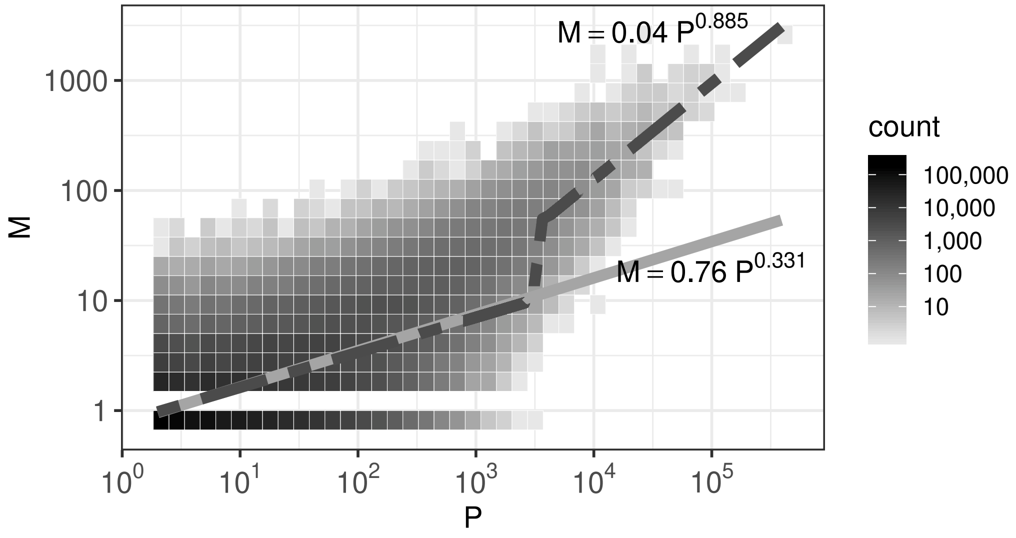

Figure 1. The relationship between the number of UpVotes and DownVotes received by users of the Stack Exchange on log-log scale. Two-dimensional joint distributions with 40 bins for P and 20 bins for M, bins colours indicate the count of observations within the bin. The line in light grey indicates the global regime $M = 0.76P^{0.331}$ obtained using log-log regression, while the darker, dotted broken line indicates local regimes of the non-linear relationship obtained using MARSplines (see Methods).

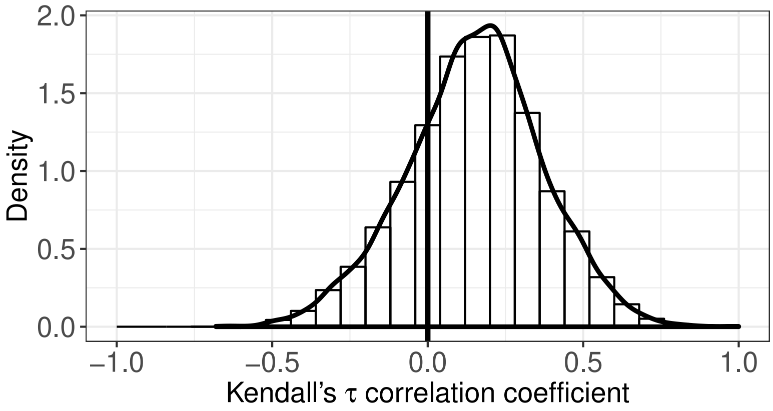

Figure 2. The distribution of Kendall's rank correlation coefficient between positive and negative citations of posts written by a selected group of 11,432 authors with significant number of negative citations. In about 76% of the cases there is a positive dependence, but generally the degree of correlation is quite small. However at a first glance, one would expect a high negative correlation, which is not the case.

Bibliometric indices

The process of aggregating data, in order to gain a more general view, e.g. while evaluating human performance, is considered a highly important matter and for many years has been a research interest of a vast number of scientists, especially in the field of SciSci, where numerous evaluative measures were proposed. In 2005, J.E. Hirsch propounded a measure of scientific achievements – the h-index (Hirsch, 2005) – which is one of the most commonly used bibliometric tools (Ciriminna & Pagliaro, 2013; Meyers & Quan, 2017).

The h-index aims to measure not only the impact but also the productivity of a given author. An author's h-index is equal to h if at least h of their papers have no fewer than h citations each, while the rest of $n - h$ papers have not more than h citations. Moreover, if none of the papers were cited, then the h-index is equal to 0.

In this contribution we shall also consider the g-index. As it is stated in (Egghe, 2006) if "citation performance of a set article is ranked in decreasing order of the number of citations that they received, the g-index is the (unique) largest number such that the top g articles received (together) at least g^2 citations". Let us notice that we can perceive the h-index as the number of papers reaching a certain level of “quality” above a threshold, that increases as the h-index increases. In that case, the g-index makes it possible for citations from highly cited papers to be employed to help papers of a lower citation score to meet this threshold. From now on we will denote the h-index of a vector X as $H(X)$ and, analogously, the g-index as $G(X)$.

Due to the fact that, in classical biliometrics, we do not distinguish the type of citations, we only observe the combined vectors $U + D$ and compute $H(U + D)$ and $G(U + D)$. Supposing we observed U and D separately or even just $U - D$, we might be able to assess scientific endeavour more fairly. Therefore, further analyses will revolve around investigating whether the aforementioned modification would have a major influence on the most popular bibliometric indices. First step shall be to calculate $H(U - D)$ and $G(U - D)$ in order to uncover to what extent taking into consideration negative citations changes the bibliometric indices of *StackOverflow* users.

It turns out that the h-index does not change in the case of 91.5% of the users. For the remaining 8.5%, whose h-index decreases, 91.6% of them experienced a reduction by one unit, only 9008 users were decreased by more than 1. Figure 3 presents the overall alteration: it illustrates that the h-index and g-index are only slightly perturbed after calculating $H(U - D)$ instead of $H(U + D)$ and $G(U - D)$ instead of $G(U + D)$.

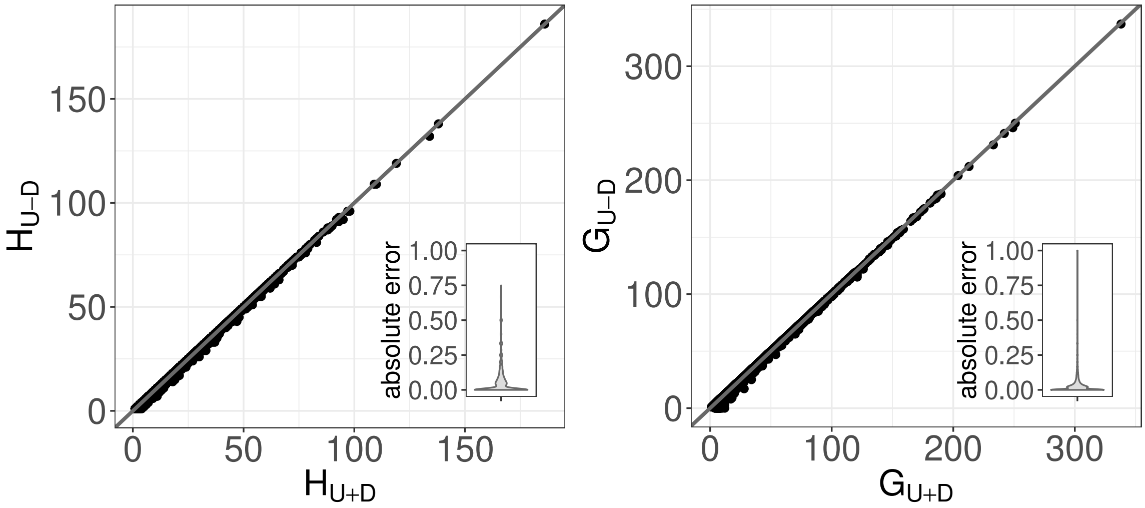

Figure 3. The h- and g- indices of $U + D$ (negative citations count as positive) vs. $U - D$ (negative citations filtered out). We observe that the impact measures do not change significantly after taking negative citations into consideration. The bottom–right insets present the absolute error between values H_{U+D} and H_{U-D} and between G_{U+D} and G_{U-D}.

Towards modelling the process of accumulating votes

Since one of the main applications of the h-index is a comparison of different fields of research or institutions (Iglesias & Pecharromán, 2007; Malesios & Psarakis, 2014), we decided to employ it to compare the two possible states of the scientific community - current, without a *dislike button* for academic papers and hypothetical - after implementing the considered modification. While facing the challenge of explaining a complex phenomenon and predicting its evolution, a descriptive, yet possibly simple, model is of the uttermost usefulness. Hence, in what follows we shall make use of bibliometric models to gain a fair understanding of the processes of accumulating positive and negative votes and, afterwards, draw more universal conclusions, specifically regarding other than online media domains such as research evaluation. The first step will consist of examining the distribution of the number of votes and and verifying whether the well-known “the rich get richer” (see (Barabási & Albert, 1999; Szell & Sinatra, 2015)) rule applies for accumulating both positive and negative votes separately.

Examining the probability distribution and the preferential attachment

Let $f(x)$ denote the probability of a post to receive x votes of a selected type (UpVotes, DownVotes or combined) or a paper to receive x citations (citations scores for scientific papers were calculated based on OpenCitations database – see Methods) and $\langle x \rangle = \int\limits_0^\infty x f(x) dx$ denote the expected value of the considered distribution. For the purpose of obtaining normalised data, we divide number of votes received by a selected post (or number of citations accumulated by a selected scientific paper) by the average and we denote it $x/\langle x \rangle$. We observe (Figure 4) that normalised values of empirical distribution $f(x/\langle x \rangle)$ rescale on the common curve, similarly as it was acknowledged for citations selected according to field, year of publication (Radicchi & Castellano, 2011), disciplines (Radicchi, Fortunato & Castellano, 2008), institutions and journals (Chatterjee, Ghosh & Chakrabarti, 2016). We see an apparent power–law trend for values $f(x/\langle x \rangle)$ higher than 1. Thus, again, the nature of the four discussed distributions is very similar to those observed for the number citations in bibliometric datasets by Price (de Solla Price, 1976), who also proposed the "cumulative advantage" (now widely known as the preferential attachment, "the rich get richer" or the Matthew Effect (Merton, 1968)) mechanism to be the possible origin of the power-law property. A similar comparison of the dynamics of online attentions and citations accumulation process has already been conducted (Néda Z, 2017).

Since power law behaviour can be different "the rich get richer" origin – see also (Perc, 2014) – we need stronger evidence that, indeed the Matthew effect is governing the evolution of process of the accumulation of both positive and negative votes in the StackOverflow network. Many techniques of measuring the Matthew effect can be found in literature (Perc, 2014). We performed a similar analyses as in (Redner, 2005): we describe the votes accumulation process by the attachment rate A_k, which denotes the estimated probability that a post with k UpVotes or DownVotes will gain another Up- or DownVote respectively. Specific information when each vote appeared is necessary.

Firstly, we distinguish 5 time slots: the year 2017 as a *reference* one and four more: 2008 – 2016, 2010 – 2016, 2012 – 2016 and 2014 – 2016 as *measuring* ones. For a selected measuring slot, we count the number of votes (positive or negative) that each post received to obtain k. Finally, to calculate A_k, we compute the mean of the number of times papers with a given k in the considered measuring time window were cited in the reference window and we receive the estimated probability that a given post with k votes will be voted on again.

The results for both positives and negatives votes are presented in figure 5 and they suggest that A_k is a strictly increasing, linear function of k, particularly, in case of positive votes for k smaller than 200, a condition that applies to 99.9% of posts; when it comes to negative votes for k smaller than 25, what again applies to 99.9% of posts. Therefore we are allowed to draw the conclusion that the “the rich get richer” rule indeed governs the two processes of receiving attentions.

Coming back to the distribution of the number attentions (figure 4), as far as fitting the rescaled curve is concerned, latest findings (Thelwall, 2016a, 2016b) suggest a Tsallis–Pareto type distribution, which surpasses previously applied power–law like distributions due to the feature of being able to grasp probability for a paper to receive the lowest scores (Thelwall, 2016c). The results of fitting Tsallis–Pareto distribution

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

to rescaled curves are presented in Figure 4 as well.

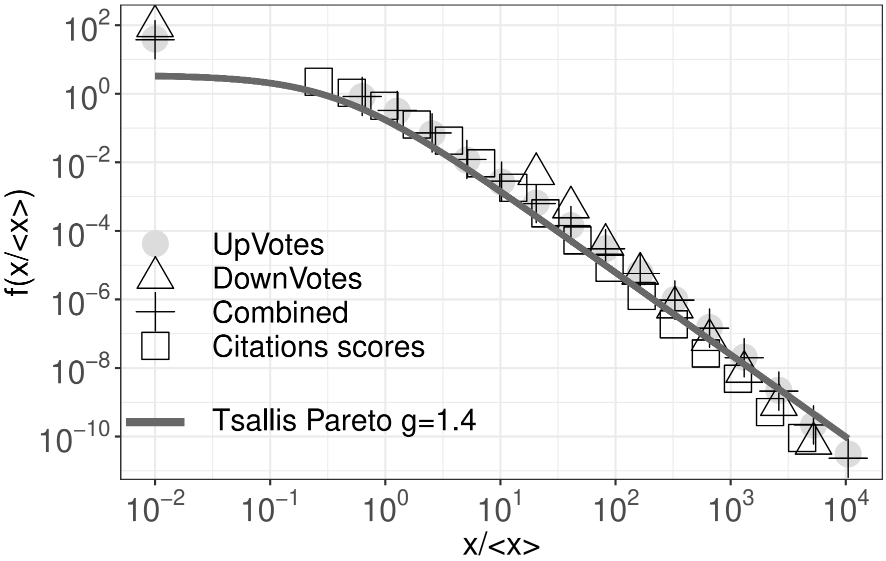

Figure 4. The probability distribution of a selected post to receive x votes (or a paper to receive x citations for squares). We present the normalised value $f(x/\langle x\rangle)$, where $\langle x\rangle$ is the mean value. The curve can be fitted with the Tsallis-Pareto distribution with $g = 1.4$ and $\langle x\rangle = 1$.

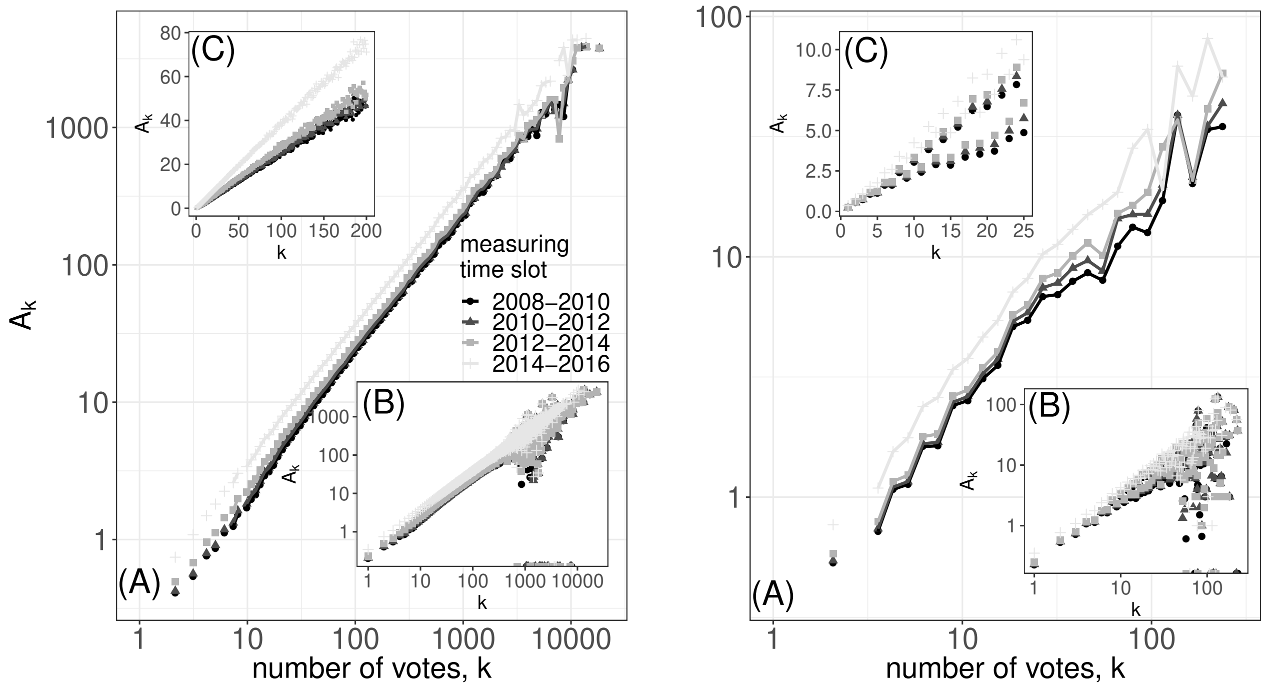

Figure 5. The attachment rate A_k is a nearly linear function of the number of citations k. The different colours indicate different year ranges for establishing k. The rate A_k for both positive and negative scores has been presented in three variants: (A) with logarithmic binning, (B) with one bin for every value of k and (C) for the first 200 values of k in case of UpVotes and 25 values in case of DownVotes – when we observe the clearest linear relationship.

The approximation of citation scores

Let us consider a decreasingly ordered vector of votes or citations X of a size N. We rank vectors' entries according to their sizes by labelling the elements. We give rank 1 to the greatest (first) element, rank k to the k-th greatest and rank N to the tiniest element. The function $N(k)$ which returns the k-th greatest element is called size-ranked distribution. It is not difficult to see (C & A, 2017) that we can obtain $N(k)$ by solving

$$\frac{k}{N} = \int\limits_{N(k)}^{\infty} f(x)\ dx, \tag{2}$$

where $f(x)$ still denotes the probability of a post to receive x votes or a paper to receive x citations.

Deciding on which probability distribution best describe the considered data, in other words, assuming the form of f, is a matter discussed in literature. One of the possible approaches is to assume (e.g. (Brown, 2009))

$$f(k) \propto k^{-\alpha},$$

therefore rank frequency distribution is Zipfian. From (2) we obtain

$$\frac{k}{N} = \frac{N(k)^{-\alpha+1}}{\alpha - 1} \tag{3}$$

and after computing $N(K)$ from (3), normalising the vector $[N(1), N(2), ...N(N)]$ and multiplying it by the total number of citations of each type (P or M) we obtain an approximation of the ordered vector of citations U, D of UpVotes and DownVotes respectively:

$$\tilde{U}_{(k)} = P\frac{k^{-c_U}}{\sum\limits_{i=1}^{N} i^{-c_U}}, \qquad \tilde{D}_{(k)} = M\frac{k^{-c_D}}{\sum\limits_{i=1}^{N} i^{-c_D}}, \tag{4}$$

for some parameters c_U, c_D, which can fit empirically to each score vector representing an individual. In this study, we decided to minimise the Root Mean Squared Logarithmic Error defined as:

$$\text{RMSLE}(X, \hat{X}) = \sqrt{\sum_{k=1}^{N} [\log(X_{(k)} + 1) - \log(\hat{X}_{(k)} + 1)]^2},$$

which is a measure that more leniently penalises large differences between the true ($X \in \{U, D\}$) and predicted ($\hat{X} \in \{\hat{U}, \hat{D}\}$) values. Figure 6 presents a typical fit of the approximation (4) to the Up- and DownVotes vectors. Unfortunately, we note that the score vectors of the individuals, unlike the cumulative ("macro-level") ones, often cannot be fit well to a straight line (on a log-log scale). Therefore an improved approach (a different assumption about f) is desired. Obviously, we employ the previously considered Tsallis-Pareto distribution (1). After performing the same steps as above we obtain:

$$\frac{k}{N} = \left(\frac{N(k)}{g-1} + 1\right)^{-g} \tag{5}$$

and from (5)

$$N(k) = \frac{(\frac{N}{k})^{\frac{1}{g}} - 1}{(g-1)\langle x \rangle}.$$

Finally, the new approximation in the form of:

$$\hat{U}_{(k)} = P\frac{(\frac{N}{k})^{\frac{1}{g_u}} - 1}{\sum_{i=1}^{N}\left[(\frac{N}{i})^{\frac{1}{g_u}} - 1\right]}, \qquad \hat{D}_{(k)} = M\frac{(\frac{N}{k})^{\frac{1}{g_d}} - 1}{\sum_{i=1}^{N}\left[(\frac{N}{i})^{\frac{1}{g_d}} - 1\right]}, \tag{6}$$

where $k = 1, 2..., N$ and g_u, g_d are parameters minimising RMSLE.

Figure 6 compares selected authors citation scores' adherence to (4) and (6). After implementing the (6) instead of (4) for positive scores, we observe a significant decrease in RMSLE error from 4.07 to 0.18, at the same time for negative citations from 4.77 to 0.12. The accuracy of the model was further enhanced by removing outliers.

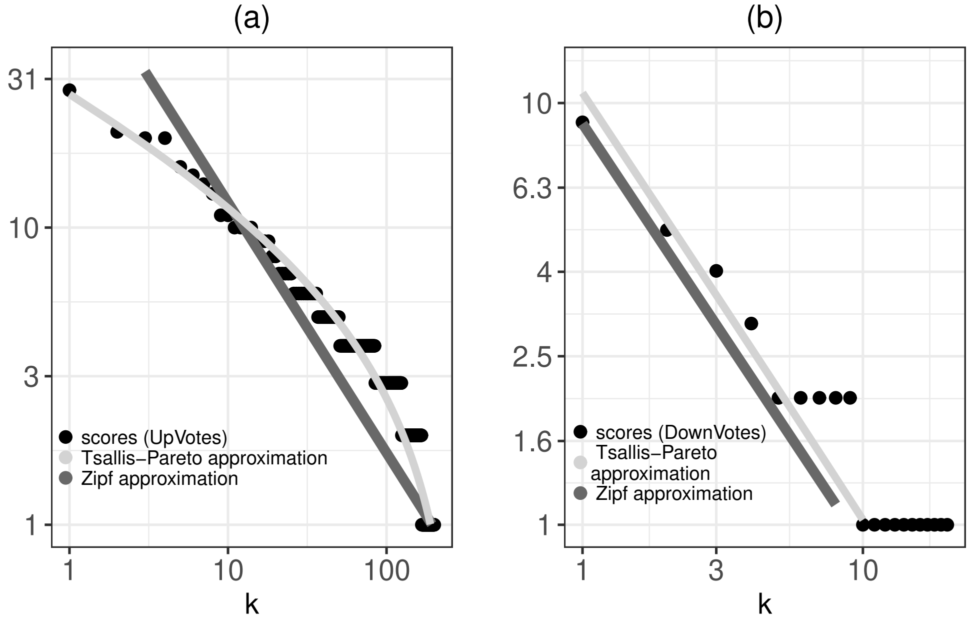

Figure 6. Comparison of fitting the approximation with rank-size distribution employing Zipf (dark grey line) and Tsallis-Pareto distribution (light grey line) for a selected author; (a) UpVotes, (b) DownVotes vector. We observe a tighter fit of the approximation employing Tsallis-Pareto to considered data.

Quantifying the impact of introducing negative citations

Equipped with the approximation derived above we are able to predict a bibliometric (citation/score vector) profile of a given author, a matter previously discussed in the literature (Ionescu & Chopard, 2013; Zogala–Siudem, Siudem, Cena & Gagolewski, 2016) and then quantify the impact of negative citations, by employing an iterative procedure which reveals by how many units the h-index of a certain scientist would decrease provided the reasons for citing a paper were be reflected on. But first we shall verify our model's capability to predict the previously mentioned quantity. Figure 7 presents the results of the comparison of the true h- index driven from data with outliers removed (based on each author's U vector) and obtained from the derived approximation for an author of a given N, P, and g. We observe a good match between these measures.

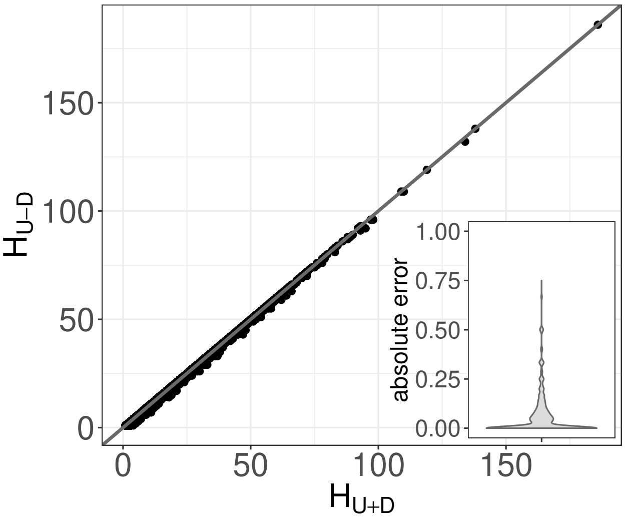

Figure 7. The h-index computed based on the approximation $\hat{H}$ compared with the true, observed ones (H). Despite the model's simplicity (it is governed by only three parameters) it predicts the indices' values quite accurately. A violin plot of the relative error between H and $\hat{H}$ (inset).

The lower bound of the h-index

Let us assume that at our disposal are only the total number of papers of a given author, the total number of positives P and the total number of negative scores M, hence exact vectors U, D are unknown. We obtain the ordered vector U employing the discussed above approximation and then we can quite effortlessly find theoretical D in the worst case scenario, which we understand is the case, when the h-index is decreased by the highest possible value after computing $H(U-D)$ instead of $H(U+D)$. The idea lies in distributing M negative citations in such a manner that the h-index is decreased utmost. Let d denote the maximum value by which we can decrease the h-index. From now on, the value $H(U)-d$ will be called *the lower bound for the h-index.* Knowing M, we shall seek the highest possible d step by step: in each iteration we distribute negative citations to vector D in such a way that the h-index is decreased by 1. Afterwards, we verify whether in the next step further decrements are possible. Initially we have $D=[0,0,...,0]$. We easily deduce that in order to decrease the h-index by one we set: $D_h = U_h - h + 1$ and for $i \in \{h+1,...,N\}$ we should have $D_i = (U_i - h + 1)\mathrm{I}_{\{U_i \geq h-1\}}$, where I denotes the indicator function, defined as:

$$\mathrm{I}_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \notin A, \end{cases}$$

and A is a set.

Similarly, to decrease the h-index by 2, we establish $D_h := D_h + 1$, $D_{h-1} = D_{h-1} - h + 2$ and for $i \in \{h+1,...,N\}$ we set $D_i := (U_i - h + 2)\mathrm{I}_{\{U_i \geq h-2\}}$, and so forth. Therefore, to decrease the h-index of an author with a positive citation vector U by d with as few negative citations given to each paper as possible, we shall set $D_h := D_h + 1$, $D_{h-1} := D_{h-1} + 1$, ..., $D_{h-d+2} = D_{h-d+2} + 1$, $D_{h-d+1} = U_{h-d+1} - h + d + 1$ and for $i \in \{h+1,...,N\}$ $D_i = (U_i - h + d)\mathrm{I}_{\{U_i \geq h-d\}}$. Consequently, we need to have at least:

$$S_d := \sum_{j=0}^{d-1} (U_{h-j} - h + (j+1) + j) + \sum_{j=h+1}^{N} \mathrm{I}_{\{U_j \geq h-d\}}(U_j - h + d) =$$

$$= \sum_{i=0}^{d-1} U_{h-i} - hd + d^2 + \sum_{j=h+1}^{N} \mathrm{I}_{\{U_j \geq h-d\}}(U_j - h + d)$$

citations to decrease the h-index by d, of course provided that the condition $S_d \leq M$ is met. After performing d steps, $M - S_d$ residue citations might be distributed randomly as they will not have any impact on the h-index. We implemented the procedure that seeks the maximal possible d.

Now, let us fix N – the number of items in a person's record. For all authors with exactly N papers, we calculate the mean value of their total number of positives votes P, denoted $\bar{P}$. Further, we take into consideration the authors with fixed P and determine the average of the total number of negative citations M (scenario I – mediocre or less pessimistic) and the maximal total number of negative citations (scenario II – rather pessimistic). Adjusting appropriate models allows us to indicate typical values of M and P and values of the parameter κ by minimising RMSLE error for a fixed value of N and use them to calculate a typical h-index using the approximation and the lower bound in the h-index in the two scenarios.

Figure 8 presents the lower bound of the h-index in the two scenarios: the first one considers mediocre users with an average number of negative citations and the second, pessimistic scenario for users with maximal numbers of negative votes, with their number of posts written taken into consideration as well as and the lower bound driven from data. We observe that introducing negative citations could be more harmful for young scientists as the line driven from the data for lower h-index values lies closer to the pessimistic scenario. Later on, it comes closer to the more optimistic one.

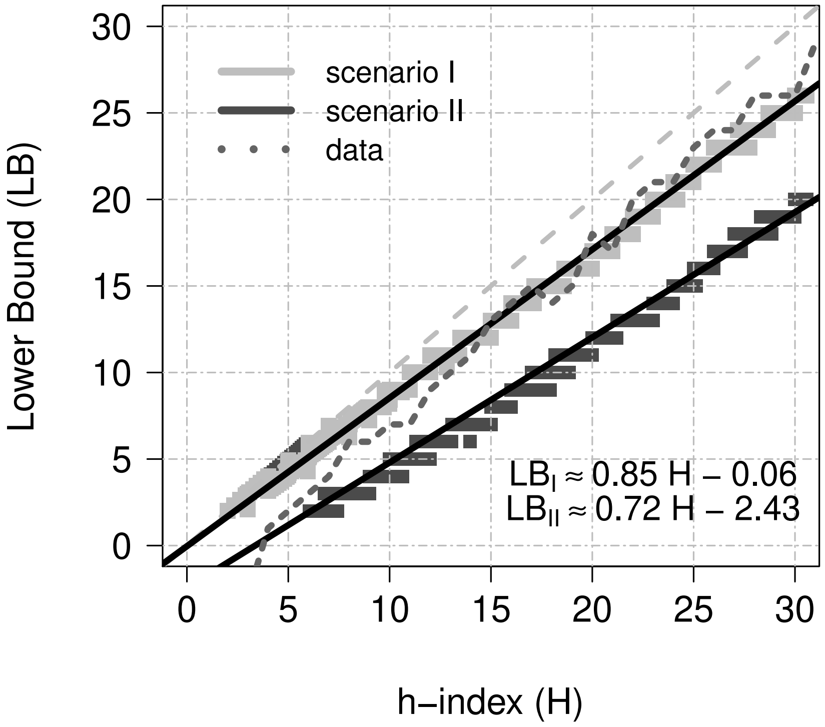

Figure 8. Lower bounds (LB_I, LB_{II}) of the h-index (H) in two scenarios and driven from data. Scenario II (more pessimistic) is more accurate for the early stage. Scenario I (mean values of total negative votes score computed for users with fixed number of positive votes) is suitable for more experienced scientists.

Discussion

Last but not least, a matter that we would like to mention is the answers' ranking system, which affects the order of answers displayed under the question. The answers are ranked according to the number of UpVotes, the number of DownVotes and the reputation of the answer's author. As a result, the answer with the highest score, therefore the "best" one, is displayed on the top and one may think that it might be the case that other users do not bother to search any further than this answer. In order to assess the impact of ranking in such a manner, we conducted a simple analysis. Firstly, we filtered out questions that received only one answer (57.3%) as this matter does not concern here. The authors of questions have the possibility to accept one of the given answers as the best, and as a consequence, this answer is displayed on the top. From now on we shall focus on 71.2% of questions with an accepted answer, what gives us 3,333,739 questions to analyse. It turns out that in 13.6% of cases the accepted answer, which is displayed on the top, did not receive the maximal number of positive votes. Therefore, a significant fraction of users surely looked further than the first answer and even expressed a different opinion than the author of the question. Although the impact of this mechanism is limited, we consider this issue a highly interesting topic for future research.

The conducted analysis of Stack Exchange data, where the inherent feature of assessing other members of one's community is present, justifies previously obtained text mining-based results (Catalini et al., 2015; Kumar, 2016). They both indicate that negative citations would not be a particularly frequent phenomenon. The processes of accumulating positive and negative votes appear to be governed by the same mechanism – "the rich get richer" rule. What is more, we cannot state that items highly positively scored obtain less negative votes. On the contrary, we observe a slightly positive correlation between positive and negative votes. It appears that the worse of all that can happen to an author, is a lack of attention rather than criticism.

All in all, the highly scored authors are repeatedly criticised and we observe even the greatest minds cannot slip away from disapproval. The employed approximation allows us to draw more universal conclusions concerning the impact of hypothetical negative citations on the scientific community. Again, surprisingly, the quantified influence appears to be of lesser importance, particularly to highly cited and therefore commonly criticised authors. Nevertheless, early career researchers and those with a smaller number of papers in their accomplishments seem to be more affected by introducing negative citations.

It appears that distinguishing between positive and negative citations could enhance the quality of scientific publications. The presented analysis proved negative citations to have further advantages. Publishers could quite straightforwardly implement the citation classification system, for instance by requiring authors to add a special mark for each item in the bibliography section. What is more, to our view, negative citations would solve the problem of information overload, caused by the exponential growth of number of papers, whereas growth of ideas covered by them is only linear (Fortunato et al., 2018). Negative citations can be seen as a tool to diminish this unfortunate trend. Unprecedented increases in computing power and the availability of data resulted in not only advancement in SciSci field, but also, in more general context, the onset of the forth technological revolution that will fundamentally alter the way we live, work, and relate to one another and consequently evaluate human performance. Therefore it is high time for a paradigm shift in research evaluation.

Methods

Data and its availability

Every so often, the Stack Exchange releases "data dumps" of all its publicly available content via archive.org (https://archive.org/download/stackexchange). The analysis is based on the files that date back to August 8, 2018.

The Initiative for Open Citations (I4OC) is a project describing itself as: *a collaboration between scholarly publishers, researchers, and other interested parties to promote the unrestricted availability of scholarly citation data and to make these data available.* We used files, containing 316,243,802 citation links, from the dump created on July 4, 2018. For details see http://opencitations.net/.

Kendall's rank correlation coefficient

Kendall's rank correlation coefficient is commonly used to measure the degree of similarity between two sets of rank data and is applied to assess the significance of the relation between two measured quantities. Let $(U_1, U_2, ..., U_n)$, $(D_1, D_2, ..., D_n)$ denote realisations of two random variables U, D, such that U_i and D_i are unique for $i = 1, ...n$. We say that a pair observations (U_i, D_i), where $i < j$ is concordant if $U_i > U_j$ and $D_i > D_j$ or $U_i < U_j$ and $D_i < D_j$. If both $U_i > U_j$ and $D_i > D_j$ or both $U_i > U_j$

and $D_i > D_j$ then we say that the pair is discordant. The Kendall τ correlation coefficient is defined as follows

$$\tau = \frac{\text{number of concordant pairs} - \text{number of discordant pairs}}{n(n-1)/2}.$$

The coefficient is approximately equal to 0, if random variables U and D are independent. The value 1 signifies the perfect agreement between the two rankings

Multivariate Adaptive Regression Splines

Multivariate Adaptive Regression Splines (MARSplines) is a non-parametric procedure, which does not require any assumptions about the relationship between dependent and independent variables. While fitting MARSplines model, dependence is build upon coefficients and so called based functions fully determined by data. General mechanism governing this method can be seen as multiple, segmented linear regression. Division points computed based on data using least squares method state where one should apply each linear model. General equitation is given:

$$y = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(X),$$

where M denotes number of functional composites of model, X are predictors, $\beta_0, \beta_1, ..., \beta_M$ estimated parameters. For details see (Friedman, 1991).

Acknowlegments

Marek Gagolewski's research was supported by the Czech Science Foundation through the project No.18-06915S. The authors would like to thank Catriona Geras and Simon James for correcting earlier versions of the manuscript. The authors are also indebted to the Reviewers as their useful comments helped to improve the manuscript significantly. The authors declare that they have no competing financial interests. A.G., G.S and M.G. designed research; A.G., G.S and M.G. performed research; A.G. analysed the data; A.G., G.S and M.G. wrote the paper.

References

Abisheva, A., Garcıa, D. & Schweitzer, F. (2016). When the Filter Bubble Bursts: Collective Evaluation Dynamics in Online Communities. *CoRR*, *abs/1602.05642*.

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

Bornmann, L. [L.] & Daniel, H.-D. (2008). What do citation counts measure? A re-view of studies on citing behavior. *Journal of Documentation*, *64*(1), 45–80.

Bornmann, L. [Lutz], de Moya Anegón, F. & Leydesdorff, L. (2010). Do scientific advancements lean on the shoulders of giants? A bibliometric investigation of the Ortega hypothesis. *PLoS One*, *5*(10), e1332.

Brown, R. J. (2009). A simple method for excluding self citations from the h-index: The b-index. *Online Information Review*, *33*(6), 1129–1136.

C, V. & A, R. (2017). Rank distributions: Frequency vs. magnitude. *PLoS ONE*, *12*(10).

Catalini, C., Lacetera, N. & Oettl, A. (2015). The incidence and role of negative citations in science. *Proceedings of the National Academy of Sciences*, *112*(45), 13823–13826.

Chatterjee, A., Ghosh, A. & Chakrabarti, B. K. (2016). Universality of Citation Distributions for Academic Institutions and Journals. *PLOS ONE*, *11*(1).

Ciriminna, R. & Pagliaro, M. (2013). On the use of the h-index in evaluating chemical research. *Chemistry Central Journal*, *7*, no. 132.

Clauset, A., Larremore, D. B. & Sinatra, R. (2017). Data-driven predictions in the science of science. *Science*, *355*(6324), 477–480.

de Solla Price, D. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the Association for Information Science and Technology*, *27*(5), 292–306.

Deville, P. et al. (2014). Career on the move: Geography, stratification, and scientific impact. *Scientific Reports*, *4*, 4770.

Dorogovtsev, S. N. & Mendes, J. F. F. (2015). Ranking scientists. *Nature Physics*, *11*, 882–883.

Egghe, L. (2006). Theory and practise of the g-index. *Scientometrics*, *69*(1), 131–152.

Fortunato, S. et al. (2018). Science of science. *Science*, *359*(6379).

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, *19*(1), 1–67.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, *102*(46), 16569–16572.

Iglesias, J. & Pecharromán, C. (2007). Scaling the h–index for different scientific isi fields. *Scientometrics*, *73*(303).

Ionescu, G. & Chopard, B. (2013). An agent-based model for the bibliometric h-index. *The European Physical Journal B*, *86*, no. 426.

Kumar, S. (2016). Structure and dynamics of signed citation networks. In *Proceedings of the 25th international conference companion on world wide web* (pp. 63–64).

Malesios, C. & Psarakis, S. Q. Q. (2014). Comparison of the h-index for different fields of research using bootstrap methodology. *Quality & Quantity*, *48*(1).

Merton, R. K. (1968). The Matthew Effect in Science. *Science*, *159*(3810), 56–63.

Meyers, M. A. & Quan, H. C. (2017). The use of the h-index to evaluate and rank academic departments. *Journal of Materials Research and Technology*, *6*(4), 304–311.

Néda Z, B. T., Varga L. (2017). Science and facebook: The same popularity law! *PLOS ONE*, *12*(7).

Perc, M. (2014). The Matthew effect in empirical data. *Journal of The Royal Society Interface*, *11*(98).

Radicchi, F. & Castellano, C. (2011). Rescaling citations of publications in physics. *Phys. Rev. E*, *83*.

Radicchi, F., Fortunato, S. & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences*, *105*(45).

Redner, S. (2005). Citation Statistics from 110 Years of Physical Review. *Physics Today*, *58*(6), 49–54.

Szell, M. & Sinatra, R. (2015). Research funding goes to rich clubs. *Proceedings of the National Academy of Sciences*, *112*(48), 14749–14750.

Thelwall, M. (2016a). Are the discretised lognormal and hooked power law distributions plausible for citation data? *Journal of Informetrics*, *10*(2).

Thelwall, M. (2016b). The discretised lognormal and hooked power law distributions for complete citation data: Best options for modelling and regression. *Journal of Informetrics*, *10*(2).

Thelwall, M. (2016c). The discretised lognormal and hooked power law distributions for complete citation data: Best options for modelling and regression. *CoRR*, *abs/1601.00473*.

Vasilescu, B. et al. (2018). Academic papers using stack exchange data. https://meta.stackexchange.com/questions/134495.

Zogala–Siudem, B., Siudem, G., Cena, A. & Gagolewski, M. (2016). Agent-based model for the h-index – Exact solution. *The European Physical Journal B*, *89*, no. 21.