A New Explanation of Some Leiden Ranking Graphs Using Exponential Functions

Leo Egghe*

Campus Diepenbeek
Universiteit Hasselt, Belgium
E-mail: leo.egghe@uhasselt.be

ABSTRACT

A new explanation, using exponential functions, is given for the S-shaped functional relation between the mean citation score and the proportion of top 10% (and other percentages) publications for the 500 Leiden Ranking universities.
With this new model we again obtain an explanation for the concave or convex relation between the proportion of top $100\theta\%$ publications, for different fractions of θ.

Keywords: Leiden ranking, exponential function, mean citation rate, S-shaped

1. INTRODUCTION

For 500 universities (from 41 countries) from the Leiden Ranking 2011/2012 one observes in Waltman et al. (2012) the relation between the mean normalized citation score (MNCS) and the proportion of top 10% publications ($PP_{top10\%}$).

Upon specifying a field, MNCS is the mean number of citations of the publications of a university in this field (normalized in several ways – see Waltman et al.). $PP_{top10\%}$ is the proportion (fraction) of the publications of a university in this field that, compared with other publications in this field, belong to the top 10% most frequently cited.

In Waltman et al. one finds an S-shaped relation between $PP_{top10\%}$ and MNCS: first convex then concave (see their Fig. 2). Allowing some other percentages, Waltman et al. find a convex relation between $PP_{top10\%}$ (as abscissa) and $PP_{top5\%}$ (as ordinate) and a concave relation between $PP_{top10\%}$ (as abscissa) and $PP_{top20\%}$ (as ordinate) – again see their Fig. 3.

In Egghe (2013) we explained all these regularities using the shifted Lotka function

$$f(n) = \frac{C}{(n+1)^{\alpha}} \qquad (1)$$

where $C > 0$, $\alpha > 1$, $n \geq 0$, which was documented in Egghe and Rousseau (2012). Here $f(n)$ is the continuous version of the number of publications with n citations. Using (1) we studied the functional relation between the non-normalized variant of MNCS, denoted MCS and $PP_{top10\%}$. Putting $x = MCS$ and $y = PP_{top10\%}$ we proved in Egghe (2013) that

$$y = 10^{-\frac{\frac{1}{x}+1}{\alpha-1}} \qquad (2)$$

where α is the exponent of Lotka in (1), where we also proved the S-shape, hereby explaining this empirical relationship in Waltman et al. (2012).

For general fractions θ we obtained in Egghe (2013)

$$y = \theta^{\frac{\frac{1}{x}+1}{\alpha-1}} \qquad (3)$$

where $x = MCS$ and $y = PP(\theta) =: PP_{top\,100\theta\%}$.

From this model we proved, for any fractions θ_1, θ_2, the following functional relation between $PP(\theta_1)$ and $PP(\theta_2)$:

$$PP(\theta_2) = PP(\theta_1)^{\frac{\ln\theta_2}{\ln\theta_1}} \qquad (4)$$

which is an explanation of the convex and concave graphs in Waltman et al.: concave if $\theta_2 > \theta_1$ and convex if $\theta_2 < \theta_1$.

In this paper the same problems are studied: explaining the S-shaped relationship between MCS and $PP(\theta)$ for any θ and the convex or concave relationships between two $PP(\theta_1)$ and $PP(\theta_2)$ as found in Waltman

et al. Now, however, we do not use the shifted Lotka function (1) but the exponential function (a very classical function)

$$f(n) = Ca^{-n} \qquad (5)$$

where $C > 0$, $a > 1$, $n \geq 0$ where the function $f(n)$ has the same meaning as explained above. With $x = MCS$ and $y = PP(\theta)$ we will prove (in the next section) that

$$y = \theta^{\frac{1}{x\ln a}} \qquad (6)$$

which is a clearly different function when compared with (3). But also this regularity explains the one found (empirically) in Waltman et al. since (6) is also S-shaped.

Remarkably, in the third section, using (6) for two fractions θ_1 and θ_2, we will reprove (4); i.e., the same regularity between any two $PP(\theta_1)$ and $PP(\theta_2)$ is found using exponential functions as when we used the shifted Lotka function (which are clearly different functions). But, at the end of the paper, we will also give two cases where (4) is not valid.

The paper closes with a conclusion and open problems section.

2. EXPLANATION OF THE RELATION BETWEEN MCS AND PP(θ)

As indicated in the introduction we use the exponential function

$$f(n) = Ca^{-n} \qquad (7)$$

denoting the continuous version of the number of publications with n citations in a field where $C > 0$, $a > 1$, $n \geq 0$. Since the field is fixed, we have also that C and a are fixed.

For a university we use the exponential function

$$\varphi(n) = C' a'^{-n} \qquad (8)$$

$(C' > 0$, $a' > 1$, $n \geq 0$) denoting the continuous version of the number of publications of this university with n

citations. Since we deal with several universities (e.g. 500 in the case of Waltman et al. (2012)) we have here that C' and a' are variables.

As noted by one referee, the fact that, per university, we use (8) with C' and a' variables, does not necessarily lead to (7) for the entire field. Therefore, formula (7) should be considered as an assumption to be valid in practice:

Denote by T the total number of publications in the entire field and by T' the total number of publications in a university in this field. We have, by definition of $f(n)$

$$T = \int_0^\infty f(n) \ dn = \frac{C}{\ln a} \qquad (9)$$

(since $a > 1$) and similarly

$$T' = \frac{C'}{\ln a'} \qquad (10)$$

Denote by A the total number of citations in the entire field and by A' the total number of citations in a university in this field. We have, by definition of $f(n)$

$$A = \int_0^\infty n \ f(n) \ dn = \frac{C}{(\ln a)^2} \qquad (11)$$

which is easily seen using partial integration, by the fact that $a > 1$ and the fact that

$$\lim_{n \to \infty} \frac{n}{a^n} = 0 \qquad (12)$$

Similarly we have

$$A' = \frac{C'}{(\ln a')^2} \qquad (13)$$

By definition of MCS, being the average number of citations per publication of a university, we have

$$MCS = \frac{A'}{T'} = \frac{1}{\ln a'} \qquad (14)$$

, using (10) and (13).

We first determine n_0 defining the top $100\theta\%$ publications in the field (for any fraction θ), by (7) :

$$\int_{n_0}^\infty C a^{-n} \ dn = \theta T \qquad (15)$$

From (15) it follows that

$$\frac{C}{\ln a} a^{-n_0} = \theta T$$

and by (9) we have

$$a^{-n_0} = \theta$$

or

$$n_0 = -\frac{\ln \theta}{\ln a} \qquad (16)$$

which is a positive number because $0 < \theta < 1$.

Then the university proportion in these top $100\theta\%$ of the papers in the field is, by (8)

$$PP(\theta) = \frac{1}{T'} \int_{n_0}^\infty C' \ a'^{-n} dn \qquad (17)$$

$$PP(\theta) = a'^{-n_0}$$

$$PP(\theta) = a'^{\frac{\ln \theta}{\ln a}}$$

(by (16))

$$PP(\theta) = e^{\frac{\ln a'}{\ln a} \ln \theta}$$

$$PP(\theta) = \theta^{\frac{\ln a'}{\ln a}}$$

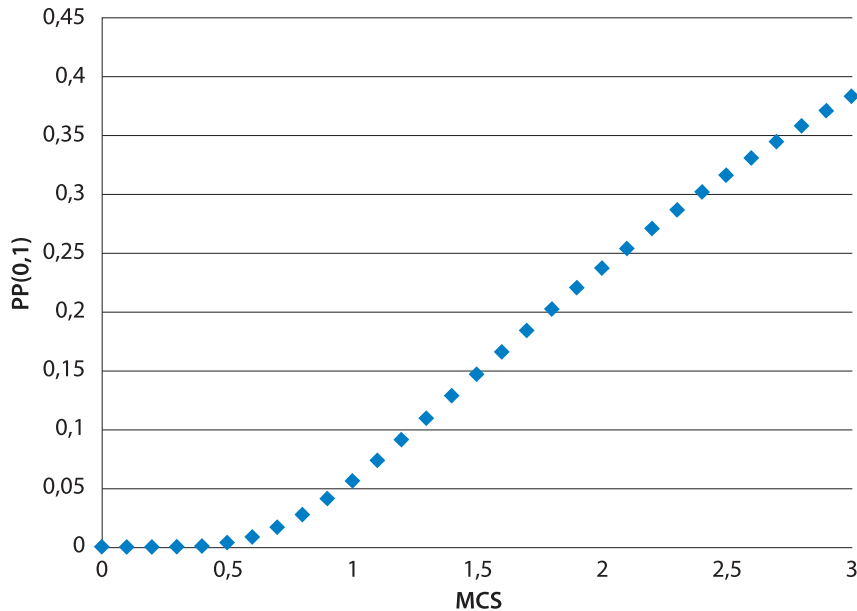$$PP(\theta) = \theta^{\frac{1}{MCS} \frac{1}{\ln a}} \qquad (18)$$

(by (14)) or the function (6). This is an increasing function (since $0 < \theta < 1$) for which ($x = MCS$)

$$\lim_{x \to \infty} PP(\theta) = 1 \qquad (19)$$

and

$$\lim_{x \to 0} PP(\theta) = 0 \qquad (20)$$

The number a is a fixed parameter (of the field). Fig. 1 is the graph of (18) for $\ln a = 0.8$ from which the S-shape is clear, and it is close to the S-shape obtained in Waltman et al. (2012). So this represents a new explanation of this regularity.

Fig. 1 Graph of (18) for ln a = 0.8

3. EXPLANATION OF THE RELATION BETWEEN ANY TWO VALUES OF PP(θ₁) AND PP(θ₂)

For any two fractions θ_1 and θ_2 we have, by (18)

$$PP(\theta_1) = \theta_1^{\frac{1}{MCS}\frac{1}{\ln a}} \qquad (21)$$

$$PP(\theta_2) = \theta_2^{\frac{1}{MCS}\frac{1}{\ln a}} \qquad (22)$$

Hence

$$\ln PP(\theta_1) = \frac{1}{MCS}\frac{1}{\ln a}\ln \theta_1$$

$$\ln PP(\theta_2) = \frac{1}{MCS}\frac{1}{\ln a}\ln \theta_2$$

from which it follows that for any two fractions θ_1 and θ_2

$$\frac{\ln PP(\theta_1)}{\ln \theta_1} = \frac{\ln PP(\theta_2)}{\ln \theta_2} \qquad (23)$$

hence

$$e^{\frac{\ln PP(\theta_1)}{\ln \theta_1}} = e^{\frac{\ln PP(\theta_2)}{\ln \theta_2}}$$

or

$$PP(\theta_1)^{\frac{1}{\ln \theta_1}} = PP(\theta_2)^{\frac{1}{\ln \theta_2}}$$

or

$$PP(\theta_2) = PP(\theta_1)^{\frac{\ln \theta_2}{\ln \theta_1}} \qquad (24)$$

which is (4).

As already remarked in the introduction, this relation is the same as the one found in Egghe (2013) where the shifted Lotka function was used, a remarkable fact!

We have derived that (since $0 < \theta_1, \theta_2 < 1$), if $\theta_2 > \theta_1$, then (by (24)), $PP(\theta_2)$ is a concave function of $PP(\theta_1)$ and that, if $\theta_2 < \theta_1$, then (by (24)), $PP(\theta_2)$ is a convex function of $PP(\theta_1)$ — see the graphs in Egghe (2013) which explain the corresponding graphs in Waltman et al. (2012).

So (24) is valid when f and φ are both shifted Lotka functions (proved in Egghe (2013)) and when f and φ are both exponential functions (proved here). Now we present two cases where (24) is not valid.

Case I

We take f to be a shifted Lotka function and φ to be an exponential function:

$$f(n) = \frac{C}{(n+1)^{\alpha}} \tag{25}$$

$(C > 0, \alpha > 2, n \geq 0)$

$$\varphi(n) = C' a'^{-n} \tag{26}$$

$(C' > 0, \alpha' > 1, n \geq 0)$

Following the method of the previous section we find, for every fraction θ

$$PP(\theta) = e^{\frac{1}{MCS}\left(1 - \left(\frac{1}{\theta}\right)^{\frac{1}{\alpha-1}}\right)} \tag{27}$$

From the method in this section we find, for any two fractions θ_1 and θ_2,

$$PP(\theta_2) = PP(\theta_1)^{\frac{1-\left(\frac{1}{\theta_2}\right)^{\frac{1}{\alpha-1}}}{1-\left(\frac{1}{\theta_1}\right)^{\frac{1}{\alpha-1}}}} \tag{28}$$

which is, clearly, not the function (24).

Case II

We take f to be an exponential function and φ to be a shifted Lotka function:

$$f(n) = Ca^{-n} \tag{29}$$

$(C > 0, a > 1, n \geq 0)$

$$\varphi(n) = \frac{C'}{(n+1)^{\alpha'}} \tag{30}$$

$(C' > 0, \alpha' > 2, n \geq 0)$

Following the methods of the previous section we find, for every fraction θ

$$PP(\theta) = \left(1 - \frac{\ln\theta}{\ln a}\right)^{-1-\frac{1}{MCS}} \tag{31}$$

From the method in this section we find, for any two fraction θ_1 and θ_2,

$$PP(\theta_2) = PP(\theta_1)^{\frac{\ln\left(1-\frac{\ln\theta_2}{\ln a}\right)}{\ln\left(1-\frac{\ln\theta_1}{\ln a}\right)}} \tag{32}$$

which is, clearly, not the function (24).

4. CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

Experimental regularities in Waltman et al. (2012) are proved mathematically in this paper. Using the exponential function we proved the S-shaped functional relation between the mean citation rate and the proportion of top $100\theta\%$ publications. We obtained a different function than in Egghe (2013) where a shifted Lotka function was used, but we obtained an S-shape in both cases.

With this new model we could reprove the function (obtained in Egghe (2013))

$$PP(\theta_2) = PP(\theta_1)^{\frac{\ln\theta_2}{\ln\theta_1}} \tag{33}$$

for the relation between two $PP(\theta)$-values. It is very remarkable that we obtain exactly the same function as in Egghe (2013) although different starting functions were used (shifted Lotka in Egghe (2013) and exponential here). We also showed that (33) explains the corresponding empirical regularities in Waltman et al. (2012).

The importance of this paper is that the assumption of a simple exponential function (5) leads to an explanation of several regularities in Waltman et al.

We state as an open problem: can the S-shape in Waltman et al. for the relation between the mean citation rate and the proportion of top $100\theta\%$ publications be proved using other starting functions (other than the shifted Lotka function and other than the exponential function)?

Also the following is an open problem: characterise the functions $f(n)$ and $\varphi(n)$ (previous section) for which (33) is valid. From Egghe (2013) and this paper, this class of functions must include the shifted Lotka function and the exponential function.

REFERENCES

Egghe, L. (2013). Informetric explanation of some Leiden Ranking graphs. *Journal of the American Society for Informetric Science and Technology*, to appear.

Egghe, L. & Rousseau, R. (2012). Theory and practise of the shifted Lotka function. *Scientometrics*, 91(1), 295-301.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C.M., Tijssen, R. J.W., van Eck, N.J., van Leeuwen, T.N., van Raan, A. F.J., Visser, M.S., & Wouters, P. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American Society for Information Science and Technology*, 63(12), 2419-2432.