Global impact measures

Peer-reviewed author version

EGGHE, Leo & Rousseau , Ronald (2022) Global impact measures. In:
SCIENTOMETRICS, 128 (1), p. 699-707.

DOI: 10.1007/s11192-022-04553-w
Handle: http://hdl.handle.net/1942/38906

Global impact measures

Leo Egghe[1] and Ronald Rousseau[2,3] *

[1] Hasselt University, 3500 Hasselt, Belgium

leo.egghe@uhasselt.be

ORCID: 0000-0001-8419-2932

[2] KU Leuven, MSI, Facultair Onderzoekscentrum ECOOM,

Naamsestraat 61, 3000 Leuven, Belgium

ronald.rousseau@kuleuven.be &

[3] University of Antwerp, Faculty of Social Sciences,

Middelheimlaan 1, 2020 Antwerp, Belgium

ronald.rousseau@uantwerpen.be

ORCID: 0000-0002-3252-2538

Abstract

We present a continuous theory of global impact measures. Such measures combine inequality (like the Lorenz theory) with productivity, leading to the notion of global impact and its measurement.

Keywords: Lorenz curve; inequality; global impact measures

* Corresponding author

1. Introduction

In this work, our investigations will lead to the introduction of a new type of measures. These measures are called global impact measures and take concentration as well as production into account. They can be considered as elements in a theory on impact as outlined in other publications (Egghe, 2021; Egghe & Rousseau, 2022a; Egghe & Rousseau, 2022b).

Many distributions studied in informetrics, such as authors and their publications (Lotka, 1926), websites and inlinks (Rousseau, 1997; Faloutsos et al., 1999), or topics and journals dealing with them (Bradford, 1934) can be described by power law relations (Egghe, 2005) or similar long-tailed distributions (Laherrère & Sornette, 1998). Moreover, these power laws have many applications in other fields such as demography (cities and their inhabitants), linguistics (words and their uses), economics (incomes in a market economy), ecology (fragmentation of forests), astronomy (initial mass functions) and many more, see (Pareto, 1895; Auerbach, 1913; Zipf, 1941, 1949; Salpeter, 1955; Newman, 2005; Saravia et al., 2018). One common characteristic of these distributions is the high concentration of items among a few sources. As such the study of concentration or inequality, with its social implications, is one of the main topics studied in our field (Rousseau et al., 2018, Section 9.5).

In the next section we introduce a dominance order in the case of a non-normalized Lorenz curve and prove an impact-concentration theorem. This then leads in the following section to the definition of global impact measures, followed by a number of practical examples. We conclude by pointing out new opportunities for further studies in the science of science.

2. Continuous dominance and the impact-concentration theorem

Let T > 0, let ***U*** be the set $\{Z: [0,T] \rightarrow \mathbb{R}^+$, Z continuous and decreasing$\}$, $\boldsymbol{U_0}$ = $\{Z \in \boldsymbol{U}; Z > 0 \text{ on } [0,T[\ \}$ and ***Z*** =$\{Z \in \boldsymbol{U};\ Z \text{ strictly decreasing}\}$. Then $\boldsymbol{Z} \subset \boldsymbol{U_0} \subset \boldsymbol{U}$.

Definition 1. The continuous Lorenz curve L(x).

Given Z in ***U***, we define the continuous Lorenz curve of Z as the graph of the function

$$[0,1] \rightarrow [0,1]: x \rightarrow \frac{\int_0^{xT} Z(s)ds}{\int_0^{T} Z(s)ds} \quad (1)$$

Definition 2. The classical Lorenz dominance order.

The dominance order on ***U*** (Marshall-Olkin-Arnold, 2011), is defined as:

if Z, Y ∈ ***U***, then $Z -\!\!<_L Y$ iff $\forall x \in [0,T]: \frac{\int_0^x Z(s)ds}{\int_0^T Z(s)ds} \leq \frac{\int_0^x Y(s)ds}{\int_0^T Y(s)ds}$ (2)

or equivalently:

$$\forall x \in [0,1]: \frac{\int_0^{xT} Z(s)ds}{\int_0^T Z(s)ds} \leq \frac{\int_0^{xT} Y(s)ds}{\int_0^T Y(s)ds}. \quad (3)$$

The relation $Z -\!\!<_L Y$ means that the continuous Lorenz curve of Y is situated above the continuous Lorenz curve of Z, providing an argument in favor of using the concave, and not the convex, form of the Lorenz curve.

Definition 3. A continuous concentration measure

A function m from ***U*** to the positive real numbers is a concentration measure if it is an order morphism from ***U***, $-\!\!<_L$ to the positive real numbers. This means that X$-\!\!<_L$ Y implies that m(X) ≤ m(Y), with equality only if X and Y have the same Lorenz curve. In practice one often requires that m(**0**) = 0, with **0** the zero function on [0,T].

Definition 4. The non-normalized Lorenz curve

Given Z in ***U***, we define the function $I_Z: [0,T] \to R^+: x \to \int_0^x Z(s)ds$. This function is concavely increasing. Its graph will be said to be the non-normalized Lorenz curve of Z.

Clearly the graph of I_Z is the graph of a cumulative function. Yet, because we use it here within a generalization of the classical Lorenz curve (Lorenz, 1905) we refer to it as a non-normalized Lorenz curve.

Definition 5. The non-normalized dominance order on ***U***

If Z, Y ∈ ***U***, then $Z -\!\!< Y$ iff $\forall x \in [0,T]: I_Z(x) = \int_0^x Z(s)ds \leq I_Y(x) = \int_0^x Y(s)ds$ (4)

Clearly, -< on ***U*** is reflexive, antisymmetric, and transitive. Hence it is a partial order (Roberts, 1979).

From the definition of the Lorenz curve of a function Z in ***U*** it follows that it is the image of its non-normalized Lorenz curve, through a linear mapping with matrix $\begin{pmatrix} 1/T & 0 \\ 0 & 1/(I_Z(T)) \end{pmatrix}$ which is a composition of a horizontal and a vertical contraction (T > 1).

Related to this observation we formulate two remarks.

Remark 1. If Z, Y ∈ ***U***, and $I_Z(T) = I_Y(T)$ then Z -< Y implies $Z -<_L Y$.

Remark 2. If Z, Y ∈ ***U*** and Z ≤ Y then Z -< Y, but the relation $Z -<_L Y$ may or may not hold.

Notation

We denote the average of Z ∈ ***U*** by $\mu_Z = \frac{1}{T}\int_0^T Z(s)ds = \frac{I_Z(T)}{T}$.

Theorem 1. The impact-concentration theorem.

$\forall Z, Y \in$ ***U₀***, with Z ≠ Y the following expressions are equivalent.

(i) Z -< Y ;

(ii) $\exists Y^* \neq Z \in$ ***U₀***, such that $\mu_Z = \mu_{Y^*}$, Z -< Y* ≤ Y and $Z -<_L Y^*$;

(iii) $\exists Y^* \neq Z \in$ ***U₀***, such that $\mu_Z = \mu_{Y^*}$, Z -< Y* -< Y and $Z -<_L Y^*$;

Proof. The implications (ii) ⇒ (iii) ⇒ (i) trivially follow from the facts that -< is transitive and that Y* ≤ Y implies Y* -< Y.

We next prove that (i) implies (ii).

If $I_Z(T) = I_Y(T)$ we may set Y* = Y, see Remark 1. We next assume that $I_Z(T) < I_Y(T)$, and hence $\mu_Z < \mu_Y$.

Define I*(x) on [0,T] as $\min(I_Y(x), I_Z(T))$. Then Z -< Y implies that $\forall x \in [0, T]$: $I_Z(x) \leq I_Y(x)$ and hence, $\forall x \in [0, T]$: $I_Z(x) \leq I^*(x) \leq I_Y(x)$. The following construction is illustrated in Fig.1.

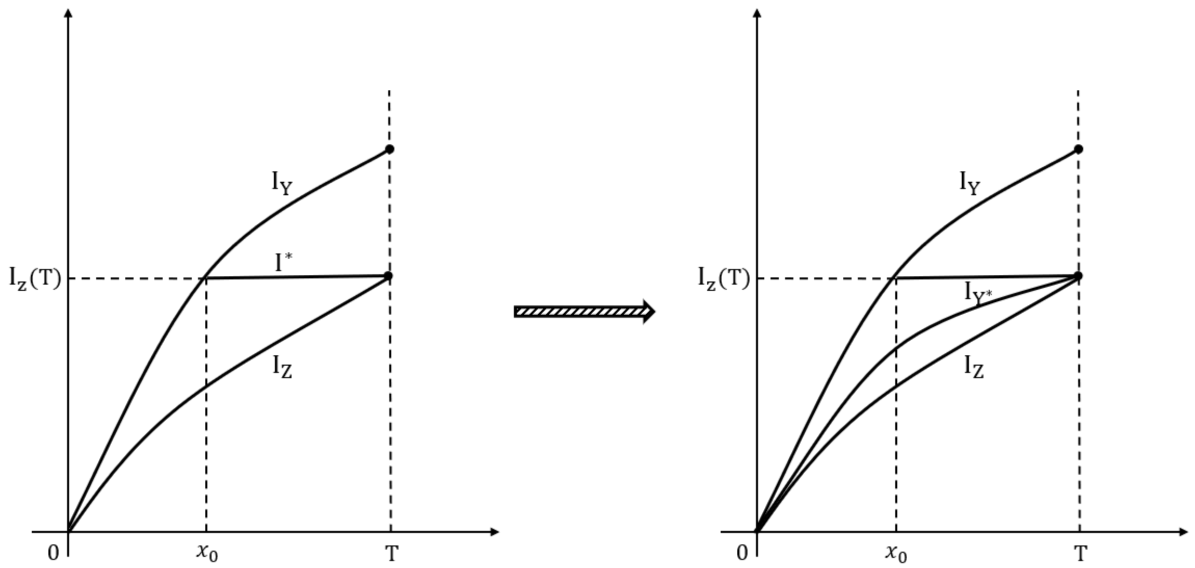

Fig. 1. Illustration of the construction of the function I_{Y^*}

As Z -< Y (Z,Y different) there exists, by continuity, a point $x_0 \in]0,T[$ such that $I_Z(x_0) < I_Y(x_0)$. As Z is strictly decreasing on [0,T] it is not zero on an interval of finite length. Hence, we see that $I_Z(x_0) < I_Z(T)$ and $I_Z(x_0) < I^*(x_0)$. The function I* has at most one kink (recall that the integral of a continuous function is differentiable). Hence, we can work around this kink by replacing I* locally by a smooth, concavely increasing curve, still denoted as I* such that $I_Z(x) < I^*(x) \leq I_Y(x)$ on [0,T] and such that I* is in no point increasing faster than I_Y. Now we define Y* as (I*)′ leading to: $I_{Y^*} = I^*$. From this, we see that Z -< Y* -< Y, actually by its construction, we even have Y* ≤ Y.

Now $I_Z(T) = I_{Y^*}(T)$ and hence $\mu_Z = \mu_{Y^*}$. By Remark 1 we then have that $L_Z < L_{Y^*}$ on]0,1[or Z $-<_L$ Y.□

Some comments on the proof of Theorem 1. In (Egghe & Rousseau, 2002) we used the standard Lorenz curve for defining the core of a scientific topic. More precisely, and in the continuous case, we can interpret I*(x) as the non-normalized Lorenz curve, corresponding to the x_0-th partial curve (using a terminology analogous to the one used in (Egghe & Rousseau, 2002)). Note that I* is just a help function as it is a cumulative function of a function that does not belong to $\boldsymbol{U_0}$ (this function is not even continuous).

Given the function Z, the function I* on the left-hand side of Fig.1 is nothing but the TOP-curve of Y, with top line y= $I_Z(T)$. This remark provides a relation between the theory of TOP-curves as developed in (Egghe et al., 2007) and the present article.

3. Global impact measures

Let $\boldsymbol{U_\mu}$ = {Z ∈ $\boldsymbol{U_0}$: $\mu_Z(T) = \mu$} and let m be a function $\boldsymbol{U_0} \to \mathbb{R}^+$. Then we have the following theorem, leading to the definition of measures of a new type.

Theorem 2

The following three expressions are equivalent:

(i) $\forall Z, Y \in \boldsymbol{U_0}$, Z≠Y: Z -< Y ⇒ m(Z) < m(Y)

(ii) $\forall Z, Y \in \boldsymbol{U_0}$: Z -< Y ⇒ m(Z) ≤ m(Y) and, for all μ > 0, if Z, Y ∈ $\boldsymbol{U_\mu}$: (Z $-<_L$ Y and Z≠Y, ⇒ m(Z) < m(Y))

(iii) $\forall Z, Y \in \boldsymbol{U_0}$: Z ≤ Y ⇒ m(Z) ≤ m(Y) and, for all μ > 0, if Z, Y ∈ $\boldsymbol{U_\mu}$ (Z≠Y), we have: (Z $-<_L$ Y ⇒ m(Z) < m(Y))

Proof.

(i) ⇒ (ii). Let Z -< Y, and Z ≠Y, then we know by (i) that m(Z) < m(Y). Of course, if Z = Y, then m(Y)=m(Z) so that always m(Z) ≤ m(Y). If now, for μ > 0, Z, Y ∈ $\boldsymbol{U_\mu}$: Z -<$_L$ Y, with Z≠Y, then $I_Z(T) = I_Y(T)$ and by Remark 1, Z -< Y, and thus m(Z) < m(Y). This proves this implication.

(ii) ⇒ (iii) is trivial as Z≤ Y implies Z -< Y.

(iii) ⇒ (i). Let Z ≠ Y and Z -< Y. By the "impact-concentration theorem", we know that there exists Y* ≠ Z in $\boldsymbol{U_0}$ such that Z -< Y* -< Y and Z -<$_L$ Y* with $\mu_Z = \mu_{Y^*}$ simply denoted as μ. Hence Z and Y* belong to $\boldsymbol{U_\mu}$. It follows from (iii) that m(Z) < m(Y*) and m(Y*) ≤ m(Y), which proves (i).

Definition 6

We say that a measure m as defined above is increasing on $\boldsymbol{U_0}$,-< if (Z -< Y ⇒ m(Z) ≤ m(Y)), and is strictly increasing if for Z≠Y: (Z -< Y ⇒ m(Z) < m(Y)).

Definition 7. Global impact measures

A function $\boldsymbol{U_0} \to \mathbb{R}^+$ such that $\forall Z,Y$ ∈ $\boldsymbol{U_0}$, Z≠Y: (Z -< Y ⇒ m(Z) < m(Y)) is called a global impact measure.

Corollary

The following expressions are equivalent

(i) m is a global impact measure on $\boldsymbol{U_0}$

(ii) m is an increasing function on $\boldsymbol{U_0}$,-< and for every μ > 0 (fixed) m is a concentration measure on $\boldsymbol{U_{0,\mu}}$

(iii) m is an increasing function on $\boldsymbol{U_0}$,≤ and for every μ > 0 (fixed) m is a concentration measure on $\boldsymbol{U_{0,\mu}}$

Note that if m is a global impact measure on $\boldsymbol{U_0}$ and m(Z) < m(Y) then it does not necessarily follow that Z -< Y.

4. Examples of global impact measures

In this section we provide some examples of global impact measures, derived for well-known concentration measures.

4.1. The generalized Gini-index

Obviously, the area under the non-normalized Lorenz curve, defined as

$$\int_0^T I_Z(x)dx \qquad (5)$$

is a global impact measure.

As we work on ***U₀*** there is no function with Gini-value equal to zero, but zero is the infimum (largest lower bound).

4.2. The length of the non-normalized Lorenz curve of Z in ***U₀*** minus T. This measure is denoted as $\mathcal{L}$(Z).

Now $\mathcal{L}(Z) = \int_0^T \sqrt{1 + \left(\frac{d}{ds}(I_Z(s))\right)^2}\, ds - T = \int_0^T \sqrt{1 + (Z(s))^2}\, ds - T$ (6)

The length of the non-normalized Lorenz curve (or this length minus T) is an increasing function on ***U₀***, ≤ and for every μ > 0 (fixed) m is a concentration measure on ***U₀,μ*** as the length of the standard Lorenz curve (or this length minus $\sqrt{2}$) is a bona fide concentration curve (Dagum, 1980).

Alternatively, we know that if f is a strictly convex, continuous, and increasing function, then Z-< Y, Z ≠ Y, implies that $\int_0^T f\big(Z(s)\big)ds < \int_0^T f\big(Y(s)\big)ds$. Taking now f(s) = $\sqrt{1+s^2}$, shows that $\mathcal{L}$ is a global impact measure. We subtract T because we require that when Z tends to zero (pointwise), $\mathcal{L}$(Z) also tends to zero.

4.3. m(Z) = $\int_0^T (Z(s))^p ds$, p > 1 (7)

This measure m is a global impact measure because f(s) = s^p (p > 1) is a strictly convex, continuous, and increasing function.

Considering now the case p = 2, we define $(\sigma_Z)^2$ as $\frac{1}{T}\int_0^T (Z(s) - \mu_Z)^2 ds$, and see that $\int_0^T (Z(s))^2 ds = (\sigma_Z)^2 + (\mu_Z)^2$. Taking μ_Z fixed, we find that the squared variance $\left(\frac{\sigma_Z}{\mu_Z}\right)^2 = 1 + \int_0^T \left(\frac{Z(s)}{\mu_Z}\right)^2 ds$ is a well-known, concentration measure on ***U₀,μ***.

4.4. The Theil measure.

We define the (generalized) Theil measure for Z in ***U₀*** as

$$Th_g(Z) = \int_0^T Z(s)\, ln(Z(s))ds \quad (8)$$

Clearly, Th_g is increasing in Z. Next, we observe that Th(Z), the analytical Theil concentration measure, defined as

$$Th(Z) = \frac{1}{T}\int_0^T \frac{Z(s)}{\mu_Z} ln\left(\frac{Z(s)}{\mu_Z}\right) ds \quad (9)$$

is equal to

$$\frac{1}{T\,\mu_Z}\int_0^T Z(s)ln\left(\frac{Z(s)}{\mu_Z}\right)ds = \frac{1}{T\,\mu_Z}\left(\int_0^T Z(s)ln(Z(s))\,ds - \int_0^T Z(s)ln(\mu_Z)\,ds\right)$$

$$= \frac{1}{T\,\mu_Z}\left(\int_0^T Z(s)ln\big(Z(s)\big)\,ds - T\,\mu_Z\,ln(\mu_Z)\right)$$

$$= \frac{1}{T\,\mu_Z}\int_0^T Z(s)ln(Z(s))\,ds - ln(\mu_Z) = \frac{Th_g(Z(s))}{T\,\mu_Z} - ln(\mu_Z).$$

Consequently: $Th_g(Z) = T.\mu_Z.\ (Th(Z) + ln(\mu_Z))$. Hence for $\mu = \mu_Z$ constant, Th_g is a strictly increasing function of the concentration measure Th. Using the corollary above, we conclude that Th_g is a global impact measure.

Alternatively, Th_g is a global impact measure because the function $s \to s\ln(s)$ is, for s > 0, strictly convex, continuous, and increasing.

5. An application

It is clear that, for all Y ∈ ***U₀*** and for x ∈ [0,T],

$$I_Y(x) = A\,L_Y\left(\frac{x}{T}\right) = \mu\,T\,L_Y\left(\frac{x}{T}\right) \qquad (10)$$

where A is the total production of all sources and μ is the average of Y over [0,T]. We will now derive an explicit formula for I_Y and L_Y in the case that Y is the Zipf curve Z.

The Zipf curve Z on]0,T] is defined as

$$Z(x) = \left(\frac{T}{x}\right)^{\beta} \qquad (11)$$

with 0 < β < 1 (Egghe, 2005) and 1/μ = 1-β (Egghe, 2005, p. 201).

The non-normalized Lorenz curve of the Z curve on [0,T], denoted as $I_Z(x)$, is:

$$I_Z(x) = T^{\beta}\int_0^x s^{-\beta}ds = (\beta \neq 1)\,T^{\beta}\left[\frac{s^{-\beta+1}}{-\beta+1}\right]_0^x = (0<\beta<1)\,\frac{T^{\beta}}{1-\beta}\,x^{1-\beta} \qquad (12)$$

Hence, $I_Z(x) = \mu\,T^{\beta}x^{\frac{1}{\mu}}$ (13)

and by (10) we have: $\mu\,T\,L_Z\left(\frac{x}{T}\right) = \mu\,T^{\beta}x^{\,1/\mu}$,

$$\text{or } L_Z\left(\frac{x}{T}\right) = \left(\frac{x}{T}\right)^{1/\mu} \qquad (14)$$

We note that formula (14) has been shown in (Egghe, 2005, p. 201) be it approximately.

6. Conclusion

We investigated a continuous theory of domination, leading to so-called global impact measures. In this context, we consider the notion of impact, as e.g., shown by a citation curve (articles ranked according to the number of received citations) as a combination of the notions of inequality (concentration) and productivity. In this way, our article extends earlier approaches, in which we mainly focused on the high producers (Egghe, 2021; Egghe & Rousseau, 2022a). This article belongs to a series of investigations studying the notion of impact, the main ideas of which were summarized in (Egghe & Rousseau, 2022b).

Equality or evenness, the opposite of inequality or concentration, plays a key role in studies of interdisciplinarity or, more generally, diversity studies (Wagner et al., 2011; Rousseau et al., 2019). Of course, concentration and diversity are also essential notions in other fields such as economics and ecology, and have widespread implications. For this reason, we expect that global impact measures, will lead to new opportunities for studies in informetrics and, more generally, the science of science.

Acknowledgment. The authors thank Li Li (National Science Library, CAS) for providing excellent figures and an anonymous reviewer for some good suggestions to improve the presentation of our results.

Conflict of interest. Leo Egghe and Ronald Rousseau belong to the distinguished reviewers board of the journal *Scientometrics* as Price Medal awardees.

Funding. No funding has been received for this investigation

Author contributions:

Leo Egghe: conceptualization, formal analysis, investigation, methodology, writing-original draft, writing-review and editing.
Ronald Rousseau: investigation, validation, writing-review and editing.

References

Auerbach, F. (1913). Das Gesetz der Bevölkerungskonzentration. *Petermanns Mitteilungen*, *59*(1), 74-76.

Bradford, S.C. (1934). Sources of information on specific subjects. *Engineering*, *137*, 85-86; reprinted in *Journal of Information Science*, *10*(4), 176-180 (1985).

Dagum, C. (1980). The generation and distribution of income, the Lorenz curve and the Gini ratio. *Economie Appliquée*, *33*(2), 327-367.

Egghe, L. (2005). *Power Laws in the Information Production Process: Lotkaian Informetrics.* Amsterdam: Elsevier. ISBN: 0-12-088753-3.

Egghe, L. (2021). Impact measures: what are they? *Scientometrics*, *127*(1), 385-406.

Egghe, L., & Rousseau, R. (2002). A proposal to define a core of a scientific subject: A definition using concentration and fuzzy sets. *Scientometrics, 54*(1), 51-62.

Egghe, L., & Rousseau, R. (2022a) Rank-frequency data and impact in a continuous model: Introducing impact bundles. *Journal of Informetrics*, *16*(3), 101297.

Egghe, L. & Rousseau, R. (2022b). The logic behind the mathematical treatment of impact. *ISSI Newsletter*, #69, *18*(1), 16-21.

Egghe, L., Rousseau, R., & Rousseau, S. (2007). TOP-curves. *Journal of the American Society for Information Science and Technology, 58*(6),777–785.

Faloutsos, M., Faloutsos, P. & Faloutsos, C. (1999). On power-law relationships of the Internet topology. In: *Applications, Technologies, Architectures, and Protocols for Computer Communication. Proceedings of the Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication* (pp. 251-262). New York (NY): ACM.

Laherrère, J. & Sornette, D. (1998). Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales. T*he European Physical Journal B - Condensed Matter and Complex Systems*, *2*(4), 525–539.

Lorenz, M. O. (1905). Methods of measuring concentration of wealth. *Publications of the American Statistical Association*, *9*, 209-219.

Lotka, A. J. (1926). The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, *16*(12), 317-323.

Marshall, A.W., Olkin, I. & Arnold, B.C. (2011). *Inequalities: Theory of majorization and its applications.* Springer, New York.

Newman, M.E.J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*(5), 323-351.

Pareto, V. (1895). La legge della domanda. *Giornale degli Economisti* (2nd series) *10*, 59-68.

Roberts, F.S. (1979). *Measurement Theory.* Addison-Wesley, Reading (MA).

Rousseau, R. (1997b). Sitations: an exploratory study. *Cybermetrics, 1*(1). https://digital.csic.es/bitstream/10261/170110/1/v1i1p1.pdf.

Rousseau, R., Egghe, L., & Guns, R. (2018). *Becoming Metric-Wise. A bibliometric guide for researchers*. Kidlington: Chandos-Elsevier.

Rousseau, R., Zhang, L. and Hu, XJ. (2019). Knowledge integration: Its meaning and measurement. In: W. Glänzel, H.F. Moed, U. Schmoch, M. Thelwall (Eds.), *Springer Handbook of Science and Technology Indicators* (pp. 69-94). Cham: Springer. ISBN: 978-3-030-02510-6

Salpeter, E. (1955). The luminosity function and stellar evolution. *Astrophysical Journal, 121*, 161-167.

Saravia, L.A., Doyle, S.R., & Bond-Lamberty, B. (2018). Power laws and critical fragmentation in global forests. *Scientific Reports, 8*, 17766; DOI:10.1038/s41598-018-36120-w

Wagner, C.S., Roessner, J.D., Bobb, K., Klein, J.T., Boyack, K.W., Keyton, J., Rafols, I. and Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, *5*(1), 14-26.

Zipf, G.K. (1941). *National Unity and Disunity; the nation as a bio-social organism.* Bloomington (IN), Principia Press.

Zipf, G.K. (1949). *Human Behavior and the Principle of Least Effort.* Addison-Wesley, Cambridge (MA); reprinted: 1965: New York: Hafner.