# On the Renormalization of Scientific Citations

István Daruka

Freelance Scientist

This paper was written in 2014.

Abstract

Through a comparative example, this paper calls for the necessity of the renormalization of scientific citations due to the exponentially inflating citation horizon.

Keywords: self-amplified research, scientometrics, citation horizon, citation share, real value of citations, renormalization

In the era of bursting scientific activity citations of peer reviewed scientific articles became vitally important. The refined scientometric records, e. g. the number of citations, H-index [1], Wu-index [2] and now their predictions for the future as possible measures of scientific potential [3], can play a decisive role in attaining research funding or even determining carrier paths [4-6]. Here we characterize the worldwide publication dynamics of the past century and propose a simple renormalization scheme for scientific citations. The derived real value equivalent of citations enables a contextual comparison of scientific impacts of different eras, offering a qualitative improvement to the naïve comparison by sheer citation numbers.

Publication dynamics of research papers show a staggering exponential growth. In Fig. 1 we plotted the number of published papers using the Thomson Reuters Web of Knowledge database, including all fields of science. In the same figure we also plotted the publication data of the Inspec database, representing the fields of physics, engineering, and information technology. The latter dataset was multiplied by a factor of 11 to obtain a data collapse. The two dips on the curves correspond to the First and Second World Wars. After WWII a rather rapid, cold war era burst can be observed lasting until 1970. After 1970 until today a somewhat smaller average slope sets in. Furthermore, it is intriguing to observe that in spite of the 11-fold difference in the average number of yearly publications, both datasets show a strongly correlated publications records dynamics. This suggests a more generic growth pattern spanning through different research fields. Such an exponential increase over more than a century can be paralleled with the empirical relation established for the four decade long exponential growth of microprocessors transistor counts known as the Moore's law.

In such a proliferating maze of scientific publications there is no doubt on the necessity and importance of scientometry. However, there are long standing and even sharp toned debates among scientists about the adequacy and accuracy of scientometric methods. “What are the conceptual limits of characterization? How to distribute citations among the authors? How can one compare the records among different fields?” are but a few of the many relevant questions.

Currently, each earned citation is rendered equally to all co-authors of the considered paper. Such a measure has two pitfalls. First, this way one citation is counted in a multiple way, determined by the number of authors on the paper. This rendering does not conserve the total number of citations. Secondly, such an equal footing might not reflect the degree of involvement of the authors properly.

These problems could possibly be resolved by introducing fractional citations [7], such that its sum is equal to one and to distribute it in a weighted manner among the authors. This would lead to the conservation of the citation number of the given paper and would also be proportional to the scientific activity of the considered scientists during the research and preparation of the paper. However, such a weighting scheme might be somewhat ambiguous/cumbersome to introduce, so as a first approximation one could follow the normalized citation scheme of the Astrophysics Data System of Harvard University [8], where the citations are distributed equally among the authors in a way that their sum should make unity.

Due to the exponential expansion of the citation horizon (Fig. 1), one cannot really compare the sheer number of citations of two authors whose publication activities were carried out in different eras separated e. g. by several decades. Even though some authors claim in advance that their top cited paper will soon be overtaking some of the most cited classic papers: “in about a year Chandrashekhar's 1945 tome, which has been the most cited paper in *Review of Modern Physics* for decades, will be dethroned by a decade-old paper on network theory” [9].

In order to compare scientific impact of different eras, similarly to the economic real value calculations, one can introduce the yearly *renormalized citation share (or real value equivalent)*, being the given scientist's total number of citations during the given year divided by the total number of citations produced by all scientists of that year.

As the gross numbers of yearly citations were not available in the Thomson Reuters Web of Knowledge database, we approximated the total number of yearly citations being proportional to the number of published papers per year (taken from the Thomson Reuters Web of Knowledge database). Such an approximation assumes that the number of references per scientific paper has not changed significantly through the years. The fitted exponential curve $e^{(x-1900)/21 + 10.6}$ in Fig. 1 can be used then as a (year dependent) renormalization factor to calculate the yearly citation shares.

Actually, taking 1970 as a reference year, one could multiply the citations gained in year *x* by a factor of $\exp^{-(x-1970)/21}$ (citation share) and sum it up for all the pertaining years to get the renormalized number of citations. Note, that this way a citation from 2010 has a *real value equivalent* only ~ $e^{-1.9} = 0.15$ of the citations made in 1970, and only 0.045 that of the 1945 citations claimed by Ref. [9]. One should also renormalize both the pertaining Hirsch- and Wu-indices accordingly.

Fig. 2 shows both the original and the renormalized number of citations for the 1943 Chandrasekhar Rev. Mod. Phys. paper [10] and the Albert and Barabasi 2002 Rev. Mod. Phys. paper [11] as well. The above considered citation renormalization scheme clearly demonstrates the “Red Queen effect” introduced by the exponential inflation of the citation horizon.

The total number of citations for the Chandrasekhar paper is 5452 that seems to be overtaken by the 6368 sheer number of citations of the Albert and Barabasi paper. However, the renormalized citation number of the latter (1010) is *only one third* of Chandrasekhar's renormalized value (3164). Thus, dethroning by sheer numbers might not offer the crown of proper comparison.

The author acknowledges inspiring discussions with A. Csoto, F. Schäffler, and G. Springholz.

**References:**

[1] J. E. Hirsch, “An index to quantify an individual's scientific research output", PNAS **102,** 16569 (2005).

[2] http://arxiv.org/abs/0805.4650v1;
http://physicsworld.com/cws/article/news/2008/jun/05/whats-your-wu-index

[3] D. E. Acuna, S. Allesina, and K. P. Kording, “Future impact: Predicting scientific success", Nature **489**, 201 (2012).

[4] A. M. Petersen, M. Riccaboni, H. E. Stanley, and F. Pammolli, "Persistence and Uncertainty in the Academic Career," Proc. Natl. Acad. Sci. USA **109**, 5213-5218 (2012).

[5] A. M. Petersen, H. E. Stanley, and S. Succi, "Statistical Regularities in the Rank-Citation Profile of Scientists," Nature Scientific Reports **1**[181], 1-7 (2011).

[6] A. M. Petersen, W.-S. Jung, J.-S. Yang, and H. E. Stanley, "Quantitative and Empirical Demonstration of the Matthew Effect in a Study of Career Longevity," Proc. Natl. Acad. Sci. USA **108**, 18-23 (2011).

[7] Z. Papp, “Miért nem használunk frakcionális mutatókat a tudományos tevékenység mérésére?”, Magyar Tudomány (2012/4); Braun Tibor – Bujdosó E. – Ruff I. “A tudomány mint a mérés tárgya. Tudománymetriai kutatás Magyarországon”, Budapest, MTA Könyvtára (1981). (These references are not available in English.)

[8] http://adswww.harvard.edu/

[9] The network takeover, Albert-Laszlo Barabasi, Nature Physics **8**, 14 (2012).

[10] S. Chandrasekhar, “Stochastic Problems in Physics and Astronomy”, Rev. Mod. Phys. **15**, 1 (1943).

[11] Reka Albert and Albert-Laszlo Barabasi, “Statistical mechanics of complex networks”,Rev. Mod. Phys. **74**, 47 (2002).

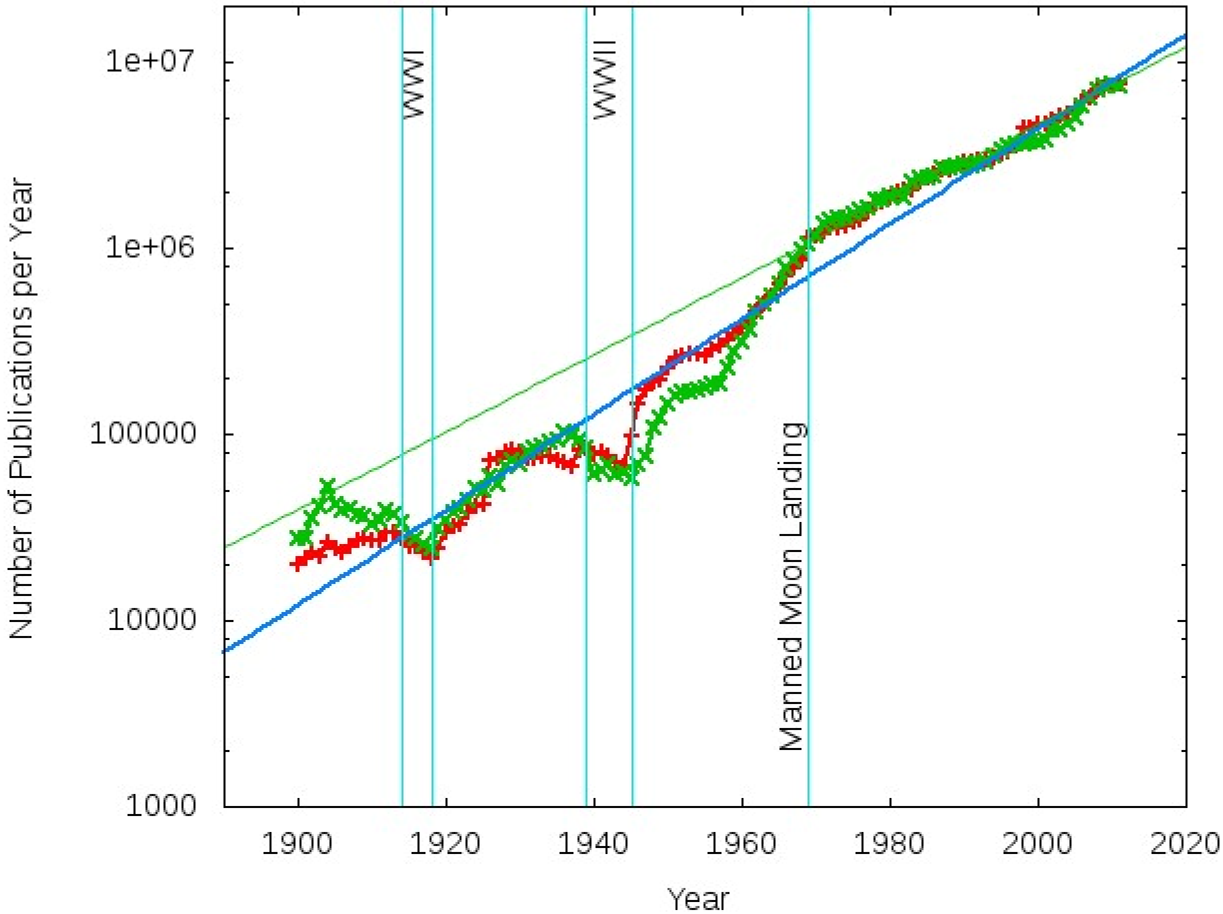


Fig. 1. Total number of scientific publications as a function of years according to the Thomson Reuters Web of Knowledge (red curve) and the Inspec (green curve) databases. The latter curve, representing the publication activity in the fields of physics, engineering, and information technology was multiplied by a factor of 11 to obtain a data collapse. Besides the strong correlation of the two curves, one can see a clear exponential signature, manifested in the more than 300-fold increment in the number of publications during the past 110 years. The two dips correspond to the First and Second World Wars. After WWII a rather rapid, cold war era burst can be observed lasting until 1970. After 1970 until today a somewhat smaller average slope sets in. Functions of $e^{(x-1900)/17 + 9.4}$ (thick blue line) and $e^{(x-1900)/21 + 10.6}$ (thin green line) were plotted.

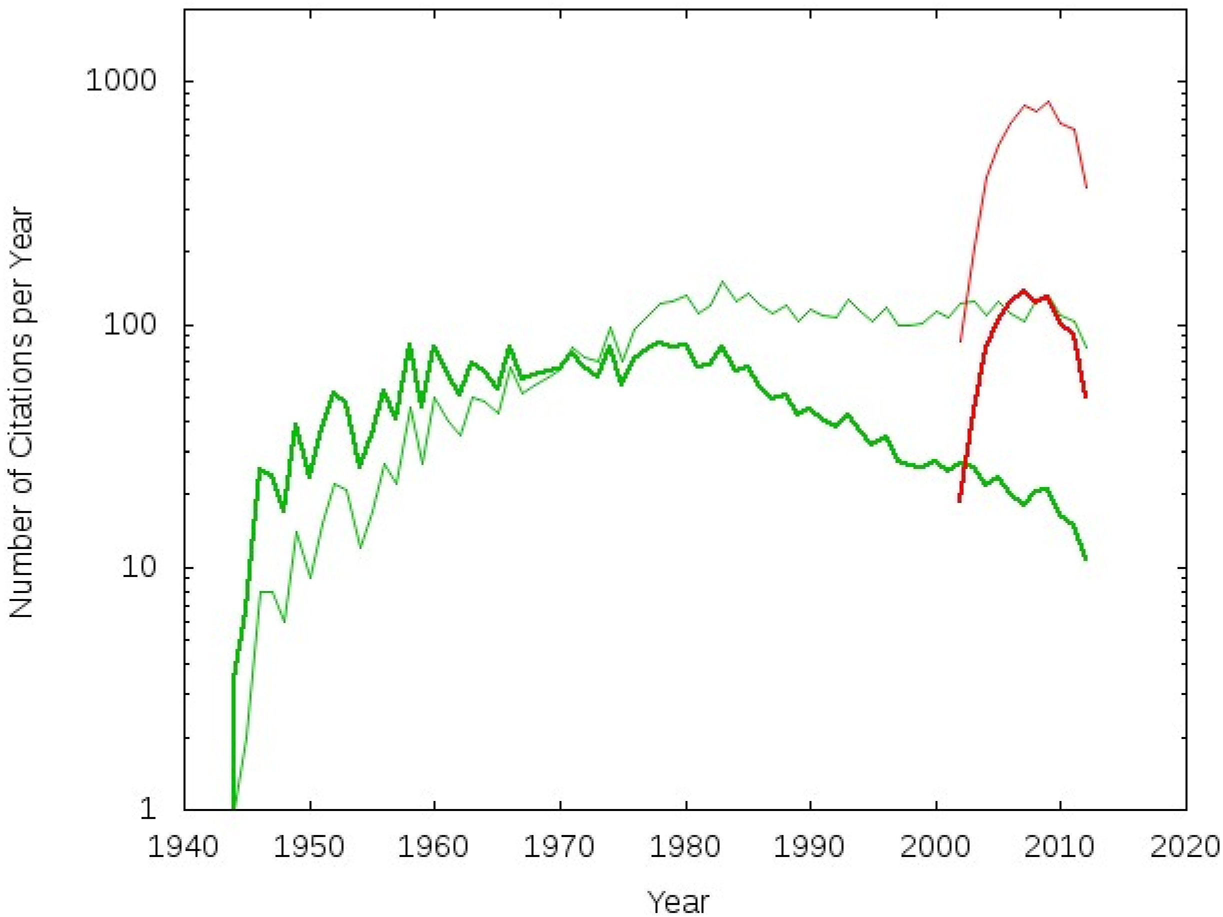


Fig. 2. The yearly citations of the 1943 Chandrasekhar Rev. Mod. Phys. paper [10] (thin green line) and that of the 2002 Albert and Barabasi paper [11] (thin red line). The renormalized number of citations are shown by the corresponding thick lines.