Loet Leydesdorff, interdisciplinariedad y diversidad

Loet Leydesdorff, interdisciplinarity, and diversity

Ronald Rousseau

Note: This article can be read in its English original version on:
https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/87483

Cómo citar este artículo.
Este artículo es una traducción. Por favor cite el original inglés:

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

Artículo recibido el 09-05-2023
Aceptación definitiva el 18-06-2023

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

Resumen

La diversidad, tal como se utiliza en los estudios de interdisciplinariedad, tiene tres componentes: variedad, igualdad y disimilitud. En 2019, Leydesdorff, Wagner y Bornmann propusieron un indicador, denominado DIV*, que operaba de forma independiente estos tres componentes y luego los combinaba. La uniformidad de Gini es un factor en esta fórmula. Un punto importante es que Leydesdorff y sus colegas rechazaron los llamados conceptos duales, es decir, conceptos que mezclan o están influenciados por al menos dos de los tres componentes básicos de la diversidad. Hace unos años, Chao y Ricotta dieron una nueva mirada a la "uniformidad" y demostraron que la medida de uniformidad de Gini, así como la curva de Lorenz, son conceptos duales, ya que están influenciados por la variedad. Por esta razón, propongo reemplazar la medida de uniformidad de Gini en DIV* con una medida de uniformidad, en realidad un perfil de uniformidad, que no está influenciado por la variedad.

Palabras clave

Perfiles de uniformidad; Diversidad; Índice de Gini; Interdisciplinariedad; Bibliometría; Indicadores; Ciencia de la ciencia; Loet Leydesdorff.

Abstract

Diversity, as used in interdisciplinarity studies, has three components: variety, evenness, and dissimilarity. In 2019, Leydesdorff, Wagner, and Bornmann proposed an indicator, denoted 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; Bibliometrics; Indicators; Science of science; Loet Leydesdorff.

1. Introducción

La interdisciplinariedad es un tema candente en la bibliometría y la ciencia de la ciencia. El término interdisciplinariedad en sí lleva a tres preguntas importantes: ¿Qué es una disciplina? ¿Cómo se puede medir la interdisciplinariedad? ¿Cómo se puede medir la diversidad en las disciplinas?

Loet Leydesdorff tenía un interés muy amplio y estudió, si puedo exagerar un poco, todos los aspectos de la ciencia de la ciencia, incluida la interdisciplinariedad. Sin embargo, aquí me enfoco en medir la diversidad y la contribución de Loet a este concepto. En mi opinión, la mejor manera de estudiar la interdisciplinariedad sigue siendo una pregunta abierta, pero la medición de la interdisciplinariedad a menudo se operacionaliza mediante el estudio de la diversidad: generalmente la diversidad de las referencias en los artículos (**Rousseau** *et al.*, 2019), pero a veces también la diversidad de los campos a los que pertenecen los autores (**Abramo** *et al.*, 2018). Otro enfoque incluye la cohesión de referencias como un aspecto de la interdisciplinariedad, además de la diversidad de referencias. Esta sugerencia proviene de **Ràfols** y **Meyer** (2010). Aquí solo menciono esta sugerencia válida sin entrar en detalles. Las aplicaciones que incluyen la cohesión en estudios interdisciplinarios se pueden encontrar en **Ràfols** y **Meyer** (2010), **Ràfols** (2014) y **Rousseau** *et al.* (2019).

2. ¿Qué es la diversidad?

Desde **Stirling** (2007) los bibliómetras y muchos otros colegas están convencidos de que la diversidad tiene tres componentes: variedad, uniformidad y disimilitud. El problema ahora es triple: ¿cómo definir estos tres componentes, cómo medirlos y cómo combinarlos?

> La diversidad tiene tres componentes: variedad, uniformidad y disimilitud. El problema es triple: ¿cómo definir estos tres componentes, cómo medirlos y cómo combinarlos?

Primero presento algo de notación. Suponga que uno tiene una situación con N ∈ ℕ, N > 1, categorías o celdas. Por el momento se trata o bien de categorías teóricas, que pueden estar vacías en un caso dado (como las categorías de la *Web of Science* en un estudio de las publicaciones de un departamento universitario), o de categorías realmente observadas (como las mariposas en el campus universitario). Una observación es una matriz $X = (x_1, x_2, \ldots, x_N)$, donde x_j denota el número de elementos en la categoría j, j=1, …, N. Dependiendo del estudio, $x_j > 0$ o $x_j \geq 0$. Siempre asumo que no todos los $x_j = 0$.

Si se proporciona N, la proporción de elementos en la categoría j se denota como $p_j = \frac{x_j}{N}$, de lo contrario, es $p_j = \frac{x_j}{n_X}$, donde n_x es el número de celdas no vacías.

La disimilitud normalizada (como sea que se mida) entre las categorías *i* y *j* se denota como $d_{ij} = d_{ji}$, con $0 < d_{ij} \leq 1$. La similitud correspondiente entre las categorías *i* y *j* es $s_{ij} = 1 - d_{ij}$.

Paso ahora a la definición de los tres componentes de la diversidad y al problema de cómo combinarlos sin perder validez ni información en cada uno de ellos.

La variedad es simplemente el número de celdas no vacías, indicadas como n_x.

La uniformidad o el equilibrio (utilizo estas dos palabras como sinónimos) pueden describirse como la distribución relativa de abundancias entre categorías (realmente presentes o que se supone que posiblemente estén presentes). Es una función del patrón de asignación de ítems a través de categorías (**Rousseau** *et al.*, 2019, p. 312). El problema, discutido en este artículo, es cómo relacionar esta descripción en palabras con una fórmula matemática aceptable.

Ràfols y **Meyer** (2010) proponen la medida de Rao-Stirling (resumido: RS) como medida de diversidad. Esta medida se define como:

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

donde en la práctica se proponen tomar $\alpha = \beta = 1$. Inspirándose en las ideas de **Jost** (2009) relacionadas con la llamada "diversidad verdadera", **Leinster** y **Cobbold** (2012) proponen el siguiente perfil de diversidad (no solo un valor, sino todo un rango de valores con parámetro *q*), donde este parámetro va de 0 a infinito (se obtienen como límites los casos q = 1 y q = ∞).

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

El caso q = 2 está relacionado con la medida RS:

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

Esta diversidad fue sugerida por **Zhang** *et al.* (2016) para aplicaciones en estudios interdisciplinarios.

A partir de ahora, sigo principalmente el razonamiento de **Leydesdorff** *et al.* (2019a), complementado con mis comentarios.

En **Leydesdorff** *et al.* (2019a) los autores propusieron modificar la medida de diversidad de Rao-Stirling en un nuevo indicador (DIV) que operacionalice de forma independiente la "variedad", el "equilibrio" y la "disparidad" y luego los combine ex-post. Estos autores señalan que en la diversidad de Rao-Stirling, dos de los tres componentes, a saber, la variedad y el equilibrio, se combinan en las definiciones (ex-ante) utilizando la medida repetida (es decir, la medida de Hirschmann-Simpson-Herfindahl), véase **Rousseau** (2018) por la razón por la que prefiero el término medida repetida. Leydesdorff, Wagner y Bornmann se refieren a tales combinaciones de variedad y uniformidad como diversidad de concepto dual.

El siguiente requisito parece natural. Cuando dos componentes dados se mantienen constantes, un aumento en el tercer componente conduciría a un aumento en la diversidad. **Rousseau** (2018) ha llamado a esto el requisito de "monotonicidad": la diversidad debe aumentar para cada uno de los tres componentes cuando los otros dos permanecen iguales.

Rousseau (2018) proporcionó un contraejemplo, mostrando que la diversidad de Rao-Stirling no cumple con este requisito de monotonicidad. De hecho, es posible que, dada la variedad y la disparidad, la diversidad no aumente monótonamente con el equilibrio. La misma conclusión es igualmente válida para la variante de "diversidad verdadera" de la diversidad de Rao-Stirling introducida por **Zhang** *et al.* (2016).

La "variedad" se puede operacionalizar de forma independiente, como el número de categorías observadas, $n_{X,}$ o como variedad relativa (limitada entre cero y uno) como n_X / N, siendo N el número total de clases disponibles. La noción de "equilibrio" se puede operacionalizar utilizando el coeficiente de Gini sin "mezclarlo" con "variedad", como se afirma en (**Nijssen** *et al.*, 1998). Dado que el coeficiente clásico de Gini (concentración) es máximamente diverso para Gini = 0 y completamente homogéneo para Gini que tiende a 1, **Leydesdorff** *et al.* (1 – Gini) de manera que se obtiene una medida de diversidad con tres factores para cada unidad de análisis X. La fórmula que propusieron es la siguiente:

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

El factor más a la derecha en esta ecuación es similar a la medida de disparidad utilizada en el caso de la diversidad de Rao-Stirling. Los otros dos factores, sin embargo, representan la variedad relativa como n_X / N , siendo N el número total de clases disponibles y el equilibrio medido como el índice de uniformidad de Gini (es decir, uno menos el coeficiente de concentración de Gini). Los autores señalan además que la variedad y la disparidad deben normalizarse para que todos los términos estén delimitados entre cero y uno.

Sin entrar en la esencia del argumento de Leydesdorff-Wagner-Bornmann, que creo que rechaza los conceptos duales, **Rousseau** (2019) hizo tres objeciones contra la fórmula (DIV). El primero fue sobre el uso del número total de categorías en (DIV). Esto excluye los casos en los que no se conoce N, como suele ser el caso en las observaciones biológicas. El segundo fue el hecho de que el tercer componente en (DIV) solo tiene en cuenta la suma total de todos los $d_{ij:}$ los valores específicos de d_{ij} no juegan ningún papel. Finalmente, la tercera objeción se refiere a la normalización de (DIV). Debido a esta normalización (DIV) no puede ser una medida de diversidad 'verdadera' en el sentido de **Jost** (2009). Recuérdese que una diversidad "verdadera" debe tener el valor N si se estudia una comunidad de N elementos igualmente abundantes, totalmente diferentes. El punto aquí es que, si una medida no es una diversidad "verdadera", no se puede discutir la diversidad en términos de porcentaje de crecimiento o disminución. Como respuesta, **Leydesdorff** *et al.* (2019b) adaptaron su fórmula (DIV) a la siguiente fórmula (DIV*):

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

Para el desarrollo posterior de este artículo, señalo el punto importante que **Leydesdorff** *et al.* (2019a; 2019b) siguió los argumentos que di en **Nijssen** *et al.* (1998), a saber, que la curva de Lorenz y, por tanto, el índice de uniformidad de Gini es una representación perfecta de la uniformidad. En ese artículo seguí las ideas de **Taillie** (1979), y estaba, por supuesto, convencido de que esto era cierto. Mi punto principal en (**Nijssen** *et al.*, 1998) fue que mostré que el índice de uniformidad de Gini y uno sobre el coeficiente de variación respetaban el orden de la curva de Lorenz. Además, proporcioné nuevas variantes del índice de Shannon y Simpson que también respetaban este orden y, por lo tanto, se consideraban medidas aceptables de uniformidad.

3. Desarrollos recientes relacionados con el concepto de uniformidad

Hace unos años, **Chao** y **Ricotta** (2019) dieron una nueva mirada a la noción de uniformidad. Cuando la variedad y la abundancia posiblemente varíen, establecen dos requisitos:

> Requisito A. Este es el criterio de falta de relación, que establece que el rango de valores que puede tomar una medida de uniformidad debe ser un intervalo fijo, independientemente de la riqueza o abundancia de especies.
>
> Requisito B. Esta es la invariancia de escala, que establece que cualquier medida de uniformidad no debe verse afectada por las unidades utilizadas. En particular, la uniformidad de los datos brutos y de las abundancias relativas debe ser la misma.

El criterio de falta de relación encaja claramente en el marco de rechazo de conceptos duales de **Leydesdorff** *et al.* (2019a). Como la uniformidad debe tomar valores en un intervalo fijo, se puede aceptar usar el intervalo [0,1].

El punto ahora es que el índice de uniformidad de Gini y cualquier medida que respete el orden de Lorenz no satisfacen el criterio de falta de relación (requisito a). En efecto, cuando N = 2, entonces el índice de equidad de Gini toma valores entre 1/2 y 1; y generalmente el índice de uniformidad de Gini toma valores entre 1/N y 1. Esto demuestra que la curva de Lorenz no es una representación perfecta de la uniformidad ya que depende de la variedad. Otra consecuencia es que el requisito de invariancia de replicación, que se origina en **Dalton** (1920) y que establece que, por ejemplo, la uniformidad de (x,y,z) es la misma que la uniformidad de (x,x,x,y,y,y,z,z,z), no es un requisito adecuado para la uniformidad.

Observo aquí un punto sutil: es el rango de los valores de uniformidad que pueden no depender de N. Uno no puede evitar usar una medida que depende de N. De hecho, la fórmula para calcular el índice de Gini claramente depende de N, y esa observación se mantendrá para todas las medidas que se sugieran para reemplazar el índice de Gini.

4. Una medida de uniformidad adecuada

Hasta ahora he despreciado la influencia de q sobre la sensibilidad de la uniformidad (siendo más o menos sensible a clases muy abundantes o raras). **Chao** y **Ricotta** (2019) aportan argumentos para rechazar medidas derivadas de funciones de distancia y, en su lugar, utilizar medidas de divergencia. En el siguiente paso, consideran cinco clases de medidas de divergencia. Estas clases y algunos casos específicos se dan en la tabla 1 de su artículo (**Chao**; **Ricotta**, 2019) al que remito al lector interesado. No repito toda la tabla, admitiendo que sería muy interesante un estudio de estos diferentes perfiles y las diferencias específicas entre ellos. Solo muestro aquí el caso que se origina en **Jost** (2010), señalado como E_3 en el artículo de Chao-Ricotta. Este perfil de uniformidad se define como:

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

Aquí se supone que se conoce N, el número de todas las categorías posibles en la situación en estudio. De lo contrario, N debe ser reemplazado por n_X. El caso q=2, con origen en **Kvålseth** (1991) es:

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

Si uno no está interesado en un perfil completo, existe una solución simple para cambiar la fórmula (5), a saber, reemplazar la medida de concentración de Gini con la llamada medida de concentración de **Pratt** (1977). De hecho, la medida de Pratt es igual a N/(N-1) veces la medida de concentración de Gini. Por tanto, cuando la medida de Gini es cero, la medida de Pratt también es cero, y cuando la medida de Gini es (N-1)/N, la medida de Pratt es uno. Además, también existe una medida de Pratt generalizada (con un parámetro r > 0), introducida en (**Egghe**; **Rousseau**, 1990). Cambiar esta medida ligeramente y pasar a la variante de diversidad conduce a un perfil de medidas uniformes:

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

5. Conclusión

Siguiendo a Chao y Ricotta, uno tiene cinco veces un número infinito (uno para cada q) de medidas de igualdad independientes (= no relacionadas), corrigiendo la medida de diversidad de Gini (y se puede agregar la variante de igualdad de la medida de Pratt generalizada). Tomando E_3 y el caso simple q=2 (Kvålseth-Jost) lleva a:

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

o (9’) con N reemplazada por n_X, dependiendo del objetivo del estudio. Como no hay conceptos duales, el requisito de monotonicidad se cumple en todos los casos.

Tenga en cuenta que la fórmula (9) es solo un ejemplo. Por el momento no tengo preferencia, excepto que siguiendo a **Leinster** y **Cobbold** (2012), y a **Chao** y **Ricotta** (2019), es mejor considerar un perfil (todos los valores de q) en lugar de un solo valor de q.

En DIV, DIV* y DIV**, los tres componentes tienen el mismo peso. Eso no es una necesidad y dado $\alpha, \beta, \gamma > 0$ uno podría definir

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

Sin embargo, no veo una buena razón para complicar aún más las cosas y prefiero el caso $\alpha = \beta = \gamma = 1$.

En este artículo, me concentré en el enfoque de Leydesdorff y propongo una corrección a mi propio trabajo y la sugerencia de Leydesdorff-Wagner-Bornmann. Sin embargo, no pretendo tener la solución definitiva para medir la diversidad en el contexto de la interdisciplinariedad. Por el momento propongo DIV** (fórmula (9) (y variantes) como la mejor medida de diversidad para ser utilizada en estudios interdisciplinarios y espero futuros desarrollos.

6. Referencias

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