# Loet Leydesdorff, interdisciplinarity, and diversity

**Ronald Rousseau**

**Nota**: Este artículo se puede leer en español en:
*https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/87483*

Recommended citation:

**Rousseau, Ronald** (2023). "Loet Leydesdorff, interdisciplinarity, and diversity". *Profesional de la información*, v. 32, n. 7, e320702.

*https://doi.org/10.3145/epi.2023.dic.02*

Article received on May 9th 2023
Approved on June 18th 2023

**Ronald Rousseau**
*https://orcid.org/0000-0002-3252-2538*
*KU Leuven, MSI, Facultair Onderzoekscentrum ECOOM*
Naamsestraat 61, 3000 Leuven, Belgium
*ronald.rousseau@kuleuven.be*
*University of Antwerp, Faculty of Social Sciences*
Middelheimlaan 1, 2020 Antwerp, Belgium
*ronald.rousseau@uantwerpen.be*

**Abstract**

Diversity, as used in interdisciplinarity studies, has three components: variety, evenness, and dissimilarity. In 2019, Leydesdorff, Wagner, and Bornmann proposed an indicator, denoted as DIV*, that independently operationalized these three components and then combined them. Gini evenness is one factor in this formula. An important point is that Leydesdorff and his colleagues rejected so-called dual concepts, i.e. concepts that mix or are influenced by at least two of the three basic components of diversity. A few years ago Chao and Ricotta took a new look at "evenness" and showed that the Gini evenness measure, as well as the Lorenz curve, are dual concepts as they are influenced by variety. For this reason, I propose to replace the Gini evenness measure in DIV* with an evenness measure, actually an evenness profile, that is not influenced by variety.

**Keywords**

Evenness profiles; Diversity; Gini index; Interdisciplinarity; Bibliometrícs; Indicators; Science of science; Loet Leydesdorff.

## 1. Introduction

Interdisciplinarity is a hot item in bibliometrics and the science of science. The term interdisciplinarity itself leads to three important questions: What is a discipline? How can interdisciplinarity be measured? How can diversity in disciplines be measured?

Loet Leydesdorff had a very broad interest and studied, if I may slightly exaggerate, every aspect of the science of science, including interdisciplinarity. Yet, here I focus on measuring diversity and his contribution to this concept. How best to study interdisciplinarity is, in my opinion, still an open question, but measuring interdisciplinarity is often operationalized by studying diversity: usually the diversity of the references in articles (**Rousseau** *et al.*, 2019), but sometimes also the diversity of the fields to which authors belong (**Abramo** *et al.*, 2018). Another approach includes cohesion of references as an aspect of interdisciplinarity, besides diversity of references. This suggestion comes from **Ràfols** and **Meyer** (2010). Here I just mention this valid suggestion without going into details. Applications including cohesion in interdisciplinarity studies can be found in **Ràfols** and **Meyer** (2010), **Ràfols** (2014), and **Rousseau** *et al.* (2019).

## 2. What is diversity?

Since **Stirling** (2007) bibliometricians and many other colleagues, are convinced that diversity has three components: variety, evenness, and dissimilarity. The problem now is threefold: how to define these three components, how to measure them, and how to combine them?

Diversity has three components: variety, evenness, and dissimilarity. The problem now is threefold: how to define these three components, how to measure them, and how to combine them?

I first introduce some notation. Assume that one has a situation with $N \in \mathbb{N}$, $N > 1$, categories or cells. For the moment these are either theoretical categories, which may be empty in a given case (such as the categories in the *Web of Science* in a study of the publications of a university department), or categories that are actually observed (such as butterflies on the university campus). An observation is an array $X = (x_1, x_2, \ldots, x_N)$, where $x_j$ denotes the number of items in category j, j=1, …, N. Depending on the study $x_j > 0$ or $x_j \geq 0$. I always assume that not all $x_j = 0$.

If N is given, the proportion of items in category j is denoted as , $p_j = \frac{x_j}{N}$ otherwise, it is , $p_j = \frac{x_j}{n_X}$ where $n_x$ is the number of non-empty cells.

The normalized dissimilarity (however measured) between categories i and j is denoted as $d_{ij} = d_{ji}$, with $0 < d_{ij} \leq 1$. The corresponding similarity between categories i and j is $s_{ij} = 1 - d_{ij}$.

I now come to the definition of the three components of diversity and the problem of how to combine them without losing validity or information in each of them.

Variety is simply the number of non-empty cells, denoted as $n_x$.

Evenness or balance (I use these two words as synonyms) may be described as the relative apportionment of abundances among categories (actually present, or assumed to be possibly present). It is a function of the pattern of the assignment of items across categories (**Rousseau** *et al*., 2019, p. 312). The problem, discussed in this article, is how to relate this description in words to an acceptable mathematical formula.

**Ràfols** and **Meyer** (2010) propose the Rao-Stirling (in short: RS) measure as a measure of diversity. This measure is defined as :

$$RS(X) = \sum_{i,j=1}^{N} d_{ij}^{\alpha} \left(p_i p_j\right)^{\beta} \qquad (1)$$

where in practice they propose to take $\alpha = \beta = 1$. Inspired by the ideas of **Jost** (2009) related to so-called "true diversity", **Leinster** and **Cobbold** (2012) propose the following diversity profile (not just one value, but a whole range of values with parameter *q*), where this parameter ranges from 0 to infinity (the cases q = 1 and q = ∞ are obtained as limits).

$${}^{q}D\ (X) = \left(\sum_{i=1}^{N} p_i \left(\sum_{j=1}^{N} (1 - d_{ij}) p_j\right)^{q-1}\right)^{1/(1-q)} \qquad (2)$$

The case q = 2 is related to the RS-measure:

$${}^{2}D\ (X) = \left(\sum_{i=1}^{N} p_i \left(\sum_{j=1}^{N} (1 - d_{ij}) p_j\right)^{1}\right)^{-1)} = \frac{1}{1 - RS(X)} \qquad (3)$$

This diversity was suggested by **Zhang** *et al.* (2016) for applications in interdisciplinarity studies.

From now on I mainly follow the reasoning in **Leydesdorff** *et al.* (2019a), complemented by my comments.

In **Leydesdorff** *et al.* (2019a) the authors proposed to modify the Rao-Stirling diversity measure into a new indicator (DIV) that independently operationalizes "variety," "balance," and "disparity" and then combines them ex-post. These authors note that in the Rao-Stirling diversity, two of the three components, namely variety, and balance, are combined in the definitions (ex-ante) using the repeat measure (i.e., the Hirschmann-Simpson-Herfindahl measure), see **Rousseau** (2018) for the reason why I prefer the term repeat measure. Leydesdorff, Wagner, and Bornmann refer to such combinations of variety and evenness as dual-concept diversity.

The following requirement seems natural. When two given components are held constant then an increase in the third component would lead to an increase in diversity. This has been called "the monotonicity" requirement by **Rousseau** (2018): diversity must increase for each of the three components when the other two remain the same.

**Rousseau** (2018) provided a counter-example, showing that the Rao-Stirling diversity does not meet this monotonicity requirement. It is, indeed, possible that for given variety and disparity, the diversity does not increase monotonically with balance. The same conclusion holds equally for the "true diversity" variant of Rao-Stirling diversity introduced by **Zhang** *et al.* (2016).

"Variety" can be independently operationalized, as the number of observed categories, $n_X$, or as relative variety (bounded between zero and one) as $n_X$/ N, with N being the total number of classes available. The notion of "balance" can be operationalized using the Gini coefficient without "co-mingling" it with "variety", as claimed in (**Nijssen** *et al.*, 1998). Since the classical Gini (concentration) coefficient is maximally diverse for Gini = 0 and fully homogeneous for Gini tending to 1, **Leydesdorff** *et al.* used (1 – Gini) so that one obtains a diversity measure with three factors for each unit of analysis X. The formula they proposed reads as follows:

$$DIV(X) = \frac{n_X}{N} * (1 - Gini) * \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{n_X} \frac{d_{ij}}{n_X(n_X - 1)}\right) \qquad (4)$$

The right-most factor in this equation is similar to the disparity measure used in the case of Rao-Stirling diversity. The two other factors, however, represent relative variety as $n_x$ / N, with N being the total number of classes available, and balance measured as the Gini evenness index (namely, one minus the Gini concentration coefficient). The authors further note that variety and disparity have to be normalized so that all terms are bounded between zero and one.

Not going into the essence of the Leydesdorff-Wagner-Bornmann argument, which I think is rejecting dual concepts, **Rousseau** (2019) made three objections against the (DIV) formula. The first was about the use of the total number of categories in (DIV). This excludes cases where N is not known, such as is often the case in biological observations. The second was the fact that the third component in (DIV) only takes the total sum of all $d_{ij}$ into account: specific $d_{ij}$ values do not play a role. Finally, the third objection refers to the normalization of (DIV). Because of this normalization (DIV) cannot be a 'true' diversity measure in the sense of **Jost** (2009). Recall that a "true" diversity must have the value N if one studies a community of N equally abundant, totally dissimilar items. The point here is that if a measure is not a "true" diversity one cannot discuss diversity in terms of percentage growth or decline. As a reply, **Leydesdorff** *et al.* (2019b) adapted their formula (DIV) to the following formula (DIV*):

$$DIV^*(X) = n_X * (1 - Gini) * \left( \sum_{\substack{i,j=1 \\ i \neq j}}^{n_X} d_{ij} \right) \qquad (5)$$

For the further developments of this article, I note the important point that **Leydesdorff** *et al.* (2019a,b) followed the arguments I gave in **Nijssen** *et al*. (1998), namely that the Lorenz curve, and hence the Gini evenness index is a perfect representation of evenness. In that article I followed the ideas of **Taillie** (1979), and was, of course, convinced that this was true. My main point in (**Nijssen** *et al*., 1998) was that I showed that the Gini evenness index and one over the coefficient of variation respected the Lorenz curve order. I moreover provided new variants of the Shannon and the Simpson index that also respected this order, and hence, were considered to be acceptable measures of evenness.

## 3. Recent developments related to the concept of evenness

A few years ago **Chao** and **Ricotta** (2019) took a new look at the notion of evenness. When variety and abundances possibly vary they state two requirements:

> Requirement A. This is the unrelatedness criterion, which states that the range of values that an evenness measure can take should be a fixed interval, regardless of species richness or abundances.
>
> Requirement B. This is scale invariance, which states that any evenness measure should not be affected by the units used. In particular evenness for raw data and for relative abundances should be the same.

The unrelatedness criterion clearly fits into the **Leydesdorff** *et al.* (2019a) framework of rejecting dual concepts. As evenness should take values on a fixed interval, one may agree to use the interval [0,1].

The point now is that the Gini evenness index and any measure respecting the Lorenz order do not satisfy the unrelatedness criterion (requirement a). Indeed, when N = 2, then the Gini evenness index takes values between 1 / 2 and 1; and generally the Gini evenness index takes values between 1/N and 1. This shows that the Lorenz curve is not a perfect representation of evenness as it depends on variety. Another consequence is that the requirement of replication invariance, which originates from **Dalton** (1920) and which states that e.g., the evenness of (x,y,z) is the same as the evenness of (x,x,x,y,y,y,z,z,z), is not a proper requirement for evenness.

I note here a subtle point: it is the range of the evenness values that may not depend on N. One cannot avoid using a measure that depends on N. Indeed, the formula to calculate the Gini index clearly depends on N, and that observation will hold for all measures that will be suggested to replace the Gini index.

## 4. A proper evenness measure

Until now I have disregarded the influence of q on the sensitivity of evenness (being more or less sensitive to highly abundant or rare classes). **Chao** and **Ricotta** (2019) provide arguments to reject measures derived from distance functions and, instead use divergence measures. In the next step, they consider five classes of divergence measures. These classes and some specific cases are given in Table 1 of their article (**Chao**; **Ricotta**, 2019) to which I refer the interested reader. I do not repeat the whole table, admitting that a study of these different profiles and the specific differences between them, would be very interesting. I just show here the case originating from **Jost** (2010), denoted as $E_3$ in the Chao-Ricotta article. This evenness profile is defined as:

$${}^{q}E_3\,(X) = \frac{\left(\sum_{j=1}^{N} p_j^q\right)^{1/(1-q)} - 1}{N-1} \qquad (6)$$

Here N, the number of all possible categories in the situation under study, is assumed to be known. Otherwise, N must be replaced by $n_X$. The case q=2, originating from **Kvålseth** (1991) is:

$${}^{2}E_3\,(X) = \frac{\left(\sum_{j=1}^{N} p_j^2\right)^{-1} - 1}{N-1} \qquad (7)$$

If one is not interested in a complete profile, there exists a simple solution to change formula (5), namely to replace the Gini concentration measure with the so-called **Pratt** (1977) concentration measure. Indeed, Pratt's measure is equal to N/(N-1) times the Gini concentration measure. Hence when the Gini measure is zero, also the Pratt measure is zero, and when the Gini measure is (N-1)/N, the Pratt measure is one. Moreover, there also exists a generalized Pratt measure (with a parameter r > 0), introduced in (**Egghe**; **Rousseau**, 1990). Changing this measure slightly and moving to the diversity variant leads to a profile of evenness measures:

$$1 - \left(\frac{1}{2(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} |p_i - p_j|^r\right)^{1/r} \qquad (8)$$

## 5. Conclusion

Following Chao and Ricotta one has five times an infinite number (one for each q) of independent (=unrelated) evenness measures, correcting the Gini diversity measure (and one may add the evenness variant of the generalized Pratt measure). Taking $E_3$ and the simple case q =2 (Kvålseth-Jost) leads to:

$$DIV^{**}(X) = n_X * \frac{\left(\sum_{j=1}^{N} p_j^2\right)^{-1} - 1}{N-1} * \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{n_X} d_{ij}\right) \qquad (9)$$

or (9') with N replaced by $n_X$, depending on the aim of the study. As there are no dual concepts the monotonicity requirement is satisfied in all cases.

Note that formula (9) is just an example. For the moment I have no preference, except that following **Leinster** and **Cobbold** (2012), and **Chao** and **Ricotta** (2019), it is better to consider a profile (all q values) instead of a single value of q.

In DIV, DIV*, and DIV** the three components are given an equal weight. That is not a necessity and for given $\alpha, \beta, \gamma > 0$ one could define

$$DIV^{**}(X) = (n_X)^{\alpha} * \left(\frac{\left(\sum_{j=1}^{N} p_j^q\right)^{1/(1-q)} - 1}{N-1}\right)^{\beta} * \left(\sum_{\substack{i,j=1 \\ i \neq j}}^{n_X} d_{ij}\right)^{\gamma} \qquad (10)$$

Yet, I do not see a good reason to complicate matters even more and prefer the case $\alpha = \beta = \gamma = 1$.

In this article, I focused on Leydesdorff's approach and propose a correction to my own work and the Leydesdorff-Wagner-Bornmann suggestion. Yet, I do not claim to have the ultimate solution for measuring diversity in the context of interdisciplinarity. For the moment I propose DIV** (formula (9)) (and variants) as the better diversity measure to be used in interdisciplinarity studies and look forward to further developments.

## 6. References

**Abramo, Giovanni**; **D'Angelo, Ciriaco A.**; **Zhang, Lin** (2018). "A comparison of two approaches for measuring interdisciplinary research output: The disciplinary diversity of authors vs. the disciplinary diversity of the reference list". *Journal of informetrics*, v. 12, n. 4, pp. 1182-1193.
*https://doi.org/10.1016/j.joi.2018.09.001*

**Chao, Anne**; **Ricotta, Carlo** (2019). "Quantifying evenness and linking it to diversity, beta diversity, and similarity". *Ecology*, v. 100, n. 12, pp. article e02852.
*https://doi.org/10.1002/ecy.2852*

**Dalton, Hugh** (1920). "The measurement of the inequality of incomes". *The economic journal*, v. 30, n. 119, pp. 348-361.
*https://doi.org/10.2307/2223525*

**Egghe, Leo**; **Rousseau, Ronald** (1990). "Elements of concentration theory". In: Egghe & Rousseau (eds.). *Informetrics 89/90*, pp. 97-137. Amsterdam: Elsevier.

**Jost, Lou** (2009). "Mismeasuring biological diversity: Response to Hoffmann and Hoffmann (2008)". *Ecological economics*, v. 68, n. 4, pp. 925-928.
*https://doi.org/10.1016/j.ecolecon.2008.10.015*

**Jost, Lou** (2010). "The relation between evenness and diversity". *Diversity*, v. 2, pp. 207-232.
*https://doi.org/10.3390/d2020207*

**Kvålseth, Tarald O.** (1991). "Note on biological diversity, evenness, and homogeneity". *Oikos*, v. 62, pp. 123-127.
*https://doi.org/10.2307/3545460*

**Leinster, Tom**; **Cobbold, Christina A.** (2012). "Measuring diversity: The importance of species similarity". *Ecology*, v. 93, n. 3, pp. 477-489.
*https://doi.org/10.1890/10-2402.1*

**Leydesdorff, Loet**; **Wagner, Caroline S.**; **Bornmann, Lutz** (2019a). "Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient". *Journal of informetrics*, v. 13, n. 3, pp. 255-269.
*https://doi.org/10.1016/j.joi.2018.12.006*

**Leydesdorff, Loet**; **Wagner, Caroline S.**; **Bornmann, Lutz** (2019b). "Diversity measurement: Steps towards the measurement of interdisciplinarity?". *Journal of informetrics*, v. 13, n. 3, pp. 904-905.
*https://doi.org/10.1016/j.joi.2019.03.016*

**Nijssen, David**; **Rousseau, Ronald**; **Van Hecke, Piet** (1998). "The Lorenz curve: A graphical representation of evenness". *Coenoses*, v. 13, n. 1, pp. 33-38.
*https://www.jstor.org/stable/43461212*

**Pratt, Allan D.** (1977). "A measure of class concentration in bibliometrics". *Journal of the American Society for Information Science*, v. 28, n. 5, pp. 285-292.
*https://doi.org/10.1002/asi.4630280508*

**Ràfols, Ismael** (2014). "Knowledge integration and diffusion: measures and mapping of diversity and coherence". In: *Measuring scholarly impact. methods and practice*. Ding, Ying; Rousseau, Ronald; Wolfram, Dietmar (eds.). Cham: Springer, pp. 169-190. ISBN: 978 3 319 10377 8
*https://doi.org/10.1007/978-3-319-10377-8*

**Ràfols, Ismael**; **Meyer, Martin** (2010). "Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience". *Scientometrics*, v. 82, n. 2, pp. 263-287.
*https://doi.org/10.1007/s11192-009-0041-y*

**Rousseau, Ronald** (2018). "The repeat rate: From Hirschman to Stirling". *Scientometrics*, v. 116, n. 1, pp. 645-653.
*https://doi.org/10.1007/s11192-018-2724-8*

**Rousseau, Ronald** (2019). "On the Leydesdorff-Wagner-Bornmann proposal for diversity measurement". *Journal of informetrics*, v. 13, n. 3, pp. 906-907.
*https://doi.org/10.1016/j.joi.2019.03.015*

**Rousseau, Ronald**; **Zhang, Lin**; **Hu, Xiaojun** (2019). "Knowledge integration: its meaning and measurement". In: *Springer handbook of science and technology indicators*. W. Glänzel; H. F. Moed; U. Schmoch; M. Thelwall (eds.), pp. 69-94.
*https://doi.org/10.1007/978-3-030-02511-3_3*

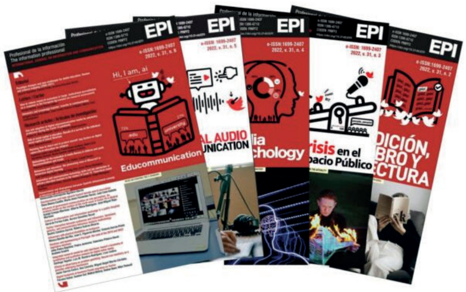
**Stirling, Andy** (2007). "A general framework for analysing diversity in science, technology and society". *Journal of the Royal Society Interface*, v. 4, n. 15, pp. 707-719.
*https://doi.org/10.1098/rsif.2007.0213*

**Taillie, C.** (1979). "Species equitability: a comparative approach". In: Grassle, Patil, Smith & Taillie (eds.). *Ecological diversity in theory and practice.* Fairland (MD): International Co-operative Publishing House, pp. 51-62. ISBN: 978 0 899740034

**Zhang, Lin**; **Rousseau, Ronald**; **Glänzel, Wolfgang** (2016). "Diversity of references as an indicator for interdisciplinarity of journals: Taking similarity between subject fields into account". *Journal of the Association for Information Science and Technology*, v. 67, n. 5, pp. 1257-1265.
*https://doi.org/10.1002/asi.23487*

REVISTA EPI

*https://www.profesionaldelainformacion.com*



***Profesional de la información*** es una revista científica sobre Comunicación, Documentación, Biblioteconomía, Informetría y Ciencias de la información. Se publica cada 2 meses (6 números al año).

e-ISSN: 1699-2407
*https://doi.org/10.3145/EPI*
SJR 2022 = 0,872 (Q1); JIF 2022 = 4,2 (Q1)