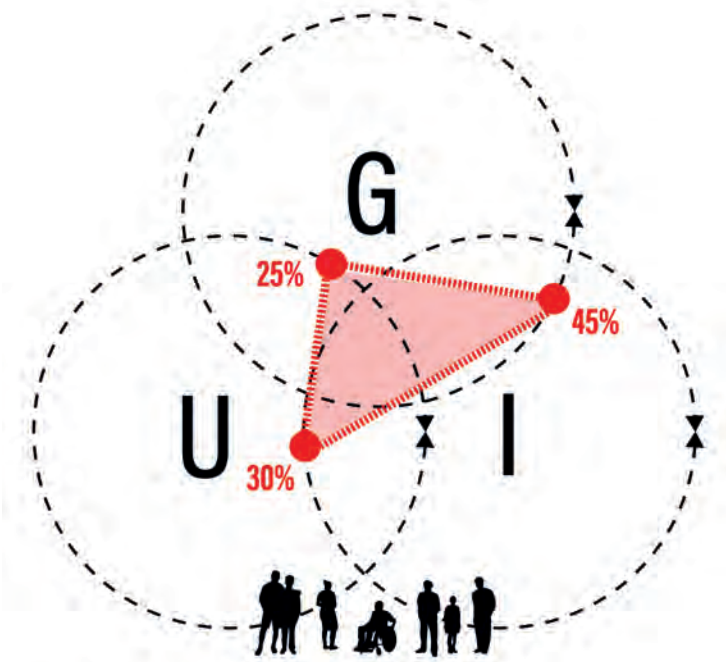
Profesional de la información
The information professional
INTERNATIONAL JOURNAL ON INFORMATION AND COMMUNICATION

e-ISSN 1699-2407
ISSN 1386-6710
CODEN: PINFF2
https://doi.org/10.3145/EPI

EPI

e-ISSN: 1699-2407
2023, v. 32, n. 7

Research articles



Tribute to Loet Leydesdorff

Profesional de la información

Bimonthly journal founded in 1992 by Tomàs Baiget and Francisca García-Sicilia

Profesional de la información is a journal published by Ediciones Profesionales de la Información S.L. VAT: ESB63664544
C/ Mistral, 36 (Cortijo del Aire)
18220 Albolote (Granada), Spain
Tel.: +34 - 608 491 521
http://www.profesionaldelainformacion.com

Cover design

Moisés Mañas
Universitat Politècnica de València
Foto de Wikipedia (https://commons.wikimedia.org/wiki/File:Loet_Leydesdorff.jpg)

Layout

Isabel Olea
isabel.iolea@gmail.com

Postal address

Profesional de la información
Apartado 13
24080 León, Spain
Tel.: +34 - 608 491 521
epi@elprofesionaldelainformacion.com

Advertising

Tel.: +34 - 608 491 521
publici@elprofesionaldelainformacion.com

Online services

Roxana Dinu: *roxadinu@gmail.com*

Online distribution

https://revista.profesionaldelainformacion.com/index.php/EPI

Legal deposit: B. 12.303-199

Papers published in EPI are approved according to the traditional double-blind "peer review" system: they are reviewed by at least two experts on the subject, from the journal's Advisory Board and/or external experts.

In order to ensure that the papers do not lose relevance, the management and reviewers of this journal make a special effort to review the articles very quickly, achieving an average acceptance or rejection time of only a few weeks.

Contents

EPI, 2023, v. 32, n. 7

Tribute to Loet Leydesdorff

RESEARCH ARTICLES

Profesional de la información is indexed in:

DATABASES

Academic search (Ebsco)
https://www.ebsco.com/products/research-databases/academic-search

Academic search premier (Ebsco)
http://www.ebscohost.com/academic/academic-search-premier

Business source (Ebsco)
https://www.ebsco.com/products/research-databases/business-source-complete

CNKI, China National Knowledge Infrastructure
https://oversea.cnki.net/index

Communication source (Ebsco)
https://www.ebsco.com/products/research-databases/communication-source

Dialnet (Universidad de La Rioja)
http://dialnet.unirioja.es/servlet/revista?&clave_revista=469

ÍnDICEs CSIC
https://indices.csic.es

Informe Académico (Cengage)
https://www.gale.com/uk/c/informe-academico

Inspec, Information services in physics, electronics and computing (IET, The Institution of Engineering and Technology)
http://www.theiet.org/publishing/inspec

ISTA, Information science and technology abstracts (Ebsco)
https://www.ebsco.com/products/research-databases/information-science-technology-abstracts

Library & Information Science Source (Ebsco)
https://www.ebscohost.com/titleLists/lls-coverage.htm

Library Science Database (ProQuest)
https://www.proquest.com/products-services/libraryscience.html

Lisa, Library and information science abstracts (CSA)
https://about.proquest.com/en/products-services/lisa-set-c

Lista, Library, information science & technology abstracts (Ebsco)
https://www.ebsco.com/products/research-databases/library-information-science-and-technology-abstracts

Scopus (Elsevier)
https://www.scopus.com/sourceid/6200180164

Social sciences citation index, Social SCI, WoS (Clarivate Analytics)
http://wokinfo.com/products_tools/multidisciplinary/webofscience/ssci

FULL TEXTS

Gestor editorial (OJS) de la revista
https://revista.profesionaldelainformacion.com/index.php/EPI/issue/archive

Academic search complete (2000-)
https://www.ebsco.com/products/research-databases/academic-search-complete

Business source complete (2000-)
https://www.ebsco.com/products/research-databases/business-source-complete

Communication source (2000-)
https://www.ebsco.com/products/research-databases/communication-source

Digitalia Publishing (2011-)
https://www.digitaliapublishing.com/ejournal/105

e-Libro (2011-)
https://e-libro.net/libros/mostrarresultadobusqueda.aspx?titulo=El%20profesional%20de%20la%20informaci%C3%B3n

Ebscohost Electronic Journals Service (2000-)
http://ejournals.ebsco.com/direct.asp?JournalID=105302

Informe Académico, Cengage (2000-)
https://www.gale.com/uk/c/informe-academico

Lista. Library, information science & technology abstracts with full text (2000-)
http://www.ebscohost.com/thisTopic.php?marketID=1&topicID=584

Profesional de la información (1992-)
http://profesionaldelainformacion.com/contenidos.html

Scipedia (2007-)
https://www.scipedia.com/sj/epdli

INDEXES

CiteScore
https://www.scopus.com/sourceid/6200180164

CiteScore 2022 = 6,1 **Q1**

Dialnet Métricas
https://dialnet.unirioja.es/metricas/idr/revistas/469

IDR = 2,05 **Q1**

Dimensions
https://app.dimensions.ai/discover/publication?search_mode=content&and_facet_source_title=jour.1138422

ERIH Plus
https://kanalregister.hkdir.no/publiseringskanaler/KanalTidsskriftInfo.action?id=449601

Google Scholar Citations
https://scholar.google.com/citations?user=zv5d9O0AAAAJ

Google Scholar Metrics
https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=es

Journal Citation Reports (JCR)
https://clarivate.com/products/journal-citation-reports

Impact Factor 2022 = 4,2 **Q1**

Latindex (Catálogo)
https://www.latindex.org/latindex/ficha?folio=23967

MIAR
https://miar.ub.edu/issn/1386-6710

SCImago Journal & Country Rank (SJR)
https://www.scimagojr.com/journalsearch.php?q=6200180164&tip=sid

Scimago Journal Rank 2022 = 0,872 **Q1**

SOCIAL MEDIA

http://twitter.com/revista_EPI
http://www.facebook.com/elprofesionaldelainformacion
http://www.linkedin.com/company/el-profesional-de-la-informacion-epi
https://www.instagram.com/revista_epi

EDITORIAL MANAGING SYSTEM

Profesional de la información
https://revista.profesionaldelainformacion.com/index.php/EPI

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 ∈ 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, …, 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 :

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

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*):

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

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:

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

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:

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

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:

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

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 α, β, $\gamma > 0$ 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

Toward the consolidation of a multi-metric-based journal ranking and categorization system for computer science subject areas

Abdul Hameed; Muhammad Omar; Muhammad Bilal; Han Woo Park

Recommended citation:

Hameed, Abdul; **Omar, Muhammad; Bilal, Muhammad**; **Park, Han Woo** (2023). "Toward the consolidation of a multi-metric-based journal ranking and categorization system for computer science subject areas". *Profesional de la información*, v. 32, n. 7, e320703.

https://doi.org/10.3145/epi.2023.dic.03

Manuscript received on 19th May 2023
Accepted on 29th June 2023

Abdul Hameed
https://orcid.org/0000-0003-3235-7779

The Islamia University
Department of Computer Science
Bahawalpur, Pakistan
abdulhameedattari@gmail.com

Muhammad Omar
https://orcid.org/0000-0002-7071-5760

The Islamia University
Department of Computer Science
Bahawalpur, Pakistan
m.omar.nazeer@gmail.com
(co-first author, and co-corresponding author)

Muhammad Bilal
https://orcid.org/0009-0004-4757-3153

The Islamia University
Department of Computer Science
Bahawalpur, Pakistan
m_bilalcsiub@yahoo.com

Han Woo Park
https://orcid.org/0000-0002-1378-2473

YeungNam University
No. 257 Humanities Hall
280, Dae-dong, Gyeongsan-si
38541 Gyeongsangbuk-do, South Korea
hanpark@ynu.ac.kr
(corresponding author)

Abstract

The evaluation of scientific journals poses challenges owing to the existence of various impact measures. This is because journal ranking is a multidimensional construct that may not be assessed effectively using a single metric such as an impact factor. A few studies have proposed an ensemble of metrics to prevent the bias induced by an individual metric. In this study, a multi-metric journal ranking method based on the standardized average index (SA index) was adopted to develop an extended standardized average index (ESA index). The ESA index utilizes six metrics: the CiteScore, Source Normalized Impact per Paper (SNIP), SCImago Journal Rank (SJR), Hirsh index (H-index), Eigenfactor Score, and Journal Impact Factor from three well-known databases (*Scopus*, *SCImago Journal & Country Rank*, and *Web of Science*). Experiments were conducted in two computer science subject areas: (1) artificial intelligence and (2) computer vision and pattern recognition. Comparing the results of the multi-metric-based journal ranking system with the SA index, it was demonstrated that the multi-metric ESA index exhibited high correlation with all other indicators and significantly outperformed the SA index. To further evaluate the performance of the model and determine the aggregate impact of bibliometric indices with the ESA index, we employed unsupervised machine learning techniques such as clustering coupled with principal component analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE). These techniques were utilized to measure the clustering impact of various bibliometric indicators on both the complete set of bibliometric features and the reduced set of features. Furthermore, the results of the ESA index were compared with those of other ranking systems, including the internationally recognized *Scopus*, SJR, and *HEC Journal Recognition System* (*HJRS*) used in Pakistan. These comparisons demonstrated that the multi-metric-based ESA index can serve as a valuable reference for publishers, journal editors, researchers, policymakers, librarians, and practitioners in journal selection, decision making, and professional assessment.

Keywords

Journal rankings; Research evaluation; Indicators; Scientific journals; Metrics; Algorithms; Machine learning; Cluster analysis; Principal component analysis (PCA); t-distributed stochastic neighbor embedding (t-SNE); Cross tabulation; ESA index.

Competing interests

The authors declare no competing interests.

Acknowledgements:

This work is related to the first author's PhD thesis.

1. Introduction

Evaluating research quality is a complex task that can significantly impact multiple decisions, such as improving the tenured track or basic pay scale systems (TTS/BPS) service structure to enhance research quality, determining hiring decisions, allocating research funding, conducting promotions, and awarding scholarly degrees.

Various evaluation systems have been developed for this purpose. Research standards can be evaluated through qualitative methods (**Wical**; **Kocken**, 2017), quantitative methods (**Beliakov**; **James**, 2011), or a hybrid approach that combines both methods (**Hsu** *et al.*, 2015). Additionally, a meta-approach for predicting journal quality has also been proposed (**Saarela**; **Kärkkäinen**, 2020). Conventionally, most journals assess the quality of a publication through a peer review process by experts in the relevant field of research (**Morris** *et al.*, 2009).

In the present era of information technology, various contemporary systems for ranking journals have been adopted by different organizations. *Web of Science* (*WoS*), *SCImago Journal & Country Rank*, and *Scopus* are the few examples of the numerous groups and for-profit institutions that maintain sizable publishing datasets that allow for the computation of citations and other potential journal influence statistics. Several well-defined bibliometric indicators have been developed for ranking journals, such as the Impact Factor (IF), Eigenfactor (EF) Score, Hirsh index (H-index), SCImago Journal Rank (SJR), Source Normalized Impact per Paper (SNIP), and CiteScore. Each metric has its own strengths and weaknesses. The IF is one of the most widely used indicators for ranking journals. However, the use of an individual indicator does not ensure reliable results (**Setti**, 2013). The main problem with citation-based indicators such as the IF, is the dissimilarity in citation practices among different disciplines. For instance, mathematical studies tend to receive fewer citations than biology research (**Ferrer-Sapena** *et al.*, 2016). To address this issue, bibliometric analysis has been applied to assess the influence of published work and their potential to enhance a journal's reputation (**Perera**; **Wijewickrema**, 2018).

Recently, there has been growing interest in the use of machine learning algorithms to automatically categorize journals, although this approach is not yet widely adopted (**Abbas** *et al.*, 2019). These models are capable of operating independently without the need for human intervention, which offers the potential for objectivity. In addition, automated classification procedures are typically less expensive to implement as compared to expert-based classification procedures. Moreover, machine learning algorithms have the advantage of being able to consider all available quality indicators, unlike citation-based indicators which are limited in scope. This implies the utilization of all the feasible bibliometric indicators (**Perera**; **Wijewickrema**, 2018). This is necessary because journal ranking is a multicriteria decision problem.

Based on the journal impact index, SJR ranks journals in each subject category into quartiles ranging from one to four (Q1-Q4). Q1 represents the top 25% SJR distribution, Q2 denotes the middle-high SJR distribution (25%-50%), Q3 indicates the middle-low SJR distribution (50%-75%), and Q4 refers to the lowest SJR distribution (the bottom 25%) (**Mañana-Rodríguez**, 2015).

The *Higher Education Commission* (*HEC*) of Pakistan typically categorizes journals into four groups to ensure research quality: W, X, Y and Z, where W indicates the highest standard, and Z indicates the lowest. The *HEC Journal Recognition System* (*HJRS*) is a recently developed online system for recognizing journals. On introducing HJRS, the *HEC* removed the Z category. The categorization system designed by the *HEC* of Pakistan only recognizes research journals that fall into the W, X, and Y categories with full implementation starting in July 2020. The *HEC* asserts that this system assesses the quality of publications using internationally acclaimed parameters (**Mubarak**; **Seemee**, 2021).

This study proposes a data-driven methodology for automatically categorizing computer science journals based on various features (metrics). The research questions for this study are as follows:

(1) Can we adopt a multi-metric-based scientific journal ranking system to develop an index known as the extended standardized average index (ESA index) to combine a number of bibliometric indices that can yield more robust and aggregated journal rankings?

(2) What is the impact of multiple bibliometric features on journal ranking? To what extent does the ESA index correlate with other bibliometric indicators?

(3) To implement a cluster analysis of the considered bibliometric indices (in conjunction with the ESA index (seven indices)) against a reduced set of indices to assess the stability of the corresponding journal ranking and categorization system.

(4) Can the ESA index function as an authentic and reliable medium for classifying the quality tiers of the *Scopus* Quartiles, *SJR* Best Quartiles, and *HJRS* Categories?

The remainder of this article is organized as follows. In Section 2, the related work and background informations regarding various bibliometric indicators, their definitions, and their advantages and disadvantages are discussed. Section 3 explains the materials and methods. Section 4 presents the experimental observations obtained from each method. Section 5 summarizes the concluding remarks and presents the conclusions.

2. Related work

The rankings of academic publications have an impact on various players in academia. Scientists consider potential venues for their research based on the rankings, departments assess their productivity using these rankings, and funding success may also be influenced by them (**Wical**; **Kocken**, 2017). The impact of academic publication rankings is not limited to academia alone, as it also affects the non-academic world. This includes publishers who want to assess the reputation of their journals, professional bodies, practitioners, and funding agencies. The application of scientometric methods in science and technology studies (STS) (**Wyatt**; **Milojevič**; **Park**; **Leydesdorff**, 2017) has significant implications for research quality. Numerous countries have implemented journal assessment standards to encourage and incentivize national academic institutions and research centers to actively contribute to the knowledge base in their respective fields (**Holmberg**; **Park**, 2018; **Saarela** *et al.*, 2016). While some rely on qualitative assessments through peer review (**Wical**; **Kocken**, 2017), others use quantitative metrics (**Yuen**, 2018), and few utilize hybrid (**Allen** *et al.*, 2009) or meta-ranking approaches (**Ennas** *et al.*, 2015). When assessing scholarly output, the quantity and quality of publications should be considered (**Zhu**; **Park**, 2022). Assessment techniques are established to gather evidence and information that can be used to evaluate different aspects of research and make informed decisions.

It is difficult to design and evaluate a system that aims to translate research materials into monetary rewards. One could contend that if an evaluation criterion based on quantitative measurements is relatively straightforward, it can have negative consequences. Since its introduction, the IF has been commonly utilized as a quantitative research method. However, there are many restrictions related to its misapplication (**Dellavalle** *et al.*, 2007). Therefore, other indices such as the SJR (**González-Pereira** *et al.*, 2010), H-index (**Lacasse** *et al.*, 2011), Eigenfactor (**Bergstrom**, 2007), CiteScore, (**James** *et al.*, 2018), and SNIP (**Moed**, 2010) have become popular for research evaluation.

Another indicator that measures the article effect is the *Altmetric*. It is based on Internet attention (**Holmberg**; **Park**, 2018). The *Altmetric* score is a metric that measures online attentions received by scholarly articles based on mentions in news publications, blog comments, tweets, and social media posts. The *Altmetric* score is often used to identify publications that have attracted a lot of attentions on the Internet (**Holmberg**; **Park**, 2018).

The literature on journal quality evaluation can be classified into four categories: conventional *subjective* ranking (qualitative approach), which is based on the opinions of experts in a specific discipline; *objective* ranking (quantitative approach), which is based on citations; *hybrid* ranking (hybrid approach), which is a combination of subjective and objective rankings; and *meta* ranking approach, which automatically ranks journals using artificial intelligence.

2.1. Qualitative approach for journal ranking

A qualitative or survey-based approach involves ranking journals based on their perceived quality and reputation by receiving feedback from qualified experts to rank journals in a specific domain (**Allen** *et al.*, 2009; **Walters**, 2017). There are two main drawbacks. First, this measure is effective only at the time it is used. This is because in a dynamic research field the top-ranking journals and popular subjects change over time (**Duan** *et al.*, 2018). Second, the ranking lists produced by survey-based methods become increasingly less trustworthy for lower-ranking journals.

2.2. Quantitative approach for journal ranking

Using quantitative approaches, journals are assessed according to their size (number of publications), influence, and number of citations (**Leydesdorff**; **Park**, 2017). These techniques are utilized to evaluate the journal quality, although these capture only a few features of quality and are simple to compute. However, it should be noted that quantitative factors are occasionally unrelated to the qualitative factors. For instance, the mere fact that a paper is published in a journal with a high volume of publications does not guarantee its quality (**Fersht**, 2009; **Tsai**, 2014). The main features of the quantitative metrics used in this study are summarized in Table 1.

2.3. Hybrid methods for journal ranking

Professionals help in decision-making to overcome the inherent drawbacks of using an individual index while maintaining the advantages of utilizing various indices and providing a distinctive aggregate score. Quantitative approaches are straightforward, unbiased, and current methods. However, survey-based approaches can incorporate qualitative data that are difficult to measure, and provide a tiered structure that aids in the creation of guidelines. An increasing number of journal rating experts consider that combining journal bibliometrics with professional assessment of journal quality is the best overall approach (**Tüselmann** *et al.*, 2015).

Business schools frequently use the A*ssociation of Business Schools* (*ABS*) *Academic Journal Guide* from among several journal ranking lists produced using hybrid techniques (**Morris** *et al.*, 2009). To develop this journal guide, members of the *ABS Scientific Council* have provided various measures such as the IF, SNIP, and SJR for each journal. After con-

Table 1. Summary of the main features of journal impact indicators provided in *WoS*, *SCImago Journal & Country Rank*, and *Scopus*

Characteristic	*WoS*		*SCImago Journal & Country Rank*		*Scopus*	
	JIF*	EF*	H-index*	SJR*	CS*	SNIP*
Calculation methodology	Ratio of citations and publications	Based on Eigenvector centrality	Ratio of citations and publications (h citations from h papers)	Citations network-based	Ratio of citations and publications	The ratio of publications to citations, normalized by citation densities across disciplines
Publication window (years)	2/5	5	h	3	4	3
Citation window (years)	1	1	1	1	4	1
Journal self-citations	Yes	No	Yes	limited up to 33%	Yes	Yes
Normalized by papers count in the journal (size independent)	Yes	No	Yes	Yes	Yes	Yes
Normalized by fields/disciplines	No	No	No	Not directly, Thematic closeness based between journals	No	Yes
Normalized by reputation (weighted)	No	Yes	No	Yes	No	No
Applicability	Only for *JCR* journals	Only for *JCR* journals	for journals in Google Scholar	for all sources (including journals, conference proceedings, book series and trade publications)	for all sources (including journals, conference proceedings, book series and trade publications)	for all sources (including journals, conference proceedings, book series and trade publications)
Availability	Requires a subscription to *JCR*	Requires a subscription to *JCR*	Free, (no subscription required)	Free, (no subscription required)	Free, (no subscription required)	Free, (no subscription required)
Limitations and drawbacks	Different types of documents included in numerator and denominator, potentially manipulable, Short citation window (for 2-Year JIF); not normalized for fields/disciplines	Inconvenient numerical value, decreasing with new journals as included in the database; not normalized for fields/disciplines	Differing citation practices of articles in different fields	Complex calculation, difficult to interpret	Not normalized for disciplines	Impact per paper but indicates impact of average articles in a journal (not for a specific article)

*JIF, Journal Impact Factor; EF, Eigenfactor; H-index, Hirsh index; SJR, SCImago Journal Rank; CS, CiteScore; SNIP, Source Normalized Impact per Publication.

sultation with their individual academic communities, they were instructed to group each publication into one of the following five categories: 4* for elite journals, 4 for top journals, 3 for highly regarded journals, 2 for good standard journals, and 1 for modest journals.

2.4. Meta-ranking approach for journal ranking

Recently, the concept of automatically ranking journals using machine-learning techniques has attracted significant attention (**Halim**; **Khan**, 2019). Because machine-learning based algorithms can operate without human intervention, these appear to be more objective. The *HJRS* is based on the meta-ranking approach and considers all the available quality indicators, unlike citation-based indicators that only consider a limited set of explanatory features. This is advantageous because ranking academic journals involves multiple criteria and decision-making factors. In various studies, machine-learning techniques such as regularized logistic regression, gradient boosting, and random forest have been used to predict journal quality (**Saarela**; **Kärkkäinen**, 2020). As shown in Table 2, several studies have demonstrated that machine-learning techniques provide better results than the qualitative, quantitative, and hybrid approaches adopted earlier. The principal component analysis (PCA) by (**Bollen** *et al.*, 2009) indicates the multidimensionality of the different impact indicators, (**Ennas** *et al.*, 2015) used various statistical and machine learning techniques to formalize an approach that ranks journals from different dimensions, thereby characterizing the aspects of research quality. The ensemble simple linear regression model by (**Duan** *et al.*, 2018) performed better for the interdisciplinary journals. A few studies on journal rankings are summarized in Table 2.

Table 2. Summary of journal ranking studies using machine learning techniques

Work	Study Purpose	Variables	Techniques	Findings
(**Bollen** *et al.*, 2009)	Evaluating research impact through citations and usage data sets.	39 bibliometric indicators	PCA	The Principal components show 92% of the variances between the correlations of journal rankings by 37 impact measures
(**Tüselmann** *et al.*, 2015)	Handling missing values and journals by DEA	Impact Factor, ABS, ABDC, VBH, CNRS	Random Forest, DEA	Treatment of missing data through imputation and better classification of journals through Random Forest
(**Tsai**, 2014)	Ranking computer science journals using IF and H-index	IF, 5-IF, H-index,	CombSUM	Find a better correlation between the impact factor and H-index of computer science journals
(**Ennas** *et al.*, 2015)	A data-driven methodology using different methods from statistics and machine learning to combine various indices to create an aggregate rating.	IF, 5-IF, SJR, H-index, Immediacy Index, Eigenfactor Score. Article Influence, SNIP, IPP	SVR, CombSUM, Borda Count and PCA	SVR and PCA outperformed well in ranking journals
(**Fernández-Cano**; **Fernández-Guerrero**, 2017)	EM journals were subjected to a multivariate evaluation based on seven highly linked evaluation variables to produce a factor-based meta-index	IF, H-index, SJR and two altmetric scores (3 months and any time)	Cronbach's alpha	The length of time (number of years) that each journal has been published would be a significant factor regarding the H-index that should be taken into consideration.
(**Duan** *et al.*, 2018)	A data-driven method used to rank MIS journals not included in ABS list	2-year IF, 5-year IF, EF, AI, SNIP, SJR, ABS, ABDC, VHB, CNRS, FNEGE	MLR, ESLR, SVM, NN	ESLR achieves the best performance among various data-driven methods and generates reasonable ranking for new journals, top journals and interdisciplinary journals
(**Perera**; **Wijewickrema**, 2018)	Investigates the relationship among four journal rankings	IF, Eigenfactor, H-index and SJR	Pearson correlation coefficient, Hierarchical clustering, PCA, Kaiser-Meyer-Olkin (KMO) test and Bartlett's test	Results indicate that a higher correlation was found between IF and SJR
(**Halim**; **Khan**, 2019)	Data Science framework to automatically categorize journals	19 features (IF, CiteScore, SNIP, H-index, SJR, Eigenfactor, article influence, immediacy index, cited half-life, publisher, website, country, age, open access, citations, percentile, peer review, number of articles published yearly) and acceptance rate	Feature selection (MI, mRMR, SD) Clustering (k-means, k-medoids), Classification (ANN, KNN) Cluster validation using DBI, DI, SC, CHI	Top nine features (CiteScore, H-index, SJR, SNIP, cited half-life, Eigenfactor, article influence, total citations, percentile) four clusters identified, Average accuracy (ANN) 92.86%
(**Saarela**; **Kärkkäinen**, 2020)	Automated rankings based on the analysis of bibliometric indicators through the expert score ranking and through data analysis and machine learning techniques	Features used (Rank, Title, publications, volume, type, start year, Norway Score, Denmark Score, SJR, IPP, SNIP, Panel, Sherpa/ Romeo Code, Publisher	SMOTE, Logistic regression, random forest and gradient boosting	High correlation found between citations- and expert-based rankings system.
(**Feng** *et al.*, 2020)	Identified the most important and contributing features for categorizing journals through unsupervised Laplacian score	2-Year IF, 5-Year IF, CiteScore, SNIP, SJR and H-index with two class labels (discipline and quantile)	Laplacian score, spectral clustering, *k*-NN, BPNN and subjective method (questionnaire used)	Based on experimental results IF, CiteScore, and H-index are the best features and by the voting method based on a seven-point Likert scale, Impact Factor and H-index got higher votes.

PCA: Principal Component Analysis, DAE: Data envelopment analysis, ABDC: Australian Business Deans Council, ABS: Association of Business Schools, VHB: Association of University of Business in German-Speaking Countries, CNRS: Centre National de la Recherche Scientifique, IF: Impact Factor, SNIP: Source Normalized Impact per Publication, SJR: SCImago Journal Rank, SVM: Support Vector Machine, SVR: Support Vector Regression, NN: Neural Network, BPNN: Back Propagation Neural Network, MLR: Multicollinearity problem, ESLR: Ensemble Simple Linear Regression

3. Methodology

The proposed methodology is in line with the intellectual recommendations of Leydesdorff's research group, namely, the use of scientometric methods in science and technology studies (STS) (**Wyatt**; **Milojevič**; **Park**; **Leydesdorff**, 2017). This study proposed a data-driven methodology to develop a novel multi-metric-based scientific impact measure called the ESA index for ranking and categorizing journals based on various bibliometric impact measures. The main objective was to propose an automated approach for categorizing journals using machine learning techniques in various computer science disciplines. Various bibliometric indicators used for this purpose were as the CiteScore, SNIP, SJR, H-index, Eigenfactor Score, and Journal Impact Factor. Each bibliometric measure has its advantages and drawbacks, and the rankings it produces can vary significantly depending on the specific metric used and the criteria for ranking. The integration of current bibliometric indicators is a potential strategy to compensate the limitations of individual indicators. A multi-dimensional space constructed using different impact measures was used to assess the journals. First, a multi-metric based ESA index was proposed for ranking and categorizing academic journals. The proposed ESA index was developed from multiple impact measures, which combines the impact of each bibliometric measure. Thereby, it functions as an alternative to various indicators for academic journal quality assessment. The journals in various disciplines of computer science were analyzed and categorized using various well-known bibliometric features (the CiteScore, SNIP, SJR, H-Index, Eigenfactor Score, and Journal IF). Consequently, we formulated a data-driven methodology to determine the impact of the ESA index with other bibliometric indicators using machine learning techniques. For this purpose, we first applied k-means clustering to the full featured dataset (seven bibliometric features). We then applied two dimensionality reduction techniques (PCA and t-distributed stochastic neighbor embedding (t-SNE)) to determine the reduced set of features. Subsequently, we applied k-means clustering to a reduced set of features. The clustering results of the proposed model were compared and validated using the two most commonly used and currently available benchmarks: (1) *SJR* Best Quartiles and (2) *Scopus* Quartiles. The proposed methodology for the preliminary investigation of journal evaluations is presented in Figure 1.

The proposed framework utilizes unsupervised machine-learning approaches such as clustering and dimensionality reduction for journal evaluation. The following section discusses the various modules of the proposed system.

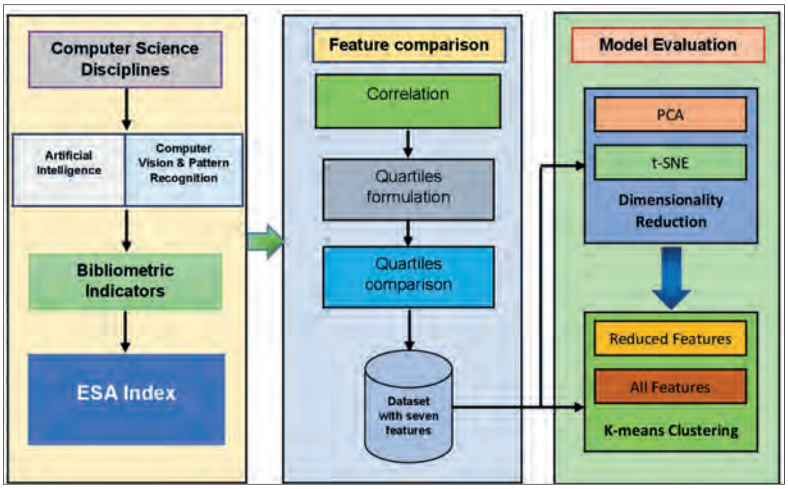


Figure 1. Block diagram of the proposed research methodology

3.1. Dataset collection

The dataset consists of all the available journals indexed in *Scopus*, *Web of Science* (*WoS*), *SJR*, or *HJRS*. The dataset was extracted from various computer science disciplines to evaluate journal quality in their respective fields. A dataset currently available for 2021 was used in this study. The extracted features included are ISSN, Journal Title, CiteScore, SNIP, SJR, H-Index, Eigenfactor Score, and Journal IF. The *Scopus* Quartiles, *SJR* Quartiles, and *HJRS* journal categories were utilized to evaluate the validity of the proposed model. For a comprehensive analysis, datasets from two disciplines of computer science (314 journals of artificial intelligence, and 106 journals of computer vision and pattern recognition) were extracted with various bibliometric features. Various journal categories such as journal quartiles Q1-Q4 were extracted from the *SJR* and *Scopus* databases. Three journal categories (W, X, and Y) were extracted from the *HJRS*. A merging technique using outer join was applied on the collected dataset to ensure that journals indexed in any of the well-known databases such as *Scopus*, *Web of Science* (*WoS*), and *SJR* were included in the dataset. The distributions of journals in the *Scopus*, *SJR*, and *HJRS* categories are shown in Figures 2, 3, and 4, respectively.

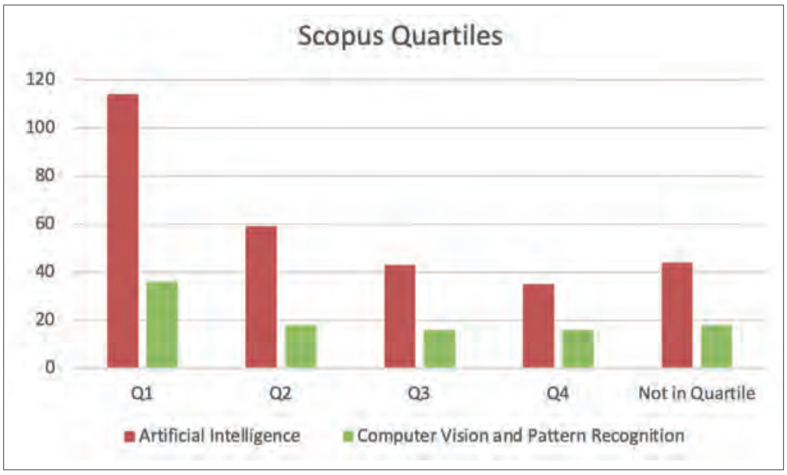


Figure 2. Distribution of journals with *Scopus* Quartiles

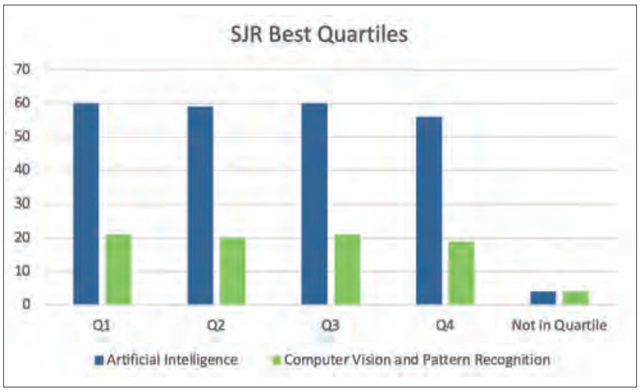


Figure 3. Distribution of journals with SJR Best Quartiles

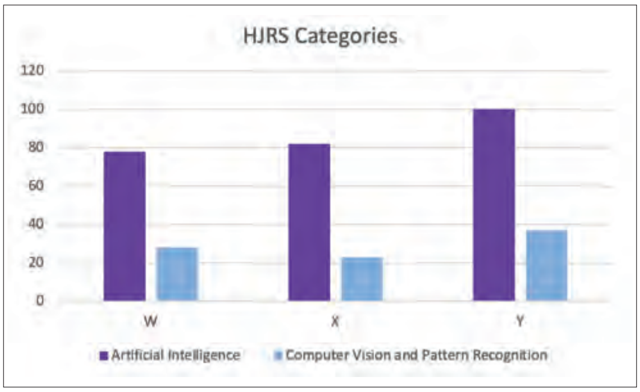


Figure 4. Distribution of Journals with HJRS categories

3.2. Preprocessing

After finalizing the dataset, standardization was implemented to preprocess the dataset for experimental use. The missing values of certain bibliometric indicators were filled with zero, and the dataset was standardized using a standard scalar. The standard scalar shifted the data of all the features in the range 0-1.

After removing the values with missing quartiles, 217 journals of artificial intelligence and 74 journals of computer vision and pattern recognition were used for data analysis, journal ranking, and categorization.

3.3. ESA index: Multi-metric based meta-ranking approach for journal ranking

Different bibliometric measures generally provide different journal rankings. This can cause ambiguities in the decision-making process. Therefore, an ESA index was introduced in this study using a feature engineering technique. The index is multi-metric because it combines various bibliometric features to propose a simple and reliable metric for ranking academic journals.

This study presented an approach to develop an ESA index as an alternative multi-metric impact indicator for evaluating academic journals. The aim was to contribute to multiple scientific impact measures such as the CiteScore, SNIP, SJR, H-index, Eigenfactor Score, and Journal Impact Factor. These bibliometric indicators were combined to develop a new metric that is simple and multi-metric-based for journal evaluation. Various impact indicators have their advantages and shortcomings. Therefore, it is necessary to use multiple indicators rather than an individual one to evaluate the journal quality. However, owing to different calculation criteria, various impact metrics generally yield different evaluation results. Journal articles with a high IF do not necessarily have a high CiteScore and vice versa. Therefore, researchers would select only one of these journals as a reference for article submission. To develop an alternative, simple, and reliable metric for various impact indicators, we adopted a concept from the method used in (**Hsu** *et al.*, 2015).

The ESA index can be calculated as:

- Normalization: Compute the normalized value/score of each journal's research impact metric from the total score. The metrics used are the CiteScore, SNIP, SJR, H-index, Eigenfactor Score, and Journal Impact Factor(IF).
- Average percentage: Calculate the average of the input features and determine the percentage.

For a given set of journals $D = \{x_i\}_i^n$ (where $x_i = (x_i^1, \dots x_i^d) \in \mathbb{R}^d$ represents *d* index values for the ith journal, e.g. $x_i^1, x_i^2, x_i^3, x_i^4, x_i^5$ and x_i^6 may, respectively represent the CiteScore, SNIP, SJR, H-index, IF, and Eigenfactor Score for a journal in the set *D*), the ESA index is calculated as follows:

(1) First, calculate the normalized value of each indicator as

(i) Calculate the normalized value of CiteScore as

$$nCS = \frac{CSi}{\sum_{i=1}^{n} CSi} \quad (1)$$

(ii) Calculate the normalized value of the SNIP as

$$nSNIP = \frac{SNIPi}{\sum_{i=1}^{n} SNIPi} \quad (2)$$

(iii) Calculate the normalized value of SJR as

$$nSJR = \frac{SJRi}{\sum_{i=1}^{n} SJRi} \quad (3)$$

(iv) Calculate the normalized value of the H-index as

$$nHI = \frac{HIi}{\sum_{i=1}^{n} HIi} \quad (4)$$

(v) Calculate the normalized value of the IF as

$$nIF = \frac{IFi}{\sum_{i=1}^{n} IFi} \quad (5)$$

(vi) Calculate the normalized value of the Eigenfactor Score as

$$nEF = \frac{EFi}{\sum_{i=1}^{n} EFi} \quad (6)$$

(2) Calculate the average value and percentage as

$$\text{ESA Index} = \frac{nCS+nSNIP+nSJR+nHI+nIF+nEF}{\#features} \times 100 \quad (7)$$

The prefix n represents the normalized value of the indicator, the subscript i represents the ith value of the journal, and the features represent the bibliometric indicators. Figure 5 shows the mean values of the various bibliometric indicators used in this study.

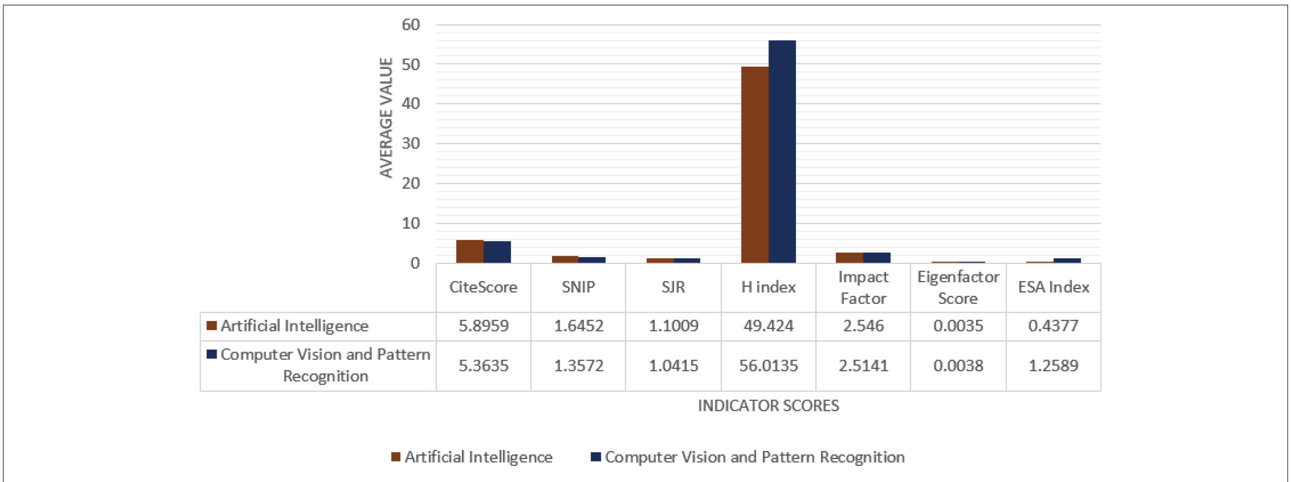


Figure 5. Mean value of various indicators in computer science disciplines

3.4. Dimensionality reduction

Dimensionality reduction addresses the problem of distinguishing valuable low-dimensional data from high-dimensional data. It represents high-dimensional data as the principal components. In this study, PCA and t-SNE were implemented on a bibliometric dataset to determine the most contributing reduced set of features.

3.4.1. Principal component analysis (PCA)

A popular multivariable statistical technique called PCA transformation employs PCA for feature extraction and dimensionality reduction in pattern analysis. By retaining significant information, plainly describing the dataset, and analyzing the observations, it aims to extract significant information from the data and reduce the dataset amount. PCA was employed in this study to reduce the dataset to a new feature space.

3.4.2. t-Distributed stochastic neighbor embedding (t-SNE)

The t-SNE algorithm is a novel method of multi-dimensional scaling. This technique is popular because it scales high-dimensional data to low-dimensional data. In this study, this technique was applied to data points (journals) that convert high-dimensional Euclidean distances between data points (journals) into conditional probabilities that represent similarities among journals.

3.5. Un-supervised evaluation models

3.5.1. Cluster analysis

To identify disconnected groupings in the collected dataset, we used an unsupervised machine learning technique known as k-means clustering. Using an unlabeled dataset, k-means clustering was used to group similar journals. The dataset was divided into groups using the k-means method, and these groups were represented by K variables. In this study, k-means clustering was used to identify clusters based on similar features. Various evaluation measures were used to determine the optimum number of clusters. This verified the percentage of variance as a function of the number of clusters. Based on the pre-evaluated cluster number, the journals were grouped into various numbers of clusters using Euclidean distance.

3.5.2. Clustering performance evaluation measures

A clustering algorithm helps to categorize the data. The quality of the clustering results can be assessed using various metrics used for the evaluation.

Internal evaluation measures

It is feasible to determine the clustering structure quality without access to external data owing to the internal validation methods. The internal measures are based on information from the input data during clustering. Here, rather than using a ground truth label from the external world, we employed the silhouette coefficient (SC) score, Calinski-Harabasz index (CHI), and Davies-Bouldin index (BDI) for internal cluster validation to assess the cluster quality.

Determining the optimal value of K

The elbow method was applied to determine the optimal number of clusters. It examines how the number of groups affects the proportion of the explained variation. The proportion of variation explained by clusters is plotted against the number of clusters. The first clusters would contribute a substantial amount of information. However, eventually the marginal gain would reduce and the graph would adopt an angle. The cluster nodes begin the calculations based on predetermined cluster numbers and are split into clusters based on the predetermined value. The Euclidean distance is used to group the cluster elements into a predetermined number of clusters.

Silhoutte coefficient (SC) score

The SC assessment metric was used to assess clustering outcomes. The dissimilarity of a data point or node from other cluster members as well as its similarity to all other points or nodes within its cluster were verified using this clustering validation measure. The SC value lies within [-1, 1]. A higher SC value denotes effective clustering, whereas values near 0 or -1 denote ineffective clustering.

Calinski-Harabasz index (CHI)

The CHI is a measure of cluster validity. It is used to evaluate clustering quality. The index is based on the technique used to determine the ratio of between and within-cluster variances. It measures the separation between clusters and their compactness. A higher index value indicates better clustering results.

Davies-Bouldin index (DBI)

Davies-Bouldin index (DBI) is used to evaluate the clustering performance. It verifies the inter and intra-cluster similarities of the nodes in clusters based on sample-specific dimensions. The DBI value lies within [0, + ∞]. A value closer to zero indicates a better clustering.

External evaluation measures

In the cluster validation process, the external ground-truth label is an additional piece of information incorporated via the external validation approach. When external data are available and there are few true labels in the dataset, an external technique can be used. The effectiveness of the clustering observations was assessed using externally provided data through external validation metrics. In this study, several external validation metrics were used to evaluate the clustering results using available external ground truth data.

Adjusted Rand score (ARI)

The adjusted Rand index (ARI) is an external clustering performance evaluation measure. It was used to validate the clustering results with external ground truth labels. In this section, the Scopus and SJR Best Quartiles are used as external class labels for comparison with the clustering labels. The lowest and highest possible values of ARI are -1 and 1, respectively.

Adjusted mutual information (AMI) score

The AMI score is a measure of the similarity between two clusters in a dataset. It considers the fact that the mutual information score which measures the amount of information shared by two clusterings, can be biased toward clustering with many small clusters. The AMI score is used here for clustering comparison. The value of Adjusted mutual information ranges from 0 to 1. the value 0 implies dissimilarity and 1 implies most similar clusters.

Homogeneity, completeness, and V-measure (HCV)

The homogeneity measures the purity of each cluster with respect to a single class. A clustering result satisfies homogeneity if all its clusters contain only data points that are members of a single class. The homogeneity score ranges from zero to one, with one indicating perfect homogeneity. The completeness measures the extent to which a class is represented by a single cluster. A clustering result satisfies completeness if all the data points that are members of a given class are assigned to the same cluster. The completeness score ranges from zero to one, with one indicating perfect completeness. The V-measure is the harmonic mean of homogeneity and completeness. It provides a single score that balances both the measures. The V-measure score ranges from zero to one, with one indicating perfect agreement between the clustering and true labels. The V-measure is a commonly used metric for clustering evaluations because it considers both homogeneity and completeness.

Fowlkes-Mallows (FM) score

The FM score is a measure of the similarity between two clusters in a dataset. This approach is based on the concepts of precision and recall. The score ranges from zero to one, with one indicating perfect agreement between the two clusters, and zero indicating no agreement beyond chance.

Cross tabulation

Cross-tabulation places categorical data in a table and then summarizes it by aligning the labels of two classes/categories with each other. Each column of the table contains the number of data members of a class belonging to the data members of another class. It can determine the frequency (either in a raw number or in proportional form) of the values that fall into the groups that the cell is planned to illustrate. Many statistical tests (the majority of which adhere to the chi-squared distribution) can then be performed using the summary data displayed in a cross-tabulated form. In this study, cross-tabulation was used to compare the *Scopus* Quartiles, *SJR Best* Quartiles, and *HJRS* categories with various journal categories observed in the proposed framework.

4. Results and discussions

Various impact indicators such as the CiteScore, SNIP, SJR, H-index, Eigenfactor Score, and Journal Impact Factor were combined to develop a multi-metric indicator called ESA index for ranking journals. We developed and utilized the index to identify the effects of multiple features using various machine learning techniques. The *Python* libraries *Scikit-learn*, *Matplotlib*, and *Seaborn* were used for these experiments. The experiments were performed using an Intel® Core TM i5 Intel(R) Core (TM) i5-5200U CPU @ 2.20 GHz 2.20 GHz.

4.1. Correlation of ESA index with other bibliometric indices

To analyze the ESA index, Spearman's correlation between various bibliometric indicators (i.e., the CiteScore, SNIP, SJR, H-index, Eigenfactor Score, and Journal IF) was calculated. Table 3 presents the correlation of artificial intelligence journals. It shows that the ESA index has the highest correlation with the SJR, and a higher correlation with the CiteScore than with the other bibliometric indicators. Table 4 presents the correlation of computer vision and pattern recognition journals. It shows that the ESA index has the highest correlation with the CiteScore, and a higher correlation with the SJR than the other bibliometric indicators. A strong correlation is observed between the ESA index and various bibliometric indicators.

Table 3. Spearman rank correlation between various bibliometric indicators (Artificial Intelligence)

	CS	SNIP	SJR	H-index	IF	EF	ESA Index
CS	1						
SNIP	0.86	1					
SJR	0.93	0.92	1				
H-index	0.66	0.56	0.63	1			
IF	0.71	0.62	0.68	0.75	1		
EF	0.66	0.59	0.66	0.80	0.96	1	
ESA Index	0.92	0.88	0.93	0.81	0.84	0.84	1

Table 4. Spearman rank correlation between various bibliometric indicators (Computer Vision and Pattern Recognition)

	CS	SNIP	SJR	H-index	IF	EF	ESA Index
CS	1						
SNIP	0.92	1					
SJR	0.97	0.96	1				
H-index	0.68	0.58	0.64	1			
IF	0.74	0.69	0.73	0.65	1		
EF	0.68	0.62	0.68	0.71	0.96	1	
ESA Index	0.96	0.91	0.95	0.78	0.84	0.82	1

*CS, CiteScore; SNIP, Source Normalized Impact per Publication; SJR, SCImago Journal Rank; H-index, Hirsh index; IF, Impact Factor; EF, Eigenfactor; ESA Index, Extended Standardized Average Index

4.2. Data analysis of ESA index with benchmark journal quartiles/categories

Table 5 presents a comparison of SA quartiles with *Scopus* quartiles, *SJR Best* quartiles and *HJRS* categories for artificial intelligence journals, and Table 6 shows that of ESA quartiles for artificial intelligence journals using the ARI, AMI score, homogeneity, completeness, V-measure (HCV), and FM score. It shows that various evaluation metrics (while comparing different quartiles) show better results for the ESA index than for the SA index. It can be observed that the *HJRS* has a higher evaluation measure value than the *SJR* and *Scopus* Quartiles. The comparison results of SA quartiles and the ESA quartiles for computer vision and pattern recognition subject area are presented in Tables 7 and 8, respectively. Tables 9 and 10 present the cross-tabulation results for the journals of artificial intelligence and those of computer vision and pattern recognition, respectively.

Table 5. Comparison of SA index with *Scopus*, *SJR Best* Quartiles and *HJRS* Categories (Artificial Intelligence)

Quartiles	ARI	MI	HCV	FM
Scopus Quartiles	0.1755	0.1781	0.1897	0.4394
SJR Best Quartiles	0.1925	0.2312	0.2282	0.4199
HJRS Category	0.3951	0.3992	0.4261	0.5915

Table 6. Comparison of ESA index with *Scopus*, *SJR Best* Quartiles and *HJRS* Categories (Artificial Intelligence)

Quartiles	ARI	MI	HCV	FM
Scopus Quartiles	0.3059	0.3298	0.3446	0.5225
SJR Best Quartiles	0.3674	0.4630	0.4513	0.5412
HJRS Category	0.6081	0.6164	0.6689	0.7329

Table 7. Comparison of SA index with *Scopus*, *SJR Best* Quartiles and *HJRS* Categories (Computer Vision and Pattern Recognition)

Quartiles	ARI	MI	HCV	FM
Scopus Quartiles	0.1949	0.2125	0.2521	0.4340
SJR Best Quartiles	0.2488	0.2985	0.3190	0.4501
HJRS Category	0.4202	0.4338	0.4879	0.6004

Table 8. Comparison of ESA index with *Scopus*, *SJR Best* Quartiles and *HJRS* Categories (Computer Vision and Pattern Recognition)

Quartiles	ARI	MI	HCV	FM
Scopus Quartiles	0.2712	0.3265	0.3587	0.4876
SJR Best Quartiles	0.3599	0.4881	0.4902	0.5320
HJRS Category	0.6138	0.6226	0.6822	0.7343

Table 9. Cross tabulations of original journal categories and ESA Index Quartiles in Artificial Intelligence Journals

		ESA-Q1	ESA-Q2	ESA-Q3	ESA-Q4	Total
***Scopus*-Q**	**Q1**	71	36	7	0	114
	Q2	7	25	26	1	59
	Q3	1	12	27	3	43
	Q4	0	1	16	18	35
***SJR*-Q**	**Q1**	55	5	0	0	60
	Q2	23	32	4	0	59
	Q3	0	33	27	0	60
	Q4	0	1	34	21	56
HJRS	**-**	0	1	0	3	4
	W	70	8	0	0	78
	X	8	63	11	0	82
	Y	0	5	58	37	100

Table 10. Cross tabulations of original journal categories and ESA Index Quartiles in Computer Vision and Pattern Recognition Journals

		ESA-Q1	ESA-Q2	ESA-Q3	ESA-Q4	Total
***Scopus*-Q**	**Q1**	22	12	2	0	36
	Q2	3	9	6	0	18
	Q3	0	4	11	1	16
	Q4	0	1	7	8	16
***SJR*-Q**	**Q1**	18	3	0	0	21
	Q2	8	12	0	0	20
	Q3	0	9	12	0	21
	Q4	0	0	12	7	19
HJRS	**-**	0	1	0	3	4
	W	23	5	0	0	28
	X	3	18	2	0	23
	Y	0	1	22	14	37

Journal categorization using ESA index

As a comprehensive structure for journal categorization using the ESA index, after calculating this index from six bibliometric indicators, we applied k-means clustering to the dataset with seven features: CiteScore, SNIP, SJR, H-index, Eigenfactor Score, Journal IF, and ESA index. K-means clustering was applied to the full dataset with the seven features. Furthermore, a reduced set of features was obtained through PCA and t-SNE dimensionality reduction techniques. In this section, we demonstrate the experimental observations obtained using the datasets from two aspects. First, we analyze the effect of multiple bibliometric features (the seven features) and with the reduced set of features.

Clustering results on dataset with seven bibliometric features

K-means clustering was performed on the dataset for *k* ranging from 2 to 15. Different k values were evaluated because the number of clusters were unknown. For each cluster, various cluster evaluation metrics including the Silhouette Coefficient Score, Calinski-Harabasz score, and Davies-Bouldin Index was computed. This enabled the determination of the value of k at which most cluster validity indices provide the best results. Figure 6 shows the elbow method, Silhouette Coefficient Score is represented in Figure 7, the Calinski-Harabasz score is shown in Figure 8, and the Davies-Bouldin Index is represented in Figure 9. This helps us determine the optimal number of clusters and internal clustering validation results. We selected four clusters for comparison with the *Scopus* Quartiles and *SCImago* Best Quartiles (Q1-Q4). Various experiments were conducted using different k values. Figure 10 presents the k-means clustering results for (a) artificial intelligence and (b) computer vision and pattern recognition.

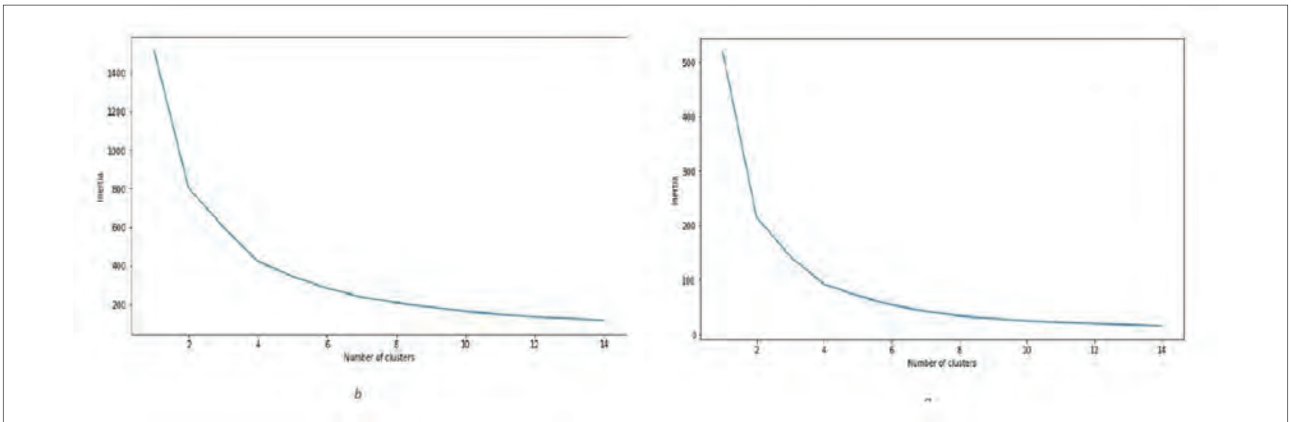


Figure 6. Elbow method (a) Artificial Intelligence (b) Computer Vision and Pattern Recognition

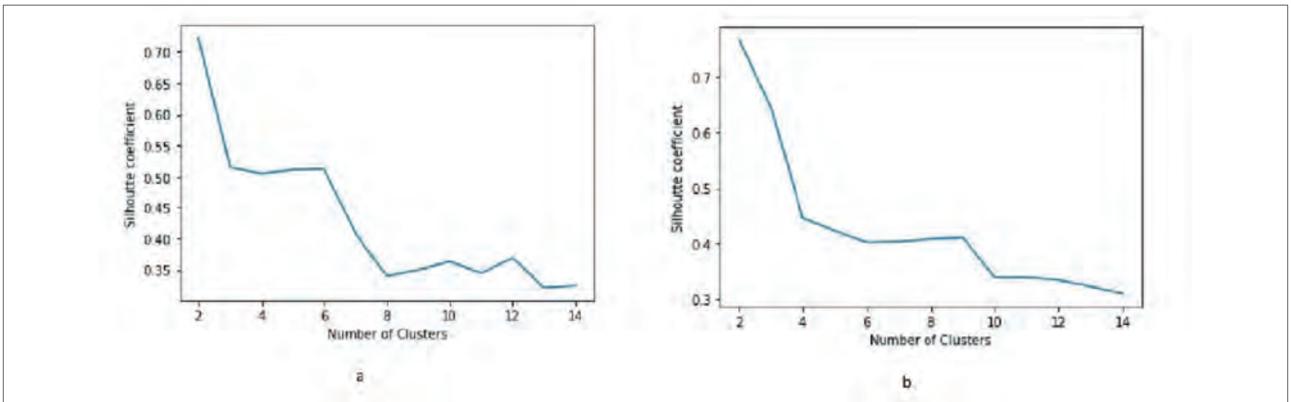


Figure 7 Silhouette Coefficient Score (a) Artificial Intelligence (b) Computer Vision and Pattern Recognition

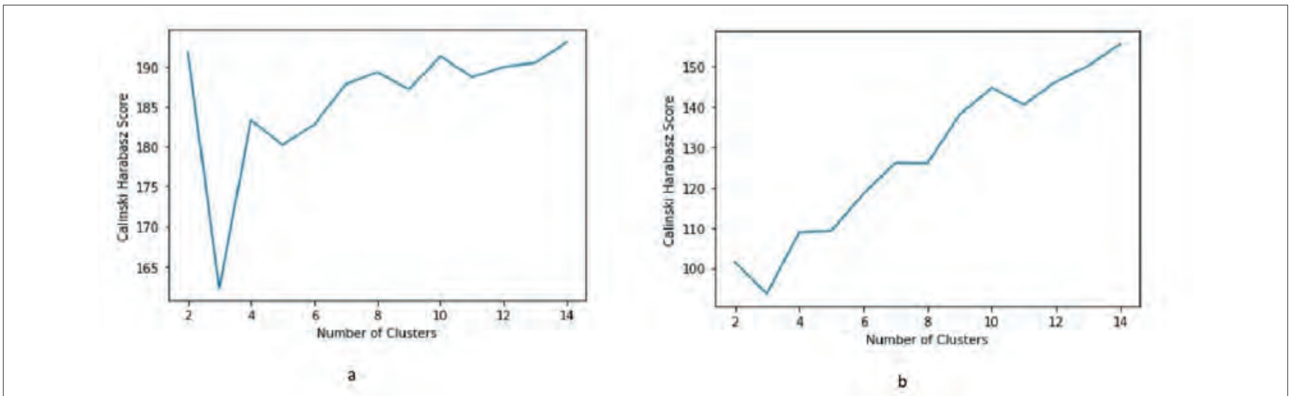


Figure 8. Calinski-Harabasz Score (a) Artificial Intelligence (b) Computer Vision and Pattern Recognition

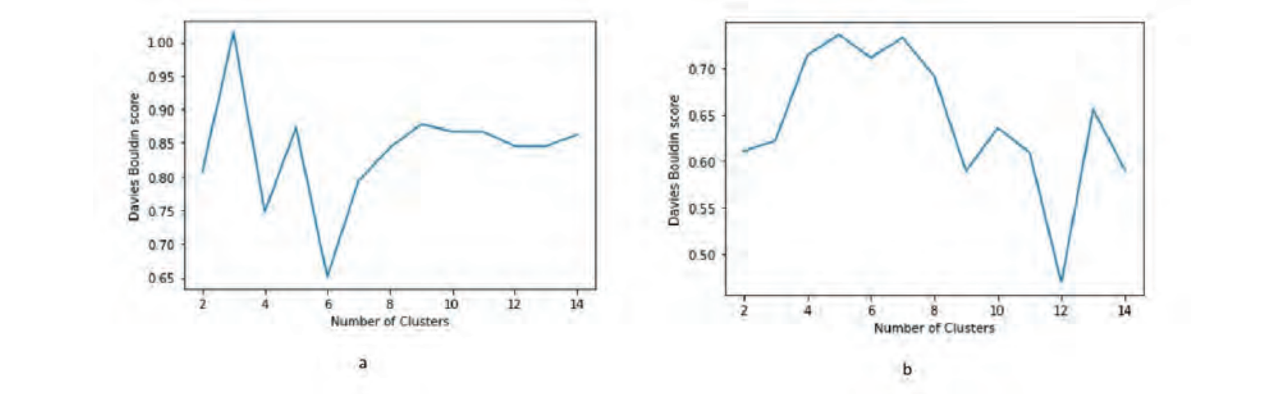


Figure 9. Davies Bouldin method (a) Artificial Intelligence (b) Computer Vision and Pattern Recognition

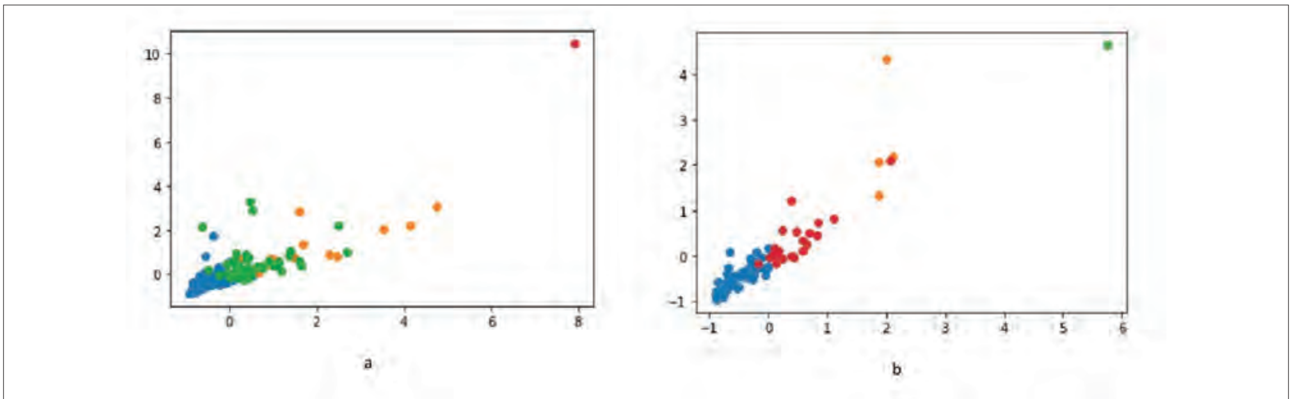


Figure 10. k-means clustering when k = 4 (a) Artificial Intelligence (b) Computer Vision and Pattern Recognition

Various clustering validity indices were calculated by applying the seven features as input variables. The *Scopus* Quartiles and *SJR* Best Quartiles were used as target labels. Therefore, in this case, four clusters were selected because the proposed experimental results could be compared with *Scopus* Quartiles and *SJR* Best Quartiles.

4.2.1. Comparison of ESA index with the existing benchmarks with all the features

The results of the proposed model were compared with the quartiles of journals from *Scopus* and *SJR*, which are the available standards worldwide. Different evaluation measures were used to measure journal performance.

Internal evaluation

The *Scopus* Quartiles and *SJR* Best Quartiles (*SCImago Journal & Country Rank*) provided in this study were used as benchmarks. *Scopus* categorize journals into four quartiles Q1-Q4. Here, Q1 is the top-ranking group. It is followed by Q2, Q3 and Q4 which is the lowest category. The *SCImago Journal & Country Rank* also classifies journals into four categories: Q1-Q4. The categories of each journal used as the input dataset were obtained from *Scopus* and *SJR*.

Scopus and *SJR* are two existing systems for journal categorization and rating. The current strategy differs primarily in that *Scopus* is based on set standards developed by certain statistical measures based on a single metric (i.e., it is calculated on the basis of the CiteScore, and *SCImago Journal & Country Ranks* use SJR for journal categorization). SJR is based on generic frameworks that learn automatically from data. In this experiment, six baseline features (i.e., the CiteScore, SNIP, SJR, H-index, Eigenfactor Score, and Journal Impact Factor) were used as input features. In addition, a new feature known as the ESA index was calculated and used in this analysis. The journal categories obtained in the proposed model were validated using different internal evaluation metrics. The results of this experiment are shown in Table 11 for artificial intelligence journals and in Table 12 for computer vision and pattern recognition journals.

Table 11. Internal clustering validity results for all features n = 7 (Artificial Intelligence)

No. of clusters	Silhouette score	Calinski-Harabasz score	Davies-Bouldin Index
K = 2	0.7103	191.7727	0.8072
K = 3	0.5154	162.3236	1.0136
K = 4	0.5052	183.3192	0.7478
K = 5	0.5112	180.2280	0.8713
K = 6	0.5127	184.2999	0.6512
K = 7	0.3831	187.6632	0.7859

Table 12. Internal clustering validity results for all features n = 7 (Computer Vision and Pattern Recognition)

No. of clusters	Silhouette score	Calinski-Harabasz score	Davies-Bouldin Index
K = 2	0.7658	101.5086	0.6111
K = 3	0.6852	87.9538	0.5589
K = 4	0.4457	108.7583	0.7142
K = 5	0.4226	109.2883	0.7366
K = 6	0.4015	118.3215	0.7175
K = 7	0.4071	126.0863	0.7317

External evaluation

Tables 13 and 14 present the k-means clustering validation results compared with the *Scopus* Quartiles and *SJR* Best Quartiles as ground-truth labels using different evaluation metrics for (1) artificial intelligence and (2) computer vision and pattern recognition, respectively. The *SJR* Quartiles showed relatively better results than the *Scopus* Quartiles.

Table 13. External validation of clustering labels with *Scopus* and *SCImago Journal Rank* (Artificial Intelligence) journals

Quartiles	ARI	MI	HCV	FM
Scopus-Q	0.0329	0.2450	0.2147	0.4373
SJR-Q	0.1754	0.3313	0.2724	0.4892

Table 14. External validation of clustering labels with *Scopus* and *SCImago Journal Rank* (Computer Vision and Computer Vision) journals

Quartiles	ARI	MI	HCV	FM
Scopus-Q	0.1132	0.2549	0.2526	0.4501
SJR-Q	0.2642	0.4210	0.3728	0.5277

4.3. Clustering results on the reduced dataset using PCA

Seven bibliometric indicators were selected as inputs for categorizing journals through k-means clustering. Then, PCA was applied. It transformed the seven-dimensional dataset into seven PCs. The variance explained by each PC based on the input dataset demonstrates that PCA can be used successfully in the categorization of journal datasets for dimensionality reduction because the first two PCs maintained approximately 89% of the variation for artificial intelligence journals and 94% for computer vision and pattern recognition journals.

Table 15. Internal clustering validity results for the reduced set of features PC1 and PC2 (Artificial Intelligence)

No. of clusters	Silhouette score	Calinski-Harabasz score	Davies Bouldin Index
K = 2	0.7366	241.5159	0.6785
K = 3	0.5679	217.2935	0.8123
K = 4	0.5641	271.5580	0.5888
K = 5	0.5592	293.3024	0.5085
K = 6	0.5601	305.7263	0.5595
K = 7	0.4639	338.2717	0.6228

Table 16. Internal clustering validity results for the reduced set of features PC1 and PC2 (Computer Vision and Pattern Recognition)

No. of clusters	Silhouette score	Calinski-Harabasz score	Davies Bouldin Index
K = 2	0.7874	117.7504	0.5591
K = 3	0.6772	118.1159	0.5432
K = 4	0.5362	154.2243	0.6043
K = 5	0.4805	167.1288	0.6620
K = 6	0.4836	202.6179	0.5595
K = 7	0.4950	250.6245	0.5217

We projected the k-means derived clusters onto 2D visuals after applying PCA to divide the dataset into two principal components. For the PCA, k-means was applied. By processing a lower-dimensional dataset via k-means, the score value increased from 0.50 to 0.56 (see Tables 15 and 16). A significant improvement in the capability to distinguish between clusters is observed in the 2D scatter plots. Table 17 and 18 display the external validity scores for artificial intelligence journals and computer vision and pattern recognition journals, respectively, when the *Scopus* and *SJR* Quartiles are employed as ground truth labels.

Table 17. External validation of PCA clustering labels with *Scopus* and *SCImago Journal Rank* (Artificial Intelligence) journals

Quartiles	ARI	MI	HCV	FM
Scopus-Q	0.0329	0.2450	0.2147	0.4373
SJR-Q	0.1850	0.3427	0.2812	0.4959

Table 18 External validation of PCA clustering labels with *Scopus* and *SCImago Journal Rank* (Computer Vision and Pattern Recognition) journals

Quartiles	ARI	MI	HCV	FM
Scopus-Q	0.0266	0.2523	0.2428	0.4217
SJR-Q	0.2104	0.3610	0.3160	0.5146

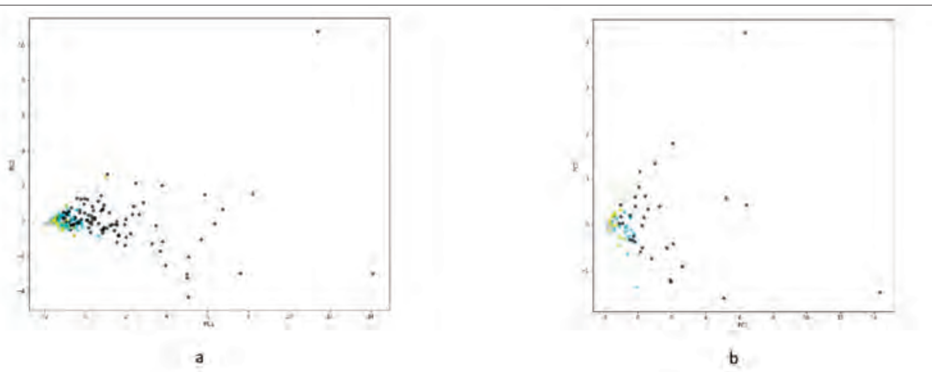


Figure 11. K-means clustering with PCA for k= 4 (a) Artificial Intelligence (b) Computer Vision and Pattern Recognition

4.4. Clustering results on the reduced dataset using t-SNE

In this section, we reduced our dataset using t-SNE and compared the k-means results with those of the PCA k-means. The dataset was reduced to two t-SNE components. The data tended to cluster into a large diffused cluster with a perplexity of 80 for artificial intelligence journals and 70 for computer vision and pattern recognition journals.

The Silhouette Coefficient Score, that we achieved by applying k-means to our two t-SNE-derived components was 0.42, whereas that we acquired by applying k-means to the two principal components of PCA was 0.56. The interpretation of t-SNE appears counterintuitive because the density of t-SNE clusters (i.e. low-dimensional space) is not proportionally related to data associations in the original (high-dimensional space) dataset. That is, although we can have good dense clusters generated by k-means, t-SNE may reveal these as broad or even numerous clusters. This is particularly so when the perplexity is excessively low. When interpreting the t-SNE plots, it is difficult to interpret the density, cluster size, number of clusters (under the same k-means cluster), and form. Although we can have numerous clusters for the same k-means cluster (particularly when the perplexity is significantly low), this has no bearing on the cluster quality. The distance and location of each k-means cluster are the key advantages of t-SNE. Although clusters that are closer together are more closely related to each other, this does not necessarily imply that clusters that are farther apart are proportionally dissimilar. Finally, we need to observe a particular level of separation between the k-means clusters, as shown by t-SNE.

Table 19. Internal clustering validity results for the reduced set of features t-SNE1 and t-SNE2 (Artificial Intelligence)

No. of clusters	Silhouette score	Calinski-Harabasz score	Davies-Bouldin Index
K = 2	0.5544	477.6020	0.6155
K = 3	0.4946	510.4948	0.6964
K = 4	0.4201	470.9313	0.8121
K = 5	0.4267	481.5732	0.8063
K = 6	0.4359	480.2519	0.7802
K = 7	0.4433	491.9881	0.7691

Table 20. Internal clustering validity results for the reduced set of features t-SNE1 and t-SNE2 (Computer Vision and Pattern Recognition)

No. of clusters	Silhouette score	Calinski-Harabasz score	Davies-Bouldin Index
K = 2	0.6244	118.4537	0.6302
K = 3	0.4878	135.8395	0.6597
K = 4	0.4794	151.3629	0.7169
K = 5	0.4968	162.5077	0.6577
K = 6	0.4944	173.9182	0.6564
K = 7	0.4984	172.9062	0.5650

We projected the k-means-derived clusters onto 2D visuals after using t-SNE to divide the dataset into two components. k-means was applied to t-SNE. By processing a lower-dimensional dataset via k-means, the score value increased from 0.50 to 0.48. This is shown in Table 19 for artificial intelligence and Table 20 shows results for computer vision and pattern recognition. A significant improvement in the capability to distinguish between clusters is observed in the 2D scatter plots. Tables 21 and 22 display the external validity scores when the *Scopus* and *SJR* Quartiles are employed as ground truth labels. Tables 21 and 22 display the external validity scores for artificial intelligence, and computer vision and pattern recognition respectively.

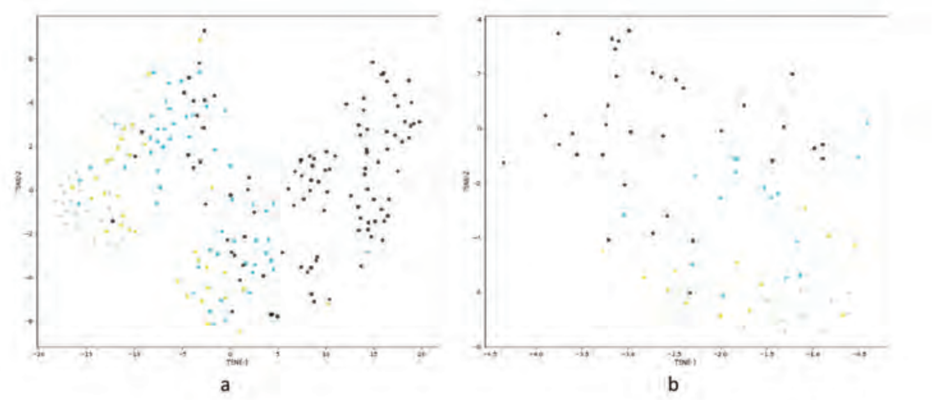


Figure 12. K-means clustering visualization with t-SNE (a) Artificial Intelligence (b) Computer vision and pattern recognition

Table 21. External validation of t-SNE clustering labels with *Scopus* and *SCImago Journal Rank* (Artificial Intelligence) journals

Quartiles	ARI	MI	HCV	FM
Scopus-Q	0.2391	0.3342	0.3655	0.4574
SJR-Q	0.3699	0.4331	0.4427	0.5265

Table 22. External validation of t-SNE clustering labels with *Scopus* and *SCImago Journal Rank* (Computer Vision and Computer Vision) journals

Quartiles	ARI	MI	HCV	FM
Scopus-Q	0.1032	0.2746	0.3021	0.3927
SJR-Q	0.3517	0.4764	0.4655	0.5448

5. Conclusions and future work

Various researchers have disapproved the evaluation of the scientific impact of a journal using an individual indicator such as the Journal Impact Factor (IF). Furthermore, the Journal IF is not only widely applied but also often misapplied. This has yielded biased and misleading results. Earlier, the HEC used the IF for research evaluations. Owing to the bias induced by individual indicators, a multi-metric journal prestige measurement system is necessary for journal quality estimation. In Pakistan, the *HEC Journal Recognition System* (*HJRS*) was launched in July 2020 to evaluate journals using proprietary JPI measures that divide journals into the W, X, and Y categories.
https://HJRS.hec.gov.pk

The *HJRS* is a multi-metric tool used to categorize journals based on the Eigenfactor Score, Article Influence (AI) Score, SCImago Journal Rank (SJR), SNIP, CD2, and H-index. However, it has few limitations: (1) CiteScore is a well-known journal-based metric launched by *Elsevier* (*Scopus*). It directly competes with the Journal Impact Factor (IF). It has not been used in the *HJRS* for journal categorization. (2) According to a few researchers, the decision-making mechanism of the *HJRS* is not satisfactory. This is because few journals that have been reported earlier in the W category have now been shifted to lower categories in the *HJRS*. This increases conflict rather than facilitating research in Pakistan. (3) Many researchers consider that HEC should reduce the threshold levels for certain categories. Therefore, the proposed study attempts to address these issues of *HJRS*. In this regard, a multi-metric-based approach was adopted from the SA index, which used two bibliometric measures: the IF and H-index. In this study, a multi-metric-based extended standardized average (ESA) index was developed using six metrics: CiteScore, SNIP, SJR, H-index, Eigenfactor Score, and Journal Impact Factor from three databases (*Scopus*, *SCImago Journal & Country Rank*, and *Web of Science*). The CiteScore was included to overcome the first issue of the *HJRS*. Second, the proposed model is not based on proprietary measures that makes the system transparent. The ESA index is strongly correlated with other well-known bibliometric indicators. Thus, this framework enhances the overall efficiency of journal ranking systems by aggregating multiple bibliometric indicators. The ESA index performed better than the SA Index and was highly correlated with all the other bibliometric indicators. Furthermore, a machine-learning based evaluation was performed on the proposed study to determine the combined impact of the ESA index with other metrics. In addition, k-means clustering coupled with dimensionality reduction techniques such as PCA and t-SNE was applied to identify hidden patterns in journal categorization. The proposed model examined the effectiveness of the journal prestige measurement system for all seven features and a reduced set of features. Based on the clustering evaluation measure and world benchmark bibliometric indices, we selected the optimum number of clusters as $k = 4$ (which indicated four clusters). The proposed model results were compared with the *Scopus* and *SCImago* Best Quartiles (Q1-Q4) and the *HJRS* Categories (W, X, and Y) using cross-tabulation. The results showed that compared with the use of the seven features for journal categorization, reduced/transformed features provided superior results with dimensionality reduction techniques such as PCA and t-SNE. It is concluded that the multi-metric ESA index can be used to facilitate the decision-making process with regard to the selection of venues for publishing research articles. Furthermore, the use of this index can also assist in predicting future performance of the selected journals.

There are several approaches to expand the scope of this study. It concentrated on computer science journals to construct the dataset. A convenient expansion would be to develop a dataset of other areas and subfields in the computer science domain (such as software, data communication and networks) and then, apply the proposed system to a new dataset to identify the patterns in other subjects. Furthermore, clustering, feature selection, and classification techniques can be used to further evaluate the framework. Other prestigious journal rankings can be used to compare the results. This would facilitate the examination of the patterns of journal popularity or decline over time.

6. Statements and declarations

The data collection, experimentation, and initial draft writing was carried out by the first author. The second author suggested the concept for the article, aided in analyzing the results, and revised the initial draft. The third author enhanced the experimental design and validated the authenticity of the experiments. The fourth author contributed to the refinement of the writing, organization of the concepts, research coordination, and professional editing.

7. References

Abbas, Naveed-Neem; **Ahmed, Tanveer**; **Shah, Syed-Habib-Ullah**; **Omar, Muhammad**; **Park, Han Woo** (2019). "Investigating the applications of artificial intelligence in cyber security". *Scientometrics*, v. 121, n. 2, pp. 1189-1211.
https://doi.org/10.1007/s11192-019-03222-9

Allen, Liz; **Jones, Ceri**; **Dolby, Kevin**; **Lynn, David**; **Walport, Mark** (2009). "Looking for landmarks: The role of expert review and bibliometric analysis in evaluating scientific publication outputs". *PLoS one*, v. 4, n. 6.
https://doi.org/10.1371/journal.pone.0005910

Beliakov, Gleb; **James, Simon** (2011). "Citation-based journal ranks: The use of fuzzy measures". *Fuzzy sets and systems*, v. 167, n. 1, pp. 101-119.
https://doi.org/10.1016/j.fss.2010.08.011

Bergstrom, Carl (2007). "Eigenfactor: Measuring the value and prestige of scholarly journals". *College & research libraries news*, v. 68, n. 5, pp. 314-316.
https://doi.org/10.5860/crln.68.5.7804

Bollen, Johan; **Van-de-Sompel, Herbert**; **Hagberg, Aric**; **Chute, Ryan** (2009). "A principal component analysis of 39 scientific impact measures". *PLoS one*, v. 4, n. 6.
https://doi.org/10.1371/journal.pone.0006022

Dellavalle, Robert P.; **Schilling, Lisa M.**; **Rodriguez, Marko A.**; **Van-de-Sompel, Herbert**; **Bollen, Johan** (2007). "Refining dermatology journal impact factors using PageRank". *Journal of the American Academy of Dermatology*, v. 57, n. 1, pp. 116-119.
https://doi.org/10.1016/j.jaad.2007.03.005

Duan, Lian; **Nasri, Farrokh**; **Paknejad, Javad** (2018). "A data-driven method for rating management information systems journals in the same scale of the Association of Business Schools Journal Guide". *Expert systems*, v. 35, n. 6.
https://doi.org/10.1111/exsy.12309

Ennas, Gianfranco; **Biggio, Battista**; **Di-Guardo, Maria-Chiara** (2015). "Data-driven journal meta-ranking in business and management". *Scientometrics*, v. 105, n. 3, pp. 1911-1929.
https://doi.org/10.1007/s11192-015-1751-y

Feng, Lin; **Zhou, Jian**; **Liu, Sheng-Lan**; **Cai, Ning**; **Yang, Jie** (2020). "Analysis of journal evaluation indicators : an experimental study based on unsupervised Laplacian score". *Scientometrics*, v. 124, pp. 233-254.
https://doi.org/10.1007/s11192-020-03422-8

Fernández-Cano, Antonio; **Fernández-Guerrero, Inés M.** (2017). "A multivariate model for evaluating emergency medicine journals". *Scientometrics*, v. 110, n. 2, pp. 991-1003.
https://doi.org/10.1007/s11192-016-2197-6

Ferrer-Sapena, Antonia; **Sánchez-Pérez, Enrique A.**; **Peset, Fernanda**; **González, Luis-Millán**; **Aleixandre-Benavent, Rafael** (2016). "The impact factor as a measuring tool of the prestige of the journals in research assessment in mathematics". *Research evaluation*, v. 25, n. 3, pp. 306-314.
https://doi.org/10.1093/reseval/rvv041
https://academic.oup.com/rev/article-abstract/25/3/306/2364555

Fersht, Alan (2009). "The most influential journals: Impact factor and eigenfactor". *Proceedings of the National Academy of Sciences of the United States of America*, v. 106, n. 17, pp. 6883-6884.
https://doi.org/10.1073/pnas.0903307106

González-Pereira, Borja; **Guerrero-Bote, Vicente P.; De-Moya-Anegón, Félix** (2010). "A new approach to the metric of journals scientific prestige: The SJR indicator". *Journal of informetrics*, v. 4, n. 3, pp. 379-391.
https://doi.org/10.1016/j.joi.2010.03.002

Halim, Zahid; Khan, Shafaq (2019). "A data science based framework to categorize academic journals". *Scientometrics*, v. 119, n. 1.
https://doi.org/10.1007/s11192-019-03035-w

Holmberg, Kim; **Park, Han Woo** (2018). "An altmetric investigation of the online visibility of South Korea-based scientific journals". *Scientometrics*, v. 117, n. 1, pp. 603-613.
https://doi.org/10.1007/s11192-018-2874-8

Hsu, Wen-Chin; **Tsai, Chih-Fong**; **Li, Jia-Huan** (2015). "A hybrid indicator for journal ranking: An example from the field of Health Care Sciences and Services". *Online information review*, v. 39, n. 7, pp. 858-869.
https://doi.org/10.1108/OIR-11-2014-0277

James, Chris; **Colledge, Lisa**; **Meester, Wim**; **Azoulay, Norman**; **Plume, Andrew** (2018). "CiteScore metrics: Creating journal metrics from the *Scopus* citation index". *ArXiv*, 16 pp.
https://onlinelibrary.wiley.com/doi/epdf/10.1002/leap.1246

Are Brazilian innovation systems innovative? Regional and sectorial decompositions of triple-helix synergies

Mariza Almeida; Igone Porto-Gómez; Loet Leydesdorff

Nota: Este artículo se puede leer en español en:
https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/87536

Recommended citation:

Almeida, Mariza; **Porto-Gómez, Igone**; **Leydesdorff, Loet** (2023). "Are Brazilian innovation systems innovative? Regional and sectorial decompositions of triple-helix synergies". *Profesional de la información*, v. 32, n. 7, e320707.

https://doi.org/10.3145/epi.2023.dic.07

Article received on July 11th 2023
Approved on August 7th 2023

Mariza Almeida ✉
https://orcid.org/0000-0001-9442-467X

Universidad Federal del Estado de Rio de Janeiro
Escuela de Ingeniería Industrial
Av. Pasteur, 458, Edificio IBIO/CCET, sala 403N
Urca, CEP 22290-255 Brazil
mariza.almeida@unirio.br

Igone Porto-Gómez
https://orcid.org/0000-0003-2865-4818

Universidad del País Vasco (*UPV/EHU*)
Departamento de Ingeniería de Expresión Gráfica y Proyectos
Plaza del Ingeniero Torres Quevedo, 1
48013 Bilbao (Vizcaya), Spain
igone.porto@ehu.eus

Loet Leydesdorff (posthumous)
https://orcid.org/0000-0002-7835-3098

University of Amsterdam
Amsterdam School of Communication Research (ASCoR)
PO Box 15793
1001 NG Amsterdam, The Netherlands

Abstract

A knowledge-based economy adds innovation as another dynamic to a political economy. Whereas a political economy is institutionalized -for example, in a nation state- the knowledge base is volatile although it leaves footprints behind by transforming the institutional layers. This transformation into new options can be measured as redundancy using the Triple-Helix indicator. The balance between historical entropy generation and the knowledge-based generation of options can be measured in terms of positive and negative contributions to the prevailing uncertainty. At what scale and in which sectors is synergy among geographical, technological, and organizational distributions of firms evident? Using mutual information in the three dimensions as an indicator, we analyse a dataset of more than 16 million firms in Brazil and compute synergy within and across states and sectors in this country in terms of bits of information. The results suggest that no synergy is generated at the national level. The political economy of the country has not (yet) been transformed into a national innovation system. At state level, synergies vary according to geographical levels and sectors due to the specifics of the states. Above-average values were found for some states in the South and Southeast Regions. Also, the political capital, Brasilia, has resulted to have no impact in the innovation system of Brazil.

Keywords

Innovation systems; Triple helix; Universities; Government; Industry; Synergies; Geography; Technology; Brazil.

1. Introduction

A number of models have been created over the last decades to explain the feedbacks among technological developments, the generation of innovations, and economic development resulting from these activities. Are innovation systems national, regional, sectorial, supra-regional? In the 1960s-1970s, government programs and projects in Latin America drew on Jorge Sabato's triangle model which proposed to promote innovation at the national level. The approach was based on the multiple and coordinated action of three key elements represented by the geometry of a triangle: government, productive structure and scientific-technological infrastructure, with the government playing a leading role in coordinating the actions of universities and the productive sector (**Sabato**; **Botana**, 1970). Although the role of the knowledge base was envisaged, the model eventually served "import substitution" as a national strategy endowed with technological capacities. Technological development remained unexplained (cf. **Nelson**; **Winter**, 1977; 1982).

Lundvall (1999) proposed to distinguish between national business systems (NBS) and national systems of innovation (NSI). The concept of national business systems is related to the constituent elements of the national system with its structural interconnections. The differences between countries are explained by the organization of the firm and the firm's behavior, due to differences in culture and formal institutions. The central aspect of this approach, however, remains the coordination of economic activities and governance, and therefore political economy.

There are three main differences between the NBS and NIS models:

(i) while NBS considers economic coordination and governance, NIS defines innovation;

(ii) NBS seeks to explain the motivations of companies and how they organize themselves whereas NIS investigates the functioning of the national economy and its performance in terms of economic development;

(iii) Different ways of using the term "system": NBS regards a system as a combination of elements in different patterns; NIS emphasizes the processes in which agents interact (**Lundvall**, 1999).

Initially, the NIS approach was based on experiences in Europe and North America, but more recently several studies have drawn on data from Latin America, African and Asian countries. NIS has been used in different contexts in developed and developing countries, considering that the main elements provide a flexible and conceptual, methodological and analytical framework. Differences among NIS in developed and developing countries have been explained in relation to four dimensions:

(i) orientations based on different needs,

(ii) the key actors and respective incentives systems are different,

(iii) institutional frameworks are less formalized in developing countries and

(iv) existing rules are also less enforceable (**Altenburg**, 2011).

University-industry-government relations are key elements of the dynamics and processes in innovation systems (**Etzkowitz**; **Leydesdorff**, 1995; 2000). In addition to an institutional network model, "triple helix" models assume that three functions are combined: wealth generation, novelty production, and governmental control. The institutional arrangements are not *sui generis*, but co-evolving with the generation of synergies in these (functional) relations.

Mutual information between geographical, organizational, and technological distributions of the firms in a region, helps to measure the interactions between the triple helix organizations. Such information measures the increase or decrease of uncertainty in the ties among the stakeholders. This methodology evaluates the difference between the information (I) generated in the relationships versus redundancy (R), which is generated through the repetitions and overlaps in the interactions between the variables analyzed (**Leydesdorff**, 2003; **Park** *et al.*, 2005; **Leydesdorff**; **Sun** 2009; **Park**; **Leydesdorff**, 2010; **Ye** *et al.*, 2013; cf. **Ulanowicz**, 1986, p. 143. The three dimensions considered are: firms, as industrial production players, university, as the main knowledge producer and Government as the main institutional stakeholder, corresponding to the 3 subsystems in an innovation system (**Edquist**, 1997).

The triple helix model explains social and economic development as occurring through interactions among universities, industries, and governments. The model can be applied to national, regional, and local environments. The complexity of a triple helix model is a result of the local trajectories observed in each region or country (**Leydesdorff**; **Etzkowitz**, 1996). A continuous process of interactions emerges at the interfaces among geographical scales, technological capacities, and organization (firms), causing an overlay of negotiations and exchanges. New options for innovations are generated in the overlaps, due to the interactions between the helixes. The triple-helix indicator enables us to measure and explain the synergy in university-industry-government relations based on the overlays of information communicated in an innovation system (**Leydesdorff**; **Fritsch**, 2006).

The objective of this study is to analyze the synergy among geographical, technological, and organizational distributions of firms in Brazil at different scales, levels. The paper is organized into the following sections, in addition to this introduction. The next section presents a review of the Brazilian innovation system. Section 3 exhibits the literature underpinning triple helix model developed to measure innovation systems. Section 4 explains the methodology that oriented the research and the main descriptive statistics of the data. Section 5 describes the results in both geographical and technological perspectives. Finally, we offer some concluding remarks about innovation performance in Brazil.

2. Is there a Brazilian innovation system?

In Brazil, a double pattern of behavior related to innovation has been observed. On the one hand, the country has achieved success in the development of some technologies, such as deep-water oil exploration carried out by *Petrobras*, the production of airplanes for regional flights, by *Embraer*, and the growth of productivity in agriculture and livestock, led by *Embrapa*. State-owned enterprises have taken part in these successful endeavors, with privatized ones appearing on the scene recently. On the other hand, the country has not built a mature innovation system with diversified interactions between research institutions and the productive sector.

Brazil is an interesting case due to the fact that other studies using different methodologies considered Brazil's innovation system as "immature" (**Albuquerque**, 2000; 2008). An immature innovation system was defined by **Albuquerque** (2000) in the following terms:

1) a large share of specific individuals in patenting activities;
2) little firm ïnvolvement in innovative activities;
3) lack of continuity in patenting activity,
4) low sophistication of inter-firm technological division, showing sectors with technological advances and other less developed ones,
5) declining role of the machinery sector, which is important for the catching up process,
6) foreign companies established in the country develop incremental innovations,
7) patents registered in Brazil are not considered very innovative by international offices.

Corroborating this analysis, the Brazilian patent ranking indicates that between 2014 and 2019, nineteen of the twenty-five largest patent depositors of products and services were from higher education institutions (*INPI*, 2021), highlighting the absence of companies to lead this process. Also, in this sense, the study by **Pacheco** (2019) considered the Brazilian innovation system "weak" due to the federal government's failure to prioritize the innovation agenda. The efforts undertaken are considered by this author as limited and disconnected from the country's general strategy.

Brazil is one of the BRICS member countries, and is also classified by the *World Bank* as an upper-middle-income country with a GDP (gross domestic product) of US$ 1,445 trillion (2020), with the economy gradually emerging after four years of recession. The country's imports last year amounted to US$ 276,032 billion, while total exports were US$ 243,739 billion.

https://data.worldbank.org/indicator/NY.GDP.MKTP.CD?locations=BR
https://data.worldbank.org/indicator/BM.GSR.TOTL.CD?locations=BR
https://data.worldbank.org/indicator/NE.EXP.GNFS.CD?locations=BR

However, twelve of Brazil's fifteen major export products are commodities and represent 71% of the country's total exports (**Canuto** *et al.*, 2013). As a consequence of this mix of products, the country has the ability to produce and export products associated with a high level of inequality (**Hartmann** *et al.*, 2017). According to the *Economic Complexity Index Brazil* is the 37th most complex economy, which corresponds with a problematic economic context (**Oreiro** *et al.*, 2020). The total population is 231 million (*OECD*, 2018) and the country remains one of the most unequal in the world where half of the population receives 10% of total household incomes, while another half holds 90% (*OECD*, 2018).

3. Operationalization

University-industry-government relations shape an ecosystem of bi and trilateral relations which can promote innovative production, prosperity for the territory and a legal framework within the innovation system. For this reason, the quality and intensity of the relationships maintained become crucial (**Leydesdorff**, 2006). In the words of **Lengyel** and **Leydesdorff** (2011, p. 6),

> "the triple helix model enables one to distinguish knowledge functions in innovation systems in addition to the two main dimensions of a political economy".

The knowledge-based economy is built on the relationships between the drivers of a political economy in terms of knowledge creation, knowledge transfer and knowledge control (**Nelson**; **Winter**, 1982; **Whitley**, 1984; 2001).

Different studies have analyzed the reduction of uncertainty at the systems level in Europe, using the triple helix indicator of synergy in the knowledge base of an economy. Examples include the Netherlands (**Leydesdorff** *et al.*, 2006), Sweden (**Leydesdorff**; **Strand**, 2013), Germany (**Leydesdorff**; **Fritsch**, 2006), Hungary (**Lengyel**; **Leydesdorff**, 2011), Norway (**Strand**; **Leydesdorff**, 2013), Spain (**Leydesdorff**; **Porto-Gómez**, 2019), and the USA (**Leydesdorff** *et al.*, 2019).

The novelty of the present study on Brazil lies in its focus on the entire economy, built around the sectorial classification made with *NACE* codes (*Nomenclature statistique des Activités économiques dans la Communauté Européenne*). While previous articles focus solely on knowledge intensive activities, this study takes into account not only high-tech innovation activities but also the medium-low and less knowledge-intensive services being performed. In this way, we might be able to gain a clearer picture of the Brazilian regions, depending on their strength.

We use three variables and their interactions to measure the performance and synergy:

(1) the geographical situation of the firms through post codes, in order to pinpoint the region;
(2) the *NACE* code in order to clarify the technological knowledge base of the firm; and
(3) the firm size by number of employees, as a measure of organizational structure.

Technology will be represented by the sector classification (*NACE Rev. 2*), organizations by the respective company sizes in terms of the number of employees, and the geographical dimension by the zip codes extracted from the address information.

4. Methodology

The (Shannon-type) information in three dimensions can be decomposed into groups as follows (**Leydesdorff**; **Strand**, 2013, p. 1895; **Theil**, 1972):

$$T = T_0 + \sum_G \frac{n_G}{N} T_G$$

where,

- T_0 is the inter-territorial uncertainty,
- T_G is the uncertainty on the geographical scale G,
- N_G is the number of firms on the specific geographical scale G,
- N is the total number of firms in the analysis.

A negative value of T_0 can be considered an indication of additional synergies at higher geographical levels.

4.1. Data and descriptive statistics

The dataset was downloaded from the *Orbis* database of *Bureau van Dijk* on November 13, 2018, using the strings: "All active companies and Companies with unknown situation combined (with a Boolean AND) with "World Region/Country/Region is country: Brazil" the total number of retrieved Brazilian companies was 21,296,980. The data were downloaded in 22 batches of 100,000 records. From this total, 15,957,292 records contained complete information on the three dimensions for the analyses: that is, zip codes, *NACE* codes and number of employees.

The geographical dimension presented in Table 1 provides the distribution of firms across the Brazilian states. There is an unequal distribution of firms across Brazilian states. The companies are concentrated in the Southeast Region. The states with the largest number of firms are São Paulo (28.4% of all firms in the sample), followed by Minas Gerais with 10.4% and Rio de Janeiro with 8.7%.

Table 1. Distribution of sampled firms across the Brazilian states.

State name	Number of firms	%	GDP trillion US$ (2018)[1]	% GDP
Acre	175,794	1.10%	4,196	0.21%
Alagoas	154,878	0.97%	14,893	0.77%
Amazonas	60,493	0.38%	27,401	1.42%
Amapá	75,068	0.47%	4,597	0.23%
Bahia	901,861	5.65%	78,349	4.10%
Ceará	509,953	3.20%	42,671	2.25%
Distrito Federal	46,961	0.29%	69,745	3.63%
Espírito Santo	348,477	2.19%	37,503	1.95%
Goiás	204,879	1.28%	53,559	2.70%
Maranhao	281,592	1.76%	26,873	1.40%
Minas Gerais	1,673,231	10.49%	168,293	8.70%
Mato Grosso do Sul	222,567	1.39%	29,278	1.52%
Mato Grosso	341,745	2.14%	37,619	1.96%
Pará	539,330	3.38%	44,169	2.30%
Paraíba	207,302	1.30%	17,619	0.91%
Pernambuco	505,102	3.17%	51,007	2.70%
Piauí	158,183	0.99%	13,788	0.71%
Paraná	1,084,840	6.80%	120,437	6.28%
Rio de Janeiro	1,399,610	8.77%	207,702	11.00%
Rio Grande do Norte	198,789	1.25%	18,330	0.95%
Rondônia	195,700	1.23%	12,293	0.64%
Roraima	75,984	0.48%	3,659	0.20%
Rio Grande do Sul	1,075,389	6.74%	125,163	6.52%
Santa Catarina	723,142	4.53%	81,626	4.25%
Sergipe	101,901	0.64%	11,500	0.60%
São Paulo	4,537,365	28.44%	605,037	31.60%
Tocantins	157,156	0.98%	9,792	0.50%
Total	15,957,292	100%	1,917	100%

Source: based on *Orbis data*, 2018, [1]IBGE (2020).

The second dimension shown in Table 2 is technology. We use the economic activity based on the four-digit sector classifications from the *NACE, the industry standard classification system used in the European Union. The current version is revision 2 and was established by Regulation (EC) No. 1893/2006. It is the European* implementation of the *UN International Standard Industrial Classification* (*ISIC*), revision 4, and allows comparison of companies according to the type of technology used by them.

The sector-based analysis focuses on the sectors of high-tech manufacturing (HTM), medium-high-tech manufaturing (MHTM) and knowledge-intensive services (KIS). Brazil has adopted the *National Classification of Economic Activities* (*CNAE*) in the production of economic statistics, which is derived from the *UN* classification *ISIC*, revision 4, enabling the use of *NACE* classification to make comparisons possible with results from previous studies on other countries. The *NACE* classification for Brazilian companies was carried out by *Orbis*.

Table 2. *NACE* classifications (Rev. 2) of high and medium-tech manufacturing industries and knowledge services. Sources: *Eurostat/OECD* (2011); cf **Laafia** (2002, p. 7) and **Leydesdorff** *et al.*, (2006, p. 186).

Manufacturing	**Services**
High-tech manufacturing (HTM) 24.4 Manufacture of pharmaceuticals, medicinal chemicals and botanical products 30 Manufacture of office machinery and computers 32 Manufacture of radio, television and communication equipment and apparatus 33 Manufacture of medical, precision and optical instruments, watches and clocks 35.3 Manufacture of aircraft and spacecraft **Medium-high-tech manufacturing (MHTM)** 24 Manufacture of chemicals and chemical products excluding excluding 24.4 Manufacture of pharmaceuticals, medicinal chemicals and botanical products 29 Manufacture of machinery and equipment N.E.C. 31 Manufacture of electrical machinery and apparatus N.E.C. 34 Manufacture of motor vehicles, trailers and semi-trailers 35 Manufacture of other transport equipment excluding 35.1 Building and repairing of ships and excluding 35.3 Manufacture of aircraft and spacecraft	**Knowledge-intensive services (KIS)** 61 Water transport 62 Air transport 64 Post and telecommunications 65 to 67 Financial intermediation 70 to 74 Real estate, renting and business activities 80 Education 85 Health and social work 92 Recreational, cultural and sporting activities Of these sectors, 59 to 63, and 72 are considered high-tech services.

Table 3 shows the distribution by number of employees. The majority of the companies are small: 58.6% have from 2 to 4 employees and 33.1% from 5 to 9 employees.

5. Findings

The country is divided into 27 states, which vary greatly in terms of size, population, geographical characteristics, GPD, economic activities and number and size of companies. Considering the analysis conducted decomposed by states, we present some of the main characteristics of each one in Table 1. The purpose is to strengthen understanding of the results obtained on synergy between the companies.

Figure 1 shows a map of Brazil with the states coloured according to their respective contribution to synergy generation in Brazil's innovation system.

Table 3. Size distribution of the firms in the sample by number of employees.

Number of employees	Number of companies	Percentage
None	0	0.0%
0-1	0	0.0%
2-4	9,218,577	57.774%
5-9	5,394,428	33.805%
10-19	658,251	4.124%
20-49	552,443	3.462%
50-99	94,690	0.593%
100-199	19,098	0.120%
200-499	12,137	0.076%
500-749	61	0.0004%
750-999	4,234	0,027%
≥1000	3,373	0,021%
Total	16,261,721	100%

Source: based on *Orbis data*, 2018.

Figure 1. Synergy generation at the level of 27 states in Brazil (NUTS2).
Source: based on *Orbis* data, 2018; using *SPSS* for the mapping.

The analyses of the results are divided into two levels: 1) the geographical perspective showing the synergy levels in the different states (NUTS 1), provinces (NUTS 2, and municipalities (NUTS 3); and 2) synergy levels in the states considering the technological activities of the firms in the sample.

5.1. Decomposition at the geographical level of states

Figure 1 shows a map of Brazil with the coloured states (NUTS2) to visualise their contribution to the generation of synergy at the national level. The total synergy of Brazil is T= - 113 mbits. 50.5% comes from economic activities in 4 states: São Paulo (-25.1114243 mbits or 22.2%), Minas Gerais (-12.99 mbits or 11.48%), Paraná (10.0360304 mbits or 8.87%) and Rio de Janeiro (8.99 mbits or 7.95%).

The differentiation between the states observed in Table 1 in terms of GDP and number of companies is reflected in the synergy between companies. The five states with the largest share of the country's GDP in 2018 are São Paulo (31. 9%), Rio de Janeiro (11%), Minas Gerais (9 %), Rio Grande do Sul (6. 6%) and Paraná (6. 3%).

The states of São Paulo, Minas Gerais and Rio de Janeiro are located in the Southeast Region and Rio Grande do Sul and Paraná in the South Region, indicating that innovation activity is more densely concentrated in those regions revealing economic inequality characteristics. Table 4 shows the main products in terms of economic value of these five states.

Table 4. Main products of the states with higher GDP and synergy (Source: *IBGE*, 2020).

States	High-tech manufacturing (HTM)	Medium high-tech manufaturing (MHTM)	Knowledge-intensive services (KIS)	Others
São Paulo	Airplanes, mobile phones and smartphones	Cars and parts and accessories	Largest financial centre in Latin America and where most national and international airlines are based	Relevant production of petroleum refining for fuel production, agricultural production of sugarcane and ethanol
Rio de Janeiro	-	Cars,parts and accessories, steel industry products	Specialized services for petroleum extraction and aircraft maintenance	It is the largest producer of petroleum in the country and also has refineries
Minas Gerais	-	Coils and steel plates	-	Mining (iron, niobium, gold),pig iron, fuel alcohol, meat production, rations for animal feed, fertilizers
Rio Grande do Sul	-	Cars and parts and accessories	-	Agricultural products (rice, tobacco), diesel fuel, rations for animal feed, fertilizers
Paraná	-	Cars	-	Petroleum, rations for animal feed, fertilizers

São Paulo is also the state with the largest population in the country, including a higher rate of urban population. The largest Brazilian universities and research centers are located in these states, in the cities of São Paulo (*University of São Paulo*), Campinas (*University of Campinas*) and São Carlos (*Federal University of São Carlos*). A considerable number of spin-offs are based in these cities, along with a high rate of PhDs residing there. The highest rate of R&D investment originating from state governments is also in this state. The average monthly salary is in the highest bracket observed in the country (400 dollars), although there is internal inequality in the distribution.

The main economic sectors in the Brazilian states mentioned in Table 1 whose GDP corresponds proportionally to less than 1% of the country's GDP include agriculture, livestock and agricultural processing: Acre, Alagoas, Amapá and Roraima. The states in this group that show other economic activities are Paraíba (textile industry, footwear and manufacture of non-metallic mineral products – cement), Amapá and Rondônia with gold and tin mining, respectively, and Sergipe (oil extraction).

5.2. Decomposition at the technological level

We decommposed the data into three different technological sectors: high-tech manufacturing, medium high-tech manufacturing, and high-tech knowledge-intensive services. Analysis was also carried out to identify each of the 27 Brazilian states' contribution to total synergy. The results are presented in Table 5. Figure 2 shows the total synergy and decreasing contributions of states. Figure 2 shows the generation of synergy by regions and (states) in descending order subdivided by each sector. The 10 states displayed, are the ones with the highest contribution. São Paulo is the leading state for innovation in Brazil, in the four groups of HTM, MHTM, HTKIS and KIS sectors.

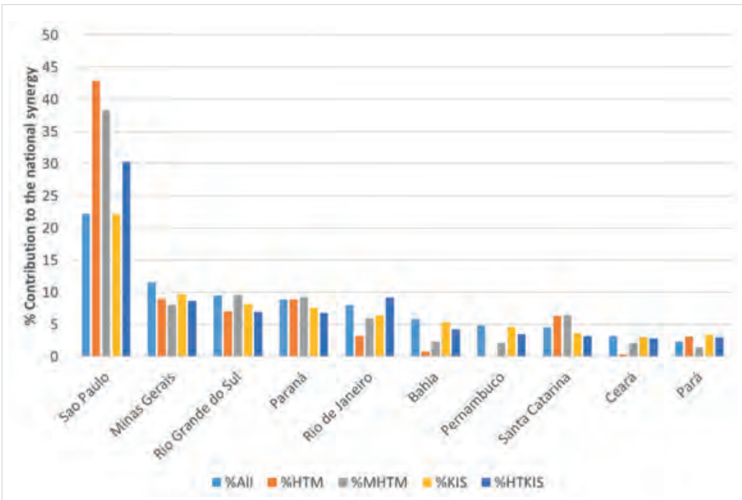


Figure 2. Contribution to national synergy. State levels.
Source: *Orbis* data.

Table 5. Synergy contribution for all 27 states and their respective economic sectors.

Federal states (a)		All sectors			High tech manufacturing				Medium-high tech manufacturing			Knowledge-intensive services			
		N of firms (b)	ΔT (mbit) (c)	% (d)	N of firms (e)	ΔT (mbit) (f)	Synergy (g)	% (h)	N (i)	ΔT (mbit) (j)	% (k)	N (l)	ΔT (mbit) (m)	Synergy (n)	% (o)
	Brazil	15,957,292	-113	100		-629,839	-629,839	100	76,477	-758,559	100	2,181,007	-202,204	-202,204	100
1	Acre	175,794	-1,1823934	1,045481	19	1,197855	0	0	471	-2,63462	0,35	18,975	-1,88065	-1,88065	0,930074
2	Alagoas	154,878	-1,14551394	1,012872	20	2,153378	0	0	396	-1,9105	0,25	17,741	-1,81105	-1,81105	0,895655
3	Amazonas	60,493	-0,66243415	0,585729	20	1,308926	0	0	88	-0,03952	0,01	19,709	-1,33696	-1,33696	0,661194
4	Amapá	75,068	-0,53461297	0,472709	3	0	0	0	209	-0,31002	0,04	6,847	-0,85084	-0,85084	0,420784
5	Bahia	901,861	-6,5648188	5,804662	218	-4,86508	-4,86508	0,772432	2,706	-17,6721	2,33	105,524	-10,8631	-10,8631	5,372351
6	Ceará	509,953	-3,60458867	3,187204	133	-2,0685	-2,0685	0,328418	2,316	-15,4491	2,04	53,631	-6,16542	-6,16542	3,049111
7	Distrito Federal	46,961	-0,34989234	0,309377	5	0	0	0	91	-0,08144	0,01	4,271	-0,49581	-0,49581	0,245201
8	Espírito Santo	348,477	-2,1866814	1,933480	78	-1,68697	-1,68697	0,267841	1,312	-7,45812	0,98	42,105	-3,88223	-3,88223	1,919955
9	Goiás	204,879	-1,92582611	1,702830	137	-0,63063	-0,63063	0,100126	1,485	-12,1284	1,60	29,009	-3,24273	-3,24273	1,603694
10	Maranhão	281,592	-1,58080918	1,397763	29	1,036982	0	0	590	-2,71377	0,36	29,875	-3,01955	-3,01955	1,493317
11	Minas Gerais	1,673,231	-12,9944568	11,489800	990	-56,437	-56,437	8,960537	7,291	-61,4073	8,10	220,154	-19,5854	-19,5854	9,685952
12	Mato Grosso do Sul	222,567	-1,72704575	1,527067	36	1,023815	0	0	661	-5,1514	0,68	25,548	-2,92628	-2,92628	1,447193
13	Mato Grosso	341,745	-2,65703323	2,349369	50	-0,08793	-0,08793	0,013961	1,220	-9,16467	1,21	38,143	-4,18102	-4,18102	2,067723
14	Pará	539,330	-2,68824462	2,376966	347	-19,6604	-19,6604	3,121492	1,635	-11,0909	1,46	61,093	-6,94257	-6,94257	3,433450
15	Paraíba	207,302	-1,30466216	1,153592	49	1,598913	0	0	733	-3,20171	0,42	22,394	-2,0493	-2,0493	1,013480
16	Pernambuco	505,102	-5,50371897	4,866430	137	3,896228	0	0	2,036	-16,428	2,17	65,825	-9,33074	-9,33074	4,614517
17	Piauí	158,183	-0,56569838	0,500195	21	2,284935	0	0	374	-1,3554	0,18	16,518	-1,6486	-1,6486	0,815318
18	Paraná	1,084,840	-10,0360304	8,873933	797	-55,831	-55,831	8,864320	7,434	-69,8901	9,21	140,575	-15,3773	-15,3773	7,604867
19	Rio de Janeiro	1,399,610	-8,99700673	7,955221	510	-20,143	-20,143	3,198121	4,402	-45,1463	5,95	232,414	-12,9227	-12,9227	6,390925
20	Rio Grande do Norte	198,789	-1,27215666	1,124850	38	0,760363	0	0	675	-4,16316	0,55	24,518	-1,90754	-1,90754	0,943372
21	Rondônia	195,700	-1,67029474	1,476887	93	-1,84425	-1,84425	0,292813	978	-7,78521	1,03	22,776	-3,07319	-3,07319	1,519847
22	Roraima	75,984	-0,72563895	0,641615	25	1,385841	0	0	125	-0,57058	0,08	13,047	-1,51357	-1,51357	0,748537
23	Rio Grande do Sul	1,075,389	-10,7514673	9,506528	719	-44,1559	-44,1559	7,010667	7,503	-72,5133	9,56	146,920	-16,4781	-16,4781	8,149244
24	Santa Catarina	723,142	-5,11843412	4,525758	532	-39,8123	-39,8123	6,321022	5,908	-48,5664	6,40	95,844	-7,33631	-7,33631	3,628173
25	Sergipe	101,901	-0,98783229	0,873449	23	0,727327	0	0	325	-1,29843	0,17	13,386	-1,82452	-1,82452	0,902315
26	São Paulo	4,537,365	-25,1114243	22,203710	3196	-270,088	-270,088	42,882080	24,851	-290,679	38,32	685,010	-44,8181	-44,8181	22,164780
27	Tocantins	157,156	-1,24691059	1,102528	18	1,319743	0	0	383	-2,4555	0,32	16,230	-2,0019	-2,0019	0,990040
			-113			-498,617	-498,617	79,16572	76,198	-711,265	93,77		-187,466	-187,466	92,711070

Source: based on *Orbis* data, 2018

In the high-tech manufacturing sectors the results for the states with the largest share of synergy are similar to those found for Brazil's overall economy. The largest share of synergy can be observed in the state of São Paulo: 42.9% of Brazil's total synergy, while Minas Gerais and Paraná contribute 8.9% and 8.9% respectively (check Table 3).

Several Brazilian states registered zero synergy in high-tech manufacturing, the majority of which are located in the North and Northeast Regions, the poorest in the country. The North Region, which comprises most of the Amazon Forest, has the lowest population density in the country, weak infrastructure (roads, telephony, electricity) and lower levels of per capita income. This region consists of seven states, five of which register zero synergy in high-tech manufacturing: Acre, Amazonas, Amapá, Roraima and Tocantins. These same states individually contribute less than 1% to the national GDP according to data presented in Table 1. The *Manaus Free Trade Zone* was installed in 1967 with the purpose of creating an industrial, commercial and agricultural center endowed with economic conditions that allow its development. In view of local factors and the great distance between the state of Amazonas and the consumer centres for their products, international companies were set up that produce televisions and communication equipment, among others. In the Northeast Region, made up of nine states, eight also have zero values for synergy in high-tech manufacturing. Of this total, five states contribute individually less than 1% of the national GDP as reported in Table 1: Alagoas, Paraiba, Piauí, Rio Grande do Norte and Sergipe. Other states with the same pattern are from the Midwest Region: Distrito Federal, Mato Grosso do Sul. None of these states contribute to the dynamics of the high-tech sector.

The greatest amount of medium-high tech manufacturing continues to take place in the same states with the largest high-tech manufacturing sector, as shown in Table 5. However, among the ten highest states, three from the Northeast Region are included (Bahia, Pernambuco and Ceará). In the last decades public policies have been developed, to boost the economy by offering incentives to companies setting up in the region.

In the case of high-tech knowledge-intensive services, the leading states are São Paulo (ΔT= -44.8 mbit), Minas Gerais (ΔT=-19.5 mbit) and Rio Grande do Sul (ΔT=-16.5 ΔT mbit). The comparison among the ten top states enables us to see that states from different regions have been developing these economic activities by taking advantage of the opportunities created by new technologies, like Bahia, Pernambuco and Ceará (Northeast), Pará (North), Santa Catarina (South).

The synergy values are significantly correlated to the numbers of firms in *all* states and sectors. The *N* of firms variess as the independent variable among the states and sectors. These results suggest that the numbers of firms and not the technological capacities are crucial for the synergy generated at each scale. Table 6, here below, shows the correlations between the number of firms and the synergy generation across Brazilian states. At the 4-digit level, the first eigenvector in this matrix accounts for almost all (97.5%) of the variance. In sum: we found no significant differences among the states in terms of distributions HTM, MHTM, KIS, HTKIS.

High-tech manufacturing

High-tech KIS

Medium-high tech manufacturing

KIS

Figure 3. Synergy generation at the level of 27 regions in Brazil (NUTS2) separated by different sectors: HTM, MHTM, HTK, KIS.
Source: based on *Orbis* data, using *SPSS* for the mapping.

Table 6. Correlations between the number of firms and synergy generation in Brazilian states and relevant sectors

		Perc_All	Perc_HTM	perc_MHTM	perc_KIS	perc_HTKIS	N_All
Perc_All	Pearson Correlation	1	.908**	.940**	.993**	.958**	.971**
	Sig. (2-tailed)		.000	.000	.000	.000	.000
	N	27	27	27	27	27	27
Perc_HTM	Pearson Correlation	.908**	1	.991**	.938**	.968**	.960**
	Sig. (2-tailed)	.000		.000	.000	.000	.000
	N	27	27	27	27	27	27
perc_MHTM	Pearson Correlation	.940**	.991**	1	.961**	.982**	.974**
	Sig. (2-tailed)	.000	.000		.000	.000	.000
	N	27	27	27	27	27	27
perc_KIS	Pearson Correlation	.993**	.938**	.961**	1	.977**	.985**
	Sig. (2-tailed)	.000	.000	.000		.000	.000
	N	27	27	27	27	27	27
perc_HTKIS	Pearson Correlation	.958**	.968**	.982**	.977**	1	.994**
	Sig. (2-tailed)	.000	.000	.000	.000		.000
	N	27	27	27	27	27	27
N_All	Pearson Correlation	.971**	.960**	.974**	.985**	.994**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	27	27	27	27	27	27

** Correlation is significant at the 0.01 level (2-tailed).

6. Discussion and conclusions

The four sectors which were distinguished by the OECD as typical for knowledge-based economic developments –(*i*) high-tech manufacturing, (*ii*) medium-high tech, (*iii*) knowledge-intensive services, and (*iv*) th subset of high-tech among these services– are not different in their contribution to the synergy in the Brazilian innovation system. When we drew the geographical maps for the four sectors, they were to our surprise virtually identical. The Pearson correlations among the distributions across the states are all larger than .9 and significant at the 0.01 level. In other words, differences in technological capacities among the sectors do not make any difference for the innovativeness of states or sectors. The knowledge-based part of the Brazilian economy is a layer which is not interacting with the remaining of the economy. The latter is a political and not a knowledge-based economy.

Two further findings arise from these results with relevance for the development of a Brazilian Innovation System.

Southeast belt around São Paulo

When comparing the results in the geographical distributions of Brazil –no matter for which sector–the synergy for the whole country are concentrated in São Paulo (22,16%), and the bordering states to São Paulo: Minas Gerais (9,68%), Rio de Janeiro (6,39%), Paraná (7,6%), and Rio Grande do Sul (8,14%). Such a strong regional effect was also found in the analysis of other nations, such as the United Sates (**Leydesdorff** *et al.*, 2019) and Spain (**Leydesdorff**; **Porto-Gómez**, 2019). In the USA, synergy was, concentrated in the north-east (around New York) and in Spain around the metropoles of Barcelona and Madrid. However, in Brazil, the remainder of the country does not participate in the knowledge-based economy. We find an absence of innovation and high-tech manufacturing in the 82% of the states of Brazil, although these states contribute 33,6% of the Brazilian GPD 2018. The states that do not participate in the knowledge-based economy are the ones with low economic development (**Haddad**, 1999; **Morais**; **Swart**; **Jordaan**, 2018), high levels of poverty and deprived productive structures. In the absence of radically new policies, these conditions will negatively affect the future growth of those states. They will not be viable in the future.

A capital without influence

Brazilian capital is Brasilia since 1960, which is located in the state Distrito Federal. Before this date, the capital was Rio de Janeiro, which maintains high levels of economic development. The relocation of the capital city to Brasilia followed a strategy to promote the economic development of the inner regions, although this plan was not achieved and Brasilia has not evolved (**Madaleno**, 1996; **Ishenda**; **Guoqing**, 2019). The absence of a network of triple-helix relations and the priority of public services in Brasilia are visible in our results. Brasilia has the lowest contributions to synergy development (0.24%) in all the sectors. Accordingly, the number of firms is also lower than for any other state in our data: 0.31% of the companies in Brazil are located in Brasilia. The Brazilian economy is based on firms responding to the needs of the public needs (*Codeplan*, 2020). This penumbra of firms earns from the political process with legal and illegal means. Considered as a political capital and not an economic capital, Brasilia has not been able to attract economic activities and has therefore not been able to promote new technological developments. This configuration is comparable with the absence of sectors other than government services to firms in Rome as the administrative capital of Italy (**Leydesdorff**; **Cucco**, 2019). In Italy, the north of the country around the Emilia-Romagna belt[1], completely overtakes the national synergy, compared to a lowest contribution of Lazio; that is, the region in which Rome is located.

Our results suggest that at the national level no synergy is generated. The Brazilian economy is anchored in agrobusiness and mineral extraction (**Petras**, 2013), with a small number of states developing economic activities with more advanced technological levels. In short periods, plans to stimulate technological development based on industrialization and knowledge-based services have been adopted, but these were primarily stimulating for the southeast and south regions of the country (**Santana** *et al.*, 2019). Precisely, in the states where 60% of the synergy can already be found. The large regional disparities in economic development, wage levels, educational and health levels, have made the North, Northeast and part of the Mid-west very marginal to the national synergy. Some of these regions have synergy values equal to zero or below 1%.

The largest contribution in terms of innovation is for the state of São Paulo, which alone contributes 22,16% of the synergy. In comparative terms it is the state with the highest values for HTM, MHT, KIS and MTKIS. In this sense, getting back to the aim of this study, we can confirm that Brazil lacks a national innovation system with interactions among geographical, technological, and organizational distributions generating innovations. A normative consequence arising from this work should points to the need to reframe the productive structure of Brazil, in order to invest in more knowledge advanced sectors, and not only in the southeastern states but also in north and center ones.

A limitation of this study is the measurement at a certain moment in time. The *Orbis* dataset employed does not offer historical series, so we have not been able to perform panel data in order to make comparisons between different political regimes such as the military dictatorship of the 70s and 80s, and the democratic period thereafter. Our data, however, is pre-Covid (2018). Covid has probably worsened the situation. It would be interesting to compare the situation in 2018 with 2008 in order to develop a historical perspective. However, data needed for this type of studies is available only during the last decade or so.

The second interest area would be in the region of São Paulo. Considering the relevance for the whole Brazilian economy, we should better understand the regional innovation system, as it was made for the Californian economy in the analysis performed by **Leydesdorff** *et al.* (2019). From an economic development perspective, we recognized that more efforts have been made in analyzing triple-helix approaches in economically advanced countries, mainly in Europe, but also in North America. Considering the lack of roads and transport connectivity between Brazil and its neighbors (**Jaimurzina** *et al.*, 2015; **Vecchio** *et al.*, 2020), one can also consider a broader perspective and analyze Brazil in its Latin-American context; for example, of *MercoSur*.

7. Note

1. The Italian case, does not correspond to the relocation of the capital but to a geographical distribution of a political capital (Rome) and an economical capital (Milano).

8. References

Albuquerque, Eduardo-da-Motta (2000). "Domestic patents and developing countries: arguments for their study and data from Brazil (1980-1995)". *Research policy*, v. 29, n. 9, pp. 1047-1060.
https://doi.org/10.1016/S0048-7333(99)00053-0

Albuquerque, Eduardo-da-Motta; **Suzigan, Wilson**; **Cário, Silvio-Antonio-Ferraz**; **Fernandes, Ana-Cristina**; **Shima, Walter**; **Britto, Jorge**; **Barcelos, Achyles**; **Rapini, Márcia-Siqueira** (2008). "An investigation on the contribution of universities and research institutes formaturing the Brazilian innovation system: preliminary results". In: *Globelics conference*, Mexico City.

Altenburg, Tilman (2011). "Building inclusive innovation systems in developing countries: challenges for IS research". In: Lundvall, B. Å., Joseph, K. J.; Chaminade, C.; Vang, J. (eds.). *Handbook of innovation systems and developing countries: building domestic capabilities in a global setting*. Cheltenham, UK, Northhampton, USA: Edward Elgar Publishing.

Bureau van Dijk (2018). *Orbis database.*
https://orbis.bvdinfo.com

Canuto, Otaviano; **Cavallari, Matheus**; **Reis, José-Guilherme** (2013). "Brazilian exports: climbing down a competitiveness cliff". *World Bank Policy Research Working Paper* (6302).

Codeplan (*Companhia de Planejamento do Distrito Federal*) (2020). Produto interno bruto do Distrito Federal 2018.
https://www.codeplan.df.gov.br/wp-content/uploads/2018/02/Relatorio-PIB-DF-2018.pdf

Edquist, Charles (1997). *Systems of innovation: Technologies, institutions and organizations*. London: Pinter.

Etzkowitz, Henry; **Leydesdorff, Loet** (1995). "The Triple Helix of University-Industry-Government relations: A laboratory for knowledge based economic development". *EASST Review*, v. 14, n. 1, pp. 11-19.
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2480085

Etzkowitz, Henry; **Leydesdorff, Loet** (2000). "The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university-industry-government relations". *Research policy*, v. 29, n. 2, pp. 109-123.
https://doi.org/10.1016/S0048-7333(99)00055-4

Eurostat/OECD (2011). *High technology and knowledge-intensive sectors*, December.
https://ec.europa.eu/eurostat/cache/metadata/Annexes/htec_esms_an3.pdf

Haddad Eduardo A. (1999). *Regional inequality and structural changes: lessons from the Brazilian experience*. Milton Park, Abingdon, Oxon; New York, NY: Routledge.

Hartmann, Dominik; **Guevara, Miguel R.**; **Jara-Figueroa, Cristian**; **Aristarán, Manuel**; **Hidalgo, César A.** (2017). "Linking economic complexity, institutions, and income inequality". *World development*, n. 93, pp. 75-93.
https://doi.org/10.1016/j.worlddev.2016.12.020

IBGE (*Instituto Brasileiro de Geografia e Estatística*) (2020). *Sistemas de Contas regionais: Brasil: 2018 / IBGE.*
https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalhes&id=2101765

INPI (*Instituto Nacional de Propriedade Intelectual*) (2021). *Ranking dos depositantes residentes*.
https://www.gov.br/inpi/pt-br/central-de-conteudo/noticias/inpi-divulga-rankings-dos-maiores-depositantes-em-2020

Ishenda, *Doris-Kokutungisa;* **Guoqing, Shi** (2019). "Determinants in relocation of capital cities". *Journal of public administration and governance*, v. 9, n. 4, pp. 200-220.
https://doi.org//10.5296/jpag.v9i4.15983

Jaimurzina, Azhar; **Pérez-Salas, Gabriel**; **Sánchez, Ricardo J.** (2015). *Políticas de logística y movilidad para el desarrollo sostenible y la integración regional*. Serie Recursos naturales e infraestructura, n. 174, Cepal, Naciones Unidas.
http://hdl.handle.net/11362/39427

Laafia, Ibrahim (2002). *National and regional employment in high tech and knowledge intensive sectors in the EU ó 1995-2000*. Statistics in focus: Science and technology, theme 9, n. 4.
https://ec.europa.eu/eurostat/en/web/products-statistics-in-focus/-/KS-NS-02-003

Lengyel, Balázs; **Leydesdorff, Loet** (2011). "Regional innovation systems in Hungary: the failing synergy at the national level". *Reg. stud*, v. 45, n. 5, pp. 677-693.
https://doi.org/10.1080/00343401003614274

Leydesdorff, Loet (1997). "The new communication regime of university-industry-government relations". In: Etzkowitz, H.; Leydesdorff, L. (eds.). *Universities and the global knowledge economy: A triple helix of university-industry-government relations*. London: Cassell Academic.

Leydesdorff, Loet (2000). "The triple helix: an evolutionary model of innovations". *Research policy*, v. 29, n. 2, pp. 243-255.
https://doi.org/10.1016/S0048-7333(99)00063-3

Leydesdorff, Loet (2003). "The mutual information of university-industry-government relations: an indicator of the triple helix dynamics" *Scientometrics*, v. 58, n. 2, pp. 445-467.
https://doi.org/10.1023/a:1026253130577

Leydesdorff, Loet (2006). *The knowledge-based economy: modelled, measured, simulated*. Boca Raton: Universal Publishers. ISBN: 1 58112 937 8

Leydesdorff, Loet; **Cucco, Ivan** (2019). "Regions, innovation systems, and the North-South divide in Italy". *El profesional de la información*, v. 28, n. 2, e280214.
https://doi.org/10.3145/epi.2019.mar.14

Leydesdorff, Loet; **Dolfsma, Wilfred**; **Van-der-Panne, Gerben** (2006). "Measuring the knowledge base of an economy in terms of triple-helix relations among 'technology, organization, and territory'". *Research policy*, v. 35, n. 2, pp. 181-199.
https://doi.org/10.1016/j.respol.2005.09.001

Leydesdorff, Loet; **Etzkowitz, Henry** (1996). "Emergence of a Triple Helix of university-industry-government relations". *Science and public policy*, v. 23, n. 5, pp. 279-286.
https://doi.org/10.1093/spp/23.5.279

Leydesdorff, Loet; **Fritsch, Michael** (2006). "Measuring the knowledge base of regional innovation systems in Germany in terms of a Triple Helix dynamics". *Research policy*, v. 35, n. 10, pp. 1538-1553.
https://doi.org/10.1016/j.respol.2006.09.027

Leydesdorff, Loet; **Porto-Gómez, Igone** (2019). "Measuring the expected synergy in Spanish regional and national systems of innovation". *The journal of technology transfer*, v. 44, n. 1, pp. 189-209.
https://doi.org/10.1007/s10961-017-9618-4

Leydesdorff, Loet; **Strand, Øivind** (2013). "The Swedish system of innovation: Regional synergies in a knowledge-based economy". *Journal of the American Society for Information Science and Technology*, v. 64, n. 9, pp. 1890-1902.
htpps://doi.org/10.1002/asi.22895

Leydesdorff, Loet; **Sun, Yuan** (2009). "National and international dimensions of the Triple Helix in Japan: University-industry-government versus international co-authorship relations". *Journal of the American Society for Information Science and Technology*, v. 60, n. 4, pp. 778-788.
https://doi.org/10.1002/asi.20997

Leydesdorff, Loet; **Wagner, Caroline S.**; **Porto-Gómez, Igone**; **Comins, Jordan A.**; **Phillips, Fred** (2019). "Synergy in the knowledge base of U.S. innovation systems at national, state, and regional levels: The contributions of high-tech manufacturing and knowledge-intensive services". *Journal of the Association for Information Science and Technology*, v. 70, n. 10, pp. 1108-1123.
https://doi.org/10.1002/asi.24182

Lundvall, Bengt-Åke (1999). "National business systems and national systems of innovation". *International studies of management & organization*, v. 29, n. 2, pp. 60-77.
https://doi.org/10.1080/00208825.1999.11656763

Madaleno, Isabel-Maria (1996). "Brasilia: the frontier capital". *Cities*, v. 13, n. 4, pp. 273-280.
https://doi.org/10.1016/0264-2751(96)00016-9

Morais, Margarida B.; **Swart, Julia**; **Jordaan, Jacob A.** (2018). "Economic complexity and inequality: Does productive structure affect regional wage differentials in Brazil?". *USE working paper series*, n. 18-11.

Nelson, Richard R.; **Winter, Sidney G.** (1977). "In search of useful theory of innovation". *Research policy*, v. 6, n. 1, pp. 35-76.
https://doi.org/10.1016/0048-7333(77)90029-4

Nelson, Richard R.; **Winter, Sidney G.** (1982). *An evolutionary theory of economic change*. Cambridge, MA, Belknap Press of Harvard University Press.

OECD (2018). *Economic surveys: Brazil 2018*.

Oreiro, José-Luis; **Mararin, Luciano-Luiz**; **Gala, Paulo** (2020). "Deindustrialization, economic complexity and exchange rate overvaluation: the case of Brazil (1998-2017)". *PSL quarterly review*, v. 73, n. 295, pp. 313-341.
https://doi.org/10.13133/2037-3643_73.295_3

Pacheco, Carlos-Américo (2019). "Institutional dimensions of innovation policy in Brazil". In: Reynolds, E. B.; Schneider, B. R.; Zylberberg, E. (eds.). *Innovation in Brazil: Advancing development in the 21st century*. New York: Routledge.

Park, Han Woo; **Hong, Heung-Deug**; **Leydesdorff, Loet** (2005). "A comparison of the knowledge-based innovation systems in the economies of South Korea and the Netherlands using triple helix indicators" *Scientometrics*, v. 65, n. 1, pp. 3-27.
https://doi.org/10.1007/s11192-005-0257-4

Park, Han Woo; **Leydesdorff, Loet** (2010). "Longitudinal trends in networks of university-industry-government relations in South Korea: The role of programmatic incentives". *Research policy*, v. 39, n. 5, pp. 640-649.
https://doi.org/10.1016/j.respol.2010.02.009

Petras, James (2013). "Brazil: Extractive capitalism and the great leap backward". *World review of political economy*, v. 4, n. 4, pp. 469-483.
https://doi.org/10.13169/worlrevipoliecon.4.4.0469

Sábato, Jorge; **Botana, Natalio** (1970). "La ciencia y la tecnología en el desarrollo futuro de América Latina". *Serie Documentos teóricos* 11. Instituto de Estudios Peruanos, Lima: Perú.
http://repositorio.iep.org.pe/bitstream/handle/IEP/1037/Sabato_Botana_ciencia-tecnologia-desarrollo-América-Latina.pdf

De-Santana, José-Ricardo; **Teixeira, André-Luiz-da-Silva**; **Rapini, Márcia-Siqueira**; **Esperidião, Fernanda** (2019). "Financiamento público à inovação no Brasil: contribuição para uma distribuição regional mais equilibrada?". *Planejamento e políticas públicas*, v. 52, Jan-Jun.
http://repositorio.ipea.gov.br/bitstream/11058/9766/1/ppp_n52_financiamento.pdf

Strand, Øivind; **Leydesdorff, Loet** (2013). "Where is synergy indicated in the Norwegian innovation system? Triple-Helix relations among technology, organization, and geography". *Technological forecasting & social change*, v. 80, n. 3, pp. 471-484.
https://doi.org/10.1016/j.techfore.2012.08.010

Theil, Henri (1972). *Statistical decomposition analysis: With applications in the social and administrative sciences*. Amsterdam, The Netherlands: North-Holland Publishing Company.

Ulanowicz, Robert E. (1986). *Growth and development: Ecosystems phenomenology*. New York: Springer.

Vecchio, Giovanni; **Tiznado-Aitken, Ignacio**; **Hurtubia, Ricardo** (2020). "Transport and equity in Latin America: a critical review of socially oriented accessibility assessments". *Transport reviews*, v. 40, n. 3, pp. 354-381.
https://doi.org/10.1080/01441647.2020.1711828

Whitley, Richard D. (1984). *The intellectual and social organization of the sciences*. Oxford: Oxford University Press.

Whitley, Richard D. (2001). *National innovation systems*. In: Smelser, N. J.; Baltes, P. B. (eds.). *International encyclopedia of the social and behavioral sciences*. Oxford: Elsevier.

World Bank (2019). *World Bank Group Country Survey, Brazil*.
https://data.worldbank.org/country/brazil

Ye, Fred Y.; **Yu, Susan S.**; **Leydesdorff, Loet** (2013). "The Triple Helix of university-industry-government relations at the country level and its dynamic evolution under the pressures of globalization". *Journal of the American Society for Information Science and Technology*, v. 64, n. 11, pp. 2317-2325.
https://doi.org/10.1002/asi.22931

Reference publication year spectroscopy (RPYS) of papers published by Loet Leydesdorff: A giant in the field of scientometrics passed away

Lutz Bornmann; Robin Haunschild

Recommended citation:

Bornmann, Lutz; **Haunschild, Robin** (2023). "Reference publication year spectroscopy (RPYS) of papers published by Loet Leydesdorff: A giant in the field of scientometrics passed away". *Profesional de la información*, v. 32, n. 7, e320701.

https://doi.org/10.3145/epi.2023.dic.01

Article received on July 20th 2023
Approved on August 26th 2023

Lutz Bornmann ✉
https://orcid.org/0000-0003-0810-7091
Administrative Headquarters of the Max Planck Society
Science Policy and Strategy Department
Hofgartenstr., 8. 80539 Munich, Germany
bornmann@gv.mpg.de
l.bornmann@fkf.mpg.de

Robin Haunschild
https://orcid.org/0000-0001-7025-7256
Max Planck Institute for Solid State Research
Heisenbergstraße 1
70569 Stuttgart, Germany
r.haunschild@fkf.mpg.de

Abstract

Loet Leydesdorff (mentioned as Loet in the following) passed away in March 2023. Our paper is dedicated to the important contributions of this exceptional researcher (in scientometrics). We investigated which studies, theories, methods, and ideas have influenced Loet's scientific work. The method reference publication year spectroscopy (RPYS) can be used to answer this and related questions. Many RPYS studies have been published regarding the historical roots of research fields, journals, and scientists. The program *CRExplorer* was specifically developed for RPYS. In this study, we used *CRExplorer* to investigate the historical roots and influential publications of Loet's oeuvre. The results demonstrate the wide range of topics in Loet's research and their fundamental meaning for the scientometric field.

Keywords

Reference publication year spectroscopy; RPYS; Citation analysis; Scientometrics; Bibliometrics; Researchers; Loet Leydesdorff.

1. Introduction

Loet Leydesdorff (mentioned as Loet in the following) passed away in March 2023. Loet was a giant in the field of scientometrics. We are not aware of any other researcher in the field of scientometrics with a similarly important contributing impact. Loet published more than 400 papers (the number that we found in *Clarivate*'s *Web of Science*, *WoS*, database). These papers focus not only on one or a few scientometric topics but have an enormous width (**Wouters**; **Wagner**, 2023). Loet dealt in his research with the history, philosophy, sociology, and economy of science –among many other disciplines, topics and themes. He was a specialist in science networks (e.g., **Leydesdorff** & **Persson**, 2010) and bibliometric indicators (e.g., **Leydesdorff** & **Bornmann**, 2012). For example, **Opthof** and **Leydesdorff** (2010) started an important discussion on the field-normalized citation score based on average citation rates –the standard indicator in bibliometrics at that time. At the end, the discussion led to a far-reaching revision of the standard indicator.

One of the authors of this paper (LB) published his first paper together with Loet in 2007. LB's first co-authored study with Loet dealt with the citation network of the chemical journal *Chimia* (**Bornmann**; **Leydesdorff**; **Marx**, 2007). This paper was followed by around 70 papers in collaboration between Loet and LB (according to the *WoS*). Many of the co-authored papers focused on the development of alternative indicators to the field-normalized citation score based on average citation rates: Loet and LB favored citation percentiles (**Leydesdorff**; **Bornmann**, 2011; **Leydesdorff**; **Bornmann**; **Mutz**; **Opthof**, 2011). This alternative was introduced already in the 1980s by Francis Narin –another giant in the scientometric field (**McAllister**; **Narin**; **Corrigan**, 1983). Another topic of Loet and LB was the development of science maps

(**Bornmann**; **Leydesdorff**, 2011; 2012). For example, they published software that could be used to generate institutional maps of science showing regional areas of high and low research performance (**Bornmann**; **Leydesdorff**; **Walch-Solimena**; **Ettl**, 2011).

The other author of this paper (RH) and Loet published their first paper together in 2015. The paper dealt with network analyses of country and reader status using data from *Mendeley* (**Haunschild**; **Bornmann**; **Leydesdorff**, 2015). The results of this collaboration were presented at the *2:AM* conference in Amsterdam (**Dinsmore**, 2015). RH had the pleasure of meeting Loet in person during the conference and enjoying Loet's hospitality. Several papers followed in collaboration between Loet and RH about network analyses and cited reference analyses. The last paper co-authored by Loet and RH reported the most influential publications in the *Web of Science* subject categories on the basis of a cited reference analysis (**Thor**; **Bornmann**; **Haunschild**; **Leydesdorff**, 2021).

In this paper, we present the results of a Reference Publication Year Spectroscopy (RPYS) with the goal of uncovering the roots of Loet's research (**Marx**; **Bornmann**; **Barth**, 2013). Previously, RPYS has been applied to other researchers, for example to Eugene Garfield by **Bornmann**, **Haunschild**, and **Leydesdorff** (2018). For applying RPYS, it is necessary to retrieve the complete set of Loet's papers. RPYS counts and visualizes the occurrences of single cited references in this set. High numbers of occurrences point to historical roots of Loet's research –especially those cited publications from early years. Although RPYS has been initially introduced by Werner Marx (WM), the method has been further developed in collaboration between Loet, WM, RH, LB, Rüdiger Mutz (RM), and especially Andreas Thor (AT). The researchers developed and published the *CRExplorer* (**Haunschild**; **Marx**; **Thor**; **Bornmann**, 2020; **Thor**; **Bornmann**; **Marx**; **Mutz**, 2018; **Thor**; **Marx**; **Leydesdorff**; **Bornmann**, 2016a, 2016b) –a software that can be used free of charge for RPYS:
https://www.crexplorer.net

2. Dataset

We downloaded the metadata (including cited references) of the papers published by Loet from the *WoS* using the search query:

AU = (leydesdorff l OR leydesclorff l OR leydesdorfl l)

We also checked his *ORCID* and *WoS Researcher Profile* records to collect the complete set of papers. However, no additional correct papers could be found here. The extracted files from the *WoS* were imported into the *CRExplorer* for further processing. The *WoS* export contains metadata of 413 papers including 17,385 cited references. We applied the clustering and merging functionalities of the *CRExplorer* to clean up the cited references dataset with respect to reference variants of the same cited publication. *CRExplorer* determines the pair-wise similarity of variants of CRs based on the Levenshtein similarity (**Thor**; **Bornmann**; **Haunschild**, 2018). To support the disambiguation process, volume and page numbers of the referenced papers have been used. After the disambiguation process, we removed all cited references that were cited less than five times to focus on publications with a substantial impact on Loet's research. The final dataset which we used for the RPYS contains metadata of 413 citing papers (from 1981 to 2023) including 742 cited references (from 1902 to 2019).

3. Methods

It is the premise of the RPYS that important publications for a certain researcher are often cited in his (or her) papers (**Bornmann**; **Marx**, 2014). The basic result of the cited references analyses using Loet's papers is a spectrogram showing the number of cited references per reference publication year. Peaks in the spectrogram are hints to possible important publications in certain reference publication years.

We analyzed the spectrogram to find relevant peaks by using the five-year median deviation. The five-year median deviation compares the number of cited references in reference publication year t (i.e., the peak in year t) with the number of cited references in bordering years: t-2, t-1, t+1, and t+2. If the peak of the median deviation is very high for year t, many cited references fall on year t –compared to the bordering years. Tukey's fences (**Tukey**, 1977) were used to support the identification of the most important peaks in the spectrogram: Important peaks were flagged based on the interquartile range of the median deviations (**Thor**; **Bornmann**; **Haunschild**, 2018).

Thor, Bornmann, Marx *et al.* (2018) developed methods that can be used to analyze cited references data further on. In this study, we used the N_TOP10 indicator for identifying landmark papers over a longer period. The indicator shows the number of cited years in which a publication belonged to the 10% most frequently referenced publications by Loet.

In this study, the spectrogram was plotted using R (*R Core Team*, 2021) with the R package 'BibPlots' (**Haunschild**, 2021). In addition to the static spectrogram presented in this paper, we produced an interactive version using the R package 'dygraphs' (**Vanderkam** *et al.*, 2018).

4. Results

4.1. Reference publication year spectroscopy

Figure 1 shows the number of cited references (grey columns) and the deviation of the number of cited references in one reference publication year from the number of cited references in bordering years (blue line). The RPYS is based on the principle that high peaks (deviations) are hints to important publications of the analyzed researcher (publications in Loet's oeuvre). Peaks in early cited reference years point to the historical roots of the researcher (rather old papers frequently cited by Loet). Peaks with significant deviations from peaks in bordering reference publication years (identified by Tukey's fences) are labeled with an asterisk and the corresponding publication year in Figure 1.

Table 1 shows the publications that are mainly responsible for the highest peaks in Figure 1: 1948, 1972/1973, 1979, 1994, 1997, and 2006. Besides the titles, the table presents the cited references counts (in other words, how often the individual publications have been cited in Loet's papers).

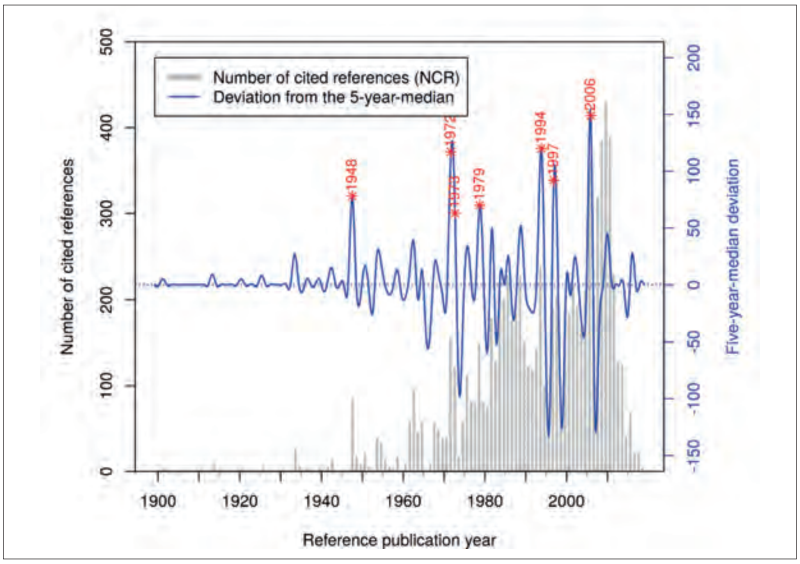


Figure 1. Number of cited references (grey columns) and median deviations of cited references (blue line). The peaks (with positive values) in the blue line show reference publication years with a significantly greater number of cited references than bordering years. An interactive version can be viewed at: *https://s.gwdg.de/uGPMP2*

Table 1. Cited references with the largest number of cited reference counts in cited reference years with the highest peaks in Figure 1

Nº	Title of the cited reference	Author, Publication year	Cited references counts
	1948		
1*	A mathematical theory of communication	(**Shannon**, 1948b)	65
2	A mathematical theory of communication	(**Shannon**, 1948a)	21
	1972		
3	Statistical decomposition analysis: With applications in the social and administrative sciences	(**Theil**, 1972)	63
4	Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies	(**Garfield**, 1972)	48
	1973		
5	The organization of complex systems	(**Simon**, 1973)	37
	1979		
6	Is citation analysis a legitimate evaluation tool	(**Garfield**, 1979b)	26
7	Citation indexing: Its theory and application in science, technology, and humanities	(**Garfield**, 1979a)	26
8	Centrality in social networks conceptual clarification	(**Freeman**, 1978/1979)	21
	1994		
9	The new production of knowledge: The dynamics of science and research in contemporary societies	(**Gibbons** *et al.*, 1994)	33
10	Tracking areas of strategic importance using scientometric journal mappings	(**Leydesdorff**; **Cozzens**; **Van-den-Besselaar**, 1994)	27
	1997		
11	Scientometrics and communication theory: Towards theoretically informed indicators	(**Leydesdorff**; **Van-den-Besselaar**, 1997)	21
12	Why words and co-words cannot map the development of the sciences	(**Leydesdorff**, 1997)	20
13	The regional world: Territorial development in a global economy	(**Storper**, 1997)	20
14	Why the impact factor of journals should not be used for evaluating research	(**Seglen**, 1997)	18
	2006		
15	Can scientific journals be classified in terms of aggregated journal-journal citation relations using the *Journal Citation Reports*	(**Leydesdorff**, 2006)	43

Note. * Two variants of the same cited reference were merged manually.

As the underlying references of the 1948 peak show (see Table 1), Loet has cited two early papers very frequently which were published by **Shannon** (1948a; 1948b). Both papers appeared with the same title and author, but in different issues of the *Bell System Technical Journal*. In his research, Loet preferred a communications view on the sciences (see here

Luhmann, 2012a; 2012b) whereby knowledge claims are usually organized into scientific papers building the archive of science (that can be bibliometrically analyzed). This view differed, e.g., from a sociological view which is focused on individuals (e.g., authors) or communities (e.g., research groups or institutions) (**Leydesdorff**, 2021). **Shannon** (1948a; 1948b) lays the basis for Loet's view by providing the corresponding mathematical foundation.

The next peak in Table 1 is visible for the years 1972 and 1973. This peak is mainly due to three publications from 1972 and one from 1973. The publication most cited by Loet below this peak is **Theil** (1972), which is a statistics book for decomposition analysis in the social sciences. Loet cited (used) (statistics from) this book very broadly such as in a paper on the European monetary system (**Leydesdorff**; **Oomes**, 1999), in a paper on networks of journal-journal citations (**Leydesdorff**, 2003), and in a paper on animations of journal maps (**Leydesdorff**; **Schank**, 2008). The other publication from 1972 (see Table 1) which Loet referenced frequently is **Garfield** (1972). Eugen Garfield is the founder of citation indexing in (large) databases (**Garfield**, 1955; 1970) for studying science. With the paper from 1972, **Garfield** (1972) proposed to evaluate scientific journals based on citation impact. In this paper, **Garfield** (1972) introduced the impact factor –one of the most popular metrics in citizen bibliometrics. The most important referenced publication from 1973 is **Simon** (1973). It is a chapter in the book "Models of discovery" and deals with several statistical topics such as the empirical Bayes approach or the aggregation of variables.

The next peak occurs in 1979. The peak in 1979 is fed by two classic publications from Eugen Garfield (see above). **Garfield** (1979b) deals with the question whether citation data can be used for research evaluation purposes. The paper picks up several critical points of citation analysis that are frequently discussed (e.g., the problem of self-citations and negative citations). It concludes that

> "when properly used, citation analysis can introduce a useful measure of objectivity into the evaluation process at relatively low financial cost" (**Garfield**, 1979b, p. 359).

Garfield (1979b) is a modification of a chapter in **Garfield** (1979a) –the next publication in Table 1.

Garfield (1979a) is a classical textbook of citation analysis entitled "Citation indexing: Its theory and application in science, technology, and humanities". It is a must-read for all scientists entering the field of bibliometrics (until today). Various important topics in the context of citation analysis are treated such as the use of citations as a search tool for literature, the use of citations in science management, the application of citations in the historical analysis of science, and the use of citations for science mapping purposes. Science mapping is also the topic of **Freeman** (1978/1979) –the following publication in Table 1. This publication deals with measures of structural centrality in social networks. The conceptualization of networks based on bibliometric data is a popular topic in bibliometrics, and a favourite with Loet (e.g., **Leydesdorff**, 2003; **Wagner**; **Leydesdorff**, 2005). **Freeman** (1978/1979) is a classic in this context.

The results in Figure 1 and Table 1 indicate that 1994 is the next peak year after 1979. As Loet's most important referenced publications in this year, we have **Gibbons** *et al.* (1994), on the one hand. The book deals with the major changes in the way research results are generated in the modern science system of today –compared to the academic science system decades ago (**Ziman**, 1996). The authors argue that the modern science system is characterized by globalization, collaboration, and competitiveness, which have led to new modes of knowledge production. Various terms have been proposed to name these new modes: post-academic science, Mode-2 (compared to Mode 1 –the academic science) or post-normal science. On the other hand, we find **Leydesdorff** *et al.* (1994) under the 1994 peak. This paper deals with the topic of identifying areas of growth or fast-changing areas in research fields. Science policy is especially interested in knowing and funding these areas, since these areas may define the research fronts in science. **Leydesdorff** *et al.* (1994) proposed a method for identifying such areas based on journal maps and applied the method to research on AIDS, superconductivity, and oncogenes.

Table 1 includes four publications for the year 1997, two of which were published by Loet himself. The four publications focus on different topics. **Leydesdorff** and **Van-den-Besselaar** (1997) deal with citation theories. Since bibliometrics is used in various contexts on a larger scale, researchers proposed concepts or theories for explaining citing decisions such as the normative and the social-constructivist theories (see **Tahamtan** & **Bornmann**, 2018, for an overview). The last proposal in this series was introduced by **Tahamtan** and **Bornmann** (2022): the Social Systems Citation Theory (SSCT). **Tahamtan** and **Bornmann** (2022) picked up from **Leydesdorff** and **Van-den-Besselaar** (1997) that a citation theory should rather focus on communications and not on cited or citing agents. In sociology, Loet's focus is rooted in Niklas Luhmann's *social systems theory* (**Luhmann**, 2012a; 2012b).

Leydesdorff (1997) –the second paper in Table 1 for 1997– also deals with communications in the network of science. Based on a set of articles from biochemistry, the study reveals that the network level –the level of the publication set– may be different from the individual paper-level perspective:

> "Words change both in terms of frequencies of relations with other words, and in terms of positional meaning from one text to another" (**Leydesdorff**, 1997, p. 418).

The differentiation between a social level that is linked but cannot be directly traced back to single individuals is a genuinely sociological perspective (**Coleman**, 1990). The next publication under the peak of 1997 is **Storper** (1997). The book can be denoted as a contribution from political science which proposes a theory of how regions worldwide have maintained their economic viability. We assume that the interesting point for Loet was the system-theoretical root of the book: The world is seen as a social system with inter-connected regional economies.

The fourth publication in Table 1 for 1997 is **Seglen** (1997). The author argues conceptually and reveals empirically that the Journal Impact Factor (JIF) should not be used as a proxy for the citation impact of single papers. The JIF is defined as the mean number of citations in one year gathered by publications appearing in the two years before. The paper was directed against the usual practice in research evaluation (at that time) of using the JIF instead of the times cited information (e.g., from the WoS) to measure citation impact. Today, many initiatives and manifestos exist against this use of the JIF such as the *Leiden manifesto* (**Hicks** *et al.*, 2015).

For the peak in 2006, we identified **Leydesdorff** (2006) as an influential paper for Loet himself (see Table 1). This paper is part of one of the most important research programs by Loet: the classification of (all) scientific journals by using data on the citation patterns of the journals from the *Journal Citation Reports* in the *Essential Science Indicators* (*Clarivate*). Loet used factor analytic methods to discover latent structures in the matrix of citation relations between the journals. One of the authors of this study (LB) was involved in follow-up studies that continued Loet's research program (e.g., **Leydesdorff**; **Bornmann**; **Wagner**, 2017; **Leydesdorff**; **Bornmann**; **Zhou**, 2016).

4.2. Publications that Loet cited very frequently over many years

Whereas Table 1 lists the publications that have been referenced by Loet very frequently, we additionally used the N_TOP10 indicator for identifying the important (most influential) publications for Loet over many citing years. Table 2 lists the cited references in our dataset that were referenced significantly more frequently than other publications in at least 10 citing years. There is only a single cited reference in Table 2 that also occurs in Table 1: **Simon** (1973).

Table 2. Cited references that belong to the 10% most frequently referenced in more citing years than other cited references. The table shows the cited references that are highly referenced in at least 10 citing years. The publication numbers (Nº) from Table 1 are continued.

Nº	Title of the cited reference	Author, Publication year	Cited references counts	N_TOP10
16	The intellectual and social organization of the sciences	(**Whitley**, 1984)	54	15
17	An algorithm for drawing general undirected graphs	(**Kamada**; **Kawai**, 1989)	68	14
18	Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity	(**Narin**, 1976)	40	13
19	Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient	(**Ahlgren**; **Jarneving**; **Rousseau**, 2003)	67	13
5	The organization of complex systems	(**Simon**, 1973)	37	10
20	Toward a structural theory of action: Network models of social structure, perception, and action	(**Burt**, 1982)	43	10
21	The static and dynamic analysis of network data using information-theory	(**Leydesdorff**, 1991)	32	10
22	The dynamics of innovation: From national systems and "Mode 2" to a Triple Helix of university-industry-government relations	(**Etzkowitz**; **Leydesdorff**, 2000)	50	10
23	The challenge of scientometrics: The development, measurement, and self-organization of scientific communications	(**Leydesdorff**, 2001)	58	10

The book "*The intellectual and social organization of the sciences*" by **Whitley** (1984) belongs to the 10% most frequently cited in 15 years of Loet's scientific career. **Whitley** (1984) conceptualizes science as a sector of society including specifically organized social systems with the goal of producing and validating knowledge. The different systems exist in particular contexts and generate knowledge in a particular way. Besides the book by **Whitley** (1984), there is another theoretically oriented publication in Table 2: "*Toward a structural theory of action: Network models of social structure, perception, and action*" published by **Burt** (1982). In this publication, **Burt** (1982) formulates basics of a structural action theory. The theory is not only based on concepts of sociological network theories, but also on classic texts from the sociology of science (published, e.g., by James S. Coleman, Robert K. Merton, and Talcott Parsons).

As Table 2 reveals, the paper by **Kamada** and **Kawai** (1989) is one of the most referenced publication by Loet over many years. The authors proposed an algorithm that can be used for an optimized layout of networks. Since Loet published many manuscripts including several types of networks based on bibliometric data, the layout of many networks was optimized based on the algorithm by **Kamada** and **Kawai** (1989). Some examples of Loet's papers using the algorithm by **Kamada** and **Kawai** (1989) are **Bornmann** and **Leydesdorff** (2015), **Bornmann**, **Wagner**, and **Leydesdorff** (2015), and **Haunschild**, **Leydesdorff**, and **Bornmann** (2020). In Table 2, we can find two other papers focusing on methods for (bibliometric) network analyses. (1) **Ahlgren** *et al.* (2003) has a specific focus on author co-citation analysis (ACA) and deals with the question whether Pearson's correlation coefficient can be used as a similarity measure in ACA. The authors conclude that this coefficient should not be used and "sets forth two natural requirements that a similarity measure applied in ACA should satisfy" (p. 550). (2) **Leydesdorff** (1991) proposes to use measures derived from information theory as a conceptual framework for multivariate analyses of bibliometric data. To empirically illustrate his approach, **Leydesdorff** (1991) used a matrix of aggregated citations among chemistry journals.

One of the most referenced publications over many years is the book "*Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*" by **Narin** (1976). Francis Narin is another giant and pioneer in scientometrics similar to Loet. **Narin** (1976) can be denoted as one of the most influential publications by Francis Narin: The book introduced bibliometrics as an assessment tool for evaluating scientific activity. The book outlines how an evaluative study should be conducted and points to typical problems in bibliometric analyses such as multiple authorship, self-citations, homonyms, and field variations in citations. Since the publication of **Narin** (1976), many following papers in bibliometrics have dealt with these problems and proposed solutions such as the introduction of field-normalized indicators tackling field variations in citations (**Bornmann**, 2019).

With more than 3,000 citations in the *WoS* (times cited), **Etzkowitz** and **Leydesdorff** (2000) is the most cited publication by Loet with many more citations than his other publications. It is also one of the most referenced papers by Loet himself over many years, as the results in Table 2 point out. **Etzkowitz** and **Leydesdorff** (2000) deal with the transformation of the science system from academic to post-academic science (see above). Academic science is traditional science whereby researchers work at universities and other publicly funded research institutions. According to **Ziman** (1998),

> "academic science was intensely individualistic. People held personal appointments earned by published contributions to knowledge. Universities and research institutes had little direct influence on their research".

Academic science can be differentiated from industrial science in the history of scientific activities: Industrial science is characterized by scientists (employed by companies) who do not undertake "their own" projects and are not free to publish their research results:

> "industrial science –from agriculture through mental medicine, and missile manufacture to zookeeping– is intimately involved in the business of daily life. The personal values and needs of customers, patients, and other users have to be taken into account" (**Ziman**, 1998).

The modern post-academic science system became visible especially since the end of the Cold War (**Etzkowitz**; **Leydesdorff**, 2000). This system is characterized by groups of researchers (**Wu**; **Wang**; **Evans**, 2019) working in projects which are funded for specific outcomes: Funders expect that project results are not only useful for science itself, but also for the economy or other sectors of the society (**Bornmann**, 2013). Research in post-academic science "stems from problems 'arising in the context of application'" (**Ziman**, 1998). In post-academic science, universities and other research-focused institutions are seen as an important player in national economic development.

The significance of **Etzkowitz** and **Leydesdorff** (2000) lies in the fact that the authors proposed an alternative to the post-academic science concept: the Triple Helix of university-industry-government. The Triple Helix is a dynamic concept that denotes the relations between three actors: university, industry, and government. The dynamic concept can be used to describe the shape of national science systems or the development of science systems in the historical context. With

> "different possible resolutions of the relations among the institutional spheres of university, industry, and government" (**Etzkowitz**; **Leydesdorff**, 2000, p. 110)

it is possible

> "to generate alternative strategies for economic growth and social transformation" (**Etzkowitz**; **Leydesdorff**, 2000, p. 110).

The book "The challenge of scientometrics: The development, measurement, and self-organization of scientific communications" (**Leydesdorff**, 2001) is listed at the end of Table 2. The first edition of the book was published in 1995; the second edition from 2001 is nearly the same as the first edition. Loet explained in the book his own paradigm of undertaking science of science studies. The paradigm is characterized by an attempt to integrate qualitative and quantitative proposals to conduct science of science studies. Loet's royal road for the integrative perspective is the information theory: "By using this method [the information theory], central problems in science studies will be addressed, both on the qualitative side (e.g., the significance of a reconstruction) and on the quantitative side (e.g., the prediction of science indicators)" (**Leydesdorff**, 2001, p. 5).

5. Discussion

The death of Loet unfortunately joins in recent deaths of other giants in the field. Henk Moed –who published fundamental important publications in scientometrics such as **Moed** (2005)– passed away in 2021 and Tibor Braun, the founder of *Scientometrics*, in 2022 (**Glänzel**; **Heeffer**, 2023). These are great losses for the scientometrics field that can be scarcely compensated. Niklas Luhmann (the founder of the *social systems theory* that was fundamental for Loet's theorizing on science) summarized his research program about a decade before his death (**Luhmann**, 2012a; 2012b). In a similar way, Loet summarized his research program two years before his death: "*The evolutionary dynamics of discursive knowledge*" (**Leydesdorff**, 2021). **Wouters** and **Wagner** (2023) identified three formative themes in this program:

> "1) the dynamics of science, technology, and innovation; 2) the scientometric operationalization and measurement of these dynamics; and 3) the Triple Helix of university-industry-government relationships" (p. 3).

The empirical results of this study can confirm these themes in Loet's program. Based on the results of our study, we would like to add the theme "quantitative research evaluation". Loet was one of the most important actors in a far-reaching debate on optimizing field-normalized citation scores in scientometrics.

With Loet, we lose a prototype of a researcher. He was extremely interested in many research topics. This was expressed not only in many publications, but also in many verbal contributions. We cannot imagine a conference or meeting, in which Loet did not present own research or was very active in the discussions with many important contributions. Even until a few weeks before his death, Loet actively participated in the *CWTS* Friday afternoon seminar:
https://www.cwts.nl/seminars/information

In research collaborations, we experienced him as a researcher who was always interested in learning new methods, techniques or approaches. In collaborative research projects, he was not interested in handing off work (e.g., the statistical analysis of the data), but in learning how to do the work himself that others were doing in a project.

For us, it was a pleasure to work with Loet in many research projects, since he had excellent ideas and a fundamental background in scientometrics. We cannot imagine a common research project that would come to nothing: The way from the idea to the paper was always characterized by Loet's inspiring contributions. But Loet's contributions were not only restricted to research projects that led to publications; he also provided programs (with source code) on his webpage (see: *https://www.leydesdorff.net*) that he developed for research projects. Both, his publications and programs will surely continue to be extremely helpful to the scientometrics community.

6. References

Ahlgren, Per; **Jarneving, Bo**; **Rousseau, Ronald** (2003). "Requirements for a cocitation similarity measure, with special reference to Pearson's correlation coefficient". *Journal of the American Society for Information Science and Technology*, v. 54, n. 6, pp. 550-560.
https://doi.org/10.1002/asi.10242

Bornmann, Lutz (2013). "What is societal impact of research and how can it be assessed? A literature survey". *Journal of the American Society of Information Science and Technology*, v. 64, n. 2, pp. 217-233.

Bornmann, Lutz (2019). "Bibliometric indicators". In: P. Atkinson, S. Delamont, A. Cernat, J. W. Sakshaug & R. Williams (eds.), *SAGE Research Methods Foundations*. Thousand Oaks, CA, USA: Sage.

Bornmann, Lutz; **Haunschild, Robin**; **Leydesdorff, Loet** (2018). "Reference publication year spectroscopy (RPYS) of Eugene Garfield's publications". *Scientometrics*, v. 114, n. 2, pp. 439-448.
https://doi.org/10.1007/s11192-017-2608-3

Bornmann, Lutz; **Leydesdorff, Loet** (2011). "Which cities produce more excellent papers than can be expected? A new mapping approach, using Google Maps, based on statistical significance testing". *Journal of the American Society For Information Science and Technology*, v. 62, n. 10, pp. 1954-1962.
https://doi.org/10.1002/asi.21611

Bornmann, Lutz; **Leydesdorff, Loet** (2012). "Which are the best performing regions in information science in terms of highly cited papers? Some improvements of our previous mapping approaches". *Journal of informetrics*, v. 6, n. 2, pp. 336-345.
https://doi.org/10.1016/j.joi.2011.11.002

Bornmann, Lutz; **Leydesdorff, Loet** (2015). "Topical connections between the institutions within an organisation (institutional co-authorships, direct citation links and co-citations)". *Scientometrics*, v. 102, n. 1, pp. 455-463.
https://doi.org/10.1007/s11192-014-1425-1

Bornmann, Lutz; **Leydesdorff, Loet**; **Marx, Werner** (2007). "Citation environment of *Angewandte Chemie*". *Chimia*, v. 61, n. 3, pp. 104-109.
https://doi.org/10.2533/chimia.2007.104

Bornmann, Lutz; **Leydesdorff, Loet**; **Walch-Solimena, Christiane**; **Ettl, Christoph** (2011). "Mapping excellence in the geography of science: An approach based on Scopus data". *Journal of informetrics*, v. 5, n. 4, pp. 537-546.
https://doi.org/10.1016/j.joi.2011.05.005

Bornmann, Lutz; **Marx, Werner** (2014). "The wisdom of citing scientists". *Journal of the American Society of Information Science and Technology*, v. 65, n. 6, pp. 1288-1292.

Bornmann, Lutz; **Wagner, Caroline**; **Leydesdorff, Loet** (2015). "BRICS countries and scientific excellence: A bibliometric analysis of most frequently cited papers". *Journal of the Association for Information Science and Technology*, v. 66, n. 7, pp. 1507-1513.
https://doi.org/10.1002/asi.23333

Burt, Ronald-Stuart (1982). *Toward a structural theory of action: Network models of social structure, perception, and action*. Cambridge, MA, USA: Academic Press. ISBN: 978 0 121471507

Coleman, James S. (1990). *Foundations of social theory*. Cambridge, MA, USA: Belknap Press of Harvard University Press. ISBN: 978 0 674312265

Dinsmore, Adam (2015). *2:AM Amsterdam 2015.*
https://altmetricsconf.wordpress.com

Etzkowitz, Henry; **Leydesdorff, Loet** (2000). "The dynamics of innovation: From national systems and ‚Mode 2' to a Triple Helix of university-industry-government relations". *Research policy*, v. 29, n. 2, pp. 109-123.
https://doi.org/10.1016/S0048-7333(99)00055-4

Freeman, Linton C. (1978/1979). "Centrality in social networks conceptual clarification". *Social networks*, v. 1, n. 3, pp. 215-239.
https://doi.org/10.1016/0378-8733(78)90021-7

Garfield, Eugene (1955). "Citation indexes for science - new dimension in documentation through association of ideas". *Science*, v. 122, n. 3159, pp. 108-111.

Garfield, Eugene (1970). "Citation indexing for studying science". *Nature*, v. 227, n. 5259, pp. 669-671.

Garfield, Eugene (1972). "Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies". *Science*, v. 178, n. 4060, pp. 471-479.

Garfield, Eugene (1979a). *Citation indexing: Its theory and application in science, technology, and humanities*. New York, NY, USA: John Wiley & Sons, Ltd.

Garfield, Eugene (1979b). "Is citation analysis a legitimate evaluation tool?". *Scientometrics*, v. 1, n. 4, pp. 359-375.

Gibbons, Michael; **Limoges, Camille**; **Nowotny, Helga**; **Schwartzman, Simon**; **Scott, Peter**; **Trow, Martin** (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. London, UK: SAGE.

Glänzel, Wolfgang; **Heeffer, Sarah** (2023). "Looking back with a heavy heart". *ISSI Newsletter*, v. 19, n. 1, pp. 1-2.
https://www.issi-society.org/media/1443/newsletter73.pdf

Haunschild, Robin (2021). *BibPlots: Plot functions for use in bibliometrics. R package version 0.0.8*.
https://cran.r-project.org/web/packages/BibPlots/index.html

Haunschild, Robin; **Bornmann, Lutz**; **Leydesdorff, Loet** (2015). "Networks of reader and country status: an analysis of Mendeley reader statistics". *PeerJ computer science*, v. 1, e32.
https://doi.org/10.7717/peerj-cs.32

Haunschild, Robin; **Leydesdorff, Loet**; **Bornmann, Lutz** (2020). "Library and information science papers discussed on Twitter: A new network-based approach for measuring public attention". *Journal of data and information science*, v. 5, n. 3, pp. 5-17.
https://doi.org/10.2478/jdis-2020-0017

Haunschild, Robin; **Marx, Werner**; **Thor, Andreas**; **Bornmann, Lutz** (2020). "How to identify the roots of broad research topics and fields? The introduction of RPYS sampling using the example of climate change research". *Journal of information science*, v. 46, n. 3, pp. 392-405.
https://doi.org/10.1177/0165551519837175

Hicks, Diana; **Wouters, Paul**; **Waltman, Ludo**; **De-Rijcke, Sarah**; **Ràfols, Ismael** (2015). "Bibliometrics: The Leiden Manifesto for research metrics". *Nature*, v. 520, n. 7548, pp. 429-431.
https://doi.org/10.1038/520429a

Kamada, Tomihisa; **Kawai, Satoru** (1989). "An algorithm for drawing general undirected graphs". *Information processing letters*, v. 31, n. 1, pp. 7-15.

Leydesdorff, Loet (1991). "The static and dynamic analysis of network data using information-theory". *Social networks*, v. 13, n. 4, pp. 301-345.

Leydesdorff, Loet (1997). "Why words and co-words cannot map the development of the sciences". *Journal of the American Society for Information Science*, v. 48, n. 5, pp. 418-427.
https://doi.org/10.1002/(SICI)1097-4571(199705)48:5<418::AID-ASI4>3.0.CO;2-Y

Leydesdorff, Loet (2001). *The challenge of scientometrics: The development, measurement, and self-organization of scientific communications*: Universal Publishers (upublish.com).

Leydesdorff, Loet (2003). "Can networks of journal-journal citations be used as indicators of change in the social sciences?" *Journal of documentation*, v. 59, n. 1, pp. 84-104.
https://doi.org/10.1108/00220410310458028

Leydesdorff, Loet (2006). "Can scientific journals be classified in terms of aggregated journal-journal citation relations using the Journal Citation Reports?". *Journal of the American Society for Information Science and Technology*, v. 57, n. 5, pp. 601-613.

Leydesdorff, Loet (2021). *The evolutionary dynamics of discursive knowledge: Communication-theoretical perspectives on an empirical philosophy of science*: Springer International Publishing.

Leydesdorff, Loet; **Bornmann, Lutz** (2011). “Integrated impact indicators (I3) compared with impact factors (IFs): An alternative research design with policy implications“. *Journal of the American Society of Information Science and Technology*, v. 62, n. 11, pp. 2133-2146.

Leydesdorff, Loet; **Bornmann, Lutz** (2012). “Percentile ranks and the integrated impact indicator (I3)“. *Journal of the American Society for Information Science and Technology*, v. 63, n. 9, pp. 1901-1902.
https://doi.org/10.1002/asi.22641

Leydesdorff, Loet; **Bornmann, Lutz**; **Mutz, R.**; **Opthof, T.** (2011). “Turning the tables on citation analysis one more time: Principles for comparing sets of documents“. *Journal of the American Society For Information Science and Technology*, v. 62, n. 7, pp. 1370-1381.
https://doi.org/10.1002/asi.21534

Leydesdorff, Loet; **Bornmann, Lutz**; **Wagner, Caroline S.** (2017). “Generating clustered journal maps: An automated system for hierarchical classification“. *Scientometrics*, v. 110, n. 3, pp. 1601-1614.
https://doi.org/10.1007/s11192-016-2226-5

Leydesdorff, Loet; **Bornmann, Lutz**; **Zhou, Ping** (2016). “Construction of a pragmatic base line for journal classifications and maps based on aggregated journal-journal citation relations”. *Journal of informetrics*, v. 10, n. 4, pp. 902-918.
https://doi.org/10.1016/j.joi.2016.07.008

Leydesdorff, Loet; **Cozzens, Susan**; **Van-den-Besselaar, Peter** (1994). “Tracking areas of strategic importance using scientometric journal mappings”. *Research policy*, v. 23, n. 2, pp. 217-229.
https://doi.org/10.1016/0048-7333(94)90054-X

Leydesdorff, Loet; **Oomes, Nienke A.** (1999). “Is the European monetary system converging to integration?”. *Social science information*, v. 38, n. 1, pp. 57-86.
https://doi.org/10.1177/053901899038001003

Leydesdorff, Loet; **Persson, Olle** (2010). “Mapping the geography of science: Distribution patterns and networks of relations among cities and institutes”. *Journal of the American Society for Information Science and Technology*, v. 61, n. 8, pp. 1622-1634.
https://doi.org/10.1002/Asi.21347

Leydesdorff, Loet; **Schank, Thomas** (2008). “Dynamic animations of journal maps: Indicators of structural changes aninterdisciplinary developments“. *Journal of the American Society for Information Science and Technology*, v. 59, n. 11, pp. 1810-1818.
https://doi.org/10.1002/asi.20891

Leydesdorff, Loet; **Van-den-Besselaar, Peter** (1997). “Scientometrics and communication theory: Towards theoretically informed indicators”. *Scientometrics*, v. 38, n. 1, pp. 155-174.
https://doi.org/10.1007/Bf02461129

Luhmann, Niklas (2012a). *Theory of society* (vol. 2). Stanford, CA, USA: Stanford University Press.

Luhmann, Niklas (2012b). *Theory of society* (vol. 1). Stanford, CA, USA: Stanford University Press.

Marx, Werner; **Bornmann, Lutz**; **Barth, Andreas** (2013). “Detecting the historical roots of research fields by Reference Publication Year Spectroscopy (RPYS)”. In: J. Gorraiz; E. Schiebel; C. Gumpenberger; M. Horlesberger; H. Moed (eds.). *14th International Society of Scientometrics and Informetrics Conference (ISSI)* (pp. 493-506). Vienne, Austria: University of Vienna.

McAllister, Paul R.; **Narin, Francis**; **Corrigan, James G.** (1983). “Programmatic evaluation and comparison based on standardized citation scores”. *IEEE Transactions on engineering management*, v. 30, n. 4, pp. 205-211.
https://doi.org/10.1109/TEM.1983.6448622

Moed, Henk F. (2005). *Citation analysis in research evaluation*. Dordrecht, The Netherlands: Springer.

Narin, Francis (1976). *Evaluative bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. Cherry Hill, NJ, USA: Computer Horizons.

Opthof, Tobias; **Leydesdorff, Loet** (2010). “Caveats for the journal and field normalizations in the CWTS (‚Leiden‘) evaluations of research performance”. *Journal of informetrics*, v. 4, n. 3, pp. 423-430.
https://doi.org/10.1016/j.joi.2010.02.003

R Core Team (2021). *R: A Language and Environment for Statistical Computing (Version 4.1.2)*. Vienna, Austria: R Foundation for Statistical Computing.
https://www.r-project.org

Seglen, Per O. (1997). “Why the impact factor of journals should not be used for evaluating research”. *British medical journal*, v. 314, n. 7079, pp. 498-502.
https://doi.org/10.1136/bmj.314.7079.497

Shannon, Claude E. (1948a). "A mathematical theory of communication". *Bell System technical journal*, v. 27, n. 4, pp. 623-656.
https://doi.org/10.1002/j.1538-7305.1948.tb00917.x

Shannon, Claude E. (1948b). "A mathematical theory of communication". *Bell System technical journal*, v. 27, n. 3, pp. 379-423.
https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Simon, Herbert A. (1973). "The organization of complex systems". In: H. A. Simon (ed.). *Models of discovery: And other topics in the methods of science* (pp. 3-27). Dordrecht: Springer Netherlands.

Storper, Michael (1997). *The regional world: Territorial development in a global economy*. New York, NY, USA: Guilford Publications.

Tahamtan, Iman; **Bornmann, Lutz** (2018). "Core elements in the process of citing publications: Conceptual overview of the literature". *Journal of informetrics*, v. 12, n. 1, pp. 203-216.
https://doi.org/10.1016/j.joi.2018.01.002

Tahamtan, Iman; **Bornmann, Lutz** (2022). "The Social Systems Citation Theory (SSCT): A proposal to use the social systems theory for conceptualizing publications and their citation links". *Profesional de la información*, v. 31, n. 4, e310411.
https://doi.org/10.3145/epi.2022.jul.11

Theil, Henri (1972). *Statistical decomposition analysis: With applications in the social and administrative sciences*. Amsterdam, The Netherlands: North-Holland Publishing Company.

Thor, Andreas; **Bornmann, Lutz**; **Haunschild, Robin** (2018). *CitedReferencesExplorer (CRExplorer) manual*.
https://andreas-thor.github.io/cre/manual.pdf

Thor, Andreas; **Bornmann, Lutz**; **Haunschild, Robin**; **Leydesdorff, Loet** (2021). "Which are the influential publications in the Web of Science subject categories over a long period of time? CRExplorer software used for big-data analyses in bibliometrics". *Journal of information science*, v. 47, n. 3, pp. 419-428.
https://doi.org/10.1177/0165551520913817

Thor, Andreas; **Bornmann, Lutz**; **Marx, Werner**; **Mutz, Rüdiger** (2018). "Identifying single influential publications in a research field: New analysis opportunities of the CRExplorer". *Scientometrics*, v. 116, n. 1, pp. 591-608.
https://doi.org/10.1007/s11192-018-2733-7

Thor, Andreas; **Marx, Werner**; **Leydesdorff, Loet**; **Bornmann, Lutz** (2016a). "Introducing CitedReferencesExplorer (CRExplorer): A program for reference publication year spectroscopy with cited references standardization". *Journal of informetrics*, v. 10, n. 2, pp. 503-515.
https://doi.org/10.1016/j.joi.2016.02.005

Thor, Andreas; **Marx, Werner**; **Leydesdorff, Loet**; **Bornmann, Lutz** (2016b). "New features of CitedReferencesExplorer (CRExplorer)". *Scientometrics*, v. 109, n. 3, pp. 2049-2051.
https://doi.org/10.1007/s11192-016-2082-3

Tukey, John W. (1977). *Exploratory data analysis*. Boston, MA, USA: Addison-Wesley Publishing Company. ISBN: 978 0 201076165

Vanderkam, Dan; **Allaire, J. J.**; **Owen, Jonathan**; **Gromer, Daniel**; **Thieurmel, Benoit** (2018). *Dygraphs: Interface to Dygraphs' interactive time series charting library. R package version 1.1.1.6*.
https://CRAN.R-project.org/package=dygraphs

Wagner, Caroline S.; **Leydesdorff, Loet** (2005). "Network structure, self-organization, and the growth of international collaboration in science". *Research policy*, v. 34, n. 10, pp. 1608-1618.
https://doi.org/10.1016/j.respol.2005.08.002

Whitley, Richard (1984). *The intellectual and social organization of the sciences*. Oxford, UK: Clarendon Press. ISBN: 978 0 199240456

Wouters, Paul; **Wagner, Caroline** (2023). "In memoriam Loet Leydesdorff (1948-2023)". *ISSI Newsletter*, v. 19, n. 1, pp. 3-4.
https://www.issi-society.org/publications/issi-newsletter

Wu, Lingfei; **Wang, Dashun**; **Evans, James A.** (2019). "Large teams develop and small teams disrupt science and technology". *Nature*, v. 566, pp. 378-382.
https://doi.org/10.1038/s41586-019-0941-9

Ziman, John (1996). "'Postacademic science': Constructing knowledge with networks and norms". *Science studies*, v. 9, n. 1, pp. 67-80.
https://doi.org/10.23987/sts.55095

Ziman, John (1998). "Why must scientists become more ethically sensitive than they used to be?". *Science*, v. 282, n. 5395, pp. 1813-1814.
https://www.science.org/doi/10.1126/science.282.5395.1813

Science overlay maps: A tribute to Loet Leydesdorff

Benjamín Vargas-Quesada; Wenceslao Arroyo-Machado; Teresa Muñoz-Écija; Zaida Chinchilla-Rodríguez

Nota: Este artículo se puede leer en español en:
https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/87524

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

Vargas-Quesada, Benjamín; **Arroyo-Machado, Wenceslao**; **Muñoz-Écija, Teresa**; **Chinchilla-Rodríguez, Zaida** (2023). "Science overlay maps: A tribute to Loet Leydesdorff". *Profesional de la información*, v. 32, n. 7, e320705.

https://doi.org/10.3145/epi.2023.dic.05

Manuscript received on 17th July 2023
Accepted on 19th August 2023

Benjamín Vargas-Quesada
https://orcid.org/0000-0001-5115-7460

Universidad de Granada
Unit for Computational Humanities and Social Sciences (*U-CHASS*)
Colegio Máximo de Cartuja
18071 Granada, Spain
benjamin@ugr.es

Wenceslao Arroyo-Machado ✉
https://orcid.org/0000-0001-9437-8757

Universidad de Granada
Unit for Computational Humanities and Social Sciences (*U-CHASS*)
Colegio Máximo de Cartuja
18071 Granada, Spain
wences@ugr.es

Teresa Muñoz-Écija
https://orcid.org/0000-0003-1311-0471

Independent researcher.
Granada, Spain
teresamunyozecija@gmail.com

Zaida Chinchilla-Rodríguez
https://orcid.org/0000-0002-1608-4478

Consejo Superior de Investigaciones Científicas (*CSIC*)
Instituto de Políticas y Bienes Públicos (*IPP*)
Albasanz, 26-28
28037 Madrid, Spain
zaida.chinchilla@csic.es

Abstract

This is a homage to Loet Leydesdorff, professor and leading scientist.. Through the combination of overlay maps, a visualization technique proposed by himself and Ismael Ràfols, together with the CAMEOs (Characterizations Automatically Made and Edited Online) proposed by Howard White, we project his scientific trajectory in five different scenarios, which turn out to be complementary. For each of the scenarios or CAMEOs, we show how he acts and interacts from the point of view of scientific research, providing the reader with online access to an interactive *VOSviewer* tool, so that he can check the information presented here, and even go deeper into the analysis. In fact, we encourage him to do so. To sum up, we can say that Loet was a brilliant scientist, a lone wolf who enjoyed collaborating with the best minds in his main research topics: scientific communication, innovation systems, bibliometrics, and science mapping; becoming in turn the reference point of these areas of research.

Keywords

Loet Leydesdorff; Science mapping; Overlay maps; CAMEOs; Network analysis; *VOSviewer*; Bibliometrics; Scientific communication; Researchers.

1. Introduction

Loet (Louis André) Leydesdorff was a renowned professor and researcher at the *University of Amsterdam*. He passed away on March 11, 2023. He was noted for his pioneering research in innovation systems, scientific communications, scientometrics and science mapping, proposing an important improvement for the latter: overlay maps (**Leydesdorff**; **Ràfols**, 2009). Overlays are a very powerful contribution integrating visualization techniques, social networks, cognitive and intellectual structure, changes over time, and benchmarking analysis, for any kind of scientific domain. In fact, overlays demonstrated right away their great potential for research policy analysis and library management (**Ràfols** *et al.*, 2010), building interactive maps (**Leydesdorff**; **Ràfols**, 2012), charting patent data (**Leydesdorff**; **Bornmann**, 2012), guesstimating interdisciplinarity (**Leydesdorff**; **Carley** *et al.*, 2013; **Leydesdorff**; **Ràfols** *et al.*, 2013), evaluating strategic intelligence in emerging technologies (**Rotolo** *et al.*, 2017), detecting and identifying emerging research fields (**Vargas-Quesada** *et al.*, 2017; **Muñoz-Écija** *et al.*, 2019), unveiling cognitive structures (**Muñoz-Écija** *et al.*, 2022), and comparing educational technologies (**Vargas-Quesada** *et al.*, 2021).

> Overlays are a very powerful contribution integrating visualization techniques, social networks, cognitive and intellectual structure, changes over time, and benchmarking analysis, for any kind of scientific domain

In this paper, we present a comprehensive homage to Loet Leydesdorff's remarkable career. Using scientific maps as a tool, we aim to synthesize the CAMEOs (Characterizations Automatically Made and Edited Online) proposed by **White** (2001) together with overlay maps –a technique Loet significantly contributed to. This approach enables us not only to reveal the various conceptual and social structures present in Loet's multifaceted research, but also to highlight the trends and visibility of his influential work across different domains. Thus, the specific objectives of this study are as follows:

1. To identify the primary research topics of Loet Leydesdorff's career, along with the trends and impacts throughout his professional journey.
2. To identify Loet Leydesdorff's main co-authors and his collaborative networks.
3. To analyze the researchers who have had an influence on Loet Leydesdorff's career, as well as those whom he has influenced.

2. Data and methods

2.1. Data

Data collection was conducted on April 27, 2023. We used *Clarivate*'s *Web of Science* (*WoS*) to search for all records of researchers with the surname "Leydesdorff" and whose first initial was "L." Two valid profiles were found: the first was verified, encompassing all production indexed in *Web of Science* (ResearcherID: E-2903-2010), and the second was unverified, with most of the production external to *Web of Science* (ResearcherID: DUT-0376-2022). We merged both profiles and downloaded all bibliographic records indexed in the *Web of Science Core Collection* in a tab-delimited file. In total, 427 bibliographic records were downloaded, representing the total research output of Loet Leydesdorff, with no documentary or temporal filtering applied.

2.2. Methods

The analysis carried out is mainly based on the social network analysis for the construction of scientific maps. Specifically, we have generated the following maps:

1. The conceptual structure, composed of the main topics and areas of research in which he published during his career.
2. The co-authors, composed of his main collaborators.
3. The citation identity, composed of the authors cited by Loet Leydesdorff and therefore on which his research is founded.
4. The citation image-makers, composed of the authors who cite Loet Leydesdorff and therefore on whom he influences.

We used *VOSviewer* (**Van Eck**; **Waltman**, 2010) for the construction of science maps. The specific details and processes involved in their elaboration are comprehensively outlined in Table 1. For each map, we have developed different overlay versions based on these networks. The purpose of these overlays is to augment the existing information with additional layers, thereby enabling a more in-depth identification and analysis of research performance and trends. These overlays essentially act as lenses that bring into focus the multifaceted aspects of Loet's research impact, thus offering a more

Table 1. Summary of scientific maps created on the production of Loet Leydesdorff

Network	Level	Data processing	Network filters
Conceptual structure Co-occurrence network	Publication	**Data:** Loet's Leydesdorff publications **Processing:** term extraction from titles and abstracts and normalization through the creation of a thesaurus	- Binary counting - Minimum 5 occurrences
Co-author Co-author network	Author	**Data:** Loet's publications **Processing:** author disambiguation through the creation of a thesaurus	- Full counting - Minimum 2 collaborations
Citation identity Citation network	Author	**Data:** Loet's publications + cited references **Processing:** the Britton Chance bibliometric analysis (**Li** *et al.*, 2014) was removed to avoid introducing noise with its references	- Minimum 3 documents
Citation image-makers Citation network	Author	**Data:** publications citing Loet's oeuvre	- Minimum 5 documents

comprehensive view of his scholarly influence and trajectory. In the case of the word co-occurrence network, the overlay map includes ad hoc indicators processed from the publications. These are the average age of the publications and the percentage of publications as the first author. In both cases, the values are normalized to range between 0% and 100%. The rest of the indicators used in overlay maps are the default calculated ones: average publication year and average normalized citations.

> In Loet's production we can distinguish three periods: 1980-2004 (preliminary development); 2005-2013 (fast development); and 2014-2022 (downward development)

All CAMEOs are accessible for interactive viewing via *VOSviewer Online*. Accompanying each figure, a link is provided to this application, enabling the exploration of the different scenarios of Loet's research trajectory and its evolution, as well as the possibility of visualizing other bibliometric indicators such as average publication year and average normalized citations, which further broaden the analytical value of each of these CAMEOs. For each map, it is specified whether the overlay can be applied via color variables (item colors) or size (item size). These options, among others included in the tool, are selectable from the drop-down panel on the left.

Due to space limitations, we do not conduct a detailed/depth analysis of each CAMEO. However, we leave this to the reader, so that she/he can take advantage of the possibilities of the tool we put at his disposal and keep discovering Loet's characterization, –and who knows, maybe she/he will find himself on those maps.

3. Results

Loet's production runs from 1980 to 2022[1] (Figure 1). We can distinguish three periods: 1980-2004 (preliminary development); 2005-2013 (fast development); and 2014-2022 (downward development). A quarter of his production is characterized by single-authorship papers (26.53%; 113 out of 426), a low level of co-authorship (2.23), and high percentage of international collaborative publications (57.41%). The correlation coefficient between average number of authors and number of publications is 0.042. Those of us who knew him know that all this says a lot about Loet's way of being and working, and all this carries over to his CAMEOs.

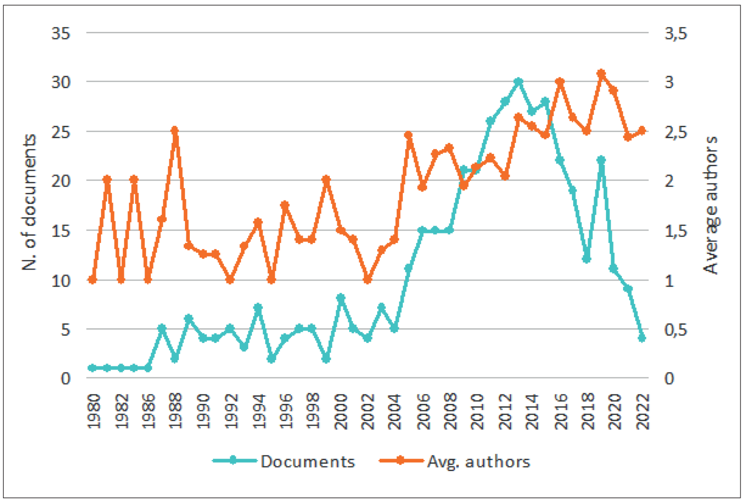


Figure 1. Loet Leydesdorff 's scientific production and co-authorship patterns.

3.1. Loet's conceptual structure

The conceptual structure, depicted through the most frequently co-occurring terms, outlines the primary areas and specific topics that Loet discusses most. With the aid of overlay maps, it is possible to discern main trends and assess the impact of these areas and topics more effectively.

Figure 2 presents the conceptual structure of research conducted by Loet. The thematic map (Figure 2a) highlights four main areas in which he has published: bibliometrics, innovation systems, science mapping, and scientific communication. This map displays a distinct division between areas focused on scientometrics and social network analysis (on the left) and areas focused on innovation systems theories and the sociology of innovation (on the right).

Within each of these topics, there exist certain topics that marked the initial trajectory of his research career, as depicted in Figure 2b. It underscores his unique contribution to the field of Communication Studies and Science and Technology Studies, with the triple helix model (**Etzkowitz**; **Leydesdorff**, 1995) (right), that aims to comprehend the relationships and collaboration between university-industry-government to understand the transformation of academic knowledge within the economy through innovation strategies. Similarly, terms with notable average age (e.g., map, graph, and citation network) stand out Loet's role as international scientific benchmark in science mapping, what enhanced the analysis and visualization of the structure and dynamics of his scientific activity (**Leydesdorff**, 1987; **Wagner**; **Leydesdorff**, 2005; **Leydesdorff**; **Ràfols**, 2009; **Ràfols** *et al.*, 2010). Other topics that have been figured from his beginnings are the design and use of indicators to predict, evaluate, and analyze scientific production, such as national performance in relation to the proportion of words in publications (**Leydesdorff**, 1990), the probabilistic entropy (**Leydesdorff**, 2003) or betweenness centrality (**Leydesdorff**, 2007a), as well as the use of various analysis units (e.g. citation) to set up citation analysis on citation-based indicators (**Amsterdamska**; **Leydesdorff**, 1989).

Connections between main topics on the left and the right are basically established through nodes titled indicator (upper-left part), network (middle part), and technology (upper-right part), all of them key and cross-cutting topics throughout Loet Leydesdorff's research career. For instance, technology is linked to the innovation systems in the upper-right of the

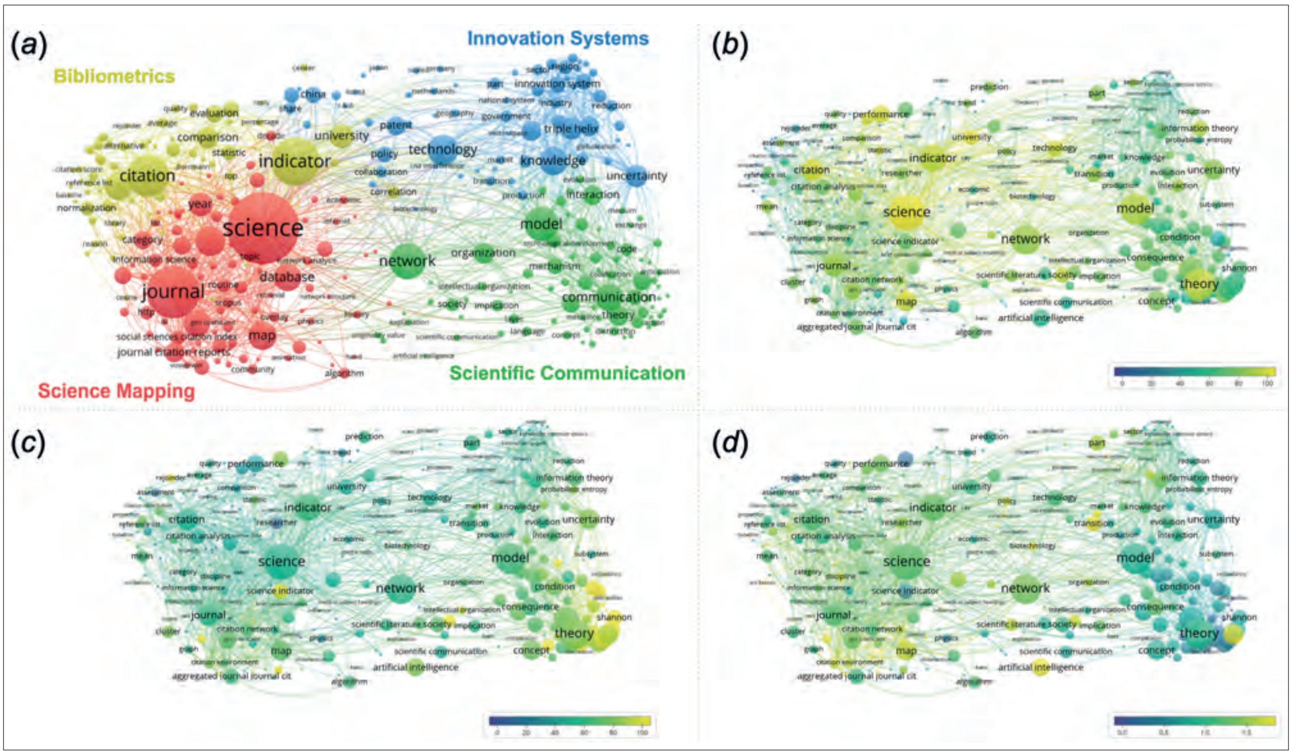


Figure 2. Loet Leydesdorff conceptual structure: (a) thematic landscape base map; (b) average age of the publications; (c) percentage of publications as the first author; (d) average normalized citations.
Visualizations available at *https://tinyurl.com/292roq5d*
Overlay map by colors (item colors -customizable at the left panel of the tool.

map, continuing with research in indicators to measure science and technology, in the upper-left part, using various methods and units –citation analysis and citation in the upper-left part; and map and journal in the lower-left part.

However, a crucial observation to make is that his performance and influence across these diverse areas does not display a consistent uniformity. Differences become evident when examining papers where he serves as the first author, understood as a proxy of leadership (**Zhou**; **Leydesdorff**, 2006) (Figure 2c), with a predominant leading role in scientific communication (we refer to **Leydesdorff**, 1994a; 1994b; 2000; 2001; 2007b; 2010; 2013; 2016; 2020; 2021). Intriguingly, this area demonstrates the lowest scholarly impact (Figure 2d), in comparison with the remaining areas where his influence is more evenly distributed. This trend shows that research areas that are more focused on theories and dogmas, in particular scientific communication, have a lower scholarly impact compared to areas concentrated on empirical data-based research, such as innovation systems and science mapping.

Regarding Loet's research (Figure 3), there is no evidence of periods in which activity is focused solely on a single topic. From its inception, his research has been highly diversified. Only in the early years, specifically in the late 1990s and early 2000s, we do observe most of the production focused on scientific communication and innovation systems. However, this period also corresponds to a lower rate of production, which could potentially accentuate these differences. From 2005 onwards, coinciding with a surge in productivity, research topics are more diversified, covering subjects across all four main areas identified in Figure 2.

For instance, by the late 1980s, the most predominant research areas were bibliometrics (e.g., evaluation, citation, statistics, citation analysis, and university) and science mapping (e.g., science and map). Nevertheless, in 1989, terms from the area of scientific communication become much more leading (e.g., intellectual organization and co-word). The same fact happens in 1990 with the area of innovation systems (e.g., information theory and prediction).

In the first half of the 1990s, research in scientific communication (e.g., model, society, network, scientific knowledge) and innovation systems (e.g., uncertainty, information theory, policy, and emergence) gain strength. Additionally, the topics within the field of science mapping remain active, while some bibliometrics topics lose prominence. It is worth noting that it is during these years when several articles on the triple helix model are published.

In the second half of the 1990s, there is a clear deepening in the study of areas of scientific communication (e.g., model, network, society, communication, theory, co-word, technological development) and innovation systems (e.g., triple helix, collaboration, technology, innovation, market, state, knowledge, competition). The area of science mapping remains active, but certain terms gain more significance (e.g., algorithm, discipline, journal, aggregate journal-journal citation, scientometrics, and scientific literature).

> Terms with notable average age (e.g., map, graph, and citation network) stand out Loet's role as international scientific benchmark in science mapping

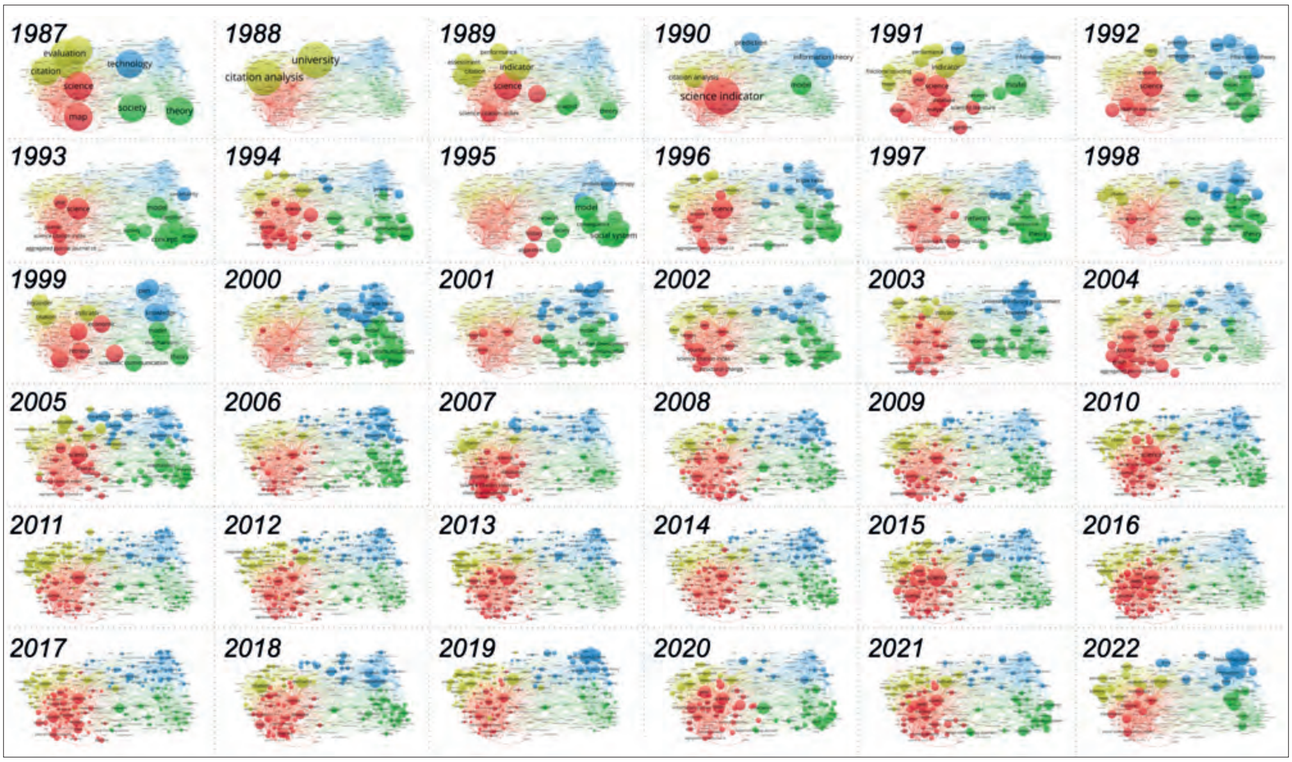


Figure 3. Overlay map of the annual evolution of Loet Leydesdorff topics between 1987 and 2022.
Visualizations available at https://tinyurl.com/24lnrkml
Overlay map by size (item size).

During the early 2000s, areas of bibliometrics, scientific communication, and innovation systems remain consistent. However, there is a notable increase in terms within the science mapping cluster between 2003 and 2004 (e.g., visualization, social network analysis, discipline, internet, cluster, classification, graph, and overlay). Moreover, in this period terms in central positions experience a growth, serving as a link between areas on the left and right (e.g., network, technology, and university).

Throughout the later years of the 2000s, the topics within science mapping continue increasing (e.g., web, journal, relation, research front, citation network, citation data, map, *Journal Citation Reports*, and *Scopus*). Bibliometrics also demonstrates an upward trend (university, indicator, citation, evaluation, normalization, ranking, and fractional counting), while the remaining areas remain stable.

Between 2011 and 2015, activity remains consistent in the four main areas. However, topics within the scientific communication field show reduced activity, while terms within bibliometrics (e.g., correlation, impact factor, and integrated impact indicator) and science mapping (e.g., animation, *VOSviewer*, subject area, overlay, interdisciplinarity) experience increased activity. This trend persists between 2016 and 2020, with a resurgence of activity in the scientific communication area (e.g., knowledge production, organization, subsystem, co-evolution) and heightened activity in certain innovation system topics (e.g., collaboration, international collaboration, patent, firm, and national system).

In the span of 2021 to 2022, the four research areas maintain their activity, with a homogeneous level of activity. It is evident that the diversification in Loet's productivity remains stable when his collaboration with other authors is high. The average collaboration indicator shows the highest values starting from 2016 when his scientific production experiences a decline.

3.2. Co-authors

Figure 4 shows Loet's social relationships, how he relates to one another, and how these co-authors can be grouped through clusters identifying groups and lines of research.

Loet publishes with 76 co-authors (Figure 4). The red cluster comprises 45 co-authors highlighting Lutz Bornmann as the most prolific partner (Figure 4a). Researchers participating in this cluster share a common research topic, scientometrics. The same pattern is observed for the two science mapping clusters, colored in blue and purple, with 9 and 6 authors. In particular, purple cluster focuses on the application of science mapping with *Scopus* dataset, as well as technical and

> "Research areas that are more focused on theories and dogmas, in particular scientific communication, have a lower scholarly impact compared to areas concentrated on empirical data-based research, such as innovation systems and science mapping"

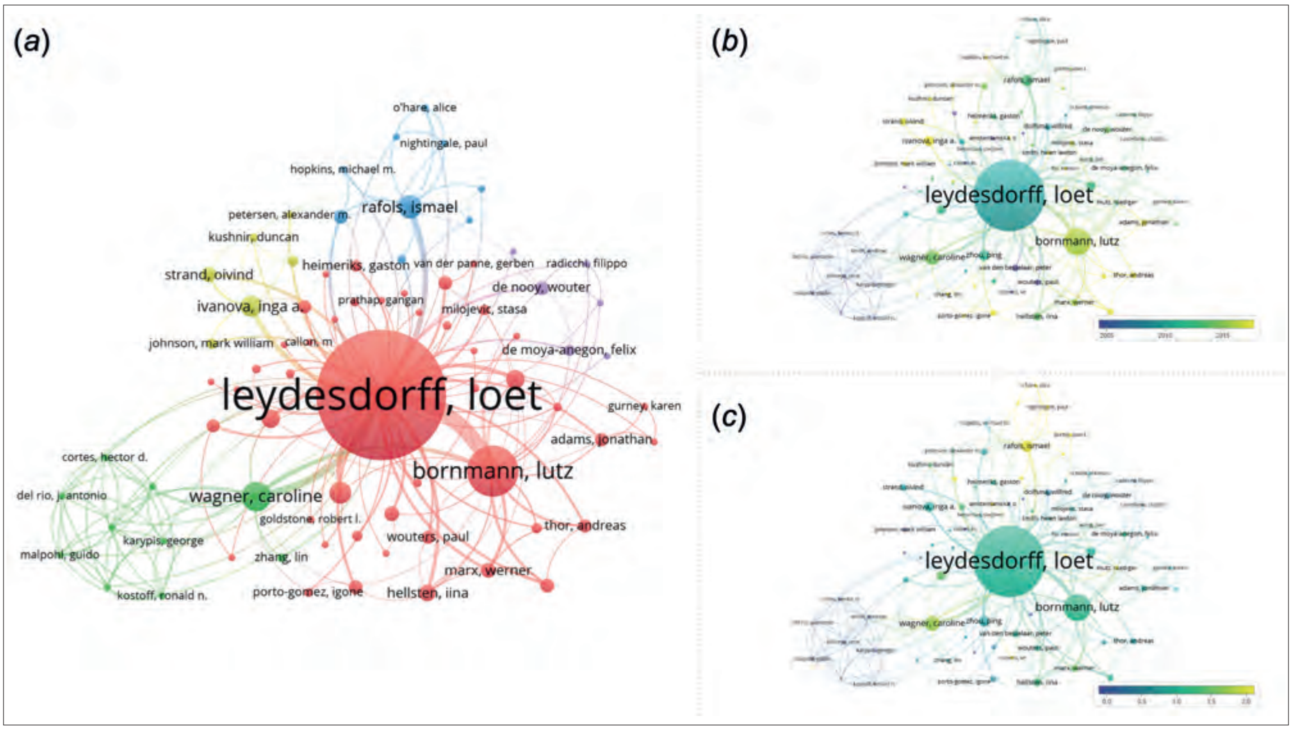


Figure 4. Focal Loet Leydesdorff co-authors: (a) co-author network base map; (b) average publication year; (c) average normalized citations. Visualizations available at *https://tinyurl.com/29kc3yp4*
Overlay map by colors (item colors).

theoretical attributes of network analysis. Here, Ismael Ràfols and Wouter de Nooy emerge as principal contributors, respectively. Similarly, the green cluster is dominated by Caroline Wagner. The focal areas within this co-author group revolve around science policy and international collaboration. Lastly, the yellow cluster displays a reduced number of collaborators –only 6– who developed research in the area of the innovation systems.

When considering the average publication year, Lutz Bornmann is also a noteworthy figure (Figure 4b). Besides being the most prolific partner, his publications are also primarily concentrated in recent years. They published their first paper as co-author in 2007 (**Bornmann** *et al.*, 2007). Together they have published more than 75 publications in different areas of scientometrics such as field normalized indicator or network analysis. Inga Ivanova is another of the most active collaborators since 2014 (**Leydesdorff**; **Ivanova**, 2014) with 15 publications in the innovation systems, whose publication output continued up until 2021. As it is clearly visible, Caroline Wagner is also another of the top co-author (28 publications), presenting the distribution of publications enhanced stability over the years (2003-2022).

Nevertheless, it is worth noting the scholarly impact of publications carried out in conjunction with Ismael Ràfols (2.23) in the blue cluster, which have generated the greatest impact (ranged from 0.84 to 3.32), as depicted in Figure 4c. The same applies in the yellow cluster to Henry Etzkowitz, co-author alongside Loet of a few papers, which have achieved utmost average impact (2.56).

3.3. Loet's citation identity

The citation identity shows those from whom Loet consumes and uses scientific information, who are his reference points/benchmarks, to whom he pays tribute with his citations, and how they relate to each other.

The citation network of Loet Leydesdorff primarily revolves around three main areas, with two additional smaller communities that focus on more specific topics (Figure 5). The foremost area encompasses traditional studies of bibliometrics and citation analysis, represented by the red cluster. Here, we can be found leading authors in the field such as Ben R. Martin, Wolfgang Glänzel, and Ronald Rousseau. Closely associated with this is the green cluster, comprising the most recent studies in scientometrics. Notable authors in this area include Lutz Bornmann, Ludo Waltman, and Mike Thelwall. The third significant area involves network studies, represented by the blue cluster, where authors such as Ismael Ràfols, Alan Porter, and Kevin Boyack are prominent. It is upon the work of these authors that Loet's research is founded.

Nonetheless, we cannot overlook the rest of clusters even though the sizes are considerably smaller. On the one hand, the scientific knowledge derived from the purple cluster has guided him in advancing research in international collaboration, public policy, identification

> Leydesdorff effect on leading scientists in the field of science and technology studies, in particular the scholarly communication, including Félix De-Moya-Anegón, Cassidy Sugimoto and Vincent Larivière is remarkable

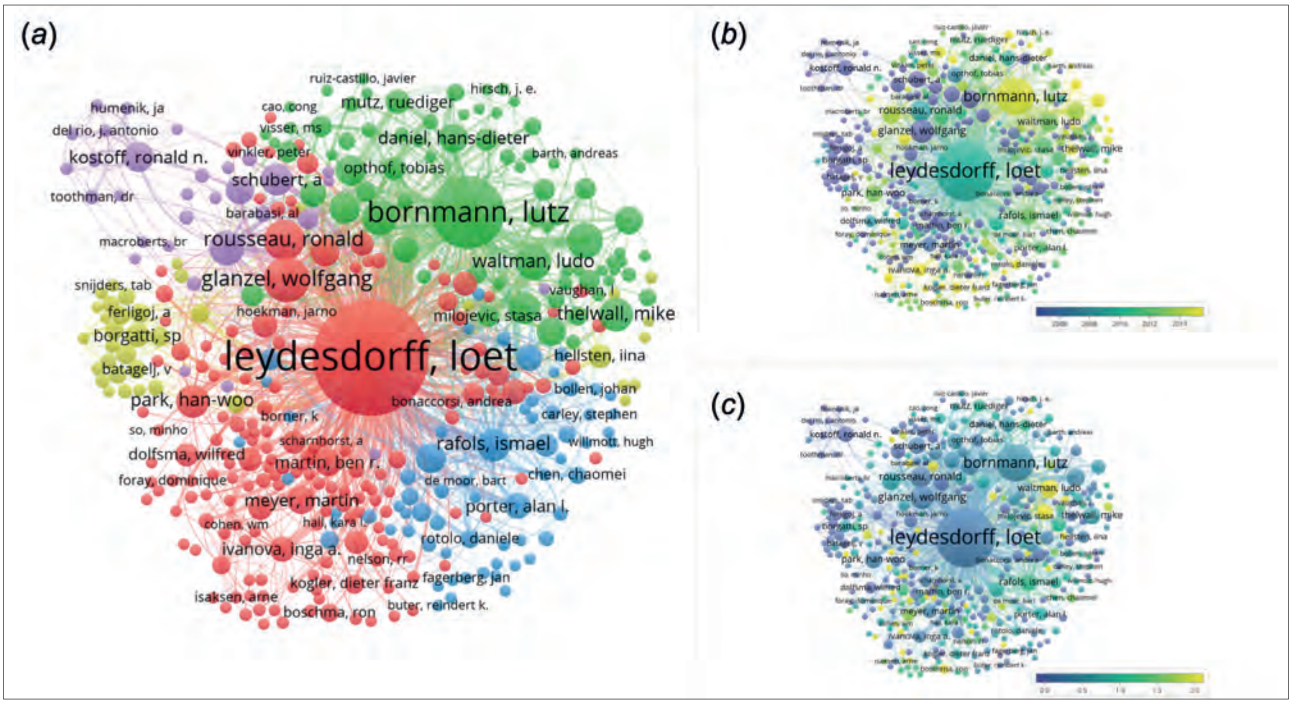


Figure 5. Loet Leydesdorff's citation identity: (a) citation network base map; (b) average publication year; (c) average normalized citations.
Visualizations available at *https://tinyurl.com/26uvs7kg*
Overlay map by colors (item colors).

of trends and delineation of research domains. Among these authors Caroline Wagner, András Schubert, and Ronald Kostoff stand out. On the other hand, authors in the yellow cluster have served as the sources of knowledge upon which Loet has relied for the development of research in social networks analysis and computational models. Prominent authors in these themes are Vladimir Batagelj, and Stephen Borgatti.

3.4. Loet's citation image-makers

This network shows just the opposite of the previous one. It reveals who consumes and uses the scientific information produced by Loet, how they relate and group themselves into invisible colleges.

In this case, the network of co-authors generated from the references of the publications citing Loet Leydesdorff (Figure 6) reflects similarities with respect to what was seen previously. The authors of bibliometrics and data analysis occupy a relevant space (red cluster), in which Loet Leydesdorff and Lutz Bornmann have a predominant presence. His effect on leading scientists in the field of science and technology studies, in particular the scholarly communication, including Félix De Moya Anegón, Cassidy Sugimoto and Vincent Lariviere (purple cluster, upper part) is also remarkable, holding central positions. Within the same cluster more focused on the scientific knowledge representation (lower part), renowned authors are displayed such as Ronald Rousseau, Wolfgang Glänzel, Ying Ding, Richard Klavans, and Kevin Boyack. The influence that Loet has in the field of technology policy assessment and emerging technology identification (blue cluster) is evident, with notable authors such as Alan Porter, Ismael Ràfols, Jan Youtie, and Philip Shapira standing out, among others. Likewise, his scientific endeavors have had a profound impact on development of indicators for analyzing knowledge-based innovation systems and scientific collaboration, as evidenced by the citations from Caroline Wagner, Han Woo Park, and Giovanni Abramo.

Lastly, there are two clusters with authors from other knowledge domains. The yellow cluster represents the influence on authors whose research careers are centered around Economy within Business Science, with a specific focus on entrepreneurship and innovation. The smaller one, the light blue cluster, comprises computer science researchers who study methods and techniques for analyzing and making decisions based on information and knowledge.

In terms of average normalized citations, researchers from the red cluster, along with those from the yellow and light blue clusters, exhibit the highest values (Figure 6c). By contrast, the average publication year of publications that cited Loet's output shows homogeneity among the various clusters comprising the network (Figure 6b). Sociology has been the pivotal axis of Loet's research, acting as a connecting bridge between his research on innovation systems and models and areas of scientometrics and social network analysis. This demonstrates the heterogeneity and transcendence/scope of the network, as Loet receives citations from diverse knowledge domains.

Sociology has been the pivotal axis of Loet's research, acting as a connecting bridge between his research on innovation systems and models and areas of scientometrics and social network analysis

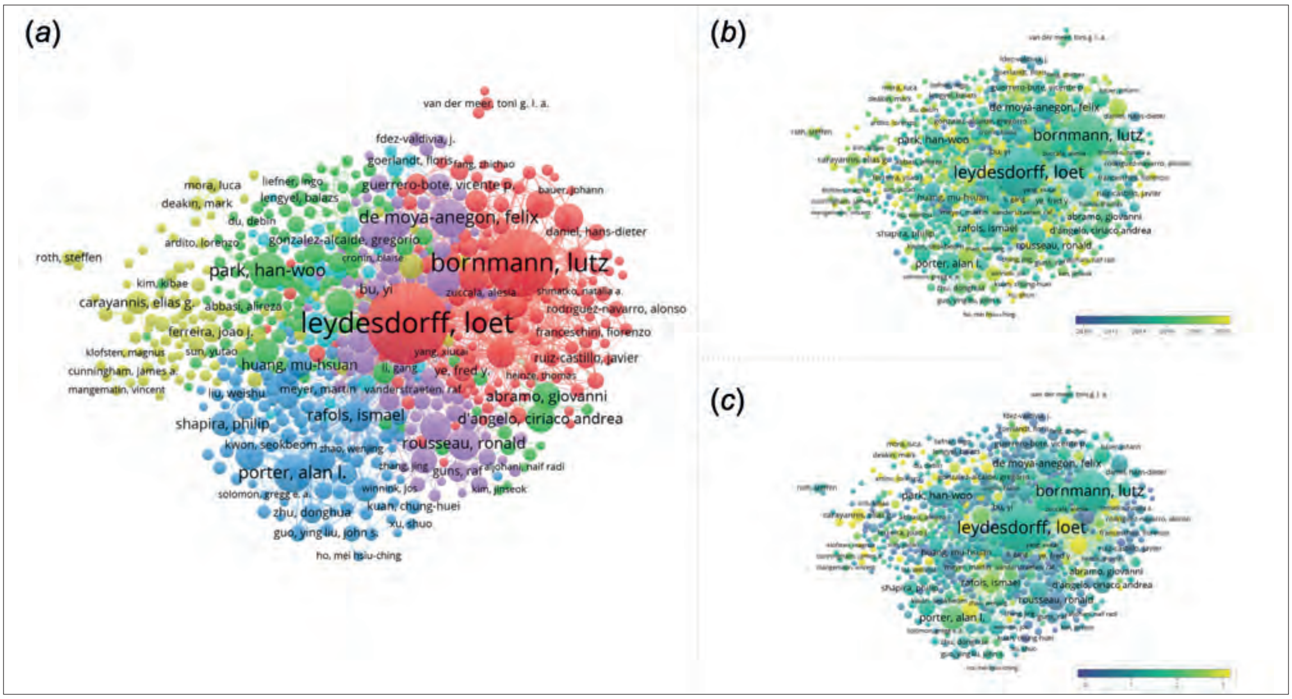


Figure 6. Loet Leydesdorff's citation image-makers: (a) citation network base map; (b) average publication year; (c) average normalized citations. Visualizations available at *https://tinyurl.com/28lzkco3*
Overlay map by colors (item colors).

Additionally, this emphasizes how his research encompasses multiple knowledge areas and how his cross-discipline intellectual perspectives exert significant influence within the scientific community.

4. Limitations

We distinguish three different periods based on the identification of main research topics and the number of co-authors at international level as a starting point for analyzing Loet's oeuvre. Notwithstanding, none of the criteria selected allow us to determine these periods with exactitude. Indeed, we attempted to utilize the evolution of the topics on which Loet has published throughout his career. However, no distinct periods are discernible in his production based on this analysis (Figure 7). Other approaches could be applied in further research to explore the possibility of stablishing these periods in a more robust way (**Glänzel**; **Abdulhayoğlu**, 2018).

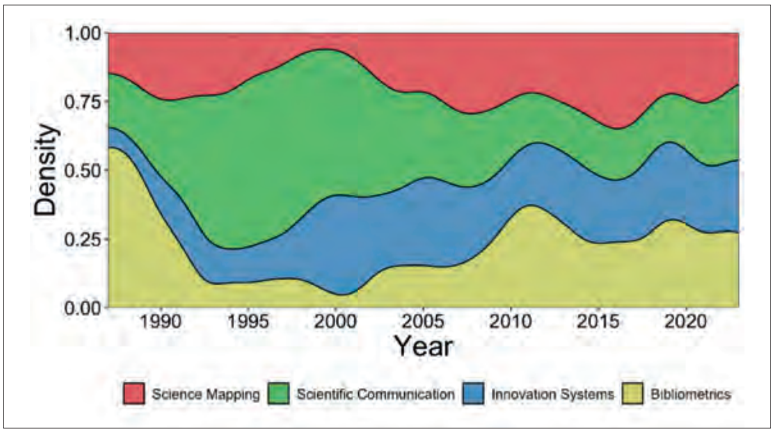


Figure 7. Annual evolution of topics published by Loet Leydesdorff by field

From a bibliometric perspective, a CAMEO allows us to see how a focal author, Loet in this case, performs in different scenarios. It is true that these CAMEOs are based on information extracted from *Web of Science*. Even if the number of publications does not significantly differ from those obtained in *Scopus* (three more documents), it is possible to enrich some CAMEOs using other sources such as *Dimensions*, *OpenAlex*, or *Google Scholar*. Nonetheless, depending on the data source used, CAMEOs based on citation would be greatly different, reduced, or even non-existent. On the other hand, we justify the absence of the CAMEO citation image because it is derived from the co-citation map based on Loet Leydesdorff as focal and the information provided coincides to a large extent with the co-authorship and citation maps.

5. Conclusions

This is a tribute to Loet Leydesdorff. Using overlay maps, a visualization technique developed by him, and CAMEOs, we show how he acts and interacts in different scenarios with science and its different actors.

Loet has always been a character. His research career has been characterized by a high number of solo papers, international collaboration, and thematic diversity. Only at the beginning, a greater focus on scientific communication and innovation systems is observed, but over time, as his productivity and scientific collaboration increases, his research topics diversify, encompassing four main subjects: scientific communication, innovation systems, bibliometrics, and science mapping.

From the point of view of scientific collaboration, Loet maintains strong rapport with his colleagues. Despite his great scientific productivity, he has distinguished himself by working alone and maintaining a small group of collaborators throughout his research life. Lutz Bornmann, Caroline Wagner, and Ismael Ràfols, the latter with whom he developed the overlay maps, stand out. The Citation Identity CAMEO reveals the authors from whom Loet draws inspiration for his research. In our field, particularly noteworthy are Ben R. Martin, Wolfgang Glänzel, and Ronald Rousseau, concerned with citation analysis; Lutz Bornmann, Ludo Waltman, and Mike Thelwall, with the most recent studies in scientometrics; and Ismael Ràfols, Alan Porter, and Kevin Boyack, with science mapping. Citation Image-Makers CAMEO shows the authors who consume Loet's scientific literature to generate new knowledge. Curiously, although this CAMEO would be the antagonist of the previous one, its results are very similar. In other words, the authors who are inspired by Loet and use him as a benchmark in their research are the same ones he turns to for reference points.. We can state that there is a very strong feedback process between his co-authors, those he cites and those who cite him, highlighting again Lutz Bornmann, Caroline Wagner, Alan Porter, and Ismael Ràfols.

> Over time, as Leydesdorff's productivity and scientific collaboration increases, his research topics diversify, encompassing four main subjects: scientific communication, innovation systems, bibliometrics, and science mapping

In the coming years, when we attend a congress or a bibliometric meeting, we will miss that at the end of any communication, the Chairman says: answers, questions... Loet?

Sit tibi terra levis, Loet.

6. Note

The only article from 2023, which is a letter authored by the *Distinguished Reviewers Board of Scientometrics*, has been omitted as it distorts the production picture for that year (**Abramo** *et al.*, 2023).

7. References

Abramo, Giovanni; **Aguillo, Isidro F.**; **Aksnes, Dag W.**; **Boyack, Kevin**; **Burrell, Quentin L.**; **Campanario, Juan-Miguel**; **Chinchilla-Rodríguez, Zaida**; **Costas, Rodrigo**; **D'Angelo, Ciriaco-Andrea**; **Harzing, Anne-Wil**; **Jamali, Hamid R.**; **Larivière, Vincent**; **Leydesdorff, Loet**; **Luwel, Marc**; **Martin, Ben**; **Mayr, Philipp**; **McCain, Katherine W.**; **Peters, Isabella**; **Ràfols, Ismael**;… **Waltman, Ludo** (2023). "Retraction of Predatory publishing in Scopus: Evidence on cross-country differences lacks justification". *Scientometrics*, v. 128, n. 2, pp. 1459-1461.
https://doi.org/10.1007/s11192-022-04565-6

Amsterdamska, Olga; **Leydesdorff, Loet** (1989). "Citations: Indicators of significance?". *Scientometrics*, v. 15, pp. 449–471.
https://doi.org/10.1007/BF02017065

Bornmann, Lutz; **Leydesdorff, Loet**; **Marx, Werner** (2007). "Citation environment of Angewandte Chemie". *Chimia*, v. 61, n. 3, pp. 104−109.

Etzkowitz, Henry; **Leydesdorff, Loet** (1995). "The Triple Helix - University-Industry-Government relations: A laboratory for knowledge based economic development". *EASST review*, v. 14, n. 1, pp. 14–19.

Glänzel, Wolfgang; **Abdulhayoğlu, Mehmet-Ali** (2018). "Garfield number: On some characteristics of Eugene Garfield's first and second order co-authorship networks". *Scientometrics*, v. 114, n. 2, pp. 533-544.
https://doi.org/10.1007/s11192-017-2623-4

Leydesdorff, Loet (1987). "Various methods for the mapping of science". *Scientometrics*, v. 11, n. 5-6, pp. 295–324.
https://doi.org/10.1007/bf02279351

Leydesdorff, Loet (1994a). "The evolution of communication". *International journal systems research and information science*, v. 6, pp. 219−230.

Leydesdorff, Loet (1994b). "Uncertainty and the communication of time". *Systems research*, v. 11, n. 4, pp. 31−51.

Leydesdorff, Loet (1990). "The prediction of science indicators using information theory". *Scientometrics*, v. 19, n. 3-4, pp. 297–324.
https://doi.org/10.1007/bf02095353

Leydesdorff, Loet (2000). "Luhmann, Habermas and the theory of communication". *Systems research and behavioral science: The official journal of the International Federation for Systems Research*, v. 17, n. 3, pp. 273−288.
https://doi.org/10.1002/(SICI)1099-1743(200005/06)17:3<273::AID-SRES329>3.0.CO;2-R

Leydesdorff, Loet (2001). *A sociological theory of communication: The self-organization of the knowledge-based society*. Universal Publishers. ISBN: 1 58112 695 6

Leydesdorff, Loet (2003). "The mutual information of university-industry-government relations: An indicator of the Triple Helix dynamics". *Scientometrics*, v. 58, n. 2, pp. 445–467.
https://doi.org/10.1023/a:1026253130577

Leydesdorff, Loet (2007a). "Betweenness centrality as an indicator of the interdisciplinarity of scientific journals". *Journal of the American Society for Information Science and Technology*, v. 58, n. 4, pp. 1303–1319.
https://doi.org/10.1002/asi.20614

Leydesdorff, Loet (2007b). "Scientific communication and cognitive codification: Social systems theory and the sociology of scientific knowledge". *European journal of social theory*, v. 10, n. 3, pp. 375−388.
https://doi.org/10.1177/136843100708070

Leydesdorff, Loet (2010). "The communication of meaning and the structuration of expectations: Giddens' 'structuration theory' and Luhmann's 'self-organization'". *Journal of the American Society for Information Science and Technology*, v. 61, n. 10, pp. 2138−2150.
https://doi.org/10.1002/asi.21381

Leydesdorff, Loet (2013). "Sociological and communication-theoretical perspectives on the commercialization of the sciences". *Science & education*, n. 22, pp. 2511-2527.
https://doi.org/10.1007/s11191-012-9458-4

Leydesdorff, Loet (2016). "Information, meaning, and intellectual organization in networks of inter-human communication". In: Sugimoto, Cassidy (ed.). *Theories of informetrics and scholarly communication*. De Gruyter, pp. 280–303.
https://doi.org/10.1515/9783110308464

Leydesdorff, Loet (2020). "The differentia specifica of interhuman communications: Luhmann and the sociological reflection of information theory". In: M. Burgin; G. Dodig-Crnkovic (eds.). *Theoretical information studies: Information in the world*, pp. 457-469. World Scientific.

Leydesdorff, Loet (2021). *The evolutionary dynamics of discursive knowledge. Qualitative and quantitative analysis of scientific and scholarly communication*. Cham: Springer.
https://doi.org/10.1007/978-3-030-59951-5_3

Leydesdorff, Loet; **Bornmann, Lutz** (2012). "Mapping (USPTO) patent data using overlays to Google Maps". *Journal of the American Society for Information Science and Technology*, v. 63, pp. 1442–1458.
https://doi.org/10.1002/asi.22666

Leydesdorff, Loet; **Carley, Stephen**; **Ràfols, Ismael** (2013). "Global maps of science based on the new Web-of-Science categories". *Scientometrics*, v. 94, n. 2, pp. 589–593.
https://doi.org/10.1007/s11192-012-0784-8

Leydesdorff, Loet; **Ràfols, Ismael** (2009). "A global map of science based on the ISI subject categories". *Journal of the American Society for Information Science and Technology*, v. 60, n. 2, pp. 348–362.
https://doi.org/10.1002/asi

Leydesdorff, Loet; **Ràfols, Ismael** (2012). "Interactive overlays: A new method for generating global journal maps from Web-of-Science data". *Journal of informetrics*, v. 6, n. 2, pp. 318–332.
https://doi.org/10.1016/j.joi.2011.11.003

Leydesdorff, Loet; **Ràfols, Ismael**; **Chen, Chaomei** (2013). "Interactive overlays of journals and the measurement of interdisciplinarity on the basis of aggregated journal-journal citations". *Journal of the American Society for Information Science and Technology*, v. 64, n. 12, pp. 2573–2586.
https://doi.org/10.1002/asi.22946

Leydesdorff, Loet; **Ivanova, Inga A.** (2014). "Mutual redundancies in interhuman communication systems: Steps toward a calculus of processing meaning". *Journal of the Association for Information Science and Technology*, v. 65, n. 2, pp. 386−399.
https://doi.org/10.1002/asi.22973

Li, Lin Z.; **Leydesdorff, Loet**; **Nioka, Shoko**; **Sun, Nannan**; **Garfield, Eugene** (2014). "Citation analysis of the scientific publications of Britton chance in ISI citation indexes". *Journal of innovative optical health sciences*, v. 07, n. 02, 1430003.
https://doi.org/10.1142/S1793545814300031

Muñoz-Écija, Teresa; Vargas-Quesada, Benjamín; Chinchilla Rodríguez, Zaida (2022). "Unveiling cognitive structure and comparative advantages of countries in knowledge domains". *Journal of information science*, Online first.
https://doi.org/10.1177/01655515221084607

Muñoz-Écija, Teresa; Vargas-Quesada, Benjamín; Chinchilla Rodríguez, Zaida (2019). "Coping with methods for delineating emerging fields: Nanoscience and nanotechnology as a case study". *Journal of informetrics*, v. 13, n. 4.
https://doi.org/10.1016/j.joi.2019.100976

Ràfols, Ismael; **Porter, Alan L.**; **Leydesdorff, Loet** (2010). "Science overlay maps: A new tool for research policy and library management". *Journal of the American Society for Information Science and Technology*, v. 61, n. 9, pp. 1871–1887.
https://doi.org/10.1002/asi.21368

Rotolo, Daniele; **Ràfols, Ismael**; **Hopkins, Michael M.**; **Leydesdorff, Loet** (2017). "Strategic intelligence on emerging technologies: scientometric overlay mapping". *Journal of the American Society for Information Science and Technology*, v. 68, n. 1, pp. 214–233.
https://doi.org/10.1002/asi.23631

Van-Eck, Nees-Jan; Waltman, Ludo (2010). "Software survey: VOSviewer, a computer program for bibliometric mapping". *Scientometrics*, v. 84, n. 2, pp. 523-538.
https://doi.org/10.1007/s11192-009-0146-3

Vargas-Quesada, Benjamín; Chinchilla-Rodríguez, Zaida; Rodriguez, Noel (2017). "Identification and visualization of the intellectual structure in graphene research". *Frontiers in research metrics and analytics*, v. 2 (October).
https://doi.org/10.3389/frma.2017.00007

Vargas-Quesada, Benjamín; **Zarco, C.**; **Cordón, O.** (2021). "Mapping the situation of educational technologies in the Spanish university system using social network analysis and visualization". *International journal of interactive multimedia and artificial intelligence*, v. 8, n. 2.
https://doi.org/10.9781/ijimai.2021.09.004

Wagner, Caroline S.; **Leydesdorff, Loet** (2005). "Mapping the network of global science: comparing international co-authorships from 1990 to 2000". *International journal of technology and globalisation*, v. 1, n. 2, pp. 185–208.
https://doi.org/10.1504/IJTG.2005.007050

White, Howard D. (2001). "Author-centered bibliometrics through CAMEOs: Characterizations automatically made and edited online". *Scientometrics*, v. 51, n. 3, pp. 607–637.
https://doi.org/10.1023/A:1019607522125

Zhou, Ping; **Leydesdorff, Loet** (2006). "The emergence of China as a leading nation in science". *Research policy*, v. 35, n. 1, pp. 83-104.
https://www.sciencedirect.com/science/article/pii/S0048733305001794

A bibliometric perspective on the academic contributions of Loet Leydesdorff

Wenjing Xiong; Ping Zhou

Recommended citation:

Xiong, Wenjing; **Zhou, Ping** (2023). "A bibliometric perspective on the academic contributions of Loet Leydesdorff". *Profesional de la información*, v. 32, n. 7, e320706.

https://doi.org/10.3145/epi.2023.dic.06

Article received on September 14, 2023
Accepted on September 29, 2023

Wenjing Xiong
https://orcid.org/0000-0002-8837-5178

Zhejiang University
School of Public Affairs
Dept. of Information Resources Management
No. 866 Yuhangtang Road
310059 Hangzhou, China
wenjing_xiong@zju.edu.cn

Ping Zhou ✉
https://orcid.org/0000-0001-5471-5261

Zhejiang University
School of Public Affairs
Dept. of Information Resources Management
No. 866 Yuhangtang Road
310059 Hangzhou, China
pingzhou@zju.edu.cn

Abstract

The purpose of this paper is to commemorate the late scholar Loet Leydesdorff for his great academic contribution on the basis of data from *Web of Science*. In the span of more than 40 years, he had 526 publications, with the years 2004-2021 being the most productive (394 publications). His international collaborations spread widely across 36 countries, with Germany, the USA, the UK, China, Russia, and South Korea being the most significant. His most frequent collaboration partners included Lutz Bornmann (Germany), Staša Milojević (USA), Caroline Wagner (USA), Henry Etzkowitz (USA), Jonathan Adams (UK), Ronald Rousseau (Belgium), and Ping Zhou (China). With a broad and deep knowledge background, Leydesdorff's research extended across multiple disciplines and fields, but he was most active in library and information science and computer science. He made profound contributions to the study of bibliometrics, innovation systems (the Triple Helix model), and communications. Leydesdorff had a remarkable and extensive citation impact, with citations in 221 *WoS* subject categories from 120 countries. His publications in 1996, 1998, 2000, 2005, 2006, and 2012 are highly cited, and those on university-industry-government relations (the Triple Helix model) are the most cited.

Keywords

Loet Leydesdorff; Bibliometrics; Informetrics; Scientometric portrait; Academic contributions; Researchers; Library and information science; Computer science.

Funding

This study received financial support from the *National Natural Science Foundation of China* (NSFC, no. 71843012).

1. Introduction

Loet Leydesdorff was professor emeritus at the *Amsterdam School of Communications Research* (*ASCoR*) at the *University of Amsterdam*, with degrees in different disciplines (Ph.D. in sociology, M.A. in philosophy, and M.Sc. in biochemistry). He published extensively in systems theory, social network analysis, scientometrics, and the sociology of innovation. With Henry Etzkowitz, he initiated a series of workshops, conferences, and special issues about the Triple Helix of university-industry-government relations. He received the *Derek de Solla Price Award for Scientometrics and Informetrics* in 2003 and held "The City of Lausanne" Honor Chair at the *School of Economics*, *Université de Lausanne* in 2005. In 2007, he was Vice President of the *8th International Conference on Computing Anticipatory Systems* (*CASYS'07*, Liège). He has been listed as a highly cited author since 2014 (*https://clarivate.com/hcr/*), and was ranked as the 27th (world) and 1st (the Netherlands) top scientist in the social sciences and humanities on the basis of citations data collected on 21-12-2022 by *Research.com* (*https://research.com/scientists-rankings/social-sciences-and-humanities*).

> In 44 years (1980-2023), Leydesdorff had more than 500 *WoS*-indexed publications, with the years 2004-2021 being the most productive.

With profound knowledge across a wide range of disciplines, such as philosophy of science, social network analysis, communication science, informatics, and sociology, Leydesdorff was able to make innovative contributions to a variety of subjects. Academic research was his lifelong passion. Even though he suffered from illness, he managed to publish his last book, *The Evolutionary Dynamics of Discursive Knowledge: Communication-Theoretical Perspectives on an Empirical Philosophy of Science*, which integrates his major contributions to three core issues: (1) the dynamics of science, technology, and innovation; (2) the measurements of the core concepts of scientometrics; and (3) the Triple Helix of university-industry-government relations (**Leydesdorff**, 2021). If he had not been unable to read and write owing to his illness and the surgeries he underwent, he would have continued his research until the last moment of his life.

In the informetric community, numerous "scientometric portraits" of eminent scholars have been published, such as those of Judit Bar-Ilan (**Halevi**, 2020; **Orduña-Malea**, 2020), Bimal Kanti Sen (**Dutta**, 2019), Eugene Garfield (**Glänzel**; **Abdulhayoğlu**, 2018), Jan Hendrik Oort (**Koley**; **Sen**, 2018), Mike Thelwall (**Vellaichamy**; **Amsan**, 2016), Santiago Grisolía (**González-Alcaide**, 2014), Nayana Nanda Borthakur (**Hazarika**; **Sarma**; **Sen**, 2010), Sivaraj Ramaseshan (**Sangam**; **Savanur**; **Manjunath**, 2007), Dorothy Crowfoot Hodgkin (**Kademani**; **Kalyane**; **Jange**, 1999) and Ronald Rousseau (**Sun**; **Jiang**, 2012). This paper aims at drawing a scientometric portrait for Loet Leydesdorff to express our remembrance of him by tracing his academic contributions from perspectives of historical trend of publications, international collaboration, interdisciplinarity, and academic impact.

2. Data and methods

Publication data were obtained from two sources: *Web of Science* (*WoS*) from *Clarivate* and *Google Scholar*. We collected 406 publications with metadata records from *WoS*. We supplemented this with 120 records from *WoS*, *Google Scholar*, and his personal homepage, resulting in a collection of 526 scientific publications, including articles, reviews, and conference papers (search date: May 30, 2023). Software packages such as *Microsoft Excel*, *R*, and *VOSviewer* were used for descriptive statistics analysis, citation analysis, and co-occurrence analysis. The jcitnetw.exe and mode2div.exe programs, developed by Leydesdorff (**Leydesdorff**; **Wagner**; **Bornmann**, 2019), were used to analyze the interdisciplinarity of his academic contributions, which generate the indicators of Variety, Disparity, Gini coefficient, and Rao-Stirling diversity (**Leydesdorff**, 2018; **Stirling**, 2007). Variety reflects the number of distinctive categories, "1 - Gini coefficient" depicts the balance in the distribution of categories, and Disparity indicates the degree to which the categories are different (**Purvis**; **Hector**, 2000; **Ràfols**; **Meyer**, 2010). The Rao-Stirling indicator explicitly or implicitly measures the properties of integrated diversity, namely the combination of variety, balance, and disparity (**Stirling**, 1998; 2007). The interdisciplinarity of each document is calculated on the basis of the subject category distribution of its references (calculation formulas for the interdisciplinarity indicators are presented in Table 1).

Table 1. Interdisciplinarity indicators.

Indicator	Formula	Description
Variety	$Variety_c = \frac{n_c}{N}$	*N* is the number of available categories.
Balance	$Balance = 1 - Gini_c = 1 - \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\lvert x_i - x_j\rvert}{2n^2\bar{x}}$	*x* is an observed value; *n* is the number of disciplines involved in the observed value; x_i is the number of observations belonging to the i^{th} discipline.
Disparity	$Disparity_c = \sum_{i,j=1; i\neq j}^{i=n_c, j=n_c} \frac{d_{ij}}{n_c(n_c-1)}$	d_{ij} is the distance between subjects *i* and *j*, normalized by $n_c(n_c-1)$.
Rao-Stirling diversity	$RS = \sum_{i\neq j}(d_{ij})^{\alpha}(p_i p_j)^{\beta}$	$p_i = x_i/\sum x_i$, and x_i denotes the number of elements belonging to subject *I*; *α* and *β* are two parameters for adjusting the relative weights of distance d_{ij} and balance or variety $p_i p_j$.

3. Results and analysis

3.1. Publication history

Figure 1 illustrates Leydesdorff's publication history. In 44 years (1980-2023), Leydesdorff published more than 500 papers. His publishing activity can be divided into three periods according to the number of yearly publications:

1. 8 years of relatively low productivity (i.e., fewer than 4 papers per year) in 1980-1988. As an academic newcomer in this period, his main task might have been to publish by applying learned knowledge to solve academic problems. Although being a newcomer, his broad scope of knowledge in different fields, such as mathematics, statistics, informatics,

and social systems theory, was well demonstrated in his publications;

2. 15 years of growing productivity in 1989-2004 with annual number of publications ranging between 4 and 11. The significant contribution he made regarding the Triple Helix (TH) model, together with Henry Etzkowitz, came out of this period; and

3. 19 years of high productivity since 2004 (i.e., more than 10 papers per year). With the first period of knowledge accumulation and the second period of knowledge creation, Leydesdorff had laid a solid foundation for this third period of high productivity. With 28, 33, 33, and 24 publications in the years 2006, 2011, 2012, and 2020, respectively, these years were most prominent. His high yield in 2020 is the most impressive, taking into account that, by that point in time, he was already suffering from the illness that eventually ended his life. His focuses in this period include new indicators, theoretical issues, and evaluation methods.

Leydesdorff published in more than 100 journals, most of which were leading informetrics journals such as *Scientometrics* (20.1%), *Journal of the Association for Information Science and Technology* (*Jasist*, 18.3%), and *Journal of informetrics* (9.0%). In other words, most of Leydesdorff's output was in informetrics or bibliometrics. The two most important journals in bibliometrics, *Scientometrics* and *Jasist*, have published 38.4% of Leydesdorff's papers throughout his publishing history. His publishing career accompanied the growth and development of the *Journal of informetrics* since its first issue in 2008. In addition to bibliometric studies, science policy was one of Leydesdorff's research interests, which is why 3.82% of his papers were published in the leading policy-related journal *Research Policy* over the years 1984-2016 (Figure 2).

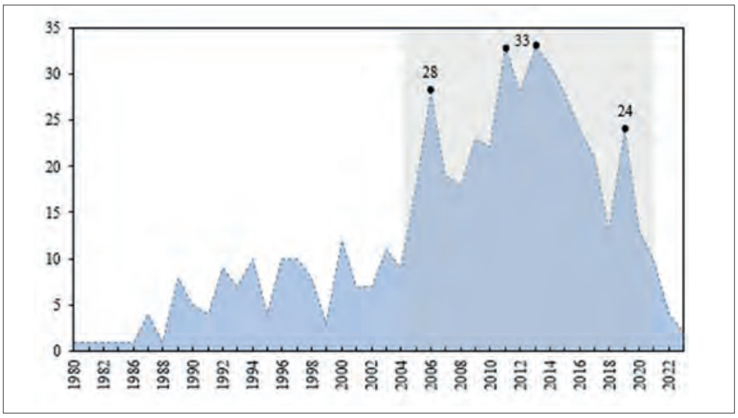


Figure 1. Annual publications of Loet Leydesdorff.

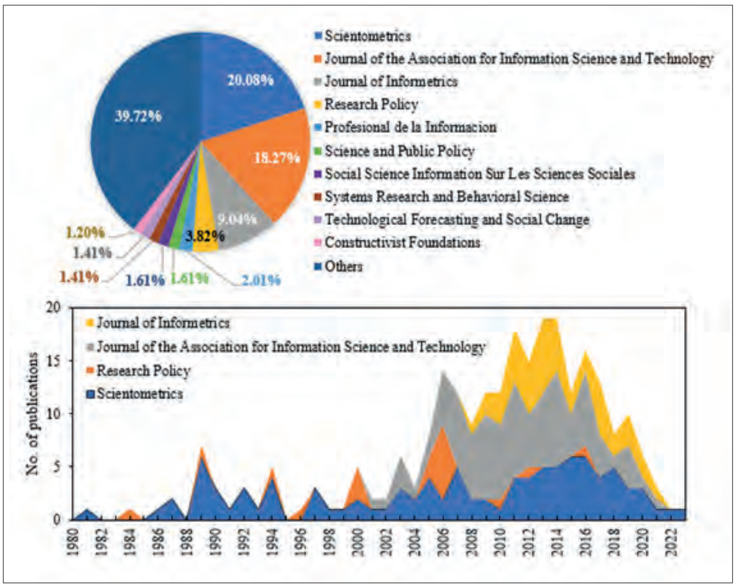


Figure 2. Publication distribution among journals.

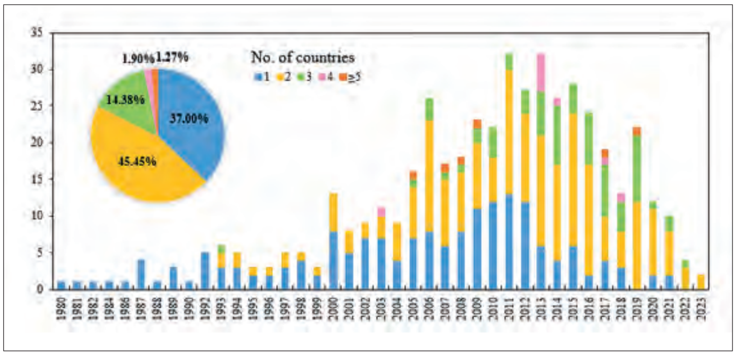


Figure 3. Annual distribution of publications with collaboration.

3.2. International collaboration

International collaboration plays a significant role in the development of science, and related topics have been explored extensively (e.g., **Luukkonen**; **Persson**; **Sivertsen**, 1992; **Katz**; **Martin**, 1997; **Freeman**, 2010; **Dusdal**; **Powell**, 2021; **Gui**; **Liu**; **Du**, 2019; **Zhou**; **Tijssen**; **Leydesdorff**, 2016; **Zhou**; **Glänzel**, 2010; **Leydesdorff**; **Wagner**, 2008; **Leydesdorff** *et al.*, 2013). International collaboration has been regarded as an important indicator when measuring the research performance of individuals and the internationalization of an organization or country. Leydesdorff's international collaborations spread across 36 countries. More than half of his publications came out of international collaborations, with most (45.5%) being with one foreign country and 14.38% with two foreign countries. Leydesdorff's international collaborations developed in step with his publication productivity, with the year 2004 serving as a divide (Figure 3). Since 2005, the number of publications stemming from international collaborations as well as the number of foreign countries collaborated with grew significantly. On average, half of his publications during 2005-2010 came out of international collaborations. Since 2011, however, most of his publications stemmed from them, with the years 2019, 2022, and 2023 being extreme cases where all of his publications resulted from internationally collaborations. A simple deduction can be made from this: growing publication productivity brought growing academic impact and thus promoted international collaboration with Leydesdorff.

With regard to the country of origin of the scholars who collaborated with Leydesdorff, Western countries including Germany, the USA, the UK, and Belgium were at the top of the list (Table 2), with Germany taking the absolute lead. Scholars from Germany included Lutz Bornmann, Robin Haunschild, and Michael Fritsch and generated 120 publications, while collaboration with scholars from the USA included Caroline Wagner, Staša Milojević, Henry Etzkowitz, Jordan Comins, and Alan Porter and produced 91 publications. Collaboration with scholars from Asia, especially China, South Korea, and Russia, and including Ping Zhou, Han Woo Park, and Inga Ivanova, was also frequent.

Table 2. Top countries and scholars collaborating with Loet Leydesdorff.

Countries (Top10)	*N*	Co-authors
Germany	120	Lutz Bornmann; Robin Haunschild; Michael Fritsch; Werner Marx; Tobias Hecking; Alexander Tekles
USA	91	Caroline Wagner; Staša Milojević; Henry Etzkowitz; Jordan Comins; Alan Porter; Alexander Petersen; Stephen Carley; Mark William Johnson; Andy Stirling
UK	72	Martin Meyer; Jonathan Adams; Daniele Rotolo; Helen Lawton Smith
China	32	Ping Zhou; Lin Zhang; Xiaojun Hu; Fred Y Ye.
Belgium	22	Ronald Rousseau; Leo Egghe; Raf Guns; Tim Engels
Spain	21	Igone Porto-Gómez; Félix De-Moya-Anegón
Russia	20	Inga Ivanova; Nataliya Smorodinskaya
South Korea	20	Han Woo Park; Jungwon Yoon; Ki-Seok Kwon
Switzerland	12	Ruediger Mutz; Hans-Dieter Daniel; Carole Probst
Italy	11	Cinzia Daraio; Simone Di-Leo; Michelina Venditti

The historical evolution of Leydesdorff's international collaborations (Figure 4) shows that his early collaborations were with USA researcher Henry Etzkowitz, with whom he created some of his most influential output –the TH model (**Etzkowitz**; **Leydesdorff**, 1995, 2000; **Leydesdorff**; **Etzkowitz**, 1996)- and with whom he explored the knowledge infrastructure of the global system or a knowledge economy. They argued that three distinguished dynamics exist in the global system or in a specific economy: the economic dynamics of the market (industries), the internal dynamics of knowledge production (universities), and the governance of their interface at different levels (government). These three sectors –university, industry, and government (UIG)– interact with each other in promoting the development of the knowledge economy. The TH model was widely accepted, and 11 TH conferences have been held globally (*https://www.leydesdorff.net/th2*).

By applying the Shannon-type information generated in the interactions among the three actors, Leydesdoff made it possible to quantify the UIG relationship (**Leydesdorff**, 2011), and thus brought about a boom of quantitative studies related to UIG relationships (e.g., **Khan**; **Park**, 2011; **Park**, 2014; **Zhang**; **Chen**; **Fu**, 2019).

The second country involved in Leydesdorff's early period of international collaborations was the UK. In 2003, Leydesdorff collaborated with Martin Meyer to explore three different sub-dynamics –economic exchanges on the market, geographical variations, and the organization of knowledge– by applying the TH model (**Leydesdorff**; **Meyer**, 2003). Since 2005, Leydesdorff's international collaboration expanded to more countries, including Germany and China. With German scholar Lutz Bornmann, Leydesdorff collaborated most frequently on a wide range of topics including citation analysis, knowledge mapping, research evaluation, and bibliometric indicators (e.g., **Leydesdorff** *et al.*, 2011; **Bornmann**;

Figure 4. Geographical and historical distribution of Leydsdorff's international collaborations.
Note: The horizontal axes are for publication years (1993-2023), and vertical axes are for the number of publications (0-12).

Leydesdorff, 2012; **Bornmann**; **Tekles**; **Leydesdorff**, 2019; **Leydesdorff**; **Bornmann**; **Wagner**, 2021). Leydesdorff had a long period of collaboration with Chinese scholar Ping Zhou, his PhD student from 2005 to 2008 (e.g., **Leydesdorff**; **Zhou**, 2005; **Zhou**; **Leydesdorff**, 2006; **Zhou**; **Leydesdorff**, 2007; **Zhou**; **Su**; **Leydesdorff**, 2010; **Zhou**; **Tijssen**; **Leydesdorff**, 2016; **Zhou**; **Leydesdorff**, 2011). An important Russian collaborator of Leydesdorff was Inga Ivanova, although the collaboration started relatively late (from the year 2014). Their collaboration topics mainly involved the TH model, innovation systems, and synergetic effects (e.g., **Ivanova**; **Leydesdorff**, 2014; **Leydesdorff**; **Ivanova**, 2016; **Leydesdorff**; **Ivanova**; **Meyer**, 2019). In 2008-2020, Leydesdorff collaborated with the Belgian scholars Rousseau Ronald, Tim Engels, and Raf Guns.

> Leydesdorff published in more than 100 journals, most of which were leading journals in informetrics such as *Scientometrics*, *Journal of the Association for Information Science and Technology*, and the *Journal of Informetrics*.

3.3. Interdisciplinary studies

When it comes to interdisciplinary research, the first issue is discipline classification, a big challenge in bibliometrics. In this paper, we adopt two classification schemes: *WoS* subject categories and topic classification. The former defines disciplinary attribution based on the publishing journals, whereas the latter defines the disciplinary attribution of publications by their research content. It is clear that the former classification is not as precise as the latter; hence, we apply both classification schemes to provide a broader and finer view of Leydesdorff's involvement in different disciplines and fields.

We first apply the broader definition –*WoS* subject categories– to map Leydesdorff's involvement in multiple disciplines. With interdisciplinary knowledge background, Leydesdorff made contributions to research topics requiring interdisciplinary knowledge. His publications involve 47 disciplines across the natural sciences, engineering, social sciences, and humanities, with most of them (82.0%) being in library and information science and computer science (Figure 5, right). Computer science took the second position in Leydesdorff's research because knowledge, especially methods, technologies, and tools, have increasingly been applied to solving problems in library and information science. In fact, the development of interdisciplinary research has resulted in increasingly newer knowledge generated at the "trading zones" (**Thagard**, 2005) of knowledge. A typical case of Leydesdorff's interdisciplinary research was to explain and simulate the development of information society by constructing mathematical models and applying computer technologies, which lasted for nearly 30 years (1995-2022) (e.g., **Leydesdorff**; **Ivanova**, 2021). Knowledge mapping was Leydesdorff's other important contribution: he wrote approximately 100 programs and made them all free access. A large number of his papers were based on his programs, especially those regarding knowledge mapping (e.g., **Leydesdorff**; **Bornmann**; **Wagner**, 2021). In addition, publishing mostly in library and information science and computer science, Leydesdorff was also involved in 36 other disciplines, for example, environmental science, the history and philosophy of science, business, economics, public administration, management science, social sciences, interdisciplinary and communication, and so on, which featured in 17.1% of his publications.

Leydesdorff's publication history in terms of disciplinary distribution (Figure 5, left) proves that he was most active in two disciplines: library and information science and computer science. His publication activities in these two disciplines were synchronous, and his research was also related to management science, social sciences, and communication science, although with relatively fewer publications.

The indicators in Table 1 are applied to measure the interdisciplinarity of Leydesdorff's research. The results (Figure 6, left) show a rapid upward trend from 1997 to 2001, followed by a long period of fluctuation. The variety of his studies fluctuated from 1997 to 2023, which implies a transformation of his research focuses. For example, the proposition and development of the TH model in 1997-2002 were based on the absorption of knowledge from different fields, such as statistics, sociology, informatics, and complex systems. During 2010-2021, Leydesdorff focused on bibliometric indicators (e.g., impact factor, diversity, citation impact, the H index, etc.) and their applications, resulting in a reduced trend in the Variety value. The Balance value declined slowly from 1997 to 2013 and then fluctuated slightly at a higher level

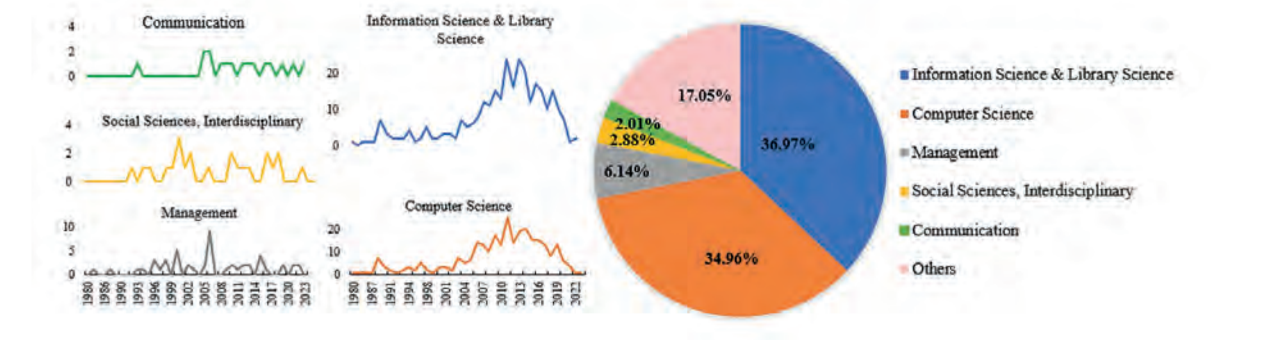


Figure 5. Publication distributions among disciplines in which Leydesdorff was most active.

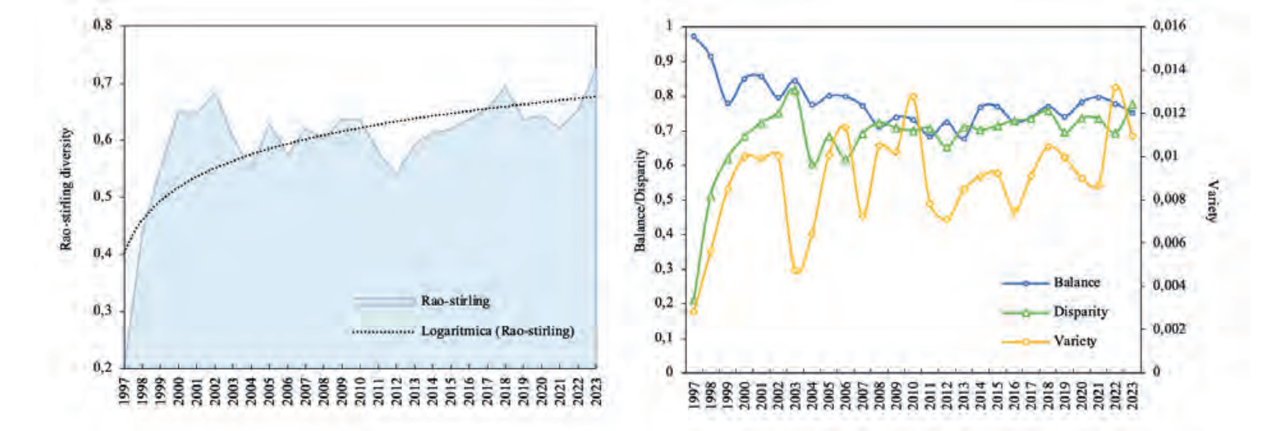


Figure 6. Annual trend of Leydesdorff's interdisciplinary studies.
Note: Detailed reference information of publications before 1997 is incomplete. Only interdisciplinarity values of publications from after 1997 are calculated.

(Balance > 0.7) in the rest of the period. The Disparity value grew rapidly during 1997-2004 (from 0.2 to 0.6) and then fluctuated slightly during the rest of the period owing to a noticeable transformation in knowledge sources. From 2003, Leydesdorff focused more on bibliometric studies, such as citation analysis, research evaluation, knowledge map, co-occurrence analysis, and social network analysis. Accordingly, the disciplines involved changed from physics, sociology, and management science to library and information science and computer science. In the last decade, his research focus had gradually fixed upon interdisciplinary research, including indicators and research evaluation, resulting in relative stable values of Disparity.

Next, we applied topic classification to define Leydesdorff's discipline/field involvement. By inputting the title, abstract, author keywords, and keywords plus of Leydesdorff's publications into *VOSviewer*, five distinct clusters were obtained: research evaluation, citation analysis, interdisciplinary study, innovation systems, and communication studies (Figure 7). Three clusters –citation analysis, interdisciplinary study, and research evaluation– are closely linked to each other because of their common knowledge foundation, with bibliometrics and citation analysis lying at their core. Thus, the three clusters can be generalized as one cluster, viz. the bibliometric cluster. The other two clusters –innovation systems (TH model) and communication studies– are related to each other but are independent from the bibliometric cluster. Leydesdorff's broad scope of knowledge is well displayed in Figure 7.

Leydesdorff was most productive in bibliometrics, a field in library and information science, and invested his energy in this field throughout his academic life. His research focuses in bibliometrics included theoretical (**Leydesdorff**; **Zhang**; **Wouters**, 2023) and methodological (**Leydesdorff**; **Ràfols**, 2012) issues, citation analysis (e.g., **Bornmann**; **Leydesdorff**,

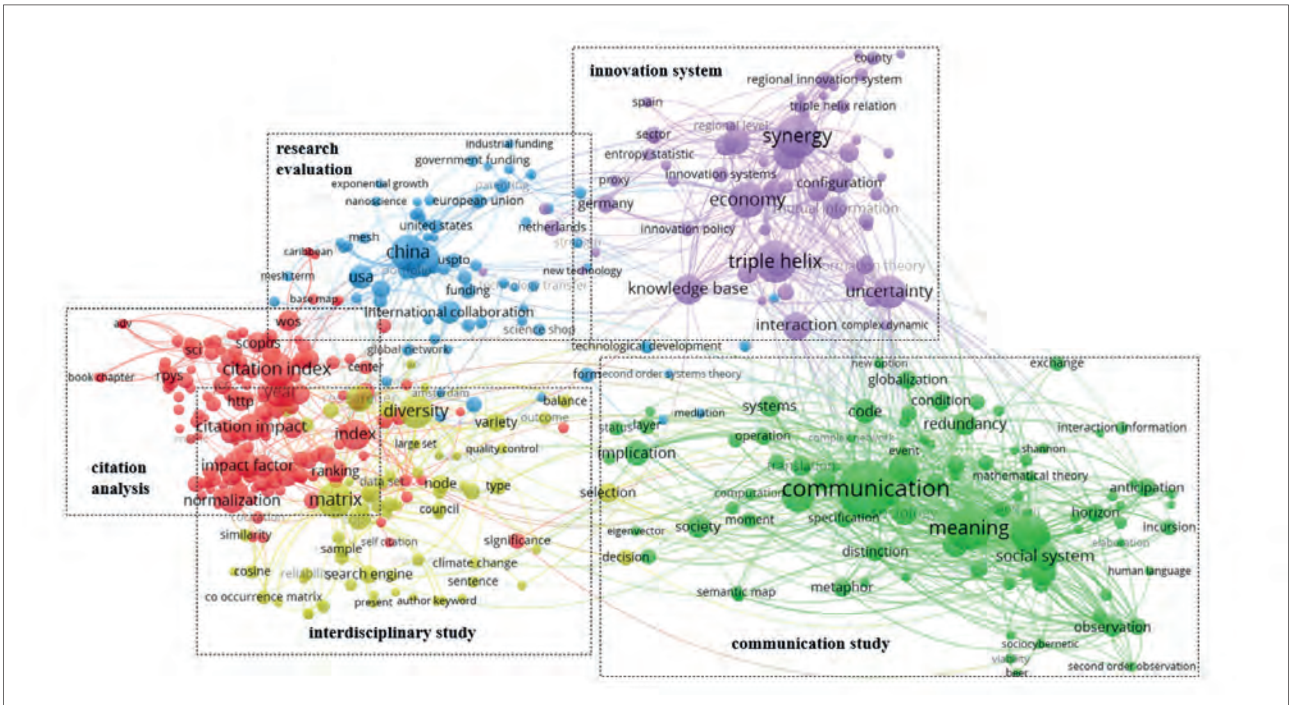


Figure 7. Co-occurrence network map of Leydesdorff's academic contributions.

2017), indicators (e.g., **Leydesdorff**; **Tekles**; **Bornmann**, 2021; **Leydesdorff**; **Bornmann**, 2021), research evaluation (e.g., **Leydesdorff**; **Milojević**, 2015), and knowledge mapping (e.g., **Chen**; **Leydesdorff**, 2014). Price (e.g., **Price**, 1965; **Price**, 1970), Garfield (e.g., **Garfield**, 1972; **Garfield**; **Merton**, 1979), and Merton (e.g., **Merton**, 1968) were Leydesdorff's important knowledge sources. With great concern placed on research evaluation, Leydesdorff reviewed classical theory such as Bradford's Law (**Bradford**, 1934), and systematically optimized existing evaluation methods (**Ràfols**; **Leydesdorff**, 2009). With his co-authors, he carried out many evaluation studies on countries, institutions, journals, and disciplines (e.g., **Zhou**; **Su**; **Leydesdorff**, 2010; **Zhou**; **Tijssen**; **Leydesdorff**, 2016; **Leydesdorff**; **Zhou**, 2014; **Leydesdorff**; **Bornmann**, 2016; **Zhou**; **Leydesdorff**, 2011; **Wagner**; **Whetsell**; **Leydesdorff**, 2017). In interdisciplinary study, Leydesdorff proposed the concept of diversity and its measurement in the paper titled "Diversity and interdisciplinarity: how can one distinguish and recombine disparity, variety, and balance?" (**Leydesdorff**, 2018). By inducing the Gini coefficient into the Rao-Stirling index, he made it possible to measure the diversity of interdisciplinary research, and thus received a high citation impact (28 citations in *WoS* and 38 citations in *Google Scholar*, retrieval date: September 13, 2023).

> Leydesdorff's international collaborations were spread widely across 36 countries, with Germany, the USA, the UK, China, Russia, and South Korea being the most significant

On the basis of the theory of information entropy (**Shannon**, 1948; **McGill**, 1954), statistical decomposition analysis methods (**Theil**, 1972), and social system theory (**Luhmann**, 1984), Leydesdorff proposed the TH model with Henry Etzkowitz and made it possible to explain and quantify interactions among industries, universities, and governments in a communication system of knowledge economy (e.g., **Leydesdorff**; **Zhou**, 2014; **Park**; **Leydesdorff**, 2010; **Kwon** *et al.*, 2012; **Leydesdorff**; **Sun**, 2009; **Leydesdorff**; **Strand**, 2013). The TH model has been widely accepted in multiple disciplines and fields, for example, bibliometrics, management science, sociology, and public administration (e.g., **Buligina** *et al.*, 2014; **Kim**; **Park**, 2012).

Time-slice analysis of research topics helps to trace the historical evolution of Leydesdorff's research focuses (Figure 8). In 1980-1990, his interests were theoretical issues in communication science, statistics, social systems, self-organization, and information entropy. His research in this period laid the foundation for his significant contributions in later years. In 2000-2007, his research involved theoretical and application issues of the innovation system (TH model), citation analysis, and knowledge mapping. In 2008-2015, his focuses included topics relevant to synergistic effects based on the TH model and bibliometric topics such as indicators and research evaluation. In 2016-2023, his interests were in innovation systems and applications of the TH model (i.e., university–industry–government interactions and multiple sectors in knowledge production). With the increasing need to tackle complex scientific problems, studies on interdisciplinary research attracted Leydesdorff's attention. He studied the interdisciplinarity of different objects and entities such as publications, journals, countries, regions, disciplines, fields, authors, organizations, and so on by conducting citation analysis, co-occurrence analysis, and social network analysis.

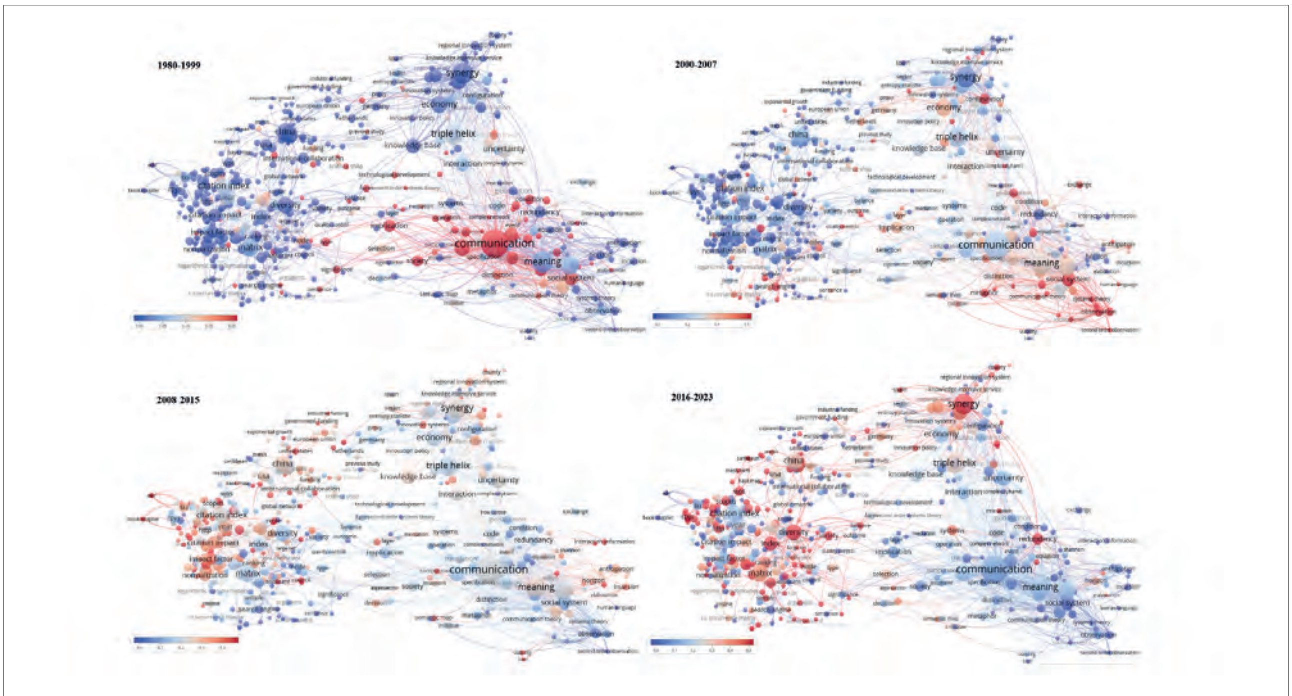


Figure 8. Time-slice analysis of research topics.
Note: The terms marked with red nodes occur more frequently than those marked with blue.

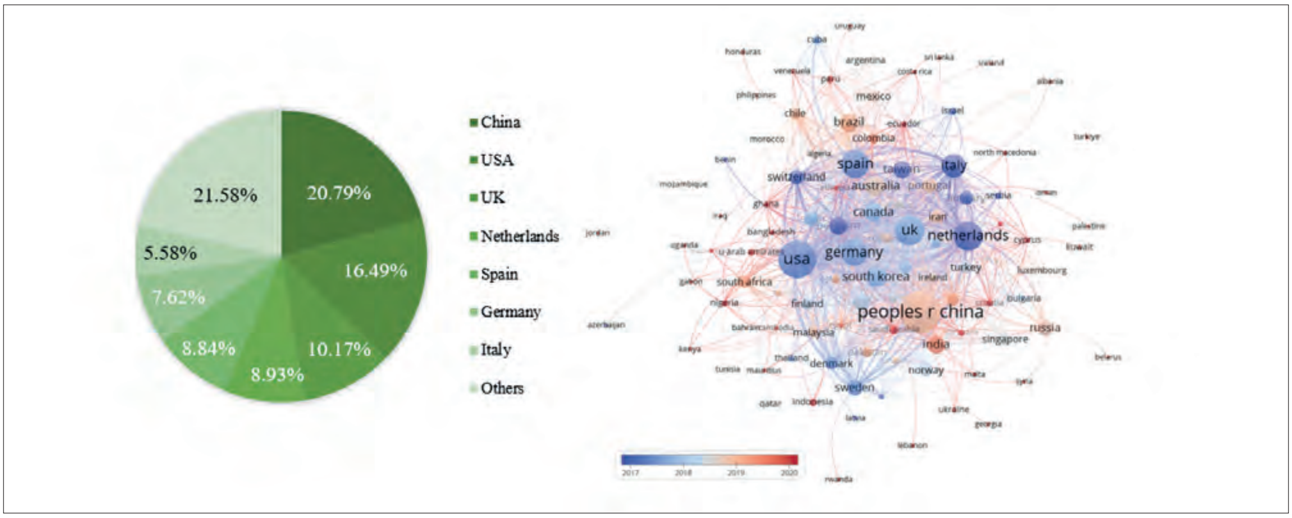


Figure 9. Country distribution of citing publications of Leydesdorff's research.

3.4. Citation impact

Citation impact is an important dimension that measures scholarly contribution. In this section, we use citations from each publishing year, citing discipline, and country/region, to respectively measure Leydesdorff's impact on said publishing years, disciplines, and countries/regions.

Leydesdorff's citation impact can be seen across 120 countries, with China, the USA, the UK, and Spain being the most significant (Figure 9). Among the citing publications, China has 1,639, accounting for 20.8%, followed by the USA (1,312 publications, 16.4%) and the UK (721 publications, 10.2%). In the early years, Leydesdorff's impact was mainly seen in North America (e.g., the USA and Canada), Europe (e.g., Italy, Spain, Germany, Switzerland, Sweden, Denmark), South Korea, and Australia. In the middle and late periods, his academic impact spread gradually to China, Russia, India, Singapore, Africa (e.g., South Africa), and South America (e.g., Brazil, Chile).

Leydesdorff's citation impact is spread across 221 disciplines and fields. His publications are cited most frequently in computer science (21.99%) and library and information science (20.39%), followed by environmental science (5.94%), management science (5.70%), and business (3.91%) (Figure 10). The time distribution of the citing disciplines shows that Leydesdorff's early impact was on library and information science, computer science, and electrical and electronic engineering, and then expanded to education and educational research, public health, green and sustainable development sciences and technology, and environmental science. This phenomenon further confirms the interdisciplinary characteristics of Leydesdorff's academic contributions.

In 42 years (1982-2023), Leydesdorff's most influential contributions were made in 1996, 1998, 2000, 2005, 2006, and 2012 (Figure 11). Given that Leydesdorff published more than one paper per year, we only mention the paper receiving the most citations in the focal year. In 1996, he and Etzkowitz first published the paper introducing the Triple Helix model to illustrate the relationship between universities, industries, and government (**Leydesdorff**; **Etzkowitz**, 1996). It should be noted that the TH model was first announced at the 1995 ESST conference (**Etzkowitz**; **Leydesdorff**, 1995), but published in a journal in 1996 (**Leydesdorff**; **Etzkowitz**, 1996). This paper had 1,731 citations.

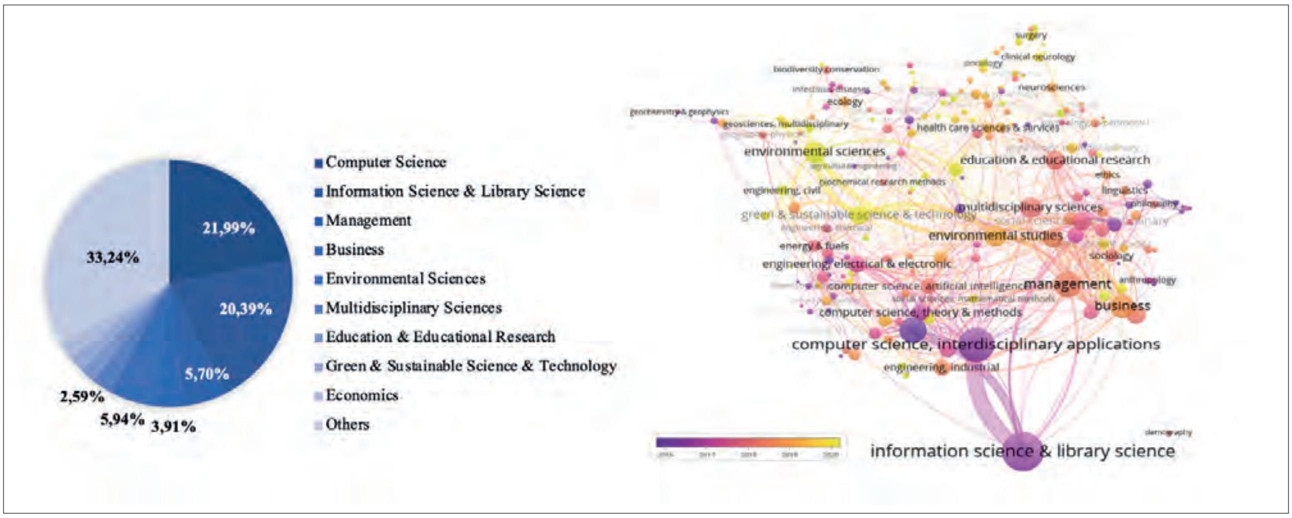


Figure 10. Disciplinary distribution of citing publications of Leydesdorff's research.

The 1998 publication (**Leydesdorff**; **Etzkowitz**, 1998, 1,397 citations) introduced the topics discussed at the second Triple Helix conference held in New York the same year. The 2000 publication (**Etzkowitz**; **Leydesdorff**, 2000) enriched the TH model with theoretical foundations and thus had the highest number of citations (3,382). The 2005 publications with high citations co-authored with Caroline Wagner (then Leydesdorff's PhD student) (**Wagner**; **Leydesdorff**, 2005, 1,349 citations) focused on principles and mechanisms of international collaboration in scientific research. Although 2006 was also the year that Leydesdorff produced papers with high citations, the count was an accumulative result, with three publications being the most representative (**Zhou**; **Leydesdorff**, 2006, 379 citations; **Leydesdorff**; **Vaughan**, 2006, 265 citations; **Leydesdorff**; **Meyer**, 2006, 202 citations).

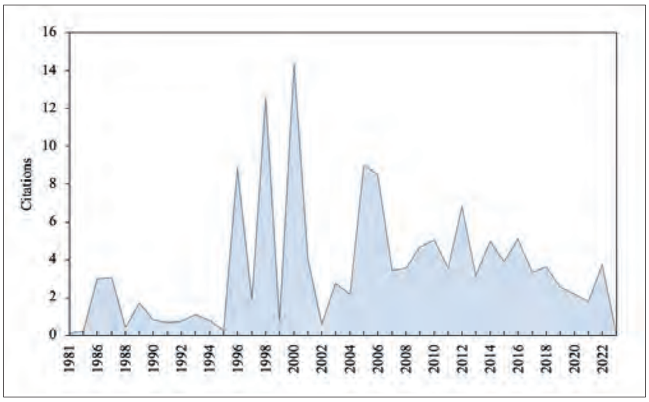


Figure 11. Number of citations in publication years (retrieval date: May 30, 2023).

4. Summary

By tracing the academic contribution of Loet Leydesdorff from different perspectives, the current paper provides a quantified portrait of him.

In 44 years (1980-2023), Leydesdorff published more than 500 *WoS*-indexed publications, with the years 2004-2021 being the most productive, with more than 10 publications per year. More than half of his papers were published in informetrics journals such as *Scientometrics*, *Jasist*, and *Journal of Informetrics* as well as the management journal *Research Policy*.

Leydesdorff was an active practitioner of international collaboration, with nearly half of his publications being co-authored. The scholars who collaborated with him were spread across 36 countries, including Germany, the USA, the UK, China, Russia, and South Korea. The year 2005 was when his international collaboration began to proliferate. In later years (2019, 2022, and 2023), almost all his publications featured international collaborations. It is no exaggeration to say that he is an internationally distinguished scholar. The importance of international collaboration in promoting science development is well illustrated by Leydesdorff's practices and achievements. International collaboration contributed to his academic achievement, and he also contributed to his international partners' achievements through collaboration.

On the basis of *WoS* journal categories, Leydesdorff published in, and thus had impact across, multiple disciplines and fields. His publications engage with 47 different disciplines and fields, including library and information science, computer science, management science, communication science, business, economics, and public administration. Topic clustering identified five disciplines/fields that Leydesdorff engaged with the most: research evaluation, citation analysis, interdisciplinary study, innovation systems (TH model), and communication studies. Using bibliometrics to generalize research evaluation, citation analysis, and interdisciplinary study because of their connections to each other, three distinct disciplines/fields are obtained (bibliometrics, innovation system (TH model), communication studies) with which Leydesdorff engaged most frequently.

Leydesdorff's far-reaching impact is unparalleled. His work was cited in more than 120 countries and 221 *WoS* subject categories. China, the USA, the UK, and Spain were the top four countries in which citations of his work are found. The disciplines citing Leydesdorff most frequently are library and information science and computer science. Over the course of 44 years, most of Leydesdorff's publications amounted to a high citation impact, which is remarkable given his extreme productivity. Another unusual phenomenon is the number of highly cited papers he produced through international collaboration on various research topics, with those on the TH model for innovation systems being the most influential.

The current study is based on *WoS* data without inclusion of Leydesdorff's four books: *A sociological theory of communication: The self-organization of the knowledge-based society* (2000); *The challenge of scientometrics: The development, measurement, and self-organization of scientific communications* (2001); *The knowledge-based economy modeled, measured, simulated* (2006), and

The evolutionary dynamics of discursive knowledge (2021). A more comprehensive profile of Leydesdorff's academic contribution might be carried out by including his books for both the quantitative and qualitative study of his output, in addition to the publications already referenced in this paper.

> With profound knowledge in a wide range of disciplines, such as philosophy of science, social network analysis, communication science, informatics, and sociology, Leydesdorff was able to make innovative contributions to a variety of subjects.

5. References

Bornmann, Lutz; **Leydesdorff, Loet** (2012). "Statistical tests and research assessments: A comment on Schneider". *arXiv preprint arXiv*: 1210.3477.

Bornmann, Lutz; **Leydesdorff, Loet** (2017). "Skewness of citation impact data and covariates of citation distributions: A large-scale empirical analysis based on Web of Science data". *Journal of informetrics*, v. 11, n. 1, pp. 164-175.
https://doi.org/10.1016/j.joi.2016.12.001

Bornmann, Lutz; **Tekles, Alexander**; **Leydesdorff, Loet** (2019). "How well does I3 perform for impact measurement compared to other bibliometric indicators? The convergent validity of several (field-normalized) indicators". *Scientometrics*, v. 119, n. 2, pp. 1187-1205.
https://doi.org/10.1007/s11192-019-03071-6

Bradford, Samuel C. (1934). "Sources of information on specific subjects". *Engineering*, n. 137, pp. 85-86.

Buligina, Ilze; Sloka, Biruta; Dzelme, Juris; Tora, Ginta (2014). *Triple Helix for VET: New challenges for public administration in Latvia*. Economic Science for Rural Development. ISBN: 978 9934 8466 2 5

Chen, Chaomei; **Leydesdorff, Loet** (2014). "Patterns of connections and movements in dual-map overlays: A new method of publication portfolio analysis". *Journal of the Association for Information Science and Technology*, v. 65, n. 2, pp. 334-351.
https://doi.org/10.1002/asi.22968

Dusdal, Jennifer; Powell, Justin J. (2021). "Benefits, motivations, and challenges of international collaborative research: a sociology of science case study". *Science and public policy*, v. 48, n. 2, pp. 235-245.
https://doi.org/10.1093/scipol/scab010

Dutta, Bidyarthi (2019). "Biobibliometric portrait of BK Sen: A librarian, information scientist and scientometrician". *Malaysian journal of library & information science*, v. 24, n. 1.
https://doi.org/10.22452/mjlis.vol24no1.1

Etzkowitz, Henry; **Leydesdorff, Loet** (1995). "The Triple Helix of University-Industry-Government relations: A laboratory for knowledge based economic development". *EASST Review*, v. 14, n. 1, pp. 11-19.

Etzkowitz, Henry; **Leydesdorff, Loet** (2000). "The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university–industry–government relations". *Research policy*, v. 29, n. 2, pp. 109-123.
https://doi.org/10.1016/S0048-7333(99)00055-4

Freeman, R. Edward (2010). *Strategic management: A stakeholder approach*. London: Cambridge University Press. ISBN: 978 0 521 15174 0
https://doi.org/10.1017/CBO9781139192675

Garfield, Eugene (1972). "Citation analysis as a tool in journal evaluation: Journals can be ranked by frequency and impact of citations for science policy studies". *Science*, v. 178, n. 4060, pp. 471-479.
https://doi.org/10.1126/science.178.4060.471

Garfield, Eugene; Merton, Robert K. (1979). *Citation indexing: Its theory and application in science, technology, and humanities* (vol. 8). New York: Wiley.

Glänzel, Wolfgang; Abdulhayoğlu, Mehmet Ali (2018). "Garfield number: On some characteristics of Eugene Garfield's first and second order co-authorship networks". *Scientometrics*, v. 114, pp. 533-544.
https://doi.org/10.1007/s11192-017-2623-4

González-Alcaide, Gregorio (2014). "Scientometric portrait of biochemist Santiago Grisolía: Publication productivity, collaboration patterns, and citation analysis". *Research evaluation*, v. 23, n. 2, pp. 150-165.
https://doi.org/10.1093/reseval/rvu003

Gui, Qinchang; Liu, Chengliang; Du, Debin (2019). "Globalization of science and international scientific collaboration: A network perspective". *Geoforum*, v. 105.
https://doi.org/10.1016/j.geoforum.2019.06.017

Halevi, Gali (2020). "The scientific legacy of Judit Bar-Ilan". *Scientometrics*, v. 123, n. 3, pp. 21-32.
https://doi.org/10.1007/s11192-020-03439-z

Hazarika, Tilak; Sarma, Dipak; Sen, B. K. (2010). "Scientometric portrait of Nayana Nanda Borthakur: a biometeorologist". *Annals of library & information studies*, v. 57, n. 1, pp. 21-32.

Ivanova, Inga A.; **Leydesdorff, Loet** (2014). "Rotational symmetry and the transformation of innovation systems in a Triple Helix of university–industry–government relations". *Technological forecasting and social change*, v. 86, pp. 143-156.
https://doi.org/10.1016/j.techfore.2013.08.022

Kademani, Basavaraj Shivappa.; Kalyane, V.; Jange, Suresh (1999). "Scientometric portrait of nobel laureate Dorothy Crowfoot Hodgkin". *Scientometrics*, v. 45, n. 2, pp. 233-250.
https://doi.org/10.1007/BF02458435

Katz, J. Sylvan; Martin, Ben R. (1997). "What is research collaboration?". *Research policy*, v. 26, n. 1.
https://doi.org/10.1016/S0048-7333(96)00917-1

Khan, Gohar-Feroz; Han Woo Park (2011). "Measuring the triple helix on the web: Longitudinal trends in the university-industry-government relationship in Korea". *Journal of the American Society for Information Science and Technology*, v. 62, n. 12, pp. 2443-2455.
https://doi.org/10.1002/asi.21595

Kim, Minjeong; Han Woo Park (2012). "Measuring Twitter-based political participation and deliberation in the South Korean context by using social network and Triple Helix indicators". *Scientometrics*, v. 90, n. 1, pp. 121-140.
https://doi.org/10.1007/s11192-011-0508-5

Koley, Susanta; Sen, B. K. (2018). "Biobibliometric portrait of the astronomer Jan Hendrik Oort". *Annals of Library and Information Studies*, v. 64, n. 4, pp. 217-228.

Kwon, Ki-Seok; Han Woo Park; So, Minho; Leydesdorff, Loet (2012). "Has globalization strengthened South Korea's national research system? National and international dynamics of the Triple Helix of scientific co-authorship relationships in South Korea". *Scientometrics*, v. 90, n. 1, pp. 163-176.
https://doi.org/10.1007/s11192-011-0512-9

Leydesdorff, Loet (2001). *A sociological theory of communication: The self-organization of the knowledge-based society*. u-Publish.com: Universal-Publishers. ISBN: 1 58112 695 6

Leydesdorff, Loet (2001). *The challenge of scientometrics: The development, measurement, and self-organization of scientific communications*. u-Publish.com: Universal-Publishers. ISBN: 1 58112 681 6

Leydesdorff, Loet (2006). *The knowledge-based economy: Modeled, measured, simulated*. u-Publish.com: Universal-Publishers. ISBN: 1 58112 937 8
https://doi.org/10.3395/reciis.v1i1.36pt

Leydesdorff, Loet (2011). "'Structuration' by intellectual organization: the configuration of knowledge in relations among structural components in networks of science". *Scientometrics*, v. 88, n. 2, pp. 499-520.
https://doi.org/10.1007/s11192-011-0397-7

Leydesdorff, Loet (2018). "Diversity and interdisciplinarity: how can one distinguish and recombine disparity, variety, and balance?". *Scientometrics*, v. 116, pp. 2113-2121.
https://doi.org/10.1007/s11192-018-2810-y

Leydesdorff, Loet (2021). *The evolutionary dynamics of discursive knowledge: Communication-theoretical perspectives on an empirical philosophy of science*. Cham: Springer Nature. ISBN: 978 3 030 59950 8
https://doi.org/10.1007/978-3-030-59951-5

Leydesdorff, Loet; **Bornmann, Lutz** (2016). "The operationalization of 'fields' as WoS subject categories (WCs) in evaluative bibliometrics: The cases of 'library and information science' and 'science & technology studies'". *Journal of the Association for Information Science and Technology*, v. 67, n. 3, pp. 707-714.
https://doi.org/10.1002/asi.23408

Leydesdorff, Loet; **Bornmann, Lutz** (2021). "Disruption indices and their calculation using web-of-science data: Indicators of historical developments or evolutionary dynamics?" *Journal of informetrics*, v. 15, n. 4, p. 101219.
https://doi.org/10.1016/j.joi.2021.101219

Leydesdorff, Loet; **Bornmann, Lutz**; **Mutz, Rüdiger; Opthof, Tobias** (2011). "Turning the tables on citation analysis one more time: Principles for comparing sets of documents". *Journal of the American Society for Information Science and technology*, v. 62, n. 7, pp. 1370-1381.
https://doi.org/10.1002/asi.21534

Leydesdorff, Loet; **Bornmann, Lutz**; **Wagner, Caroline S.** (2021). "Improved clusterings and visualizations of 11,359 journals in the JCRs 2015". *Scientometrics*, v. 126, pp. 5353-5354.
https://doi.org/10.1007/s11192-021-03938-7

Leydesdorff, Loet; **Etzkowitz, Henry** (1996). "Emergence of a Triple Helix of university-industry-government relations". *Science and public policy*, v. 23, n. 5, pp. 279-286.

Leydesdorff, Loet; **Etzkowitz, Henry** (1998). "The triple helix as a model for innovation studies". *Science and public policy*, v. 25, n. 3, pp. 195-203.

Leydesdorff Loet; Ivanova, Inga A. (2016). "'Open innovation' and 'triple helix' models of innovation: can synergy in innovation systems be measured?". *Journal of open innovation: technology, market, and complexity*, v. 2, n. 3, p. 11.
https://doi.org/10.1186/s40852-016-0039-7

Leydesdorff, Loet; **Ivanova, Inga A.** (2021). "The measurement of 'interdisciplinarity' and 'synergy' in scientific and extra-scientific collaborations". *Journal of the Association for Information Science and Technology*, v. 72, n. 4, pp. 387-402.
https://doi.org/10.1002/asi.24416

Leydesdorff Loet; Ivanova, Inga A.; Meyer, Martin (2019). *Synergy in innovation systems measured as redundancy in triple helix relations*. Cham: Springer handbook of science and technology indicators. ISBN: 978 3 030 02510 6

Leydesdorff, Loet; **Liwen, Vaughan** (2006). "Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment". *Journal of the American Society for Information Science and technology*, v. 57, n. 12, pp. 1616-1628.
https://doi.org/10.1002/asi.20335

Leydesdorff, Loet; **Meyer, Martin** (2003). "The Triple Helix of university-industry-government relations". *Scientometrics*, v. 58, pp. 191-203.
https://doi.org/10.1023/A:1026276308287

Leydesdorff, Loet; **Meyer, Martin** (2006). "Triple Helix indicators of knowledge-based innovation systems: Introduction to the special issue". *Research policy*, v. 35, n. 10, pp. 1441-1449.
https://doi.org/10.1016/j.respol.2006.09.016

Leydesdorff, Loet; **Ràfols, Ismael** (2012). "Interactive overlays: A new method for generating global journal maps from Web of Science data". *Journal of informetrics*, v. 6, n. 2, pp. 318-332.
https://doi.org/10.1016/j.joi.2011.11.003

Leydesdorff, Loet; **Staša Milojević** (2015). "The citation impact of German sociology journals: Some problems with the use of scientometric indicators in journal and research evaluations". *Soziale Welt*, v. 66, n. 2, pp. 193-204.
https://doi.org/10.5771/0038-6073-2015-2-193

Leydesdorff, Loet; **Strand, Øivind** (2013). "The Swedish system of innovation: Regional synergies in a knowledge-based economy". *Journal of the American Society for Information Science and Technology*, v. 64, n. 9, pp. 1890-1902.
https://doi.org/10.1002/asi.22895

Leydesdorff, Loet; **Sun, Yuan** (2009). "National and international dimensions of the Triple Helix in Japan: University-industry-government versus international co-authorship relations". *Journal of the American Society for Information Science and Technology*, v. 60, n. 4, pp. 778-788.
https://doi.org/10.1002/asi.20997

Leydesdorff, Loet; **Tekles**, **Alexander**; **Bornmann, Lutz** (2021). "A proposal to revise the disruption indicator". *Profesional de la información*, v. 30, n. 1, e300121.
https://doi.org/10.3145/epi.2021.ene.21

Leydesdorff, Loet; **Wagner, Caroline S.** (2008). "International collaboration in science and the formation of a core group". *Journal of informetrics*, v. 2, n. 4, pp. 317-325.
https://doi.org/10.1016/j.joi.2008.07.003

Leydesdorff, Loet; **Wagner**, **Caroline S.**; **Bornmann, Lutz** (2019). "Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient". *Journal of informetrics*, v. 13, n. 1, pp. 255-269.
https://doi.org/10.1016/j.joi.2018.12.006

Leydesdorff, Loet; **Wagner, Caroline S.; Park, Han Woo; Adams, Jonathan** (2013). "International collaboration in science: The global map and the network". *arXiv preprint* arXiv:1301.0801.
https://doi.org/10.3145/epi.2013.ene.12

Leydesdorff, Loet; **Zhang, Lin; Wouters, Paul** (2023). "Trajectories and regimes in research versus knowledge evaluations: Contributions to an evolutionary theory of citation". *Profesional de la información*, v. 32, n. 1, e320103.
https://doi.org/10.3145/epi.2023.ene.03

Leydesdorff, Loet; **Zhou, Ping** (2005). "Are the contributions of China and Korea upsetting the world system of science?". *Scientometrics*, v. 63, pp. 617-630.
https://doi.org/10.1007/s11192-005-0231-1

Leydesdorff, Loet; **Zhou, Ping** (2014). "Measuring the knowledge-based economy of China in terms of synergy among technological, organizational, and geographic attributes of firms". *Scientometrics*, v. 98, pp. 1703-1719.
https://doi.org/10.1007/s11192-013-1179-1

Luhmann, Niklas (1984). *Soziale systeme: grundriss einer allgemeinen theorie*. Suhrkamp.

Luukkonen, Terttu; Persson, Olle; Sivertsen, Gunnar (1992). "Understanding patterns of international scientific collaboration". *Science, technology, & human values*, v. 17, n. 1, pp. 101-126.
https://doi.org/10.1177/016224399201700106

McGill, William J. (1954). "Multivariate information transmission". *Psychometrika*, v. 19, n. 2, pp. 97-116.
https://doi.org/10.1007/BF02289159

Merton, Robert K. (1968). "The Matthew effect in science: The reward and communication systems of science are considered". *Science*, v. 159, n. 3810, pp. 56-63.
https://doi.org/10.1126/science.159.3810.56

Orduña-Malea, Enrique (2020). "Crossing the academic ocean? Judit Bar-Ilan's oeuvre on search engines studies". *Scientometrics*, v. 123, n. 3, pp. 1317-1340.
https://doi.org/10.1007/s11192-020-03450-4

Park, Han Woo (2014). "Transition from the triple helix to N-tuple helices? An interview with Elias G. Carayannis and David FJ Campbell". *Scientometrics*, v. 99, n. 1, pp. 203-207.
https://doi.org/10.1007/s11192-013-1124-3

Park, Han Woo; **Leydesdorff, Loet** (2010). "Longitudinal trends in networks of university-industry-government relations in South Korea: The role of programmatic incentives". *Research policy*, v. 39, n. 5, pp. 640-649.
https://doi.org/10.1016/j.respol.2010.02.009

Price-de-Solla, Derek J. (1965). "Networks of scientific papers". *Science*, v. 149, n. 3683, pp. 510-515.
https://doi.org/10.1126/science.149.3683.510

Price-de-Solla, Derek J. (1970). "Citation measures of hard science, soft science, technology, and nonscience". *Communication among scientists and engineers*, v. 1, 3r22.

Purvis, Andy; Hector, Andy (2000). "Getting the measure of biodiversity". *Nature*, v. 405, n. 6783, pp. 212-219.
https://doi.org/10.1038/35012221

Ràfols, Ismael; Leydesdorff, Loet (2009). "Content-based and algorithmic classifications of journals: Perspectives on the dynamics of scientific communication and indexer effects". *Journal of the American Society for Information Science and Technology*, v. 60, n. 9, pp. 1823-1835.
https://doi.org/10.1002/asi.21086

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

Sangam, S.; Savanur, Kiran; Manjunath, M. (2007). "Communication and collaborative research pattern of Sivaraj Ramaseshan: A scientometric portrait". *Scientometrics*, v. 71, n. 2, pp. 217-230.
https://doi.org/10.1007/s11192-007-1670-7

Shannon, Claude E. (1948). „A mathematical theory of communication". *Bell System technical journal*, v. 27, n. 3, pp. 379-423.
https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Stirling, Andy (1998). "On the economics and analysis of diversity". SPRU Electronic working paper.
http://www.sussex.ac.uk/Units/spru/publications/imprint/sewps/sewp28/sewp28.pdf

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

Sun, Junwei; Jiang, Chunlin (2012). "Ronald Rousseaus academic thought and the academic exchange with Chinese colleague". *Journal of modern information*, v. 32, n. 3, pp. 4-15.

Thagard, Paul (2005). "Being interdisciplinary: Trading zones in cognitive science". *Interdisciplinary collaboration: An emerging cognitive science*, pp. 317-339.

Theil, Henri (1972). *Statistical decomposition analysis: With applications in the social and administrative sciences*. Amsterdam, The Netherlands: North-Holland Publishing Company.

Vellaichamy, A.; Amsan, E. (2016). "Scientometric portrait of Mike Thelwall". *Library philosophy and practice*, p. 1487.

Wagner, Caroline S.; **Leydesdorff, Loet** (2005). „Network structure, self-organization, and the growth of international collaboration in science". *Research policy*, v. 34, n. 10, pp. 1608-1618.
https://doi.org/10.1016/j.respol.2005.08.002

Wagner, Caroline S.; Whetsell, Travis A; **Leydesdorff, Loet** (2017). "Growth of international collaboration in science: revisiting six specialties". *Scientometrics*, v. 110, pp. 1633-1652.
https://doi.org/10.1007/s11192-016-2230-9

Zhang, Yi; Chen, Kaihua; Fu, Xiaolan (2019). "Scientific effects of Triple Helix interactions among research institutes, industries and universities". *Technovation*, v. 86, pp. 33-47.
https://doi.org/10.1016/j.technovation.2019.05.003

Zhou, Ping; Glänzel, Wolfgang (2010). "In-depth analysis on China's international cooperation in science". *Scientometrics*, v. 82, n. 3, pp. 597-612.
https://doi.org/10.1007/s11192-010-0174-z

Zhou, Ping; Leydesdorff, Loet (2006). "The emergence of China as a leading nation in science". *Research policy*, v. 35, n. 1, pp. 83-104.
https://doi.org/10.1016/j.respol.2005.08.006

Zhou, Ping; **Leydesdorff, Loet** (2007). "The citation impacts and citation environments of Chinese journals in mathematics". *Scientometrics*, v. 72, n. 2, pp. 185-200.
https://doi.org/10.1007/s11192-007-1713-0

Zhou, Ping; **Leydesdorff, Loet** (2011). "Fractional counting of citations in research evaluation: A cross-and interdisciplinary assessment of the Tsinghua University in Beijing". *Journal of informetrics*, v. 5, n. 3, pp. 360-368.
https://doi.org/10.1016/j.joi.2011.01.010

Zhou, Ping; Su, Xinning; Leydesdorff, Loet (2010). "A comparative study on communication structures of Chinese journals in the social sciences". *Journal of the American society for Information science and Technology*, v. 61, n. 7, pp. 1360-1376.
https://doi.org/10.1002/asi.21343

Zhou, Ping; Tijssen, Robert; Leydesdorff, Loet (2016). "University-industry collaboration in China and the USA: A bibliometric comparison". *PloS one*, v. 11, n. 11, e0165277.
https://doi.org/10.1371/journal.pone.0165277

Creating a collection of publications categorized by their research guarantors into the *Scopus ASJC* scheme

Jesús M. Álvarez-Llorente; Vicente P. Guerrero-Bote; Félix De-Moya-Anegón

Nota: Este artículo se puede leer en español en:
https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/87486

Recommended citation:

Álvarez-Llorente, Jesús M.; **Guerrero-Bote, Vicente P.**; **De-Moya-Anegón, Félix** (2023). "Creating a collection of publications categorized by their research guarantors into the *Scopus ASJC* scheme". *Profesional de la información*, v. 32, n. 7, e320704.

https://doi.org/10.3145/epi.2023.dic.04

Manuscript received on 12th September 2023
Accepted on 10th October 2023

Jesús M. Álvarez-Llorente ✉
https://orcid.org/0000-0002-4901-3457

Universidad de Extremadura
Departamento de Ingeniería de Sistemas Informáticos y Telemáticos
Plazuela Ibn Marwan
06001 Badajoz, Spain
llorente@unex.es

Vicente P. Guerrero-Bote
https://orcid.org/0000-0003-4821-9768

Universidad de Extremadura
Departamento de Información y Comunicación
Plazuela Ibn Marwan
06001 Badajoz, Spain
guerrero@unex.es

Félix De-Moya-Anegón
https://orcid.org/0000-0002-0255-8628

SCImago Research Group
18220 Granada, Spain
felix.moya@scimago.es

Abstract

Given the need in Scientometrics to get beyond merely classifying scientific production based on the classification of the journals in which it is published, there have been many attempts to classify papers directly. Little has been done, however, to check how reliable the results are. In this work, a collection of publications was generated which we call an Author's Assignation Collection (AAC) comprising 13449 papers referenced in the *Scopus* database and classified by their research guarantor with fractional weighting in terms of *Scopus*'s own *ASJC* scheme. The methodological approach taken is described, and the collection's representativeness is evaluated and compared with the journal-based classification. There stand out both the great number of papers assigned by their research guarantors to more than one category (at times with even the same weight) and how frequently authors assigned categories which were not assigned to the journals in which their paper was published.

Keywords

Scientometry; Scientific classification; Scientific production; Paper-by-paper classification; Author's Assignation Collection; Authors; Research guarantors; Representation of papers; 2020; Test documentary collection; *Scopus*; *ASJC*.

Funding

Grant project PID2020-115798RB-I00 funded by *Ministerio de Ciencia e Innovación* of Spain (*Micin*), *Agencia Estatal de Investigación* (*AEI*) / 10.13039/501100011033.

1. Introduction

Scientometric studies require a reliable bibliographic database that covers the research papers published in the main scientific journals. The papers also need to be classified by discipline so that the progress of each discipline can be quantified. This classification is needed not only to quantify the research in each discipline but also to normalize the impact in that discipline, since publication and citation habits vary from one discipline to another (**Althouse** *et al.*, 2009; **Lancho-Barrantes**; **Guerrero-Bote**; **Moya-Anegón**, 2010; **Opthof**; **Leydesdorff**, 2010; **Bornmann**; **Leydesdorff**, 2017; **Bornmann**; **Tekles**; **Leydesdorff**, 2019).

What have until now been most used (**Gómez-Crisóstomo**, 2011; **Wang**; **Waltman**, 2016) are the classifications of the bibliographic databases themselves, *Scopus*'s *ASJC* (*All Science Journal Classification*) (*Elsevier*, 2023), and the *JCR* categories. Using scientific journals as scientometric units is quite common, and they have also been used to visualize the structure of science (**Leydesdorff**; **Moya-Anegón**; **Guerrero-Bote**, 2010; 2015; **Hassan-Montero**; **Guerrero-Bote**; **Moya-Anegón**, 2014). These journals' classifications are extended to the research papers they publish. The classification systems need to include some multidisciplinary categories, and many journals are assigned to various categories because they publish work corresponding to more than one. Nevertheless, not all the work a journal publishes are from all the categories to which it is assigned, indeed, quite the contrary is the case. All of this leads to great imprecision in both quantifying and normalizing the impact.

> Not all the work a journal publishes are from all the categories to which it is assigned, indeed, quite the contrary is the case

Numerous attempts have been made to improve these classification systems, and have generally aimed at classifying individual papers according to their own characteristics rather than those of the journal they belong to. Among these characteristics there stand out those based either on citation networks (direct citation, co-citation, bibliographic coupling, etc.) or text analysis (frequency of terms, etc.). **Šubelj**, **Van-Eck** and **Waltman** (2016) provide a discussion of these methods.

Some of these approaches are based on using automatic clustering systems to generate a new category classification scheme in which to distribute publications (**Klavans**; **Boyack**, 2005, 2006; **Waltman**; **Van-Eck**, 2012; **Janssens**; **Glänzel**; **De-Moor**, 2008; **Janssens** *et al.*, 2009). The results produced by these systems tend to change greatly as new literature is introduced into the classification, and have a randomness factor that can lead to disparate outcomes each time the procedure is restarted, even with the same sets of starting publications. Unfortunately, many bibliometric studies require classifications that are persistent and stable over time, even after the addition of new publications as they arise, so this type of classification is not usually widely accepted by the scientific community.

Other systems try to reorganize publications maintaining the category scheme of the journals, but also considering each publication's reference network to estimate the most precise category in which to assign it. This is done by **Glänzel**, **Schubert** & **Czerwon** (1999) and **Glänzel**, **Thijs** & **Huang** (2021), for example, to categorize articles from *WoS* multidisciplinary journals, and by **Milojević** (2020) to uniquely assign *WoS* publications. However, all of these works assume to be valid only certain assignations of papers to the category of their journal, and these are then used as trivial cases (starting points) with which to recursively solve the path of the citation network. This means that they are based not on the total number of references but on a smaller set. The classifications obtained through these methods have either not been evaluated or have been evaluated only in a very basic way.

In this paper, we describe our generation of a collection of documents from among those indexed by *Scopus*, and which their corresponding authors have classified using *Scopus*'s own *ASJC* scheme. The aim is for it to serve, with its limitations, as a possible further way to evaluate classification algorithms and the *Scopus* journal-based classification. Throughout this work and in future work, we shall use the term Author's Assignation Collection (AAC) to refer to this collection comprising the set of documents plus their classification.

For this, thousands of corresponding authors as research guarantors were surveyed (**De-Moya-Anegón** *et al.*, 2013) for them to determine the most appropriate category or categories in which to classify their works. We shall select a sufficiently large and representative sample from *Scopus*, and we shall have to answer some research questions about the responses obtained, such as:

Has the response obtained been homogeneous by country, proportionately distributed by subject, and adequately representative of all journals?

How many categories do the authors assign?

How do they distribute the weight among the different categories?

To what extent do these distributions coincide with the assignations that arise from the database's journal classification?

> The classifications obtained through these methods have either not been evaluated or have been evaluated only in a very basic way

2. Method and data

For this work, we used an April 2022 snapshot of the *Scopus* database (to which *SCImago* has access by agreement with its owner, the company *Elsevier*).

Scopus is known as the world's largest scientific database. It appeared in 2004 (**Hane**, 2004; **Pickering**, 2004) as an alternative to the *Thomson Reuters Web of Science* (*WoS*), covering most of the journals included in *WoS* and more (**Guerrero-Bote**; **Moya-Anegón**, 2012), and providing metadata on scientific documents and on citation links between these documents (**Guerrero-Bote** *et al.*, 2021).

The *Scopus* database uses the *ASJC* classification. This classifies journals into 27 subject areas, one of which is Multidisciplinary, which is where clearly multidisciplinary journals such as *Science* or *Nature* are classified. The other 26 subject areas are subdivided into 311 specific subject areas or categories, but each of those 26 subject areas has a miscellaneous category: Agricultural and Biological Sciences (miscellaneous), Arts and Humanities (miscellaneous), Biochemistry, Genetics and Molecular Biology (miscellaneous), etc.

In order to specify this classification a little more precisely, we made a "fractional assignation" of the journals to the categories. This fractional assignation consists of the fact that, if a journal is ascribed to 5 categories, each of those 5 affiliations is weighted by 1/5. We also eliminated both the Multidisciplinary subject area and the miscellaneous categories, distributing the weight among the corresponding categories. We were left with 26 subject areas and 285 categories or specific subject areas. The weight assigned to an affiliation to the Multidisciplinary subject area is divided among the 285, and the weight assigned to the miscellaneous categories is divided among the rest of the categories of the same subject area. In this way, there are journals that have the same weight for all the categories to which they belong and others that have different weights. This is a consequence of direct assignation, of possible assignation to a miscellaneous category, and of possible assignation to the Multidisciplinary subject area.

Although some classifications have forced each work to be assigned to a single category (**Milojević**, 2020; **Waltman**; **Van Eck**, 2012), there are currently many studies on multidisciplinarity (**Zhang**; **Rousseau**; **Glänzel**, 2016; **Huang** *et al.*, 2021; **Thijs**; **Huang**; **Glänzel**, 2021), so that we have considered allowing authors to assign more than one category. To do so, the research-guarantor corresponding authors (**De-Moya-Anegón** *et al.*, 2013) are asked to assign up to 5 categories for each work, indicating the percentage for which the work would belong to each category. They are asked to assign as few categories as possible and that, in so far as possible, the assigned categories be from the categories that *Scopus* assigns to the journal in which the work was published. One must bear in mind that, when authors submit a paper, they do so knowing the scope of the journal and the categories assigned to it, and the review process that the papers follow is oriented to the said scope and category.

The survey is done by email (Figure 1 shows an example).

Figure 1. Example of an email sent.

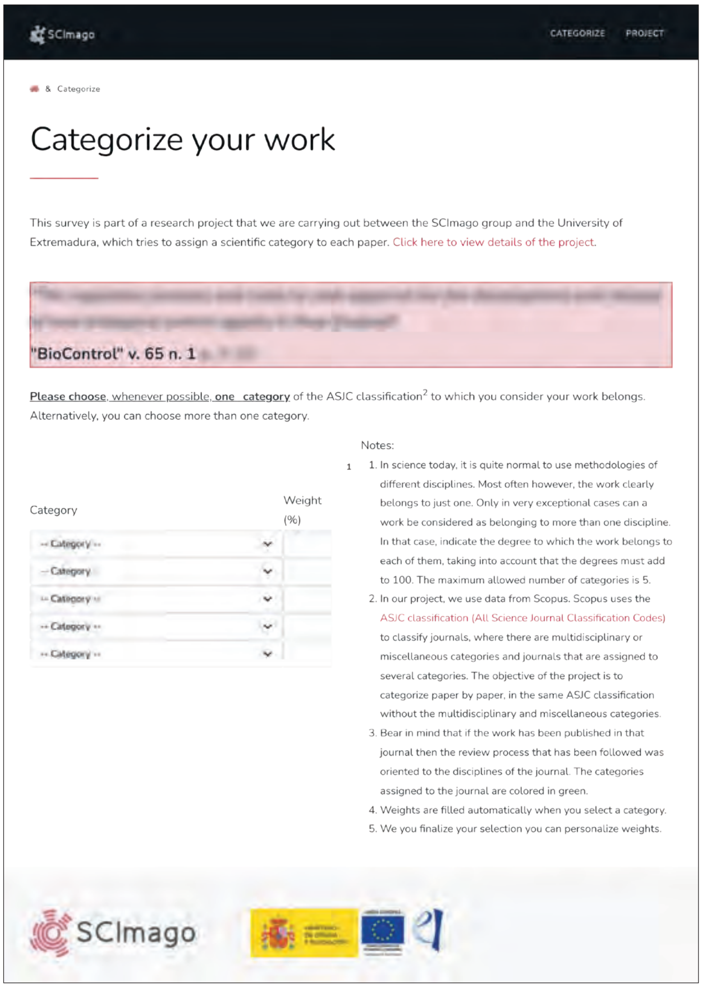


Figure 2. Example of the information collection form.

The survey form can be accessed in the links. It is as shown in Figure 2. Note that in both the email and the form the author is asked to assign a single category whenever possible, and better if it is one of the journal's categories.

To generate the sample, we start from the list of active *Scopus* journals for which the SJR is calculated (**Guerrero-Bote**; **Moya-Anegón**, 2012). We decided to focus on a recent year even if it were not the most recent, which is why 2020 was chosen. In that year, the SJR is calculated for 34 169 journals. It is considered that a sample of 15 000 works could be both feasible and more than sufficient.

For all the journals to be represented based on their size, it was decided that the sample would include one paper for every 200 papers a journal published, in particular, the integer part of dividing by 200 the number of papers published. This led to a total of 8751 papers representing 3271 journals. The rest of the journals publish fewer than 200 papers. For these journals, one paper is taken from those which are most prestigious, as measured by having an SJR greater than 0.6. Of these, there were 6338, making a total of 15 089 papers.

Once the selection of the journals had been made and the corresponding number of their papers to be taken established, all the papers published in them in 2020 that included the corresponding author's email in the database were taken. To avoid bombarding an author with several emails, we kept only one work per author.

We then assigned a random order number to each paper from each journal. Taking those papers with a random order number less than or equal to the number of papers assigned to each journal, we sent the first wave of emails on 17-10-2022.

We received 1123 responses for this first wave of 15 089 emails. This led us to launch a second wave with 13 966 emails corresponding to the following papers from each incomplete journal by its order number. For this second wave we obtained 1017 responses, so that there were still 12 949 left, proceeding in the same way with a third wave, and so on.

We ended the survey on 14-01-2023, having received 13 449 responses, which represents 89.13% of our objective.

3. Results

To check how robust the sample of responses received was, we compared it with the total number of publications in 2020 from different perspectives. Throughout this section, we shall refer to three sets of publications as follows:

a) "Citables 2020": The set of citable *Scopus* papers of 2020.

b) "Sent": The selection of papers to whose authors the invitation to participate in the survey had been sent (a subset of Citables 2020).

c) "Received": The set of papers for which we received a response from the authors (a subset of Sent).

We chose as a first verification that of the country of affiliation of the corresponding authors of the works in the sample. Table 1 lists the percentages by country of the corresponding authors of the set of Citables 2020, of Sent, and of Received, for countries with a greater than 1% Citables 2020 percentage. Figure 3 shows a plot of these data is shown. The complete table with all the countries is given in Annex 1. In the calculation of these percentages, it had to be taken into account that there are works with multiple affiliations that may cause them to be added to more than one country.

Table 1. Percentages of affiliation by the country of the authors.

Country	% Citables 2020	% Sent	% Received
China	20.63	18.53	5.97
United States	16.25	19.19	12.59
India	5.57	4.64	5.47
United Kingdom	4.25	5.10	3.10
Germany	3.98	4.45	4.90
Russian Federation	3.29	2.21	3.61
Japan	3.28	3.18	2.23
Italy	3.17	3.02	6.56
Spain	2.41	2.74	7.32
South Korea	2.36	2.33	1.21
Canada	2.36	2.65	2.09
France	2.35	2.63	2.95
Brazil	2.33	2.12	4.11
Australia	2.25	2.56	2.85
Iran	1.84	1.31	1.78
Turkey	1.37	1.16	2.06
Indonesia	1.36	1.18	1.89
Poland	1.33	1.48	1.56
Netherlands	1.23	1.72	1.17

One observes that the choice of works for the survey (Sent) has the same proportional distribution by country as in the total of Citables 2020. The response obtained, being also quite proportional, shows some striking data, such as the low response of authors affiliated to the countries with the highest proportion of scientific output (China especially, and the United States to a lesser extent), compared with the high response rate of such countries as Italy, Spain, and Brazil.

There also stands out the difference between the percentages of Sent works and the total Citables 2020 in the cases of China and the United States. This small imbalance is probably due to the fact that, as explained above, in the case of journals with fewer than 200 articles, only those with an SJR greater than 0.6 are considered, which today is commoner in journals of the United States than in those of China.

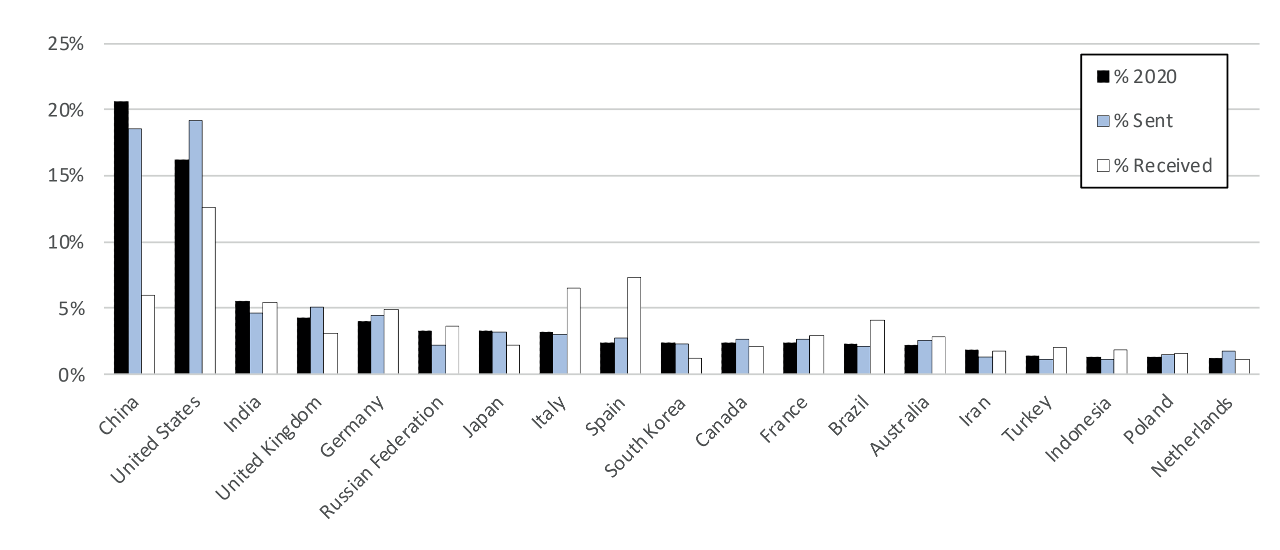


Figure 3. Plot of the affiliation percentages by country of the authors.

To check the thematic distribution of the sample, let us also compare the percentages of subject areas and categories.

Table 2 lists the percentages by subject area of the set of Citables 2020, of Sent, and of Received, for the 26 subject areas. Figure 4 shows a plot of these data.

Table 2. Assignation percentages by area.

ASJC	**Description**	**% 2020**	**% Sent**	**% Received**
1100	Agricultural and Biological Sciences	4.71	4.90	5.23
1200	Arts and Humanities	2.24	1.19	1.21
1300	Biochemistry, Genetics and Molecular Biology	5.32	6.28	6.45
1400	Business, Management and Accounting	1.51	2.31	2.37
1500	Chemical Engineering	2.05	1.94	1.72
1600	Chemistry	3.98	3.88	3.45
1700	Computer Science	8.10	6.11	6.80
1800	Decision Sciences	0.68	0.57	0.77
1900	Earth and Planetary Sciences	3.38	3.57	3.14
2000	Economics, Econometrics and Finance	0.97	1.62	2.18
2100	Energy	2.38	1.97	2.24
2200	Engineering	10.64	8.43	7.60
2300	Environmental Science	3.95	4.45	3.94
2400	Immunology and Microbiology	1.18	1.42	2.21
2500	Materials Science	5.52	5.01	5.36
2600	Mathematics	3.95	4.49	5.60
2700	Medicine	19.81	19.78	17.33
2800	Neuroscience	1.24	1.57	1.73
2900	Nursing	1.00	0.91	1.44
3000	Pharmacology, Toxicology and Pharmaceutics	1.83	1.91	1.65
3100	Physics and Astronomy	6.51	6.43	4.37
3200	Psychology	1.32	2.25	2.51
3300	Social Sciences	6.02	7.31	8.99
3400	Veterinary	0.53	0.43	0.26
3500	Dentistry	0.42	0.45	0.34
3600	Health Professions	0.75	0.83	1.10

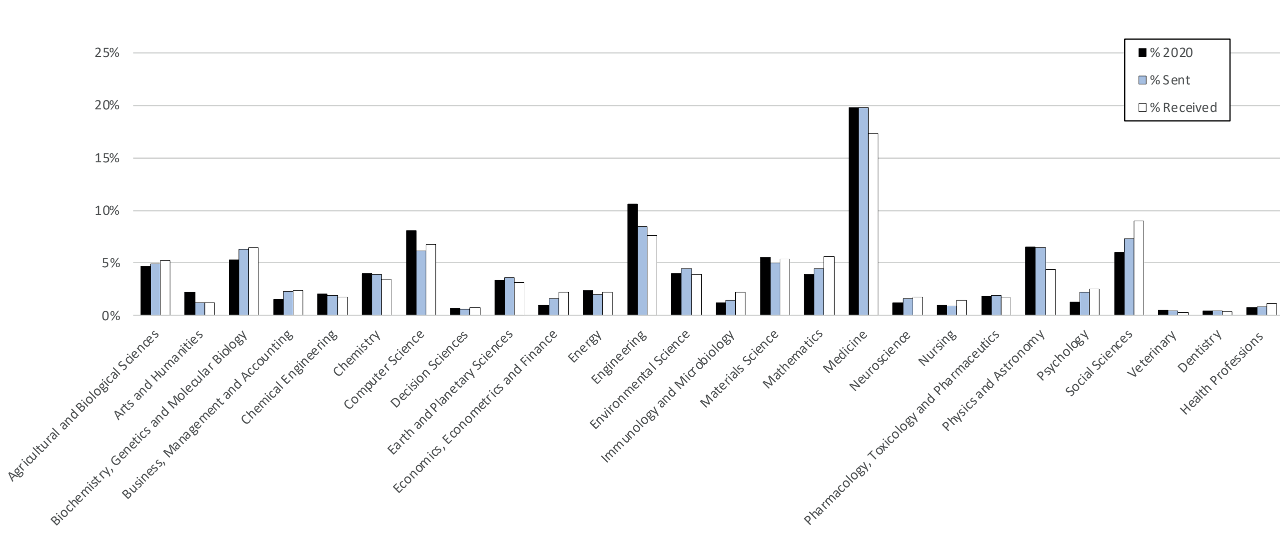


Figure 4. Graphical representation of assignation percentages by area.

As can be seen, there is very little variation in the percentages of assignation to subject areas between the sets of Citables 2020, Sent, and Received. At a finer grain, for the 285 specific subject areas, the distribution again shows little variation between Citables 2020, Sent, and Received. In this case, due to its length, we relegate the complete percentage data table to Annex 2.

Once verified that the sample used is robust in terms of its thematic variety and affiliation by country, we shall analyse the data collected, i.e., the categorizations that the authors made of their papers into the different specific subject areas. Table 3 compares the categorization percentages by number of categories, i.e., the percentage of papers that have a single category assigned, the percentage of those that have two categories assigned, etc., where:

a) Citables 2020 refers again to the set of 2020 citable papers from *Scopus*, classified with the "fractional assignation" of the journals.

b) Sample refers to the set of papers completed in the survey (Received), classified with the "fractional assignation" of the journals.

c) Survey refers to the set of papers completed in the survey (Received), but classified with the new assignation made by the authors.

The Items column indicates the number of papers in the set, and the Assignations column indicates the total number of assigned categories that accumulate those papers. As can be seen, while there is little variation between the categorization percentages of the total citable in 2020 and the sample, which serves to strengthen the validity of the sample, there is great variation from the categorization obtained from the survey.

Firstly, it called our attention that there is no greater percentage of authors who assign their works to a single category, despite the indications given in the questionnaire and in the email, considering the effort made in works such as **Milojević** (2020) or in **Waltman** & **Van-Eck** (2012) for achieving a categorization system that classifies papers into unique categories. Also striking is the high percentage of papers with 4 or more categories in the set of citable from *Scopus* and in the set of papers from the survey in the "fractional assignation". Let us recall that, as explained in the Introduction, the "fractional assignation" consists of eliminating the subject area Multidisciplinary, so that the papers assigned to that subject area are assigned to all categories (with a weight of 1/285), and miscellaneous categories are also removed from each subject area, reassigning their papers to all other categories of the subject area (with a weight of 1 divided by the number of remaining categories). Therefore, any paper originally belonging to the subject area Multidisciplinary or to any of the miscellaneous categories will, in the "fractional assignation," be assigned to a large number of categories (although with little weight in each of them).

Table 3. Categorization percentages by number of assigned categories.

Source	Items	Assignations	%1	%2	%3	%4	%>4
Citables 2020	3246022	56360548	15.57	17.81	11.79	7.77	47.05
Sample	13449	248621	15.01	18.83	11.83	7.09	47.25
Survey	13449	26141	44.85	30.70	14.40	5.39	4.68

For this reason, it may be pertinent to also analyse these percentages from the perspective of the original *Scopus* assignation, but excluding from the statistics the papers in the Multidisciplinary subject area and in any of the miscellaneous categories (both for the percentages of the original assignation and for those of the authors' assignation). Table 4 lists these values. Again, one observes little variation between the percentages of original categorization of the total of 2020 citable and of the survey. It does seem interesting that the elimination of the statistics of all these presumably multidisciplinary papers (which represent approximately 43% of those received in the survey and 42% of the total citable papers) does not seem to have substantially modified the authors' assignation percentages. It is true that by including them in the statistics, the percentage of unique assignations significantly decreases in favour of multiple ones, but with a very homogeneous distribution among those of 2, 3, 4, and more categories.

> It called our attention that there is no greater percentage of authors who assign their works to a single category, despite the indications given in the questionnaire and in the email

Table 4. Categorization percentages by number of assigned categories excluding the Multidisciplinary and miscellaneous categories.

Source	Items	Assignations	%1	%2	%3	%4	%>4
Citables 2020	1895436	4900726	26.67	30.3	19.6	12.35	11.1
Sample	7613	18798	26.52	33.0	20.4	11.6	8.5
Survey	7613	14356	47.432	29.7	13.9	4.926	4.09

Table 5 presents data similar to those of Table 3, but instead of with the number of assigned categories, with the number of winning assigned categories understood as being those in which the categories with the greatest weight have exactly the same weight.

Table 5. Categorization percentages by number of winning categories.

Source	Items	Winners	%1 w.	%2 w.	%3 w.	%4 w.	%5 w.
Citables 2020	3246022	30600572	27.95	23.72	15.39	9.67	23.28
Sample	13449	131659	28.49	25.10	15.70	9.15	21.55
Survey	13449	21037	65.58	21.25	7.26	2.97	2.93

As in Table 4, one sees that the sample reflects the total 2020 set fairly accurately. Likewise, there is a major increase in the results obtained from the survey with respect to the number of papers with a winning category, although, as in Table 4, the large number of categories the authors assign with equal weight is still striking.

Table 6 presents the range of weights the authors assign to the areas of their works, as well as the percentage of categories they assign to their works that are included among those the database assigns to the journal, which we denote by "coincidence". For example, for the first band (Bin 1), in 306 papers the authors assigned a total of 411 areas with weights between 0% and 10% (>=0 and <10). In 190 of these 411 assignations, this assignation was also found among the journal's areas, representing 46.23% coincidence.

Table 6. Assignations classified by weight, and percentages of coincidences with the respective journal.

Bin	Min wt	Max wt	Items (papers)	Assignations	Coincidences	%
1	0	9.34	306	411	190	46.23
2	10	18	998	1400	655	46.79
3	20	29	2411	5653	2706	47.87
4	30	39	2001	4176	2315	55.44
5	40	48.94	635	785	480	61.15
6	50	55	3124	5743	3833	66.74
7	60	68	444	444	336	75.68
8	70	75	682	682	531	77.86
9	80	89	555	555	452	81.44
10	90	99.99	198	198	161	81.31
11	100	100	6094	6094	5317	87.25
Total	-	-	13449	26141	16976	64.94

Figure 5 shows the number of assignations within each weight band of Table 6. One observes that most of the assignations are around the values of 20%, 30%, 50%, and 100%, which, contrasted with the data in Table 5 which indicates a very low percentage of works with many winning categories, leads us to think that authors tend to use round numbers to distribute the weight of the different categories of their works.

It seems logical to think that those author-made assignations with greater weight should have a greater likelihood of coincidence (i.e., of being among those assigned to the respective journals), and this is indeed reflected in the table. However, we find the percentage of assignations in which this is not the case to be very high, also considering that the papers published in journals in the Multidisciplinary subject area will always coincide with the journal since these journals are assigned to all categories due to the "fractional assignation." The same occurs, although with a lower probability, with the journals in the miscellaneous categories, and, furthermore, these coincidences can be understood as very weak coincidences.

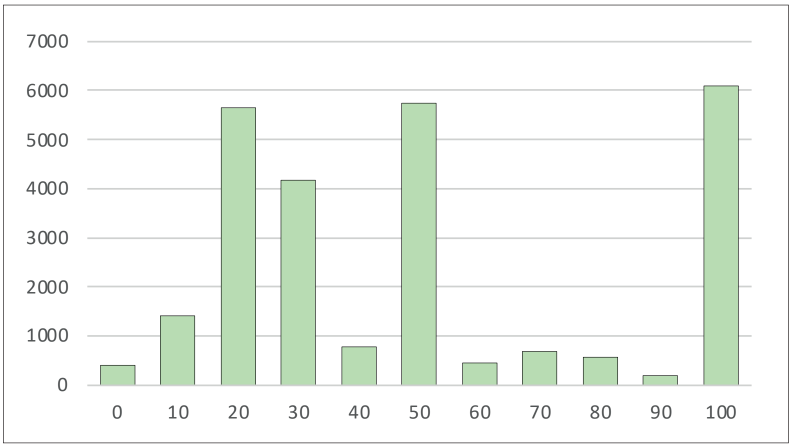


Figure 5. Number of assignations per range of assignation weights.

These cases of weak coincidence make us think that, to better estimate the coincidence, rather than treating it as a logical value, it should be weighted taking into account the concordance between the assignation percentage that the author establishes in a category (author weight) and the weight that the journal has in the category with the "fractional assignation" (journal weight). We have defined a coincidence weight that is calculated as the sum, for each coincident assignation made by the author, of the lesser of the weights – that of the author or that of the journal. Thus, if all the author assignations coincide with those of the journal and with the same weight, the final weight of the coincidence will be 100%.

Table 7 lists the percentages of coincidences and their weights according to the order of assignation. In the second row for example, for 2788 papers the authors made assignations to categories with a weight that made them come in second place (non-winning categories), with a total of 3411 assignations (i.e., in some cases there were assignations to two or more categories with the same weight, remaining in second place because there was another assignation with greater weight). On 53.00% of the occasions there was coincidence (those assignations were included among those of the journal). The average weight of the assignations made by the authors in those cases was 23.69%, although the average weight of the coincidences by item was 7.67%.

For the first row, in the 13449 cases (total), categories (sometimes divided into several) with winning weight were assigned, a total of 21041 assignations, among which in 68.33% of the cases there was coincidence. The average weight of the authors' assignations was 59.06% and the coincidence per item was 20.91%.

Table 7. Percentages of coincidences and their weights by order of assignation.

Order	Items	Assignations	Coincidences	Average author wt	Average coincidence wt
1	13449	21037	68.33	59.06	20.91
2	2788	3411	53.00	23.69	7.67
3	984	1187	47.85	14.40	4.60
4	314	400	45.00	9.65	2.79
5	106	106	42.45	7.41	1.96

The data in Table 7 indicate that the coincidence rate, as well as its weight, is greater in the first-order assignations, and that they decrease noticeably as the order number increases. Furthermore, the low average percentage of the weight of the coincidences is striking, even for the first-order assignations. Here one must take into account the effect produced by the "fractional assignation" since, in the publications of journals in the Multidisciplinary subject area, the weight of the coincidence will be extremely low –in the most favourable of cases (when its author assigns it to 5 subject areas), it will have a maximum coincidence percentage of 5×1/285 = 1.75%. Although to a lesser extent, the same is the case with the papers of each subject area's miscellaneous detailed subject area. For example, papers published in category 3301 Social Sciences (miscellaneous) of the subject area Social Sciences would be limited to 5×1/22 ≈ 22.72%.

Table 8 presents just the winning assignations according to the number of winners. For example, a single winning category was assigned in 8818 papers, with an average weight of 89.10%, which resulted in coincidence with the journal in 83.19% of the occasions with an average weight in the coincidence of 32.93%. There were two winning categories in 2858 papers (therefore, 2858×2=5716 winning assignations) with an average weight of 48.92%, in which 66.10% were coincidences with an average weight of 17.16%.

The low average percentage of the weight of the coincidences is striking, even for the first-order assignations

Table 8. Coincidence percentages and their weights according to the number of winning categories.

Winners	Items	Assignations	Coincidences	Average author wt	Average coincidence wt
1	8820	8820	83.19	89.10	32.93
2	2858	5716	66.10	48.92	17.16
3	977	2931	55.00	33.25	10.30
4	400	1600	50.50	25.00	7.12
5	394	1970	42.64	20.00	5.02

One observes that the coincidences are higher the fewer the winning categories.

In some particular case that we verified manually, we saw that the classification made by the authors is inconsistent with the references used as intellectual bases

We also investigated possible correlations of the percentage of categories assigned to journals and the percentage of winners assigned to the journal with the paper's number of references, citation, normalized citation, number of authors, with the corresponding author's prestige as measured by their number of papers or by their brute force (number of papers × average normalized citation), with the average of the authors, finding no significant correlations at 1%.

The only minimally significant correlation found was with the journal's SJR (0.067 and 0.068) in the sense that the greater the journal's SJR, the greater the probability that the categories assigned in the survey are among those assigned to the journals, which could be interpreted as that high-impact journals contain papers on subjects which (according to their authors) are more closely linked to the journal's subject area.

A significant negative correlation was also found between the percentage of assignations included in the journal and the percentage of winners included in the journal with the number of categories assigned and with the number of winners (of the order of -0.30). This reveals a certain tendency that the more categories the author assigns, the lower will be the probability that they coincide with those of the journal.

4. Conclusions

In this work, a collection of papers has been generated representative of those indexed in *Scopus* in 2020, categorized by the corresponding authors themselves as research guarantors (**De-Moya-Anegón** *et al.*, 2013) using the same *ASJC* scheme in a fractional way with up to a maximum of 5 categories, which we have named Author's Assignation Collection. The publications in the collection closely represent the thematic variety by area and category, as well as by country of affiliation of their authors, of the complete set of *Scopus* publications. However, as we have shown, authors of all nationalities did not respond equally.

The most important thing is that, despite having been explicitly urged to use few assignations, and to match them in so far as possible with those assigned by the journal, the author's responses show what is, in our opinion, a high proportion of multiple assignations that do not coincide with the journals. This deviation from the journal's theme is more notable the greater the number of assignations made by the author.

In some particular case that we verified manually, we saw that the classification made by the authors is inconsistent with the references used as intellectual bases. For example, there are some cases in which the authors assigned a paper to the Library and Information Science area without including a single reference to a paper that can be considered as from that area.

However, given the importance of the human factor involved in a survey-based methodology, we cannot state determinatively that the authors' classification is unquestionably more accurate, or that it presents better scientometric characteristics. What is certain is that there is a striking deviation between the journal-based classifications established in *Scopus* and the classification that the authors of the works themselves believe to be most appropriate, and to which, in some way, any classification should converge.

With all this, we consider that the Author's Assignation Collection (AAC) that we have created can be used as a further classification of reference for evaluating other classification systems of scientific documents collected in *Scopus* that also use the *ASJC* scheme.

The Author's Assignation Collection (AAC) that we have created can be used as a further classification of reference for evaluating other classification systems of scientific documents collected in *Scopus*

5. References

Althouse, Benjamin M.; **West, Jevin D.**; **Bergstrom, Carl T.**; **Bergstrom, Theodore** (2009). "Differences in impact factor across fields and over time". *Journal of the Association for Information Science and Technology*, v. 60, n. 1, pp. 27-34.
https://doi.org/10.1002/asi.20936

Bornmann, Lutz; **Leydesdorff, Loet** (2017). "Skewness of citation impact data and covariates of citation distributions: A large-scale empirical analysis based on Web of Science data". *Journal of informetrics*, v. 11, n. 1, pp. 164-175.
https://doi.org/10.1016/j.joi.2016.12.001

Bornmann, Lutz; **Tekles, Alexander**; **Leydesdorff, Loet** (2019). "How well does I3 perform for impact measurement compared to other bibliometric indicators? The convergent validity of several (field-normalized) indicators". *Scientometrics*, v. 119, n. 2, pp. 1187-1205.
https://doi.org/10.1007/s11192-019-03071-6

De-Moya-Anegón, Félix; **Guerrero-Bote, Vicente P.**; **Bornmann, Lutz**; **Moed, Henk F.** (2013). "The research guarantors of scientific papers and the output counting: a promising new approach". *Scientometrics*, v. 97, 421-434.
https://doi.org/10.1007/s11192-013-1046-0

Elsevier (2023). *What is the complete list of Scopus*, February 02. *Subject Areas and All Science Journal Classification Codes (ASJC)? - Scopus: Access and use Support Center*.
https://service.elsevier.com/app/answers/detail/a_id/15181/supporthub/scopus

Glänzel, Wolfgang; **Schubert, András**; **Czerwon, Hans-Jürgen** (1999). "An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis". *Scientometrics*, v. 44, n. 3, pp. 427–439.
https://doi.org/10.1007/bf02458488

Glänzel, Wolfgang; **Thijs, Bart**; **Huang, Ying** (2021). "Improving the precision of subject assignment for disparity measurement in studies of interdisciplinary research". In: W. Glänzel, S. Heeffer, P.S. Chi, R. Rousseau, *Proceedings of the 18th International Conference of the International Society of Scientometrics and Informetrics (ISSI 2021)*, Leuven University Press, pp. 453–464.
https://kuleuven.limo.libis.be/discovery/fulldisplay?docid=lirias3394551&context=SearchWebhook&vid=32KUL_KUL:Lirias

Gómez-Crisóstomo, María-Rocío (2011). *Estudio y comparación de la web of Science y Scopus (1996-2007)*. Tesis de doctorado. Universidad de Extremadura.
https://portalcientifico.uah.es/documentos/5da82cb42999523a1dc44829?lang=en

Guerrero-Bote, Vicente P.; **Chinchilla-Rodríguez, Zaida**; **Mendoza, Abraham**; **De-Moya-Anegón, Félix** (2021). "Comparative analysis of the bibliographic data sources dimensions and scopus: An approach at the country and institutional levels". *Frontiers in research metrics and analytics*, v. 5.
https://doi.org/10.3389/frma.2020.593494

Guerrero-Bote, Vicente P.; **De-Moya-Anegón, Félix** (2012). "A further step forward in measuring journals' scientific prestige: The SJR2 indicator". *Journal of informetrics*, v. 6, n. 4, pp. 674-688.
https://doi.org/10.1016/j.joi.2012.07.001

Hane, Paula J. (2004). *Elsevier announces Scopus service*, March 15.
https://newsbreaks.infotoday.com/NewsBreaks/Elsevier-Announces-Scopus-Service-16494.asp

Hassan-Montero, Yusef; **Guerrero-Bote, Vicente P.**; **De-Moya-Anegón, Félix** (2014). "Graphical interface of the Scimago Journal and Country Rank: an interactive approach to accessing bibliometric information". *El profesional de la información*, v. 23, n. 3.
https://doi.org/10.3145/epi.2014.may.07

Huang, Ying; **Glänzel, Wolfgang**; **Thijs, Bart**; **Porter, Alan L.**; **Zhang, Lin** (2021). *The comparison of various similarity measurement approaches on interdisciplinary indicators*. Feb - KU Leuven.

Janssens, Frizo; **Glänzel, Wolfgang**; **De-Moor, Bart** (2008). "A hybrid mapping of information science". *Scientometrics*, v. 75, n. 3, pp. 607–631.
https://doi.org/10.1007/s11192-007-2002-7

Janssens, Frizo; **Zhang, Lin**; **De-Moor, Bart**; **Glänzel, Wolfgang** (2009). "Hybrid clustering for validation and improvement of subject-classification schemes". *Information processing & management*, v. 45, n. 6, pp. 683–702.
https://doi.org/10.1016/j.ipm.2009.06.003

Klavans, Richard; **Boyack, Kevin W.** (2005). "Identifying a better measure of relatedness for mapping science". *Journal of the Association for Information Science and Technology*, v. 57, n. 2, pp. 251-263.
https://doi.org/10.1002/asi.20274

Klavans, Richard; **Boyack, Kevin W.** (2006). "Quantitative evaluation of large maps of science". *Scientometrics*, v. 68, n. 3, pp. 475-499.
https://doi.org/10.1007/s11192-006-0125-x

Lancho-Barrantes, Barbara S.; **Guerrero-Bote, Vicente P.**; **De-Moya-Anegón, Félix** (2010). "What lies behind the averages and significance of citation indicators in different disciplines?". *Journal of information science*, v. 36, n. 3, pp. 371-382.
https://doi.org/10.1177/0165551510366077

Leydesdorff, Loet; **De-Moya-Anegón, Félix**; **Guerrero-Bote, Vicente P.** (2010). "Journal maps on the basis of Scopus data: A comparison with the Journal Citation Reports of the ISI". *Journal of the American Society for Information Science and Technology*, v. 61, n. 2, pp. 352-369.
https://doi.org/10.1002/asi.21250

Leydesdorff, Loet; **De-Moya-Anegón, Félix**; **Guerrero-Bote, Vicente P.** (2015). "Journal maps, interactive overlays, and the measurement of interdisciplinarity on the basis of scopus data (1996–2012)". *Journal of the Association for Information Science and Technology*, v. 66, n. 5, pp. 1001-1016.
https://doi.org/10.1002/asi.23243

Milojević, Staša (2020). "Practical method to reclassify Web of Science articles into unique subject categories and broad disciplines". *Quantitative science studies*, v. 1, n. 1, pp. 183-206.
https://doi.org/10.1162/qss_a_00014

Opthof, Tobias; **Leydesdorff, Loet** (2010). "Caveats for the journal and field normalizations in the CWTS ('Leiden') evaluations of research performance". *Journal of informetrics*, v. 4, n. 3, pp. 423-430.
https://doi.org/10.1016/j.joi.2010.02.003

Pickering, B. (2004). "Elsevier prepares Scopus to rival ISI Web of science". *Information world review*, 8.

Šubelj, Lovro; **Van-Eck, Nees-Jan**; **Waltman, Ludo** (2016). "Clustering scientific publications based on citation relations: A systematic comparison of different methods". *PLoS one*, v. 11, n. 4, e0154404.
https://doi.org/10.1371/journal.pone.0154404

Thijs, Bart; **Huang, Ying**; **Glänzel, Wolfgang** (2021). *Comparing different implementations of similarity for disparity and variety measures in studies on interdisciplinarity*. FEB Research Report MSI_2103, Report No. MSI_2103.
https://lirias.kuleuven.be/retrieve/610314

Waltman, Ludo; **Van-Eck, Nees-Jan** (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the Association for Information Science and Technology*, v. 63, n. 12, pp. 2378-2392.
https://doi.org/10.1002/asi.22748

Wang, Qi; **Waltman, Ludo** (2016). "Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus". *Journal of informetrics*, v. 10, n. 2, pp. 347-364.
https://doi.org/10.1016/j.joi.2016.02.003

Zhang, Lin; **Rousseau, Ronald**; **Glänzel, Wolfgang** (2016). Diversity of references as an indicator of the 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

6. Annexes

Annex 1. Complete list of affiliation percentages by country of the authors

Country	Citables 2020 %	Sent %	Received %
China	20.63015	18.52546	5.96702
United States	16.24934	19.19230	12.58621
India	5.57112	4.64144	5.47226
United Kingdom	4.25447	5.09919	3.10345
Germany	3.98272	4.44903	4.89505
Russian Federation	3.29436	2.20958	3.60570
Japan	3.27829	3.17927	2.23388
Italy	3.16538	3.01603	6.55922
Spain	2.40927	2.74096	7.31634
South Korea	2.36397	2.32767	1.21439
Canada	2.35498	2.65483	2.09145
France	2.34842	2.62843	2.95352
Brazil	2.33221	2.11511	4.10795
Australia	2.25274	2.56106	2.84858
Iran	1.83511	1.30866	1.78411
Turkey	1.36696	1.15932	2.06147

Country	Citables 2020 %	Sent %	Received %
Indonesia	1.35858	1.17807	1.88906
Poland	1.33076	1.48093	1.55922
Netherlands	1.22525	1.71849	1.16942
Taiwan	0.96073	1.05235	1.03448
Malaysia	0.88135	0.65086	1.13193
Switzerland	0.87462	1.03220	1.19940
Sweden	0.80177	1.16071	1.05697
Egypt	0.68262	0.64183	0.79460
Mexico	0.67758	0.62793	1.24438
Portugal	0.64170	0.68906	2.00900
South Africa	0.64124	0.59390	0.71214
Saudi Arabia	0.63869	0.50846	0.67466
Belgium	0.62897	0.76269	0.55472
Pakistan	0.56870	0.39524	0.44978
Denmark	0.54454	0.75922	0.43478
Czech Republic	0.53950	0.49943	0.62969

Country	Citables 2020 %	Sent %	Received %
Austria	0.52935	0.54041	0.77961
Israel	0.50487	0.63696	0.87706
Ukraine	0.49631	0.25145	0.56222
Norway	0.49356	0.68559	0.53223
Thailand	0.49269	0.41121	0.88456
Iraq	0.47436	0.44664	0.76462
Hong Kong	0.46572	0.43969	0.39730
Greece	0.46264	0.47720	1.00450
Singapore	0.44408	0.40635	0.18741
Finland	0.42363	0.58348	0.35982
Romania	0.38178	0.34036	0.60720
Argentina	0.36531	0.40079	1.36432
Viet Nam	0.35986	0.29035	0.46477
Colombia	0.34956	0.30355	0.77211
Chile	0.34562	0.34661	0.84708
New Zealand	0.32949	0.38968	0.38231
Ireland	0.31905	0.38760	0.26237
Nigeria	0.30580	0.22506	0.59970
Hungary	0.26290	0.25284	0.59970
Morocco	0.24156	0.21116	0.34483
Slovakia	0.19884	0.16254	0.23238
Algeria	0.19669	0.14865	0.33733
Tunisia	0.18645	0.17991	0.30735
United Arab Emirates	0.18486	0.15143	0.20240
Croatia	0.18324	0.15768	0.31484
Bangladesh	0.17848	0.11600	0.21739
Serbia	0.17149	0.15907	0.34483
Bulgaria	0.15296	0.10489	0.29235
Slovenia	0.14246	0.15559	0.20990
Ethiopia	0.14049	0.18546	0.32234
Jordan	0.13938	0.09794	0.14243
Philippines	0.12230	0.09308	0.17991
Peru	0.11911	0.07224	0.16492
Kazakhstan	0.11813	0.08058	0.11994
Ecuador	0.11778	0.09655	0.18741
Lithuania	0.10099	0.10141	0.08996
Qatar	0.08538	0.06946	0.08246
Uzbekistan	0.08506	0.07224	0.12744
Lebanon	0.08295	0.08822	0.08246
Cyprus	0.07952	0.07780	0.13493
Ghana	0.07793	0.08058	0.12744
Estonia	0.06850	0.07710	0.11994
Sri Lanka	0.06427	0.05696	0.11244
Kenya	0.06264	0.07780	0.14993
Macao	0.05365	0.05418	0.01499
Latvia	0.05325	0.03612	0.03748
Nepal	0.05043	0.04446	0.07496
Cuba	0.04736	0.02084	0.05997
Kuwait	0.04521	0.04515	0.06747
Oman	0.04513	0.04098	0.11244

Country	Citables 2020 %	Sent %	Received %
Belarus	0.04356	0.02709	0.07496
Cameroon	0.03965	0.05279	0.14993
Bosnia and Herzegovina	0.03741	0.03959	0.07496
Azerbaijan	0.03680	0.01875	0.03748
Luxembourg	0.03666	0.05071	0.08996
Uruguay	0.03350	0.03543	0.10495
Uganda	0.03213	0.04723	0.04498
Tanzania	0.03092	0.03890	0.08246
Iceland	0.02700	0.03265	0.00000
Venezuela	0.02517	0.01737	0.02999
Georgia	0.02381	0.01875	0.02999
Costa Rica	0.02320	0.01806	0.02999
Bahrain	0.02297	0.01181	0.00750
Armenia	0.02253	0.01250	0.05247
Palestine	0.02216	0.01598	0.05997
Malta	0.02062	0.02084	0.05997
Zimbabwe	0.01926	0.01945	0.02249
Puerto Rico	0.01845	0.02640	0.00750
Macedonia	0.01830	0.01181	0.01499
Jamaica	0.01700	0.02362	0.02249
Sudan	0.01627	0.01806	0.02999
Syrian Arab Republic	0.01566	0.01042	0.01499
Vatican City State	0.01552	0.01042	0.01499
Senegal	0.01517	0.01598	0.03748
Yemen	0.01369	0.01528	0.03748
Botswana	0.01331	0.01598	0.03748
Brunei Darussalam	0.01262	0.00972	0.02249
Myanmar	0.01198	0.01181	0.00000
Benin	0.01114	0.01181	0.00750
Montenegro	0.01108	0.01042	0.01499
Panama	0.01096	0.01945	0.02999
Malawi	0.01047	0.00903	0.01499
Moldova	0.01038	0.00903	0.00750
Trinidad and Tobago	0.01003	0.01598	0.01499
Libya	0.00954	0.00834	0.04498
Albania	0.00940	0.01042	0.04498
Zambia	0.00934	0.01320	0.01499
Rwanda	0.00922	0.00764	0.02249
Côte d’Ivoire	0.00902	0.01598	0.02249
Burkina Faso	0.00815	0.01042	0.02249
Namibia	0.00769	0.00834	0.00000
Kyrgyzstan	0.00754	0.01111	0.01499
Mauritius	0.00751	0.00556	0.00000
Fiji	0.00737	0.00764	0.00000
Paraguay	0.00658	0.00278	0.00750
Mozambique	0.00655	0.00834	0.03748
Bolivia	0.00650	0.00556	0.01499
Democratic Republic Congo	0.00621	0.00903	0.02999
Reunion	0.00516	0.00556	0.00750
Mongolia	0.00510	0.00556	0.00750

Country	**Citables 2020 %**	**Sent %**	**Received %**
North Korea	0.00510	0.00556	0.00000
Cambodia	0.00505	0.00347	0.00750
Madagascar	0.00493	0.00695	0.02249
Kosovo (UNMIK)	0.00473	0.00417	0.00000
Afghanistan	0.00386	0.00556	0.00750
Honduras	0.00360	0.00347	0.01499
Tajikistan	0.00357	0.00278	0.00000
Dominican Republic	0.00351	0.00417	0.00750
Congo	0.00345	0.00417	0.01499
Bhutan	0.00336	0.00556	0.01499
Togo	0.00334	0.00347	0.00000
Guatemala	0.00328	0.00417	0.00000
Laos	0.00319	0.00139	0.00000
Guadeloupe	0.00313	0.00556	0.00750
French Guiana	0.00313	0.00417	0.00000
Mali	0.00270	0.00139	0.00000
Gambia	0.00261	0.00347	0.00000
Gabon	0.00261	0.00347	0.00000
Monaco	0.00247	0.00139	0.00000
Barbados	0.00244	0.00347	0.00000
Papua New Guinea	0.00215	0.00278	0.00750
New Caledonia	0.00212	0.00208	0.00750
Niger	0.00203	0.00208	0.00750
Grenada	0.00200	0.00208	0.00000
Liechtenstein	0.00197	0.00139	0.00000
Sierra Leone	0.00191	0.00208	0.01499
Saint Kitts and Nevis	0.00180	0.00000	0.00000
French Polynesia	0.00177	0.00069	0.00000
Angola	0.00177	0.00278	0.00750
Swaziland	0.00174	0.00000	0.00000
Martinique	0.00162	0.00139	0.00000
Eritrea	0.00142	0.00208	0.00000
Burundi	0.00136	0.00278	0.00750
Lesotho	0.00122	0.00000	0.00000
Guinea	0.00116	0.00069	0.00000
Mauritania	0.00113	0.00208	0.00000
Guam	0.00107	0.00139	0.00000
Somalia	0.00102	0.00278	0.01499
Greenland	0.00099	0.00139	0.00000
Bahamas	0.00096	0.00069	0.00000
El Salvador	0.00093	0.00000	0.00000
Faroe Islands	0.00087	0.00208	0.00750
Belize	0.00084	0.00069	0.00750
Haïti	0.00081	0.00139	0.00000
Maldives	0.00078	0.00069	0.00000
San Marino	0.00078	0.00000	0.00000
Seychelles	0.00078	0.00069	0.00000

Country	**Citables 2020 %**	**Sent %**	**Received %**
Suriname	0.00073	0.00208	0.00000
Guyana	0.00070	0.00069	0.00000
Nicaragua	0.00067	0.00139	0.00000
Cape Verde	0.00064	0.00069	0.00000
Guinea-Bissau	0.00061	0.00069	0.00750
Samoa	0.00058	0.00000	0.00000
Chad	0.00046	0.00000	0.00000
Liberia	0.00046	0.00000	0.00000
Central African Republic	0.00046	0.00139	0.00000
Bermuda	0.00038	0.00139	0.00000
Antigua and Barbuda	0.00038	0.00069	0.00000
Djibouti	0.00035	0.00069	0.00000
Republic of South Sudan	0.00035	0.00069	0.00000
Aruba	0.00035	0.00000	0.00000
Gibraltar	0.00029	0.00000	0.00000
Falkland Islands (Malvinas)	0.00026	0.00000	0.00000
Turkmenistan	0.00026	0.00139	0.00750
Vanuatu	0.00026	0.00000	0.00000
Virgin Islands (U.S.)	0.00026	0.00000	0.00000
Solomon Islands	0.00023	0.00000	0.00000
Equatorial Guinea	0.00023	0.00000	0.00000
Timor-Leste	0.00020	0.00000	0.00000
Curaçao	0.00020	0.00000	0.00000
Andorra	0.00020	0.00069	0.00000
Cayman Islands	0.00017	0.00139	0.00000
Federated States of Micronesia	0.00015	0.00000	0.00000
Mayotte	0.00015	0.00000	0.00000
Palau	0.00015	0.00000	0.00000
Dominica	0.00012	0.00000	0.00000
Saint Vincent and the Grenadines	0.00012	0.00000	0.00000
Virgin Islands (British)	0.00009	0.00000	0.00000
Anguilla	0.00009	0.00000	0.00000
Turks and Caicos Islands	0.00009	0.00069	0.00000
Nauru	0.00009	0.00000	0.00000
Svalbard and Jan Mayen	0.00006	0.00000	0.00000
Tonga	0.00006	0.00000	0.00000
Saint Helena	0.00006	0.00000	0.00000
Saint Lucia	0.00003	0.00000	0.00000
Comoros	0.00003	0.00000	0.00000
Norfolk Island	0.00003	0.00000	0.00000
American Samoa	0.00003	0.00000	0.00000
Northern Mariana Islands	0.00003	0.00000	0.00000
Tuvalu	0.00003	0.00000	0.00000
South Georgia and the South Sandwich Islands	0.00003	0.00000	0.00000

Annex 2. Assignation percentages by specific subject areas

ASJC	Description	% 2020	% Sent	% Received
1102	Agronomy and Crop Science	0.57	0.51	0.68
1103	Animal Science and Zoology	0.57	0.58	0.62
1104	Aquatic Science	0.41	0.44	0.39
1105	Ecology, Evolution, Behavior and Systematics	0.78	1.01	1.05
1106	Food Science	0.74	0.69	0.79
1107	Forestry	0.23	0.21	0.23
1108	Horticulture	0.23	0.19	0.16
1109	Insect Science	0.24	0.23	0.23
1110	Plant Science	0.71	0.81	0.86
1111	Soil Science	0.24	0.23	0.22
1202	History	0.48	0.18	0.19
1203	Language and Linguistics	0.03	0.03	0.13
1204	Archeology (arts and humanities)	0.12	0.14	0.08
1205	Classics	0.06	0.03	0.00
1206	Conservation	0.06	0.04	0.15
1207	History and Philosophy of Science	0.14	0.17	0.15
1208	Literature and Literary Theory	0.38	0.05	0.04
1209	Museology	0.05	0.04	0.00
1210	Music	0.09	0.06	0.02
1211	Philosophy	0.39	0.28	0.34
1212	Religious Studies	0.24	0.10	0.05
1213	Visual Arts and Performing Arts	0.22	0.06	0.05
1302	Aging	0.13	0.13	0.26
1303	Biochemistry	0.83	0.94	0.76
1304	Biophysics	0.25	0.25	0.28
1305	Biotechnology	0.40	0.40	0.50
1306	Cancer Research	0.555	0.59	1.05
1307	Cell Biology	0.54	0.67	0.77
1308	Clinical Biochemistry	0.24	0.27	0.10
1309	Developmental Biology	0.16	0.22	0.19
1310	Endocrinology	0.19	0.24	0.17
1311	Genetics	0.60	0.78	0.80
1312	Molecular Biology	0.68	0.88	0.82
1313	Molecular Medicine	0.31	0.36	0.23
1314	Physiology	0.31	0.38	0.36
1315	Structural Biology	0.13	0.16	0.15
1402	Accounting	0.11	0.21	0.27
1403	Business and International Management	0.24	0.35	0.38
1404	Management Information Systems	0.10	0.12	0.08
1405	Management of Technology and Innovation	0.22	0.30	0.27
1406	Marketing	0.18	0.32	0.43
1407	Organizational Behavior and Human Resource Management	0.13	0.26	0.32
1408	Strategy and Management	0.34	0.47	0.37
1409	Tourism, Leisure and Hospitality Management	0.14	0.21	0.22
1410	Industrial Relations	0.06	0.08	0.03
1502	Bioengineering	0.38	0.35	0.29
1503	Catalysis	0.47	0.47	0.43

ASJC	Description	% 2020	% Sent	% Received
1504	Chemical Health and Safety	0.15	0.13	0.05
1505	Colloid and Surface Chemistry	0.19	0.18	0.20
1506	Filtration and Separation	0.19	0.17	0.11
1507	Fluid Flow and Transfer Processes	0.34	0.32	0.44
1508	Process Chemistry and Technology	0.33	0.31	0.21
1602	Analytical Chemistry	0.70	0.71	0.70
1603	Electrochemistry	0.46	0.43	0.36
1604	Inorganic Chemistry	0.58	0.56	0.42
1605	Organic Chemistry	0.87	0.82	0.92
1606	Physical and Theoretical Chemistry	0.88	0.88	0.77
1607	Spectroscopy	0.50	0.48	0.29
1702	Artificial Intelligence	1.07	0.71	1.66
1703	Computational Theory and Mathematics	0.33	0.29	0.32
1704	Computer Graphics and Computer-Aided Design	0.30	0.25	0.12
1705	Computer Networks and Communications	1.20	0.76	0.81
1706	Computer Science Applications	1.38	1.09	1.51
1707	Computer Vision and Pattern Recognition	0.53	0.43	0.69
1708	Hardware and Architecture	0.53	0.33	0.10
1709	Human-Computer Interaction	0.42	0.37	0.39
1710	Information Systems	0.75	0.60	0.51
1711	Signal Processing	0.54	0.41	0.37
1712	Software	1.06	0.86	0.33
1802	Information Systems and Management	0.35	0.14	0.16
1803	Management Science and Operations Research	0.19	0.21	0.37
1804	Statistics, Probability and Uncertainty	0.14	0.22	0.24
1902	Atmospheric Science	0.39	0.45	0.42
1903	Computers in Earth Sciences	0.13	0.15	0.20
1904	Earth-Surface Processes	0.27	0.30	0.33
1905	Economic Geology	0.11	0.13	0.05
1906	Geochemistry and Petrology	0.32	0.31	0.24
1907	Geology	0.43	0.46	0.42
1908	Geophysics	0.35	0.31	0.37
1909	Geotechnical Engineering and Engineering Geology	0.49	0.41	0.47
1910	Oceanography	0.23	0.27	0.25
1911	Paleontology	0.15	0.21	0.22
1912	Space and Planetary Science	0.40	0.44	0.14
1913	Stratigraphy	0.11	0.13	0.03
2002	Economics and Econometrics	0.65	1.16	1.78
2003	Finance	0.32	0.46	0.40
2102	Energy Engineering and Power Technology	0.89	0.62	0.82
2103	Fuel Technology	0.37	0.33	0.14
2104	Nuclear Energy and Engineering	0.29	0.24	0.16
2105	Renewable Energy, Sustainability and the Environment	0.83	0.78	1.13
2202	Aerospace Engineering	0.52	0.33	0.36
2203	Automotive Engineering	0.30	0.24	0.16
2204	Biomedical Engineering	0.48	0.44	0.46
2205	Civil and Structural Engineering	0.73	0.69	0.81
2206	Computational Mechanics	0.20	0.17	0.24

ASJC	Description	% 2020	% Sent	% Received
2207	Control and Systems Engineering	1.03	0.96	0.62
2208	Electrical and Electronic Engineering	2.54	1.85	1.93
2209	Industrial and Manufacturing Engineering	0.94	0.90	0.62
2210	Mechanical Engineering	1.33	0.99	0.99
2211	Mechanics of Materials	0.86	0.67	0.42
2212	Ocean Engineering	0.27	0.26	0.19
2213	Safety, Risk, Reliability and Quality	0.56	0.27	0.22
2214	Media Technology	0.23	0.16	0.09
2215	Building and Construction	0.43	0.34	0.37
2216	Architecture	0.22	0.15	0.12
2302	Ecological Modeling	0.12	0.15	0.12
2303	Ecology	0.49	0.55	0.79
2304	Environmental Chemistry	0.43	0.47	0.44
2305	Environmental Engineering	0.42	0.40	0.59
2306	Global and Planetary Change	0.16	0.22	0.13
2307	Health, Toxicology and Mutagenesis	0.40	0.46	0.17
2308	Management, Monitoring, Policy and Law	0.40	0.54	0.39
2309	Nature and Landscape Conservation	0.23	0.28	0.16
2310	Pollution	0.51	0.57	0.38
2311	Waste Management and Disposal	0.33	0.33	0.23
2312	Water Science and Technology	0.46	0.50	0.54
2402	Applied Microbiology and Biotechnology	0.16	0.16	0.36
2403	Immunology	0.36	0.48	0.64
2404	Microbiology	0.35	0.42	0.70
2405	Parasitology	0.13	0.14	0.18
2406	Virology	0.18	0.22	0.32
2502	Biomaterials	0.47	0.48	0.55
2503	Ceramics and Composites	0.53	0.47	0.37
2504	Electronic, Optical and Magnetic Materials	1.26	1.06	0.58
2505	Materials Chemistry	0.89	0.81	1.08
2506	Metals and Alloys	0.60	0.52	0.64
2507	Polymers and Plastics	0.60	0.56	0.63
2508	Surfaces, Coatings and Films	0.59	0.54	0.45
2509	Nanoscience and Nanotechnology	0.58	0.58	1.05
2602	Algebra and Number Theory	0.23	0.31	0.48
2603	Analysis	0.25	0.32	0.73
2604	Applied Mathematics	0.78	0.92	0.86
2605	Computational Mathematics	0.25	0.25	0.29
2606	Control and Optimization	0.45	0.22	0.26
2607	Discrete Mathematics and Combinatorics	0.16	0.17	0.22
2608	Geometry and Topology	0.16	0.24	0.38
2609	Logic	0.12	0.15	0.10
2610	Mathematical Physics	0.17	0.21	0.38
2611	Modeling and Simulation	0.45	0.39	0.63
2612	Numerical Analysis	0.12	0.16	0.20
2613	Statistics and Probability	0.32	0.50	0.76
2614	Theoretical Computer Science	0.51	0.64	0.31
2702	Anatomy	0.17	0.16	0.14

ASJC	Description	% 2020	% Sent	% Received
2703	Anesthesiology and Pain Medicine	0.35	0.35	0.28
2704	Biochemistry (medical)	0.19	0.20	0.27
2705	Cardiology and Cardiovascular Medicine	0.93	0.81	0.98
2706	Critical Care and Intensive Care Medicine	0.29	0.30	0.23
2707	Complementary and Alternative Medicine	0.27	0.21	0.12
2708	Dermatology	0.40	0.44	0.40
2709	Drug Guides	0.13	0.12	0.03
2710	Embryology	0.14	0.13	0.04
2711	Emergency Medicine	0.26	0.24	0.13
2712	Endocrinology, Diabetes and Metabolism	0.45	0.48	0.46
2713	Epidemiology	0.26	0.31	0.61
2714	Family Practice	0.21	0.17	0.08
2715	Gastroenterology	0.43	0.43	0.37
2716	Genetics (clinical)	0.28	0.34	0.29
2717	Geriatrics and Gerontology	0.26	0.27	0.20
2718	Health Informatics	0.37	0.34	0.27
2719	Health Policy	0.42	0.46	0.55
2720	Hematology	0.36	0.36	0.33
2721	Hepatology	0.26	0.27	0.17
2722	Histology	0.19	0.20	0.06
2723	Immunology and Allergy	0.45	0.49	0.25
2724	Internal Medicine	0.36	0.35	0.20
2725	Infectious Diseases	0.62	0.66	0.85
2726	Microbiology (medical)	0.38	0.41	0.31
2727	Nephrology	0.25	0.25	0.24
2728	Neurology (clinical)	0.70	0.79	0.62
2729	Obstetrics and Gynecology	0.50	0.47	0.45
2730	Oncology	0.98	0.94	0.96
2731	Ophthalmology	0.47	0.45	0.45
2732	Orthopedics and Sports Medicine	0.59	0.57	0.52
2733	Otorhinolaryngology	0.35	0.33	0.20
2734	Pathology and Forensic Medicine	0.38	0.35	0.20
2735	Pediatrics, Perinatology and Child Health	0.74	0.63	0.49
2736	Pharmacology (medical)	0.57	0.55	0.29
2737	Physiology (medical)	0.29	0.29	0.14
2738	Psychiatry and Mental Health	0.74	0.95	0.88
2739	Public Health, Environmental and Occupational Health	1.05	1.10	1.23
2740	Pulmonary and Respiratory Medicine	0.42	0.45	0.37
2741	Radiology, Nuclear Medicine and Imaging	0.71	0.63	0.53
2742	Rehabilitation	0.25	0.25	0.23
2743	Reproductive Medicine	0.21	0.21	0.19
2744	Reviews and References (medical)	0.13	0.12	0.07
2745	Rheumatology	0.26	0.26	0.26
2746	Surgery	1.19	1.07	1.02
2747	Transplantation	0.22	0.21	0.11
2748	Urology	0.37	0.38	0.26
2802	Behavioral Neuroscience	0.15	0.20	0.24
2803	Biological Psychiatry	0.12	0.14	0.08

ASJC	Description	% 2020	% Sent	% Received
2804	Cellular and Molecular Neuroscience	0.20	0.27	0.29
2805	Cognitive Neuroscience	0.18	0.24	0.41
2806	Developmental Neuroscience	0.09	0.12	0.08
2807	Endocrine and Autonomic Systems	0.08	0.09	0.05
2808	Neurology	0.31	0.35	0.42
2809	Sensory Systems	0.11	0.15	0.16
2902	Advanced and Specialized Nursing	0.07	0.06	0.05
2903	Assessment and Diagnosis	0.02	0.02	0.11
2904	Care Planning	0.02	0.02	0.01
2905	Community and Home Care	0.04	0.04	0.01
2906	Critical Care Nursing	0.04	0.04	0.02
2907	Emergency Nursing	0.04	0.04	0.01
2908	Fundamentals and Skills	0.04	0.03	0.02
2909	Gerontology	0.05	0.05	0.05
2910	Issues, Ethics and Legal Aspects	0.04	0.04	0.08
2911	Leadership and Management	0.06	0.04	0.10
2912	LPN and LVN	0.03	0.03	
2913	Maternity and Midwifery	0.04	0.04	0.05
2914	Medical and Surgical Nursing	0.04	0.03	0.05
2915	Nurse Assisting	0.02	0.02	0.02
2916	Nutrition and Dietetics	0.25	0.27	0.40
2917	Oncology (nursing)	0.04	0.04	0.03
2918	Pathophysiology	0.02	0.02	0.06
2919	Pediatrics	0.04	0.04	0.14
2920	Pharmacology (nursing)	0.02	0.02	0.00
2921	Psychiatric Mental Health	0.02	0.02	0.12
2922	Research and Theory	0.03	0.02	0.11
2923	Review and Exam Preparation	0.02	0.02	
3002	Drug Discovery	0.35	0.35	0.42
3003	Pharmaceutical Science	0.58	0.56	0.56
3004	Pharmacology	0.61	0.67	0.39
3005	Toxicology	0.29	0.33	0.28
3102	Acoustics and Ultrasonics	0.45	0.42	0.22
3103	Astronomy and Astrophysics	0.61	0.67	0.68
3104	Condensed Matter Physics	1.54	1.57	0.93
3105	Instrumentation	0.90	0.70	0.27
3106	Nuclear and High Energy Physics	0.65	0.69	0.50
3107	Atomic and Molecular Physics, and Optics	1.09	1.04	0.91
3108	Radiation	0.42	0.42	0.20
3109	Statistical and Nonlinear Physics	0.42	0.47	0.34
3110	Surfaces and Interfaces	0.43	0.45	0.32
3202	Applied Psychology	0.22	0.42	0.42
3203	Clinical Psychology	0.30	0.46	0.46
3204	Developmental and Educational Psychology	0.30	0.52	0.52
3205	Experimental and Cognitive Psychology	0.16	0.28	0.37
3206	Neuropsychology and Physiological Psychology	0.11	0.17	0.17
3207	Social Psychology	0.23	0.40	0.57
3302	Archeology	0.10	0.14	0.18

ASJC	Description	% 2020	% Sent	% Received
3303	Development	0.20	0.28	0.23
3304	Education	1.23	1.79	2.22
3305	Geography, Planning and Development	0.53	0.70	0.66
3306	Health (social science)	0.22	0.24	0.41
3307	Human Factors and Ergonomics	0.05	0.06	0.11
3308	Law	0.48	0.36	0.29
3309	Library and Information Sciences	0.19	0.17	0.19
3310	Linguistics and Language	0.49	0.50	0.54
3311	Safety Research	0.10	0.06	0.07
3312	Sociology and Political Science	0.64	0.98	1.21
3313	Transportation	0.18	0.19	0.34
3314	Anthropology	0.16	0.16	0.25
3315	Communication	0.25	0.32	0.54
3316	Cultural Studies	0.33	0.17	0.17
3317	Demography	0.07	0.12	0.07
3318	Gender Studies	0.09	0.12	0.17
3319	Life-span and Life-course Studies	0.06	0.06	0.03
3320	Political Science and International Relations	0.29	0.39	0.59
3321	Public Administration	0.11	0.17	0.16
3322	Urban Studies	0.12	0.18	0.30
3323	Social Work	0.06	0.07	0.11
3399	E-learning	0.06	0.08	0.13
3402	Equine	0.17	0.13	0.04
3403	Food Animals	0.19	0.15	0.13
3404	Small Animals	0.17	0.14	0.09
3502	Dental Assisting	0.06	0.06	0.01
3503	Dental Hygiene	0.06	0.06	0.03
3504	Oral Surgery	0.12	0.13	0.15
3505	Orthodontics	0.09	0.09	0.06
3506	Periodontics	0.08	0.10	0.09
3602	Chiropractics	0.02	0.01	
3603	Complementary and Manual Therapy	0.02	0.01	0.01
3604	Emergency Medical Services	0.01	0.01	0.02
3605	Health Information Management	0.05	0.06	0.07
3606	Medical Assisting and Transcription	0.01	0.01	0.01
3607	Medical Laboratory Technology	0.04	0.03	0.06
3608	Medical Terminology	0.01	0.01	0.00
3609	Occupational Therapy	0.02	0.02	0.03
3610	Optometry	0.02	0.03	0.07
3611	Pharmacy	0.05	0.04	0.09
3612	Physical Therapy, Sports Therapy and Rehabilitation	0.21	0.23	0.21
3613	Podiatry	0.01	0.01	0.01
3614	Radiological and Ultrasound Technology	0.08	0.06	0.05
3615	Respiratory Care	0.01	0.01	0.03
3616	Speech and Hearing	0.05	0.08	0.09
3699	Sports Science	0.15	0.21	0.34

Loet Leydesdorff: bibliometric analysis and mapping of his scientific production

Audilio Gonzales-Aguilar; María-Jesús Colmenero-Ruiz; Francisco-Carlos Paletta; Lise Verlaet

Nota: Este artículo se puede leer en español en:
https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/87543

Recommended citation:

Gonzales-Aguilar, Audilio; **Colmenero-Ruiz, María-Jesús**; **Paletta, Francisco-Carlos**; **Verlaet, Lise** (2023). "Loet Leydesdorff: bibliometric analysis and mapping of his scientific production". *Profesional de la información*, v. 32, n. 7, e320709.

https://doi.org/10.3145/epi.2023.dic.09

Article received on December 01st 2023
Approved on December 17th 2023

Audilio Gonzales-Aguilar ✉
https://orcid.org/0000-0001-8693-2076

Université Paul-Valéry Montpellier III
Département de Documentation
Montpellier, France
audilio.gonzales@gmail.com

María-Jesús Colmenero-Ruiz
https://orcid.org/0000-0002-1650-1091

Universidad Complutense de Madrid
Santísima Trinidad, 37
28010 Madrid, Spain
mcolmene@ucm.es

Francisco-Carlos Paletta
https://orcid.org/0000-0002-4112-5198

Universidade de São Paulo
Departamento de Informação e Cultura
São Paulo, Brazil
fcpaletta@usp.br

Lise Verlaet
https://orcid.org/0000-0002-3909-0184

Université Paul-Valéry Montpellier III
Département de Documentation
Montpellier, France
lise.verlaet@univ-montp3.fr

Abstract

This study conducts a comprehensive bibliometric analysis and mapping of the scientific production of Loet Leydesdorff. Leveraging bibliometric techniques, the research aims to explore the breadth and impact of Leydesdorff's scholarly contributions. The analysis includes an examination of his publication patterns, citation impact, and collaborations over time. Additionally, mapping techniques will be employed to visually represent the networks and interdisciplinarity associated with Leydesdorff's work. The study provides insights into the evolution of his research interests, the influence of his contributions within the scientific community, and the interdisciplinary connections inherent in his body of work.

Keywords

Bibliometric analysis; Bibliometrics; *Biblioshiny*; *Web of Science*; *Cosma*; Text analysis; Social network analysis; Mapping; Researchers; Loet Leydesdorff.

Acknowledgements

We express our sincere thanks to Tomàs Baiget for his kind invitation to contribute an article in homage to Loet Leydesdorff.

"Citation analysis has conquered the world of science policy analysis"
(**Amsterdamska**; **Leydesdorff**, 1988)

1. Introduction

Loet Leydesdorff was a Dutch sociologist of science and science communication researcher. He received his PhD in Sociology of Science from the *University of Amsterdam* in 1979 with a thesis entitled "Dynamic and stochastic models for reciprocal citation processes" (**Leydesdorff**, 1979). His research focused mainly on the application of complex systems theory to scientific communication.

Died in March 2023, Leydesdorff is internationally recognised for his contributions to the mapping and analysis of scientific collaboration networks. In (**Leydesdorff**, 2007a) he proposed the use of "intermediate centrality" as an indicator of the interdisciplinarity of scientific journals. Furthermore, his innovative methods for visualising and analysing networks, as shown in (**Leydesdorff**; **Ràfols**, 2009), offer a unique perspective on the global structure of scientific research.

Simultaneously, Leydesdorff developed sociological models of scientific communication. His book "*A sociological theory of communication: The self-organization of the knowledge-based society*", explores the dynamics of knowledge-based societies (**Leydesdorff**, 2001). He has also contributed to understanding the dynamics of scientific research, as evidenced in (**Leydesdorff**; **Meyer**, 2010), on the decline of university patents. Together with Henry Etzkowitz, he developed the Triple Helix model, a conceptual model that describes the interaction and collaboration between universities, industries and governments in the innovation process (**Etzkowitz**; **Leydesdorff**, 1996), with great repercussions in the area. The breadth of Loet Leydesdorff's contributions to the understanding of research networks, scientific communication, innovation and the sociology of science is noteworthy.

> Leydesdorff is internationally recognised for his contributions to the mapping and analysis of scientific collaboration networks

Bibliometric analysis is a macroscopic tool for extracting and discovering knowledge from a large amount of research literature very quickly compared to a traditional systematic review. In recent years, bibliometric analysis has attracted the interest of researchers for various reasons, such as the emergence of digital technologies or bibliometric software like *VOSviewer*, *CiteSpace* and *Biblioshiny* and the development of academic databases like *Web of Science*, *Scopus* and *Google Scholar* (**Moral**-**Muñoz** *et al.*, 2020).

Bibliometric analysis uses both quantitative and qualitative methods. Quantitative methods include descriptive and performance metrics of the research output of a field (the number of publications or citations, etc.), as well as the identification of the most important research constituents (the most cited articles, the most productive sources, etc.). Qualitative methods include the analysis of scientific mapping to explore the relationships between research constituents (**Donthu**, 2021). Scientific mapping is carried out by analysing networks of textual units, with techniques such as co-word analysis, co-citation analysis and collaboration analysis (**Zupic**; **Čater**, 2015).

This paper is structured as follows: section 2 presents the methodology used. Section 3 provides the results of the bibliometric analysis. Section 4 presents the interactive mapping analysis to explore the relationships between the elements studied (co-authors, keywords, keywords plus and publications). Finally, section 5 provides a brief summary of the main results and conclusions of the work.

2. Methodology

This section describes the research methodology used. The data were obtained from the *Web of Science* (*WoS*) *Core Collection* database, with the search phrase "L. Leydesdorff OR Loet Leydesdorff":
http://www.webofscience.com

Figure 1 gives an overview of the data corpus.

Based on them, a bibliometric analysis was carried out to answer the research questions, which involves, on the one hand, a statistical analysis and, on the other hand, the visualisation of the research production of a field. This bibliometric analysis was carried out using *Bibliometrix*, an open source software, supported by the *R* environment, which provides tools for the calculation of performance metrics (**Aria**; **Cuccurullo**, 2017):
https://www.bibliometrix.org/home

Bibliometrix is integrated with *Biblioshiny*, a web interface for bibliometric network visualisation. The statistical analysis consisted of a performance analysis to show publication and citation patterns, publications, authors and countries, as well as the most cited articles. For visualisation, publications were mapped to explore topic and keyword trends through co-word analysis, co-citation clusters through co-citation analysis, and collaboration

> Together with Henry Etzkowitz, he developed the Triple Helix model, a conceptual model that describes the interaction and collaboration between universities, industries and governments in the innovation process

structure between countries through collaboration analysis. The latter results were visualised in the form of networks. The workflow steps followed for the bibliometric analysis were: study design, data collection, data analysis, and visualisation and interpretation of the results (**Zupic**; **Čater**, 2015).

The main objective is to examine and visualise Loet Leydesdorff's scientific publications from 1980 to 2023. The research questions of this study are as follows:

- What is the evolution of publications and citations?
- What are the most relevant and influential sources, countries and publications?
- What are the most common research topics and keyword trends in Leydesdorff research?
- What are the main co-citation clusters?
- What is the collaborative network of countries in social media research?

In addition, a scientific network cartography has been created, made up of this set of publications belonging to this author, with the idea of displaying it interactively on the web and with the possibility of downloading it. The software used for this has been *Cosma*, a software for the visualisation of documentary graphs (**Perret** *et al.*, 2021). In the section corresponding to the cartography we will explain in more detail its characteristics and functionalities.

The Table 1 presents a summary of the main information about the dataset. Specifically, our dataset contains 424 articles published between 1980 and 2023. These articles were published in 101 different scientific sources. The sources are made up of various types of documents: scientific journal articles, conference papers, books and book chapters, reviews, etc. The average number of years it takes for an article to be cited is approximately 0.92 and each article has on average 50.71 citations. The total number of references cited in all articles is 7486. In addition, the articles contain 3,297 author keywords and 1,470 plus (additional) keywords.

Table 1. Main information on the bibliographic dataset

Description	Results
Period	1980:2023
Sources (journals, books, etc.)	101
Documents	424
Annual growth rate	2,59
Average age of documents	13,5
Average number of citations per document	50,71
References	7486
Content of the documents	
Plus Keywords (ID)	423
Author Keywords (DE)	592
Authors	
Authors	197
Authors of single-authored documents	1
Collaboration of authors	
Single-authored documents	112
Co-authors per paper	2,27
International co-authorships %	57,78
Types of documents	
articles	314
articles; book chapters	9
article; conference proceedings	5
book review	6
proofreading	2
editorial material	23
letter	28
meeting summary	2
news	1
note	2
minutes	28
retraction	1
summary	3

Author keywords are keywords defined by authors to determine the content of their publications, while keywords plus are keywords generated by the *WoS* database from titles, keywords and abstracts of publications. Our dataset covers a number of single-authored publications of 112, and the rest were co-authored with a total of 197 co-authors. On average, each article is written by about 3 authors (i.e. authors per paper is 3.48). The collaboration rate is around 2.27 for the total set.

Figure 1 visually summarises the data presented in table 1.

Figure 1. Summary of the main information on the dataset

3. Results

This section presents and interprets the results of the bibliometric analysis based on the research questions of our study.

3.1. Evolution of publications and citations, sources, countries and most relevant publications.

Presented in this subsection are the results that answer the question: What is the evolution of Loet Leydesdorff's publications and citations on the research topics of his publications (Table 2 and Figure 2).

Table 2. Distribution of publications and citations

Year	Nº of publications	Total cites	Year	Nº of publications	Total cites	Year	Nº of publications	Total cites
1980	1	2	1996	4	124	2010	21	1.513
1981	1	3	1997	5	182	2011	26	1.323
1982	1	2	1998	5	504	2012	28	1.498
1984	1	2	1999	2	38	2013	30	1.048
1986	1	65	2000	8	3.843	2014	27	1.233
1987	5	143	2001	5	43	2015	27	978
1988	2	1	2002	4	79	2016	22	548
1989	5	255	2003	7	293	2017	19	462
1990	4	115	2004	5	220	2018	12	201
1991	4	92	2005	11	1.097	2019	21	260
1992	5	40	2006	15	1.443	2020	11	65
1993	3	126	2007	15	691	2021	9	64
1994	7	155	2008	15	1.042	2022	4	40
1995	2	12	2009	21	1.463	2023	3	4

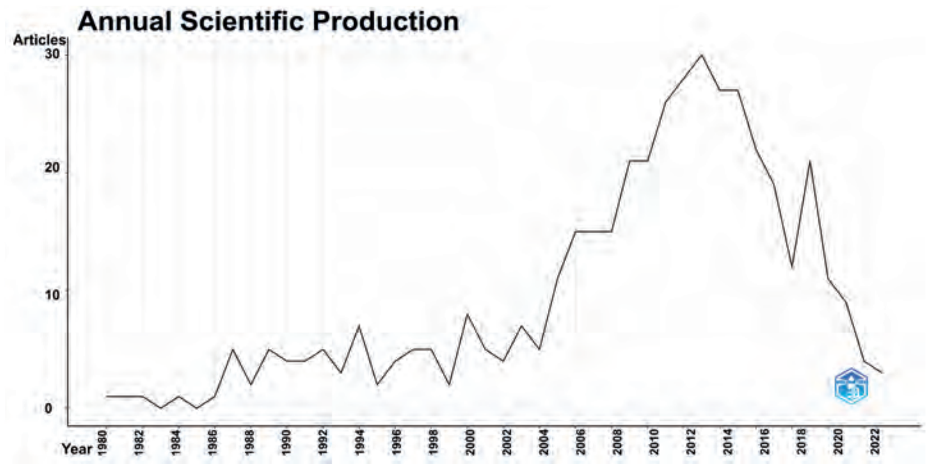


Figure 2. Annual scientific production

Table 2 shows the distribution of publications and citations for the 43-year period between 1980 and 2023. The highest number of publications is between 2006 and 2019, representing 66% (280) of the total number of publications. In the years 2020 to 2023, there was a decrease in the number of publications compared to the previous years.

Below are presented the results to answer the question: What are Leydesdorff's most relevant sources, countries and publications? Table 3 and Figure 3 present the 10 scientific journals where he published the most. These journals cover 64.1% of the total number of publications in our dataset. The top three journals that cover articles in Leydesdorff's research areas are *Scientometrics*, *Journal of the American Society for Information Science and Technology* and *Journal of informetrics*.

Table 3. Top 10 scientific journals where Loet published more

Sources	Articles
Scientometrics	93
Journal of the American Society for Information Science and Technology	55
Journal of informetrics	45
Journal of the Association for Information Science and Technology	28
Research policy	15
Profesional de la información	10
Social science information sur les sciences sociales	8
Technological forecasting and social change	7
Systems research and behavioral science	6
17th International conference on scientometrics & informetrics (ISSI2019), vol. 1	5

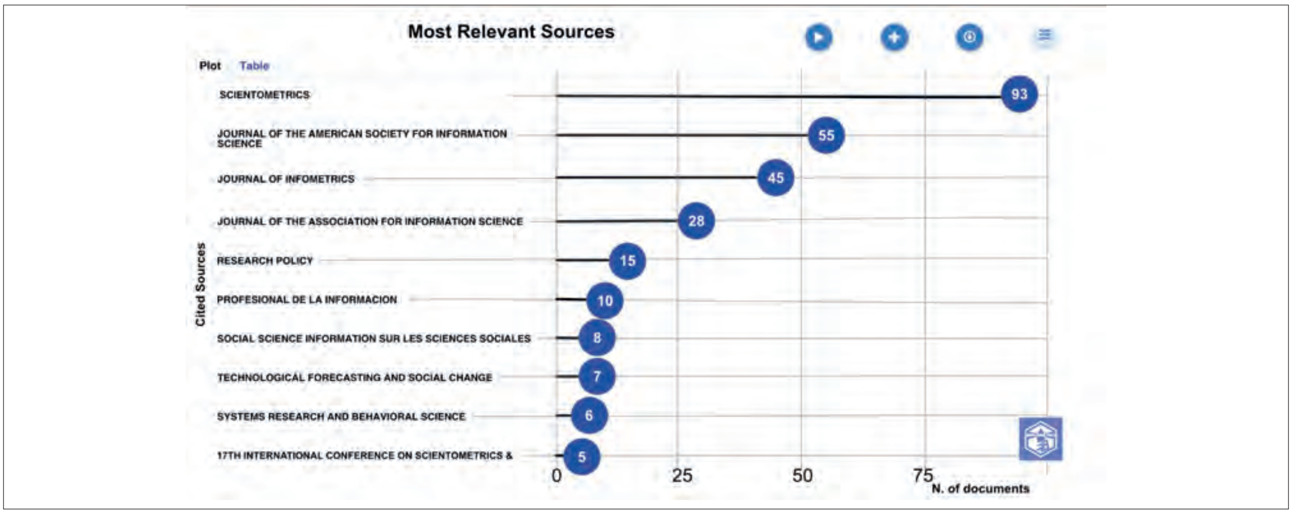


Figure 3. Top 10 scientific journals where Loet published more

Table 4 shows the top 10 most cited journals in the field. The most cited journals in the field measure the number of citations in the field received per cited reference within the reference lists of the publications in the dataset. From these results, we can see that *Scientometrics* is the most cited source among researchers. This source has been cited 1886 times. The second most cited source is *Journal of the American Society for Information Science and Technology* (1243 times), followed by *Research Policy* (887 times). This shows that these journals are the main references for publications by this author and his co-authors.

Table 4. The most cited journals in the field

Sources	Articles
Scientometrics	1886
Journal of the American Society for Information Science and Technology	1558
Research policy	887
Journal of informetrics	661
Science	236
Social networks	184
Journal of the Association for Information Science and Technology (Jasist)	173
Nature	160
Social studies of science	160
Social science information	141

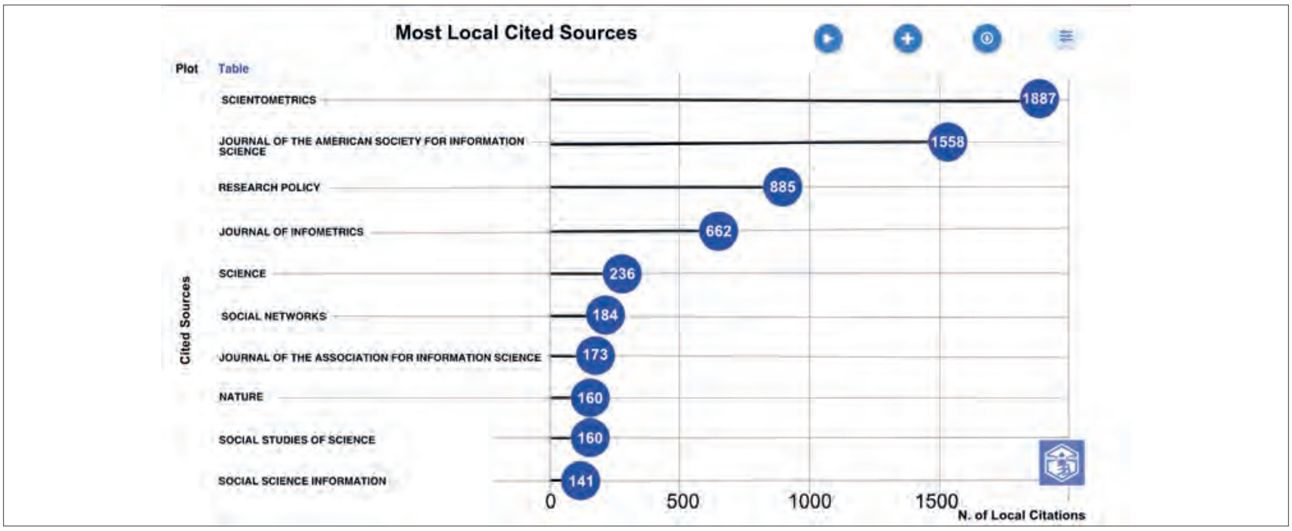


Figure 4. The most cited journals in the field.

Table 5 presents the top 10 most influential publications in terms of total citations. The total number of citations of a journal is the number of citations received by the articles published in that journal in the dataset. From these results it can be seen that the top three journals, *Journal of the American Society for Information Science and Technology*, *Scientometrics* and *Journal of informetrics*, have published many articles that have received a high number of total citations with a high h-index.

However, there are also magazines, such as *Journal of the Association for Information Science and Technology* and *Research Policy et Social Science*, which have a high number of citations with a limited number of articles published in Leydesdorff's publications.

Table 5. Top 10 most influential journals by total citations

Journals	h index	g index	m_index	Total citations	Number of publications	Start year
Journal of the American Society for Information Science and Technology	36	55	1,565	4388	55	2001
Scientometrics	34	56	0,791	3491	93	1981
Journal of informetrics	27	45	1,688	2.41	45	2008
Journal of the Association for Information Science and Technology	18	28	1,8	1102	28	2014
Research policy	14	15	0,35	5937	15	1984
Social science information sur les sciences sociales	7	8	0,219	101	8	1992
Technological forecasting and social change	7	7	0,368	254	7	2005
Profesional de la informacion	6	10	0,375	172	10	2008
Plos one	5	5	0,357	276	5	2010
Journal of data and information science	4	4	0,571	55	4	2017

Table 6 shows the data for the 20 countries where this author has published the most; SCP stands for *Single Country Publication* (refers to scientific publications originating from a single country. That is, all the main authors or affiliations are located in a single country); MCP is *Multiple Country Publication* (refers to scientific publications involving collaboration of authors or affiliations from several countries) *MCP ratio* (this is the measure that indicates the proportion of the total number of publications that are collaborative between authors or affiliations from different countries.

These results show that between the Netherlands and Germany, 299 articles have been published and 98 of them involved international collaborations. The third country with the highest number of publications is the United States, with 23 international collaborations.

However, some observations can be made about the proportion of CCMs. Countries such as Germany, China and the UK have a higher degree of international collaboration than other countries.

Table 6. The 20 most frequent countries of origin of authors and co-authors of publications

Country	Articles	SCP	MCP	Freq	MCP_Ratio
Netherlands	258	160	98	0,608	0,38
Germany	41	0	41	0,097	1,00
USA	25	2	23	0,059	0,92
China	22	0	22	0,052	1,00
United Kingdom	17	0	17	0,040	1,00
Russia	8	0	8	0,019	1,00
Korea	7	0	7	0,017	1,00
Switzerland	6	1	5	0,014	0,83
Belgium	4	0	4	0,009	1,00
Italy	3	0	3	0,007	1,00
Colombia	2	0	2	0,005	1,00
France	2	1	1	0,005	0,50
Hungary	2	0	2	0,005	1,00
India	2	0	2	0,005	1,00
Norway	2	0	2	0,005	1,00
Australia	1	0	1	0,002	1,00
Cuba	1	0	1	0,002	1,00
Finland	1	0	1	0,002	1,00
Ireland	1	0	1	0,002	1,00

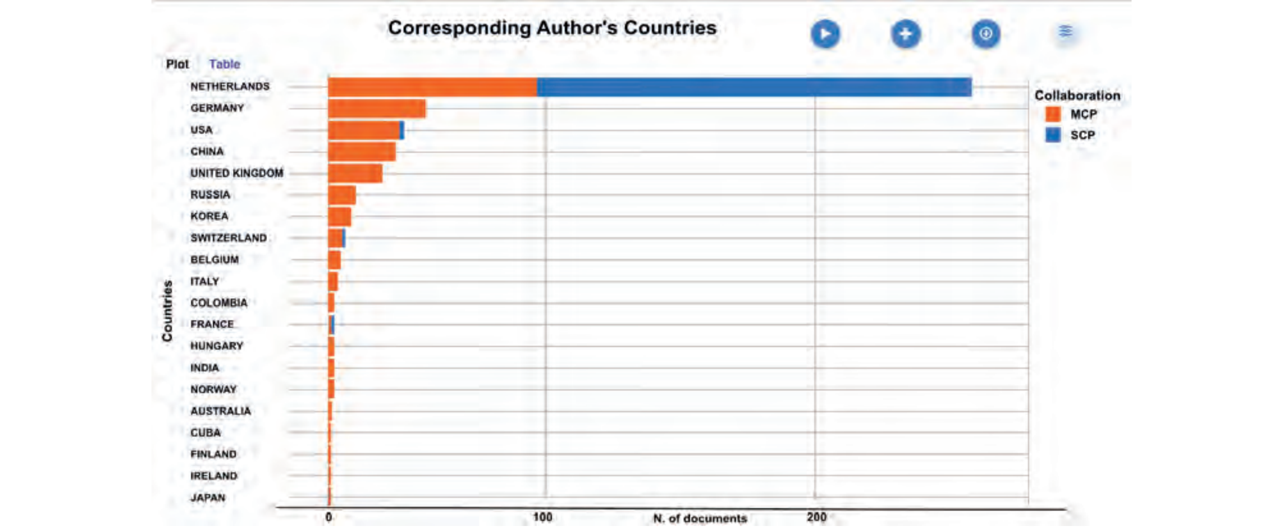


Figure 5. Top 20 countries of origin of authors and co-authors of publications

Finally, Table 7 presents the ranking of Leydesdorff's most cited articles.

Table 7. Ranking of most cited publications

Lead author – *Publication*	Title	Total citations
Etzkowitz H, 2000, *Res policy*	The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university-industry-government relations *10.1016/S0048-7333(99)00055-4*	3,491
Wagner CS, 2005, *Res policy*	Network structure, self-organization, and the growth of international collaboration in science. *10.1016/j.respol.2005.08.002*	614
Mingers J, 2015, *Eur J Oper Res*	A review of theory and practice in scientometrics. *10.1016/j.ejor.2015.04.002*	426
Leydesdorff L, 2009, *J AM SOC INF SCI TEC*-a-b-c	A global map of science based on the ISI subject categories. *10.1002/asi.20967*	393
Zhou P, 2006, *Res policy*	The emergence of China as a leading nation in science. *10.1016/j.respol.2005.08.006*	385
Leydesdorff L, 2007, *J AM SOC INF SCI TEC*	Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *10.1002/asi.20614*	347
Ràfols I, 2012, *Res policy*	How journal rankings can suppress interdisciplinary research: A comparison between Innovation Studies and Business & Management. *10.1016/j.respol.2012.03.015*	335
Leydesdorff L, 2012, *J Knowl Econ*	The Triple Helix, Quadruple Helix, …, and an N-Tuple of Helices: Explanatory models for analyzing the knowledge-based economy? *10.1007/s13132-011-0049-4*	307
Ràfols I, 2010, *J AM SOC INF SCI TEC*	Science overlay maps: A new tool for research policy and library management	290
Leydesdorff L, 2006, *J AM SOC INF SCI TEC*	Co-occurrence matrices and their applications in information science: Extending ACA to the web environment	270

As can be seen in the results of Table 7, the article "*The dynamics of innovation: from National Systems and "Mode 2" to a Triple Helix of university-industry-government relations*" (**Etzkowitz**; **Leydesdorff**, 2000) has received the highest number of citations, followed by the article "*Network structure, self-organization, and the growth of international collaboration in science*" (**Wagner**; **Leydesdorff**, 2005) y "*A review of theory and practice in scientometrics*" (**Mingers**; **Leydesdorff**, 2015). Based on the content analysis of the best research articles, three themes are identified: 3 of the 10 articles focused on the Triple Helix, 3 of the 10 focused on scientometrics, and 1 of the 10 focused on a global map of science. (**Leydesdorff**; **Ràfols**, 2009).

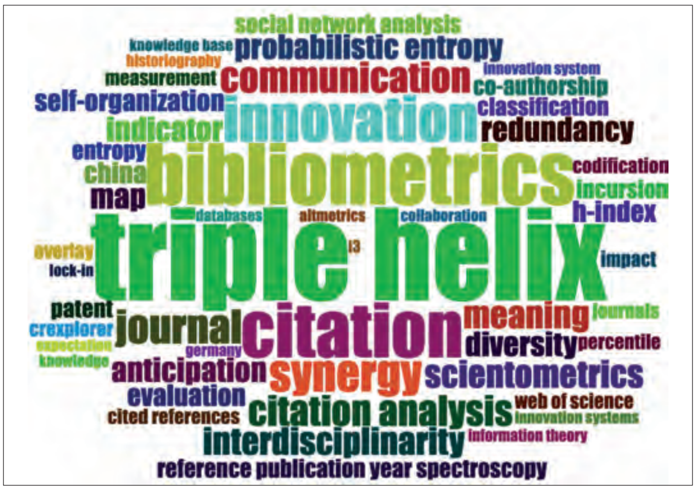


Figure 6. Word cloud based on author keywords

Figure 7. Keywords plus treemap.

3.2. Research topics and keyword trends

In this section we present the results to answer the question: what are the most common research topics and keyword trends in Leydesdorff publications? We present a thematic analysis to detect the main research topics in the field using a word cloud and a word treemap.

Figure 6 shows the word cloud for the 50 most common author keywords in the collection of publications. The size of the keyword in the figure indicates the frequency of the keyword in the dataset. As can be seen in the figure, the most frequently used words determine the content of most of the studies in the collection. More specifically, the frequent keyword "triple helix" is the main theme, as the papers in the collection address various aspects of the triple helix of knowledge production, university-industry-government relations and perspectives on innovation systems. The keywords "bibliometrics", "citation analysis", "scientometrics", "social network analysis", "cited references" and "evaluation" show

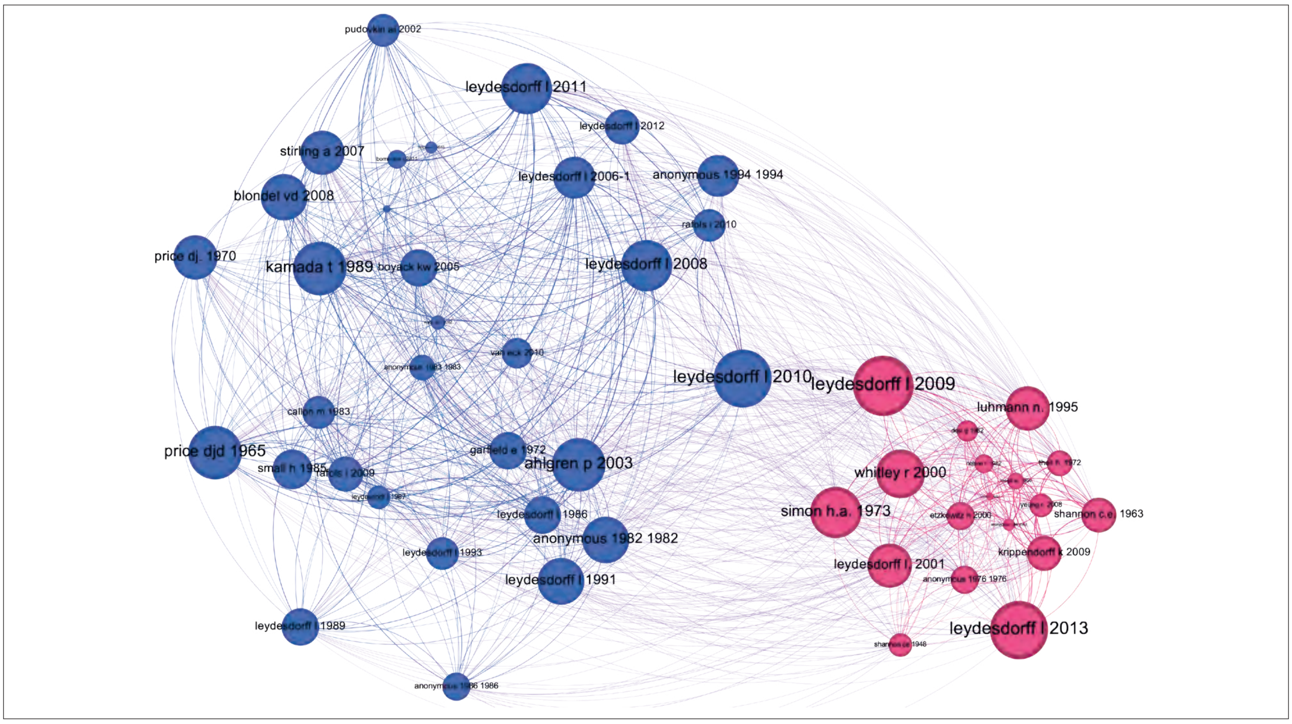


Figure 8. Leydesdorff's publication co-citation network

Figure 9. Map of collaboration between authors from different countries

their importance and represent the main methodologies according to their conceptual meaning in Leydesdorff's publications and research.

The 50 most commonly used keywords in articles are presented in the form of a Treemap of words. The Treemap in Figure 7 highlights the combination of plus keywords, indicating triple helix and bibliometrics. "Triple helix" is the most used keyword, while "lock-in" is the least used. Focusing on the keywords, other research areas that were of interest are science, indicators, network maps, innovation and interdisciplinarity.

3.3. Co-citation network

In this subsection, results are presented to answer the question: What are the main groups of co-citations related to Leydesdorff's publications?

Figure 8 shows the co-citation network, a type of network in which nodes represent scientific papers and links between nodes indicate that these papers have been cited together in the same reference work. This network was conducted with a minimum degree of co-citation equal to three and a threshold of 50 network nodes. The nodes were labelled with the first author and the year of publication of the article, while the network link is the co-citation between two documents. The node size indicates the number of citations received by the papers and the link thickness represents the strength of the co-citation links. The colour of the node shows the cluster with which the article is associated, in our case two: one related to bibliometrics and the other to scientific innovation.

3.4. Networking between countries

In this subsection, results are presented to answer the question: what is the network of collaboration between authors from different countries in Leydesdorff's research?

Figure 9 shows such a map of international collaboration. It depicts the publication output of authors from each country and the collaboration between authors from different countries. Countries with a darker colour indicate more publications than countries with a lighter colour, while the thickness of the lines represents stronger collaborations between countries.

Figure 10 shows the collaborative social network at country level in detail. The node in the network represents the country and the link between two nodes represents the cooperation between countries. The size of the country indicates the degree of cooperation, and the

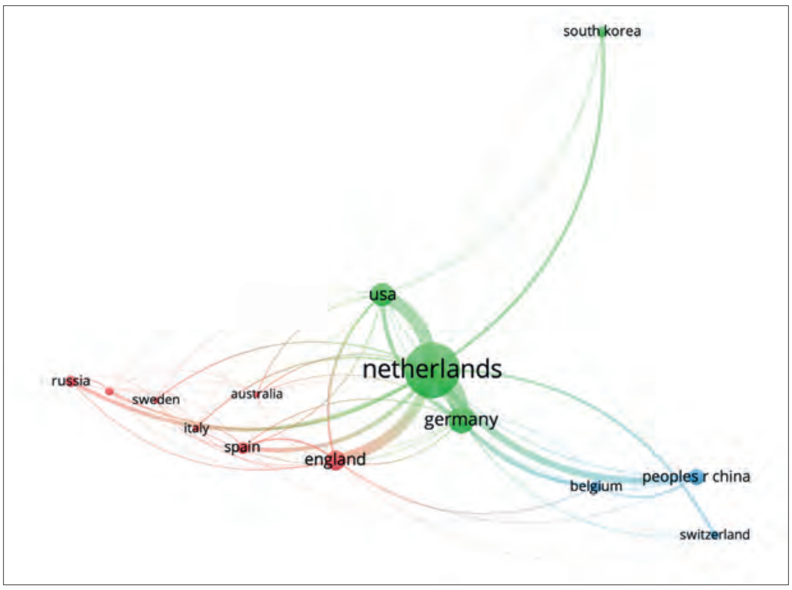


Figure 10. Cross-country collaboration network (graph generated with *VOSwiever*)

thickness of the link indicates the closeness of the collaboration between countries. It is worth highlighting the collaboration with authors from our country, consistent with the high number of publications by this author in the journal *Profesional de la información*.

4. Mapping scientific networks

The concept of "scientific network mapping" refers to the process of visualising and analysing the relationships and connections between different elements within the scientific domain, such as researchers, institutions, or subject areas. This practice uses network analysis tools to graphically represent the structure and dynamics of interactions in the scientific community.

In the context of scientific research, network mapping can reveal patterns of collaboration, identify centres of influence, and provide information on the interconnectedness between disciplines or areas of study. Methods employed include social network analysis, where nodes represent entities such as researchers or institutions, and links between nodes represent collaborative relationships, citations, or any other form of relevant interaction.

This approach helps to understand the structure and evolution of scientific communities, facilitating the identification of key research areas, the assessment of the impact of researchers and institutions, and the visualisation of interdisciplinarity in scientific production. Scientific network mapping is valuable for analysing and communicating the complexity of interactions in academia.

4.1. Leydesdorff publications on scientific network mapping

Loet Leydesdorff is recognised for his significant contributions to the mapping of scientific networks (29 publications, see Table 8). He has developed innovative methodologies and tools to analyse and visualise the structure of scientific collaborations at different scales. Here are some key points about his work in this field:

1. *Intermediation centrality*: In his 2007 article, Leydesdorff proposes the use of "betweenness centrality" as an indicator of the interdisciplinarity of scientific journals. Betweenness centrality measures the frequency with which a node is on the shortest path between two other nodes in a network. He applied this measure to assess how scientific journals can serve as bridges between different disciplines, providing a means to map interdisciplinary connections (**Leydesdorff**, 2007a).

2. *Social network analysis*: Leydesdorff used social network analysis methods to explore collaborations between researchers and institutions. These analyses can visually map connections and relationships between actors in the scientific domain, revealing network structure and key points of collaboration.

3. *Bibliometric indicators*: In addition to mapping collaborations, Leydesdorff has developed bibliometric indicators to assess the impact and visibility of researchers and institutions. These indicators can be used to understand the distribution of collaborations and the influence of publications within the scientific network.

4. *Network visualisation*: In his 2009 paper, Leydesdorff presented visualisation methods for graphically representing scientific networks. These visualisations offer an intuitive insight into the distribution of collaborations and the overall structure of the scientific research network. (**Leydesdorff**; **Ràfols**, 2009).

Table 8. Leydesdorff's most relevant publications related to "mapping scientific networks"

Thematic area	Subject - Cite
Patent portfolio analysis of cities: statistics and maps of technological inventiveness	Cities and knowledge-based economy (**Kogler**; **Heimeriks**; **Leydesdorff**, 2018)
Betweenness and diversity in journal citation networks as measures of interdisciplinarity - A tribute to Eugene Garfield	Interdisciplinary journal ranking (**Leydesdorff**; **Wagner**; **Bornmann**, 2018)
Mapping patent classifications: portfolio and statistical analysis, and the comparison of strengths and weaknesses	Cooperative patent classifications (CPC) (**Leydesdorff**; **Kogler**; **Yan**, 2017)
Journal portfolio analysis for countries, cities, and organizations: Maps and comparisons	*Web of Science* data for portfolio analysis (**Leydesdorff**; **Heimeriks**; **Rotolo**, 2016)
Journal maps, interactive overlays, and the measurement of interdisciplinarity on the basis of *Scopus* data (1996-2012)	Global map of science using *Scopus* (**Leydesdorff**; **De-Moya-Anegón**; **Guerrero-Bote**, 2015)
International collaboration in science: the global map and the network	Global network of international co-authorship (**Leydesdorff**; **Wagner**; **Park**; **Adams**, 2013)
Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks	Preferential attachment in coauthorship networks (**Abbasi**; **Hossain**; **Leydesdorff**, 2012)
Mapping (*USPTO*) patent data using overlays to *Google Maps*	Patent-based *Google Maps* (**Leydesdorff**; **Bornmann**, 2012)
Mapping excellence in the geography of science: An approach based on *Scopus* data	Mapping centers of excellence worldwide (**Bornmann**; **Leydesdorff**; **Walch-Solimena**; **Ettl**, 2011)
'Meaning' as a sociological concept: A review of the modeling, mapping and simulation of the communication of knowledge and meaning	Discursive knowledge and communication (**Leydesdorff**, 2011)

Thematic area	Subject - Cite
The semantic mapping of words and co-words in contexts	Measuring semantics using latent semantic analysis (**Leydesdorff**; **Welbers**, 2011)
Science overlay maps: A new tool for research policy and library management	Science overlay maps for benchmarking (**Rafols**; **Porter**; **Leydesdorff**, 2010)
Mapping the geography of science: Distribution patterns and networks of relations among cities and institutes	Overlaying scientific networks on geographic maps (**Leydesdorff**; **Persson**, 2010)
Maps on the basis of the *Arts & Humanities Citation Index*: The journals *Leonardo* and *Art journal* versus "digital humanities" as a topic	Mapping *Arts & Humanities Citation Index* (*A&HCI*) (**Leydesdorff**; **Salah**, 2010)
Journal maps on the basis of *Scopus* data: A comparison with the *Journal Citation Reports* of the *ISI*	Comparing *Scopus* and *Journal Citation Reports* (**Leydesdorff**; **De-Moya-Anegón**; **Guerrero-Bote**, 2010)
Knowledge linkage structures in communication studies using citation analysis among communication journals	Mapping communication studies (**Park**; **Leydesdorff**, 2009)
Dynamic animations of journal maps: Indicators of structural changes and interdisciplinary developments	Dynamic analysis of structural change in sciences (**Leydesdorff**; **Schank**, 2008)
Korean journals in the *Science Citation Index*: What do they reveal about the intellectual structure of S&T in Korea?	South Korea's research output (**Park**; **Leydesdorff**, 2008)
Betweenness centrality as an indicator of the interdisciplinarity of scientific journals	Centrality measures in journal citation networks (**Leydesdorff**, 2007a)
Mapping interdisciplinarity at the interfaces between the science citation index and the social science citation index	Combining journal citation reports (**Leydesdorff**, 2007)
Clustering methodologies for identifying country core competencies	Mexican science and technology literature (**Kostoff**; **Del-Río**; **Cortés**; **Smith**; **Smith**; **Wagner**; **Leydesdorff**; **Karypis**; **Malpohl**; **Tshiteya**, 2007)
Mapping the *Chinese Science Citation Database* in terms of aggregated journal-journal citation relations	Mapping *Chinese Science Citation Database* (**Leydesdorff**; **Jin**, 2005)
Mapping the *Chinese Science Citation Database*	Mapping *Chinese Science Citation Database* (alternative abstract) (**Leydesdorff**; **Bihui**, 2004)
Clusters and maps of science journals based on bi-connected graphs in *Journal Citation Reports*	Decomposing journal-journal citation matrix (**Leydesdorff**, 2004)
Why words and co-words cannot map the development of the sciences	Word co-occurrence analysis in biochemistry (**Leydesdorff**, 1997)
Mapping change in scientific specialties: A scientometric reconstruction of the development of artificial intelligence	Emergence of artificial intelligence as a discipline (**Van-den-Besselaar**; **Leydesdorff**, 1996)
Tracking areas of strategic importance using scientometric journal mappings	Indicators for tracking emerging developments (**Leydesdorff**; **Cozzens**; **Van-den-Besselaar**, 1994)
Various methods for the mapping of science	(**Leydesdorff**, 1987)

4.2. The *Cosma* software

Cosma[1] (*https://cosma.arthurperret.fr*) was developed as part of the *ANR HyperOtlet* programme, which aimed to represent Paul Otlet's social network in the form of an interactive graph, known as *Otletosphere* (*https://hyperotlet.huma-num.fr/otletosphere*). *Cosma* was developed by Guillaume Brioudes, Olivier Le Deuff and other collaborators in 2021.

It is a free experimental research tool, published under a free licence, which offers an innovative way to visualise and explore documentary networks.

Main features of *Cosma*:

1. Graph visualisation: allows the visualisation of an interactive documentary graph. The nodes of the network represent countries and the links between the nodes represent the cooperation between the authors of these countries.

2. Navigation functionalities: *Cosma*'s interface is divided into three zones. A panel on the left contains navigation functions such as search, index and display filters. On the right, a panel displays the selected record with a bibliography automatically generated from the sources cited in the text of the record.

3. Data export: Unlike most visualisation tools, *Cosma* inverts the usual logic. The application part, called cosmograph, is a simple creation form. The created export, an HTML file called cosmoscope, constitutes the actual visualisation interface. This HTML file can be exported, used and shared independently.

4. Discussion support: *Cosmoscope* files can be shared, making them a support for discussion in the context of a research or teaching assignment.

5. Recognition of categorisation: *Cosma* recognises, from a previous analysis, the categorisation of the cards and associates graphic codes (colours, layouts) and interactions (filtering of displayed elements) to them.

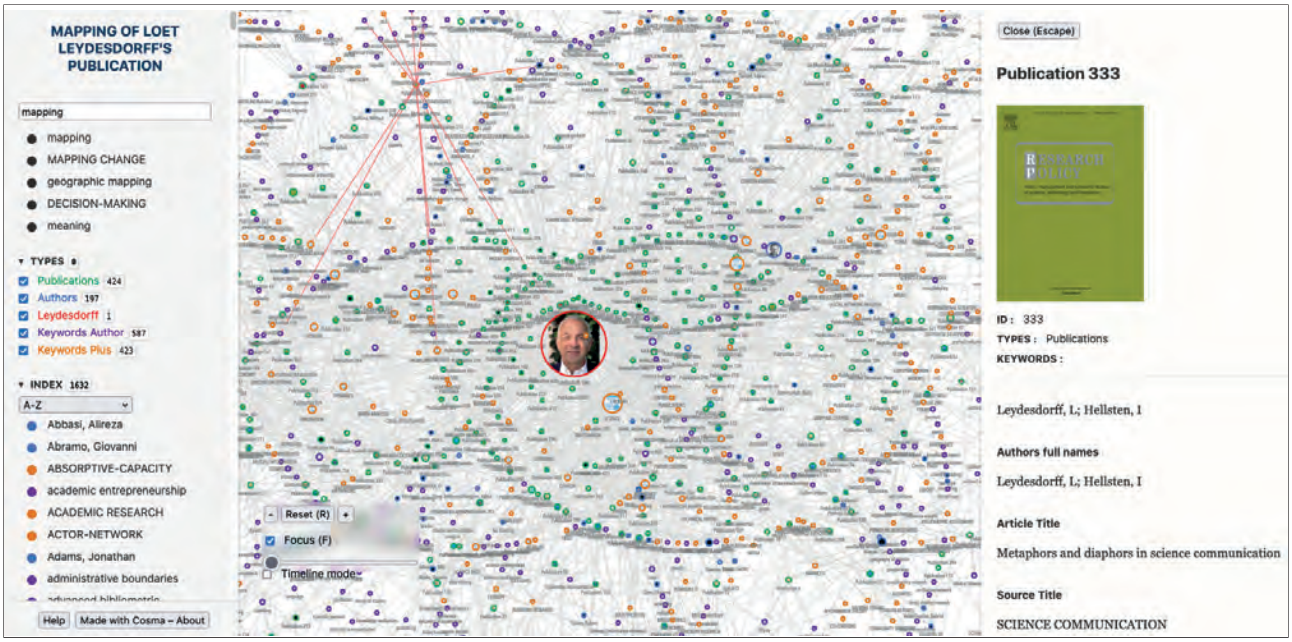


Figure 11. *Cosma* interface
https://cosma.arthurperret.fr

4.3. Stages of analysis of Loet Leydesdorff's mapping of scientific networks

Scientific mapping attempts to find representations of the intellectual connections within the dynamically changing system of scientific knowledge (**Small**, 1997). In other words, scientific mapping shows the structural and dynamic aspects of scientific research (**Cobo** *et al.*, 2011). In this paper we have developed a mapping of the publications of L. Leydesdorff who, throughout his career, has been one of the pioneers of this type of research.

The scientific mapping we have produced is a visual representation of how disciplines, fields, documents, co-authors, author words and index words (*keywords plus*) relate to each other in Leydesdorff publications.

In this section we describe the different stages of analysis of this mapping:

(a) *Source of data*: For this mapping, publication data were retrieved from the *Web of Science* (*WoS*) *Core Collection* database. Our dataset contains 424 articles published between 1980 and 2023.

(b) *The units of analysis*: From the 424 documents we selected the references of the publications, the co-authors, the title, the abstract. In addition, we selected the original keywords of the documents (author keywords) and the indexing keywords provided by the database (*ISI Keywords Plus*) as words to be analysed. We also took the *WoS* subject categories, the Orcid of the co-authors and the DOI of the publication.

(c) *Data pre-processing*: A scientific mapping analysis cannot be applied directly to data retrieved from bibliographic sources, i.e. a pre-processing of the retrieved data is necessary. Thus, the pre-processing of the data has been as follows:

- The correction and addition of the names of certain co-authors, and the inclusion of their photo in the database.
- Detection of duplicate elements and misspellings of co-authors and author words and index words.
- The division of the articles into different time sub-periods, in order to analyse the evolution of Leydesdorff's research (see table).

(d) *Standardisation process*: The *Cosma* software uses a specific visualisation algorithm to graphically represent the relationships between documentary elements. This algorithm is called a "*cosmograph*".

(e) *Methods of analysis:* Before applying the visualisation, *Cosma* performs a preliminary analysis to categorise the document files. This analysis determines the assignment of graphic codes such as colours and layouts.

(f) *Display algorithm*: force-layout algorithm, which *Cosma* incorporates.

As a final result, our Leydesdorff cosmoscope is available for use at this URL:
http://metroteach.com/Leydesdorff/index.html

It can be downloaded by clicking on the link Made with *Cosma* - About at the bottom left of the interface.
http://metroteach.com/Leydesdorff/cosmoscopio.html

The interface allows switching to a timeline visualisation (Table 9) to show the changes that occur with this variable (by clicking on the Timeline mode checkbox).

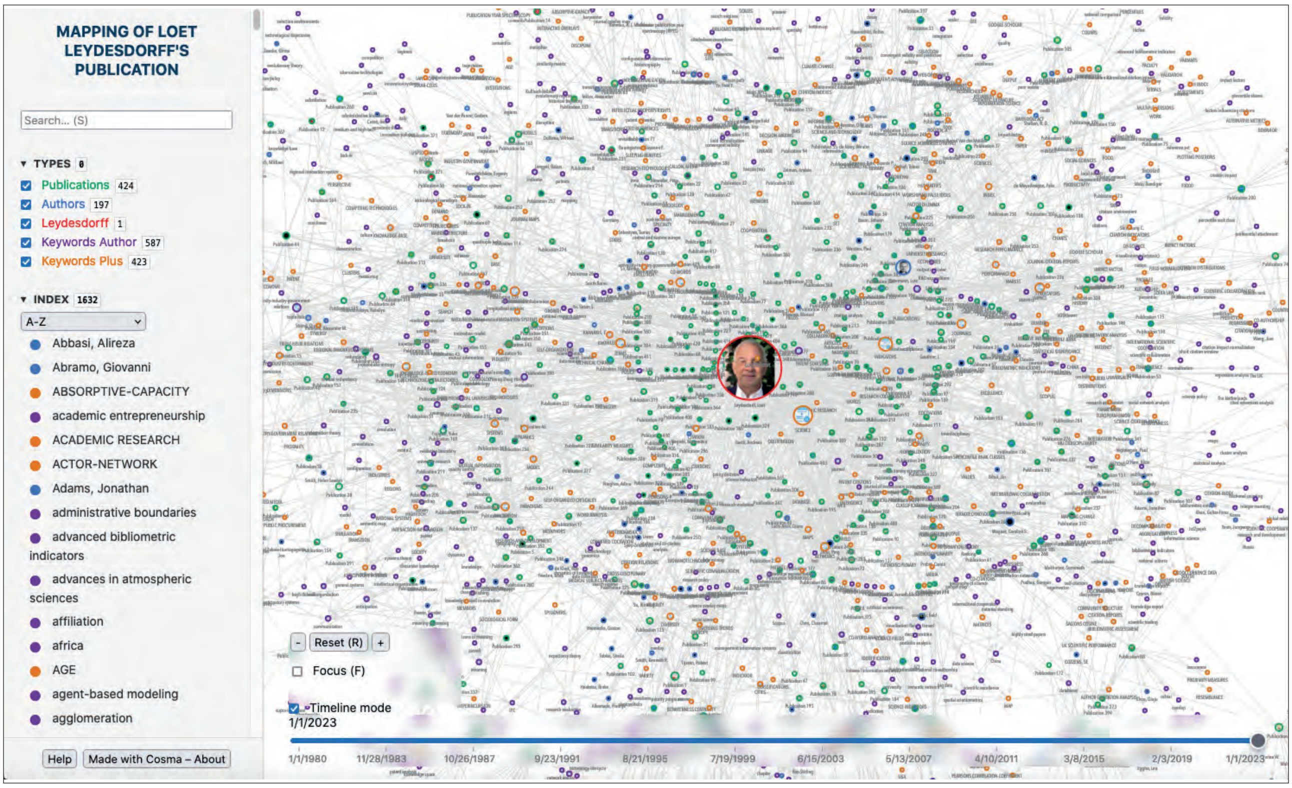


Figure 12. Scientific network mapping of Leydesdorff's publications.
http://metroteach.com/Leydesdorff/index.html

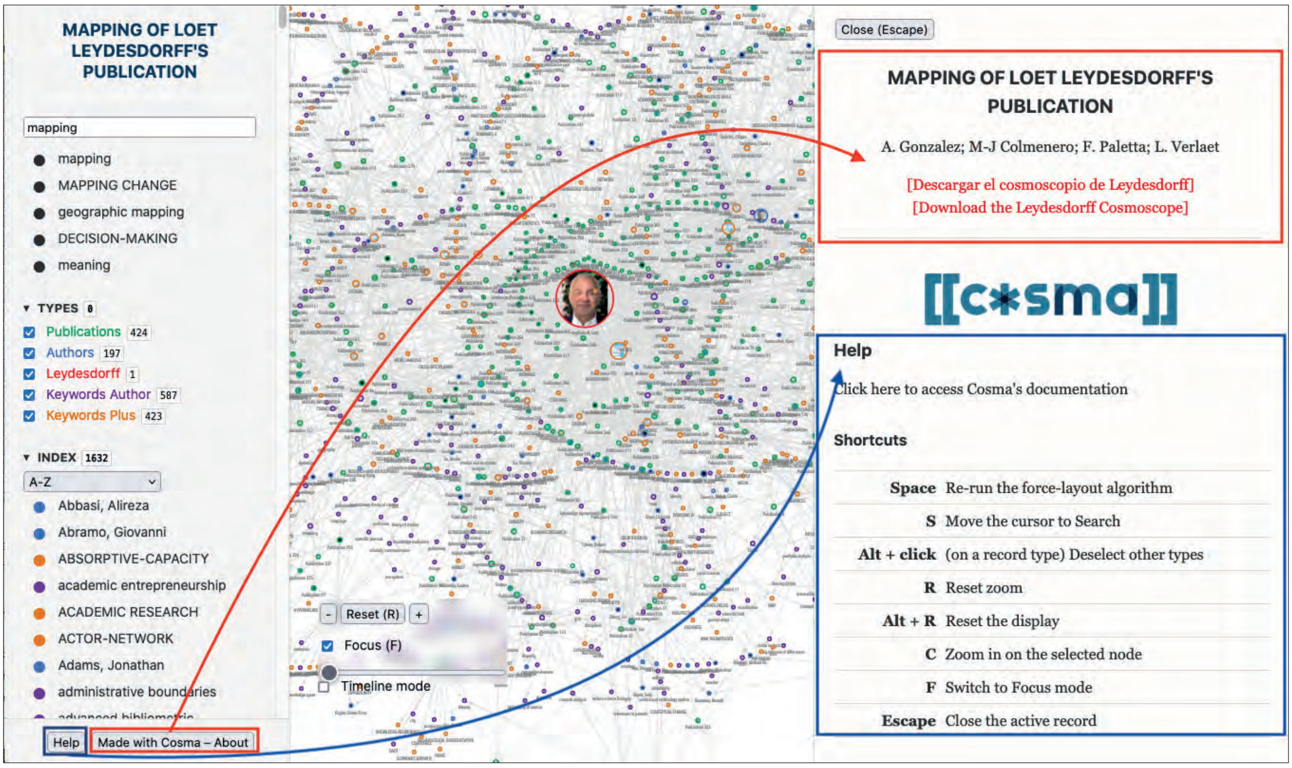


Figure 13. Discharge line of the Leydesdorff cosmoscope.
http://metroteach.com/Leydesdorff/cosmoscopio.html

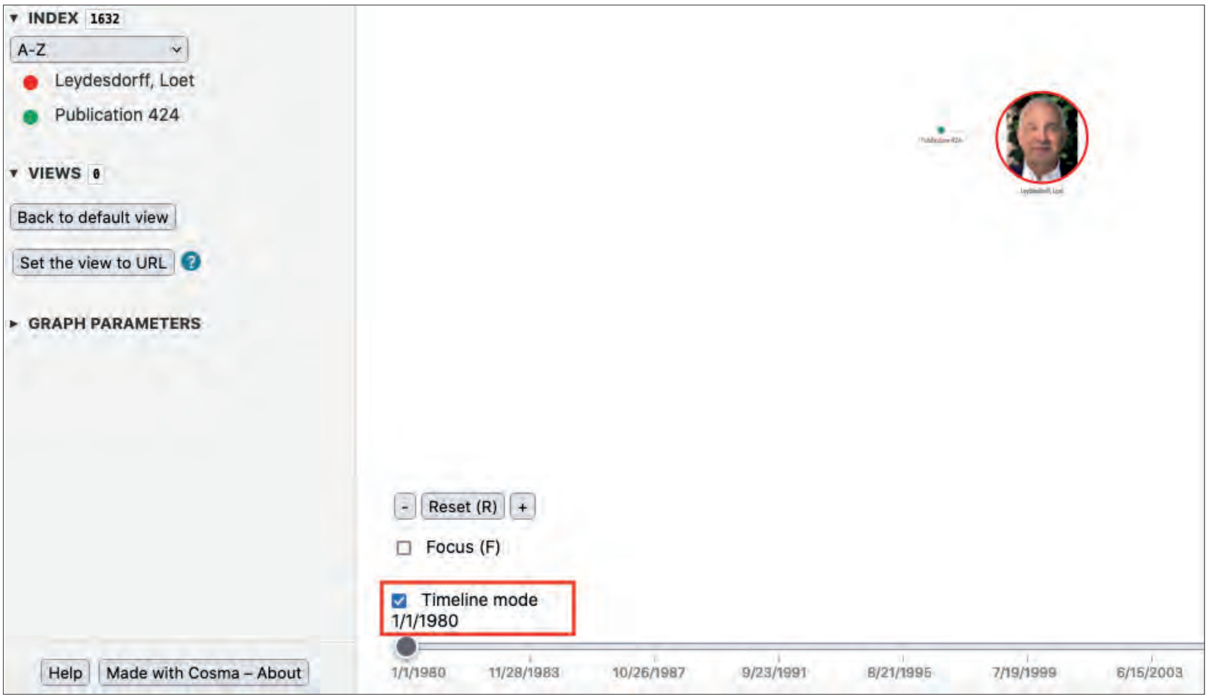
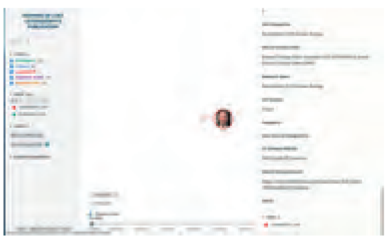
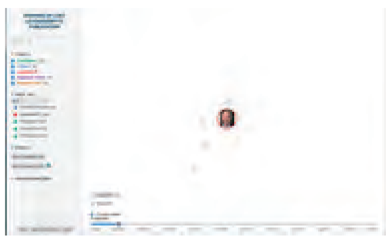
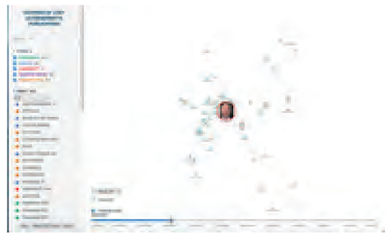
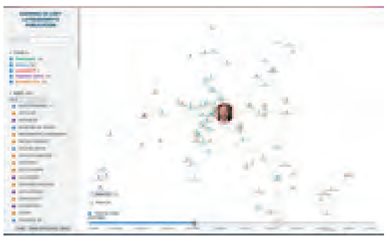
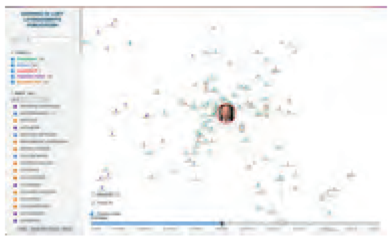
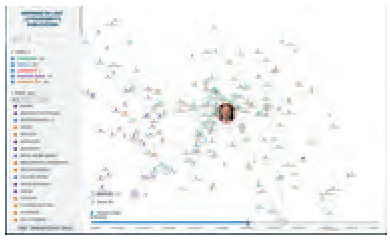
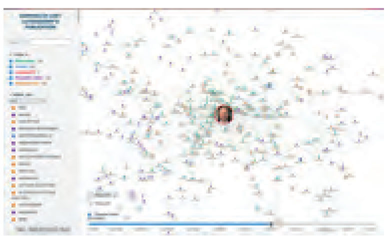
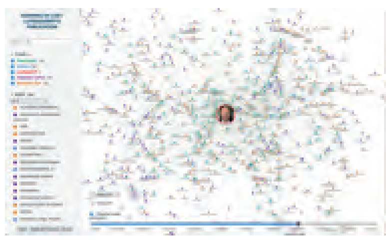
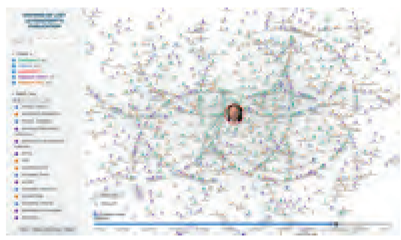
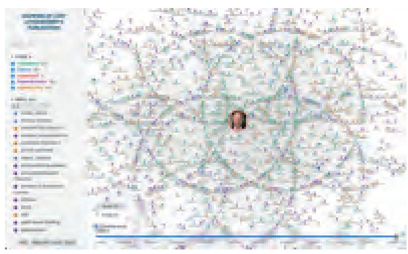
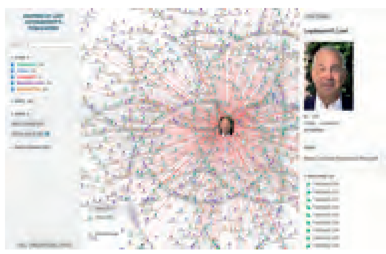
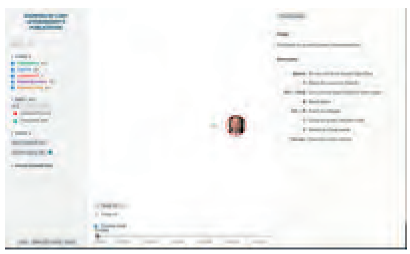


Figure 14. Leydesdorff cosmoscope "timeline" option

Table 9. Timeline of publications

		
1980	1983	1987
		
1995	1999	2003
		
2007	2011	2015
		
2019	2023	Ayuda de *Cosma*

5. Conclusions

A bibliometric analysis of Leydesdorff's publications collected in *WoS* has been carried out and a scientific mapping of his papers, including co-authors, their countries of origin and keywords (author and indexing), has also been produced.

The literature review on Loet Leydesdorff highlights his significant contributions to the sociology of science, bibliometric analysis, science communication and scientific network mapping. His work continues to influence and guide research in these areas, demonstrating the breadth of his intellectual impact.

For the mapping of scientific output we have included different methods of analysis (although some of them are common), which allowed us to discover the different facets of the knowledge investigated by Leydesdorff. Since the visualisations are different in each of them, different views of the field can be generated to help interpret and analyse the results. This cooperation between tools leads to a positive synergy, which allows extracting the knowledge hidden behind the data.

Our work has been intended as a small tribute to this recently deceased scientist, who has contributed to the advancement of several areas of study. We believe that a more complete scientific mapping analysis of Leydesdorff's work should be carried out, including the publications that *WoS* does not include, in order to compile all the knowledge, important theoretical contributions and the different perspectives (intellectual, social or conceptual) of the author's different areas of research.

6. Note

1. The name *Cosma* comes from Cosmas Rosellius, a Florentine Dominican monk, author of a *Thesaurus artificiosæ memoriæ* (1579), which translates as "treasure of artificial memory". A feature of Rossellius' book is the mnemonic verses given to help memorize orders of places, whether orders of Hell, or the order of the signs of the zodiac.
https://archive.org/download/thesaurusartifi00padogoog/thesaurusartifi00padogoog.pdf

7. References

Abbasi, Aliresa; **Hossain, Liaquad**; **Leydesdorff, Loet** (2012). "Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks". *Journal of informetrics*, v. 6, n. 3.
https://doi.org/10.1016/j.joi.2012.01.002

Aria, Massimo; **Cuccurullo, Corrado** (2017). "Bibliometrix: An R-tool for comprehensive science mapping analysis". *Journal of informetrics*, v. 11, n. 4.
https://doi.org/10.1016/j.joi.2017.08.007

Bornmann, Lutz; **Leydesdorff, Loert**; **Walch-Solimena, Christiane**; **Ettl, Christoph** (2011). "Mapping excellence in the geography of science: An approach based on Scopus data". *Journal of informetrics*, v. 5, n. 4.
https://doi.org/10.1016/j.joi.2011.05.005

Cobo, Manuel J.; **López-Herrera, Antonio-Gabriel; Herrera-Viedma, Enrique**; **Herrera, Francisco** (2011). "Science mapping software tools: Review, analysis, and cooperative study among tools". *J. Assoc. Inf. Sci. Technol.*, v. 62, pp. 1382-1402.

Donthu, Naveen; **Kumar, Satish**; **Mukherjee, Debmalya**; **Pandey, Nitesh**; **Lim, Weng-Marc** (2021). "How to conduct a bibliometric analysis: An overview and guidelines". *Journal of business research*, v. 133.
https://doi.org/10.1016/j.jbusres.2021.04.070

Etzkowitz, Henry; **Leydesdorff, Loet** (1996). "A triple helix of academic-industry-government relations: Development models beyond 'capitalism versus socialism'". *Current science*, v. 70, n. 8.

Etzkowitz, Henry; **Leydesdorff, Loet** (2000). "The dynamics of innovation: from National Systems and 'Mode 2' to a Triple Helix of university-industry-government relations". *Research policy*, v. 29, n. 2.
https://doi.org/10.1016/S0048-7333(99)00055-4

Kogler, Dieter-Franz; **Heimeriks, Gaston**; **Leydesdorff, Loet** (2018). "Patent portfolio analysis of cities: statistics and maps of technological inventiveness". *European planning studies*, v. 26, n. 11.
https://doi.org/10.1080/09654313.2018.1530147

Kostoff, Ronald N.; **Del-Río, J. Antonio**; **Cortés, Héctor D.**; **Smith, Charles**; **Smith, Andrew**; **Wagner, Caroline**; **Leydesdorff, Loet**; **Karypis, George**; **Malpohl, Guido**; **Tshiteya, Rene** (2007). "Clustering methodologies for identifying country core competencies". *Journal of information science*, v. 33, n. 1.
https://doi.org/10.1177/0165551506067124

Leydesdorff, Loet (1979). "*Dynamic and stochastic models for reciprocal citation processes*" (Thèse de doctorat, Université d'Amsterdam).

Leydesdorff, Loet (2007). "Betweenness centrality as an indicator of the interdisciplinarity of scientific journals". *Journal of the American Society for Information Science and Technology*, v. 58, n. 9.
https://doi.org/10.1002/asi.20614

Leydesdorff, Loet (2001). "*A sociological theory of communication; The self-organization of the knowledge-based Society*". Universal Publishers / upublish.com, USA.
https://www.universal-publishers.com/m/book.php?method=ISBN&book=1581126956

Leydesdorff, Loet (1987). "Various methods for the mapping of science". *Scientometrics*, v. 11, n. 5-6.
https://doi.org/10.1007/BF02279351

Leydesdorff, Loet (1997). "Why words and co-words cannot map the development of the sciences". *Journal of the American Society for Information Science*, v. 48, n. 5.

Leydesdorff, Loet (2004). "Clusters and maps of science journals based on bi-connected graphs in Journal Citation Reports". *Journal of documentation*, v. 60, n. 4.
https://doi.org/10.1108/00220410410548144

Leydesdorff, Loet (2007). "Mapping interdisciplinarity at the interfaces between the science citation index and the social science citation index". *Scientometrics*, v. 71, n. 3.
https://doi.org/10.1007/s11192-007-1694-z

Leydesdorff, Loet (2007). "Betweenness centrality as an indicator of the interdisciplinarity of scientific journals". *Journal of the American Society for Information Science and Technology*, v. 58, n. 9.
https://doi.org/10.1002/asi.20614

Leydesdorff, Loet (2011). "'Meaning' as a sociological concept: A review of the modeling, mapping and simulation of the communication of knowledge and meaning". *Social science information sur les sciences sociales*, v. 50, n. 3-4.
https://doi.org/10.1177/0539018411411021

Leydesdorff, Loet; **Bihui, Jin** (2004). "Mapping the Chinese science citation database". In: *University of Amsterdam*, v. 41.
https://doi.org/10.1002/meet.1450410157

Leydesdorff, Loet; **Bihui, Jin** (2005). "Mapping the Chinese Science Citation Database in terms of aggregated journal-journal citation relations". *Journal of the American Society for Information Science and Technology*, v. 56, n. 14.
https://doi.org/10.1002/asi.20209

Leydesdorff, Loet; **Bornmann, Lutz** (2012). "Mapping (USPTO) patent data using overlays to Google Maps". *Journal of the American Society for Information Science and Technology*, v. 63, n. 7.
https://doi.org/10.1002/asi.22666

Leydesdorff, Loet; **Cozzens, Susan**; **Van-den-Besselaar, Peter** (1994). "Tracking areas of strategic importance using scientometric journal mappings". *Research policy*, v. 23, n. 2.
https://doi.org/10.1016/0048-7333(94)90054-X

Leydesdorff, Loet; **De-Moya-Anegón, Félix**; **Guerrero-Bote, Vicente P.** (2010). "Journal maps on the basis of Scopus data: A comparison with the Journal Citation Reports of the ISI". *Journal of the American Society for Information Science and Technology*, v. 61, n. 2.
https://doi.org/10.1002/asi.21250

Leydesdorff, Loet; **De-Moya-Anegón, Félix**; **Guerrero-Bote, Vicente P.** (2015). "Journal maps, interactive overlays, and the measurement of interdisciplinarity on the basis of Scopus data (1996-2012)". *Journal of the Association for Information Science and Technology*, v. 66, n. 5.
https://doi.org/10.1002/asi.23243

Leydesdorff, Loet; **Heimeriks, Gaston**; **Rotolo, Daniele** (2016). "Journal portfolio analysis for countries, cities, and organizations: Maps and comparisons". *Journal of the Association for Information Science and Technology*, v. 67, n. 3.
https://doi.org/10.1002/asi.23551

Leydesdorff, Loet; **Kogler, Dieter-Franz**; **Yan, Bowen** (2017). "Mapping patent classifications: portfolio and statistical analysis, and the comparison of strengths and weaknesses". *Scientometrics*, v. 112, n. 3.
https://doi.org/10.1007/s11192-017-2449-0

Leydesdorff, Loet; **Meyer, Martin** (2010). "The decline of university patenting and the end of the Bayh-Dole effect". *Scientometrics*, v. 83, n. 2.
https://doi.org/10.1007/s11192-009-0001-6

Leydesdorff, Loet; **Persson, Olle** (2010). "Mapping the geography of science: Distribution patterns and networks of relations among cities and institutes". *Journal of the American Society for Information Science and Technology*, v. 61, n. 8.
https://doi.org/0.1002/asi.21347

Leydesdorff, Loet; **Ràfols, Ismael** (2009). "A global map of science based on the ISI Subject Categories". *Journal of the American Society for Information Science and Technology*, v. 60, n. 2.
https://doi.org/10.1002/asi.20967

Leydesdorff, Loet; **Salah, Alkim-Almila-Akdag** (2010). "Maps on the basis of the *Arts & Humanities Citation Index*: The journals *Leonardo* and *Art Journal* versus 'digital humanities' as a topic". *Journal of the American Society for Information Science and Technology*, v. 61, n. 4.
https://doi.org/10.1002/asi.21303

Leydesdorff, Loet; **Schank, Thomas** (2008). "Dynamic animations of journal maps: Indicators of structural changes and interdisciplinary developments". *Journal of the American Society for Information Science and Technology*, v. 59, n. 11.
https://doi.org/10.1002/asi.20891

Leydesdorff, Loet; **Wagner, Caroline S.**; **Bornmann, Lutz** (2018). "Betweenness and diversity in journal citation networks as measures of interdisciplinarity - A tribute to Eugene Garfield". *Scientometrics*, v. 114, n. 2.
https://doi.org/10.1007/s11192-017-2528-2

Leydesdorff, Loet; **Wagner, Caroline S.**; **Park, Han-Woo**; **Adams, Jonathan** (2013). "International collaboration in science: the global map and the network". *Profesional de la informacion*, v. 22, n. 1.
https://doi.org/10.3145/epi.2013.ene.12

Leydesdorff, Loet; **Welbers, Kasper** (2011). "The semantic mapping of words and co-words in contexts". *Journal of informetrics*, v. 5, n. 3.
https://doi.org/10.1016/j.joi.2011.01.008

Mingers, John; **Leydesdorff, Loet** (2015). "A review of theory and practice in scientometrics". *European journal of operational research*, v. 246, n. 1.
https://doi.org/10.1016/j.ejor.2015.04.002

Moral-Muñoz, José A.; **Herrera-Viedma, Enrique**; **Santisteban-Espejo, Antonio**; **Cobo, Manuel J.** (2020). "Software tools for conducting bibliometric analysis in science: An up-to-date review". *Profesional de la información*, v. 29, n. 1.
https://doi.org/10.3145/epi.2020.ene.03

Park, Han Woo; **Leydesdorff, Loet** (2008). "Korean journals in the Science Citation Index: What do they reveal about the intellectual structure of S&T in Korea?". *Scientometrics*, v. 75, n. 3.
https://doi.org/10.1007/s11192-007-1862-1

Park, Han Woo; **Leydesdorff, Loet** (2009). "Knowledge linkage structures in communication studies using citation analysis among communication journals". *Scientometrics*, v. 81, n. 1.
https://doi.org/10.1007/s11192-009-2119-y

Perret, Arthur; **Brioudes, Guillaume**; **Borel, Clément**; **Le Deuff, Olivier** (2023). *Cosma* (2.2.1).
https://doi.org/10.5281/zenodo.5412314
https://cosma.arthurperret.fr

Ràfols, Ismael; **Porter, Alan L.**; **Leydesdorff, Loet** (2010). "Science overlay maps: A new tool for research policy and library management". *Journal of the American Society for Information Science and Technology*, v. 61, n. 9.
https://doi.org/10.1002/asi.21368

Van-den-Besselaar, Peter; **Leydesdorff, Loet** (1996). "Mapping change in scientific specialties: A scientometric reconstruction of the development of artificial intelligence". *Journal of the American Society for Information Science*, v. 47, n. 6.
https://doi.org/10.1002/(SICI)1097-4571(199606)47:6<415::AID-ASI3>3.0.CO;2-Y

Wagner, Caroline S.; **Leydesdorff, Loet** (2005). "Network structure, self-organization, and the growth of international collaboration in science". *Research policy*, v. 34, n. 10.
https://doi.org/10.1016/j.respol.2005.08.002

Zupic, Ivan; **Čater, Tomaž** (2015). "Bibliometric methods in management and organization".
https://journals.sagepub.com/doi/10.1177/1094428114562629

New map of the research published in *Profesional de la Información* (2006-2023)

Pablo Guerrero-Castillo; María-Victoria Nuño-Moral; Vicente P. Guerrero-Bote; Félix De-Moya-Anegón

Nota: Este artículo se puede leer en español en:
https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/87540

Recommended citation:

Guerrero-Castillo, Pablo; **Nuño-Moral, María-Victoria**; **Guerrero-Bote, Vicente P.**; **De-Moya-Anegón, Félix** (2023). "New map of the research published in *Profesional de la Información* (2006-2023)". *Profesional de la información*, v. 32, n. 7, e320708.

https://doi.org/10.3145/epi.2023.dic.08

Manuscript received on September 18th 2023
Accepted on October 29th 2023

Pablo Guerrero-Castillo
https://orcid.org/0009-0000-9495-6553

Universidad de Extremadura
Facultad de Ciencias de la Documentación y la Comunicación
06071 Badajoz, Spain
pablogc@unex.es

María-Victoria Nuño-Moral
https://orcid.org/0000-0002-5447-4813

Universidad de Extremadura
Facultad de Ciencias de la Documentación y la Comunicación
06071 Badajoz, Spain
mvnunmor@unex.es

Vicente P. Guerrero-Bote ✉
https://orcid.org/0000-0003-4821-9768

Universidad de Extremadura
Facultad de Ciencias de la Documentación y la Comunicación
06071 Badajoz, Spain
guerrero@unex.es

Félix De-Moya-Anegón
https://orcid.org/0000-0002-0255-8628

SCImago Research Group
18220 Granada, Spain
felix.moya@scimago.es

Abstract

In 2006, *Profesional de la Información* (*EPI*) began to be indexed by international scientific literature databases and is currently one of the leading Spanish journals in Library & Information Science and in Communication. Research fields can be characterized and analysed based on the patterns of keywords used in the publications. One of the most used techniques for this is co-word analysis. This technique is used in the present study to examine the structure of the research published in *EPI*. The journal's two-fold spirit in Library & Information Science and in Communication is revealed, comprising six main thematic areas. Since no poor behaviour is seen in any of these areas, it can be concluded that, in becoming part of *WoS* and *Scopus*, *EPI* has entered a virtuous cycle that has led it to successfully expand its thematic scope, and to attain levels of impact and excellence superior to those of its origins.

Keywords

Scientometrics; Co-word analysis; Knowledge maps; Thematic analysis; Scientific journals; Library & Information Science; Communication; *Scopus*; *Profesional de la Información*.

1. Introduction

Profesional de la Información (*https://www.profesionaldelainformacion.com*) is a scientific journal also known as *EPI*. Its original title, dating back to 1992, was *Information World en Español* (*IWE*), a newsletter that published news and reports. In 1998, its title was changed to *El Profesional de la Información* and, in response to the demand of most of its subscribers, it began to publish peer-reviewed articles. Its indexing in the *Institute for Scientific Information* (*ISI*)'s *Social Sciences Citation Index* database (*Web of Science*) and in *Elsevier*'s *Scopus* began in 2006. In 2020, the editors decided to eliminate the article "El" (masculine article) to avoid what could appear to be gender discrimination. In 2023, *EPI* ceased its subscriber model to become a 100% open access journal. Despite the existence of other scientific journals in Spain in the field of Library and Information Science, *EPI* has played a prominent role ever since its creation. It has become the

Spanish publication with the greatest impact in this field, and has expanded its scope to the area of Communication.
In scientometrics, information from large scientific literature databases is used to analyse research quantitatively and qualitatively, and to examine it thematically. As **Neff** & **Corley** (2009) state, research fields can be characterized and analysed based on the keyword patterns used in their publications. One of the techniques most used in thematic analysis is co-word analysis (**Callon** *et al.*, 1986; 1991). In co-word networks, the nodes are keywords and the links are weighted based on the documents in which the pair of keywords forming the link occur together. These networks are subjected to procedures designed to detect the most closely related groups of keywords, thus revealing the thematic structure of the research (**Romo-Fernández**; **Guerrero-Bote**; **De-Moya-Anegón**, 2013; **Blázquez-Ruiz**; **Guerrero-Bote**; **De-Moya-Anegón**, 2016; 2017; **Olmeda-Gómez**; **Ovalle-Perandones**; **Perianes-Rodríguez**, 2017; **Faraji** *et al.*, 2022). Co-word analysis is sometimes used to study the thematic structure of scientific journals (**Romo-Fernández**; **Guerrero-Bote**; **De-Moya-Anegón**, 2013; **López-Robles** *et al.*, 2019), of research in specific periods (**Herrera-Viedma** *et al.*, 2020), and of such fields as Food Science (**Romo-Fernández**; **Guerrero-Bote**; **De-Moya-Anegón**, 2016; 2017), Library and Information Science in Spain (**Olmeda-Gómez**; **Ovalle-Perandones**; **Perianes-Rodríguez**, 2017), Intellectual capital (**Faraji** *et al.*, 2022), Communication in Spain (**Segado-Boj**; **Gómez-García**; **Díaz-Campo**, 2022), Entrepreneurship (**Lechuga-Sancho**; **Martínez-Fierro**; **Ramos-Rodríguez** 2023), and data-driven scientific research (**Velasco-López** *et al.*, 2023).

Studies of co-word networks are known as co-word analyses. However, not all co-word analyses use the same method. Some, such as those that use the *SciMAT* program (**López-Robles** *et al.*, 2019; **Herrera-Viedma** *et al.*, 2020; **Segado-Boj**; **Gómez-García**; **Díaz-Campo**, 2022; **Velasco-López** *et al.*, 2023; **Lechuga-Sancho**; **Martínez-Fierro**; **Ramos-Rodríguez**, 2023) are based on the strategic diagrams defined by **Callon** *et al.* (1986; 1991). Others use more visual methods based on clustering or community detection algorithms, together with layout algorithms that allow the co-word network to be viewed and navigated (**Romo-Fernández**; **Guerrero-Bote**; **De-Moya-Anegón**, 2013; **Blázquez-Ruiz**; **Guerrero-Bote**; **De-Moya-Anegón**, 2016; 2017; **Olmeda-Gómez**; **Ovalle-Perandones**; **Perianes-Rodríguez**, 2017; **Faraji** *et al.*, 2022).

In the past, it was the senior researchers who knew the intellectual structure of a discipline, usually that of their own field of study. But this structure was neither formal nor included in any support. Instead, it was a subjective structure that the researcher had formed mentally as a result of the deep knowledge they had of their discipline. It thus suffered from conservatism, bias, and subjectivity (**Bornmann**, 2011; **Irvine** *et al.*, 1985). Carrying out this type of research therefore involves a more objective revelation of the structure of scientific fields which can be readily assimilated by both novel and senior researchers.

The principal objective of the present study was to establish the intellectual structure of the journal *EPI* based on the analysis of the keywords present in those papers it has published which are collected in international scientific literature databases. This led to such specific research questions as:

- How has the journal evolved since being included in the international databases?
- How many sub-areas make up the main structure of *EPI*?
- How do they relate to each other?
- Which topics are the most central and which the most specialized?
- What is the scientific impact of each topic and how has it evolved?
- What are the keyword burst periods?

2. Method and data

The records corresponding to the articles published by *Profesional de la Información* were downloaded from both *WoS* and *Scopus* on 21 September 2023.

As can be seen in Figure 1, the production in *WoS* and in *Scopus* is coincident, with the former totaling 1,774 documents and the latter 1,809. Because of this slightly greater completeness of data, we chose to use that of *Scopus*. Of these *Scopus* documents, the vast majority were articles (1,663) and the rest reviews (111). The original language was Spanish in 1,305 documents (72%) and English in 479 (26.5%).

The Author Keywords were extracted, giving in total 6,864 keywords, with a total of 15,806 occurrences. Not all the records had Author Keywords –only 1,774. The extracted keywords were unified by first applying

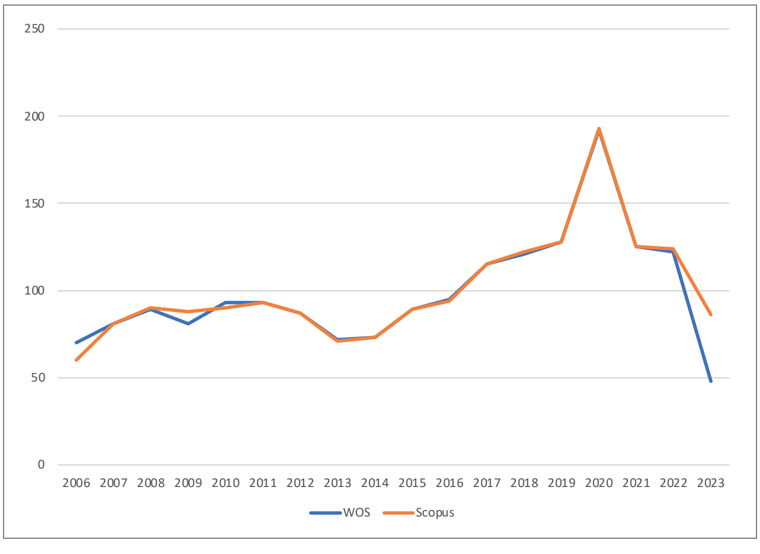


Figure 1. Scientific production published in *Profesional de la Información* as registered in *WoS* and *Scopus*.

Porter's stemming algorithm (1980) which reduced them to the root, and then, in order not to be left with just the root, choosing the commonest form. This unification left a total of 6,277 keywords. Since this number is difficult to cover and introduces excessive noise from keywords which occur in just a few documents (**Romo-Fernández**; **Guerrero-Bote**; **De-Moya-Anegón**, 2013; **Blázquez-Ruiz**; **Guerrero-Bote**; **De-Moya-Anegón**, 2016; 2017), in order to get a manageable number of keywords we selected those that appeared in more than 8 documents. This left a total of 279 keywords appearing in 1,616 documents, which represented 89% of the total documents and 92.7% of the keyword-containing documents. Figure 2 shows the evolution of this percentage during the period studied. As can be seen, the journal has good representation, with only a small gap in the first part of the period.

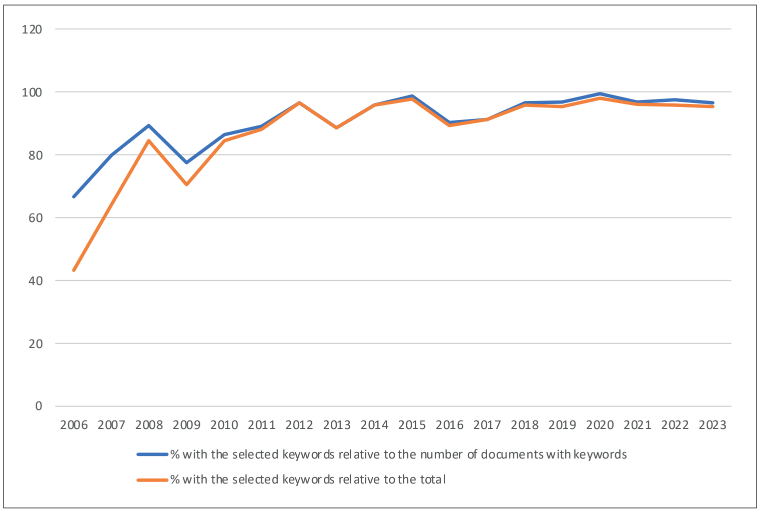


Figure 2. Percentage representation of the 279 keywords that appear in more than 8 documents relative to the total number of articles and to those that have keywords.

If the links between keywords were weighted by the number of co-occurrences, the keywords with the most occurrences would also be those with the links of greatest weight. To avoid this, the links between the keywords were normalized. The idea is to give greater weight to links between keywords that appear together more than expected, i.e., weight each link with the ratio between the proportion of the co-occurrences it represents and the probability that the two keywords co-occur according to their number of appearances.

The occurrences of each keyword were given a weight, firstly, by dividing by the number of keywords in the article because co-occurrence in an article that has many keywords is not the same as in one that has few. The average number of keywords per paper was 9 with a standard deviation of 3.67, meaning that there was considerable variation. In this way, each co-occurrence was weighted by the product of the weights of each keyword which, since it was the same document, is the inverse of the square of the number of keywords that occur in the document. And secondly, by dividing the sum of the weights of the co-occurrences by the sum of the weights of the co-occurrences in which each keyword participates separately, and then multiplying by twice the total sum of the network's co-occurrence weights. In this way, the weight of each link represents the ratio between the proportion of the co-occurrence weights it represents and the probability that the two keywords co-occur.

To make the co-word map and grouping, we used the *SCImago Graphica* tool (**Hassan-Montero** *et al.*, 2022). This uses **Clauset** (2004)'s community identification algorithm and the LinLog algorithm (**Noack**, 2007) to generate the layout. The latter uses an energy model which generates layouts that are strongly coherent with the communities identified (**Noack**, 2009). Although these communities are often called clusters, the concept is different since communities are not groupings of similar objects that are formed by evaluating their characteristics. Instead, they are usually formed by removing the links that participate in more geodesics, thus forming groups of frequently co-occurring words.

We used the burst algorithm developed by **Jon Kleinberg** (2003) which detects when certain terms become fashionable in a discourse and then fade away. We applied it to both the keywords and the communities. The algorithm generates a table with the bursty periods of the most frequent words, indicating the length, the strength, and the time interval in which the burst occurs.

To show the evolution of the journal itself and of the different communities, the following indicators were used:

- Ndoc: Number of documents published in scientific journals included in the *Scopus* database.
- %Int: Percentage of documents in whose byline appear authors from different countries.
- Normalized Impact (NI): Average normalized citation received by each document, understanding this to be the ratio between the citations received by the document and the average citations of the documents of the same type, year, and category (**Rehn**; **Kronman**, 2008).
- %Excellence: Percentage of documents that are among the 10% most cited of the same year, type, and category (**Bornmann** *et al.*, 2012).
- %Excellence1: Percentage of documents that are among the 1% most cited of the same year, type, and category.
- Authors: Average number of authors that appear in the byline of the articles.
- %ARC: Annual percentage rate of change calculated from the slope of the regression line, dividing it by the average of the indicator in the period and multiplying by 100. This indicator is designed to show the average evolution of other indicators in a period.

As the data for the year 2023 are incomplete, they were not taken into account for the calculation of the %ARC. Furthermore, for the citation-based indicators (NI and the %Exc indicators), the 2022 data were not taken into account as they are still not sufficiently stable.

3. Results

The only parameter in **Clauset**'s (2004) community identification algorithm is the number of communities. After trying several possibilities, we chose the value 6. On increasing this number by 1, the algorithm splits one of the previous communities into two, and so on successively. To generate a second level, we established 27 communities. Of the 27 communities, 8 contained a single keyword which was included in its neighbouring community. The community structure obtained in this way is presented in Table 1. Each community has been manually labeled. Since these communities are generated through co-occurrence links, they are keywords that co-occur frequently. While in many cases they are semantically related, in others it is hard to see beforehand any relationship that makes them frequently co-occur.

Table 1. Two-level community structure obtained with the parameters 6 and 27. In parentheses, the number of documents in which they occur.

C 1 Informetrics
C 1.1 Bibliometrics-Scientometrics
Bibliometrics (50), Transparency (39), Research evaluation (23), Bibliometric indicators (22), Citation analysis (18), Accountability (16), Wikipedia (13), Scientific output (13), Scientometrics (13), Science communication (10), Google Scholar (9), Research projects (9)
C 1.2 Altmetrics
Open access (52), Innovation (50), Universities (47), Indicators (38), Trends (31), Altmetrics (25), Impact (20), Metrics (18), Citations (18), Patents (11)
C 1.3 Scholarly Communication
Spain (217), Scholarly communication (37), Scientific production (24), Web of Science (23), Metadata (20), Rankings (20), Scopus (19), Review article (17), Information science (14), Interviews (14), Review (12), Communication research (9), China (9)
C 2 Health Social media
Social media (204), Content analysis (42), Health information (35), Health communication (20), Engagement (17), TikTok (9)
C 3 Social Networks
C 3.1 Entertainment networks
Social networks (192), Television (62), Facebook (50), Public libraries (45), YouTube (28), Instagram (27), Marketing (24), Audiovisual documentation (17), Users (11), Public opinion (11), Personalization (11), Social networking sites (10), Infotainment (9), Citizen participation (9)
C 3.2 Political networks
Twitter (131), Political communication (95), Academic libraries (46), Elections (28), Latin America (25), Political information (12), Populism (12), WhatsApp (12), Agenda-setting (9)
C 4 Communication
C 4.1 Data & automation
Audiences (58), Open data (34), News (25), Artificial intelligence (22), Gender (21), Mobile devices (19), Europe (16), Open government (15), Software (15), Video (15), Smartphones (15), Algorithms (15), Women (13), Automation (11), Programming (11), Apps (10), Public sphere (10), Democracy (10), AI (10), Applications (9), Gender gap (9), Stereotypes (9), TV (9)
C 4.2 Journalism
Journalism (131), Media (100), Internet (87), Digital journalism (57), Digital media (53), Journalists (38), Online media (31), Online journalism (30), Business models (28), Cybermedia (26), Digital press (25), Electoral campaigns (18), Usability (17), Political parties (17), Multimedia (17), Convergence (16), Influencers (16), Information architecture (15), Cyberjournalism (15), Digital newspapers (15), Online newspapers (15), Audiovisual (14), Methodology (13), Professionals (13), E-learning (12), Social web (12), User experience (11), New media (11), Colombia (11), Content (11), Policies (11), Crisis (11), Digital libraries (10), Cinema (10), Library cooperation (9), Networking (9), History (9), Polarization (9)
C 4.3 Disinformation
Disinformation (47), Fake news (40), Surveys (31), Mass media (25), Press (21), Fact-checking (19), Credibility (18), Post-truth (15), Trust (13), Catalonia (13), Hoaxes (12), Misinformation (11)
C 4.4 Corporate communication
Corporate communication (37), Organizational communication (35), Public relations (35), Interactivity (31), Companies (14), Standards (13), Management (12), Corporate social responsibility (12), Reputation (12), Internal communication (11), CSR (10), Organizations (10), Recommendations (9), Storytelling (9)
C 5 Information
C 5.1 Information Research
Research (62), Scientific journals (30), Evaluation (29), Social sciences (16), Quality (14), Websites (13), Visibility (11)
C 5.2 Information Stores
Newspapers (52), Libraries (46), Digitization (35), Branding (18), Documentation (17), Archives (17), Transmedia (13), Photography (12), Geolocation (9), El País (9)
C 5.3 Information Management
University libraries (33), Evolution (26), Privacy (23), Radio (23), Digital communication (22), Information sources (21), Knowledge management (20), Information professionals (19), Podcasting (19), Information management (18), Strategies (17), Content management (15), Information technologies (14), Document management (14), Museums (14), Audio communication (13), Platforms (12), Science (11), Publications (11), Literature review (11), Peer review (10), Classifications (10), Immersive journalism (10), Data journalism (9), Sustainability (9), Analysis (9)

C 5.4 Information Systems
Web 2.0 (66), Information retrieval (23), Semantic web (23), Library and information science (23), Web (23), Information visualization (22), Social network analysis (19), Profession (19), Websites (17), Ontologies (15), Spanish universities (14), ICT (14), Collaboration (14), Blogs (13), Public information (13), Information systems (12), Network analysis (12), Intranets (11), Statistics (11), City councils (10), Competitive intelligence (9), E-government (9), Cloud computing (9), Data visualization (9), Information design (9)
C 5.5 Search Engines & Books
Google (18), Ebooks (17), Publishing (17), Book (16), SEO (14), Search engines (13), Reading (12)
C 5.6 Information Professionals
Communication (104), Professional profiles (30), Technology (30), Higher education (26), Librarians (25), Education (20), Information literacy (18), Health (17), Information scientists (9)
C 5.7 Information Subjects
Covid-19 (81), Pandemics (66), Coronavirus (60), Big data (36), Information (30), Databases (29), Repositories (25), Media literacy (22), Scientific communication (20), Public administration (19), Framing (18), Scientific publication (17), Future (16), Crisis communication (16), Data (15), Politics (15), Training (14), Digital humanities (13), Institutional communication (13), Audiovisual communication (12), Perception (12), Skills (12), Adolescents (12), SARS-CoV-2 (12), Library services (11), European Union (11), Hate speech (11), Feminism (11), Challenges (11), Research data (10), Open source (10), Machine learning (10), Children (10), Activism (10), Knowledge (9), Digital divide (9), Risk (9), Power (9), Health crisis (9)
C 6 Advertising
Advertising (52), Participation (36), Ethics (31), Self-regulation (12), Governance (12)

The structure comprises six first-level communities, two that are quite small labeled *C 6 Advertising* and *C 2 Health Social media*, two medium sized labeled *C 1 Informetrics* and *C 3 Social Networks*, and two larger ones labeled *C 4 Communication* and *C 5 Information*. At the second level, the large communities have been subdivided into 4 and 7 communities respectively, the medium ones into 3 and 2 respectively, and the small ones left undivided.

Figure 3 is a map of the co-words. They are coloured based on the second-level community to which they belong. We have coloured not just the nodes but also the minimum convex hull so that the area which each community covers can be seen. For those of second-level included in one of first-level, we have chosen different shades of the same colour.

One observes in the figure that on the left are the communities more related to communication, while on the right are those more related to *Library and Information Science* and social networks.

Within this structure, the informetrics communities are at the top right, and Social Networks at the bottom right. The communities included within *C 5 Information* are found in the central part, and act as mortar holding all the other communities together. Of these, the one that occupies the largest area is that labeled *C 5.7 Information Subjects*.

At centre right is the community labeled *C 2 Health Social media* next to Social Networks. This is explained by the weight in them of some social media such as *TikTok*. Just above is the one labeled *C 4.4 Corporate communication*, which can be explained by the importance of Social Media in corporate communication.

The upper right appears dominated by the community labeled *C 4.2 Journalism.* From the centre downwards there begins to appear the one labeled *C 4.1 Data & automation,* and in the lower central area the one labeled *C 4.3 Disinformation*, curiously by the side of Social Networks.

At bottom right, relatively isolated, is *C 6 Advertising*.

Figure 4 shows an enlarged view of the upper right corner where the Informetrics communities are found. The largest node is Spain, which was one of the nodes isolated in the second level. Furthermore, Spain is the geographical domain that has been most studied in the informetrics studies published in *EPI*.

Figure 5 shows an enlarged view of the lower right corner. The node corresponding to Social Media is seen to be the largest of the C 2 community, and that of Social Networks the largest of the C 3 community.

Figure 6 shows an enlarged view of the lower left corner. The community labeled as *C 4.3 Disinformation* can be seen in full as well as its interaction with terms of other communities such as Artificial Intelligence or Democracy.

Figure 7 shows an enlarged view of the upper left corner of the co-word map. The first part contains words also related to Informetrics. In the central part, there already start to appear the largest nodes, nodes which are Journalism related.

Table 2 presents scientometric indicators of the journal and its communities. As indicators, we chose the number of documents, the normalized citation, the percentage of international collaboration, the percentage of excellence (papers included in the top 10% most cited in their categories, document types, and year), the percentage of excellence 01 (articles included in the top 1% most cited of their categories, document types, and year), and the average number of authors.

The first row of the table includes the data of the entire journal, a total of 1809 documents registered in *Scopus*, with an annual growth rate of more than 4.5%. The normalized impact is greater than the mean (1), and also its growth rate during the period is more than 10%, indicative of the journal's good evolution. The case is similar with the excellence parameters. The percentage of excellence (top 10% most cited) is very close to 30%, almost three times more than the mean, with a close to 15% growth during the period, and the same is the case with the percentage of excellence 01.

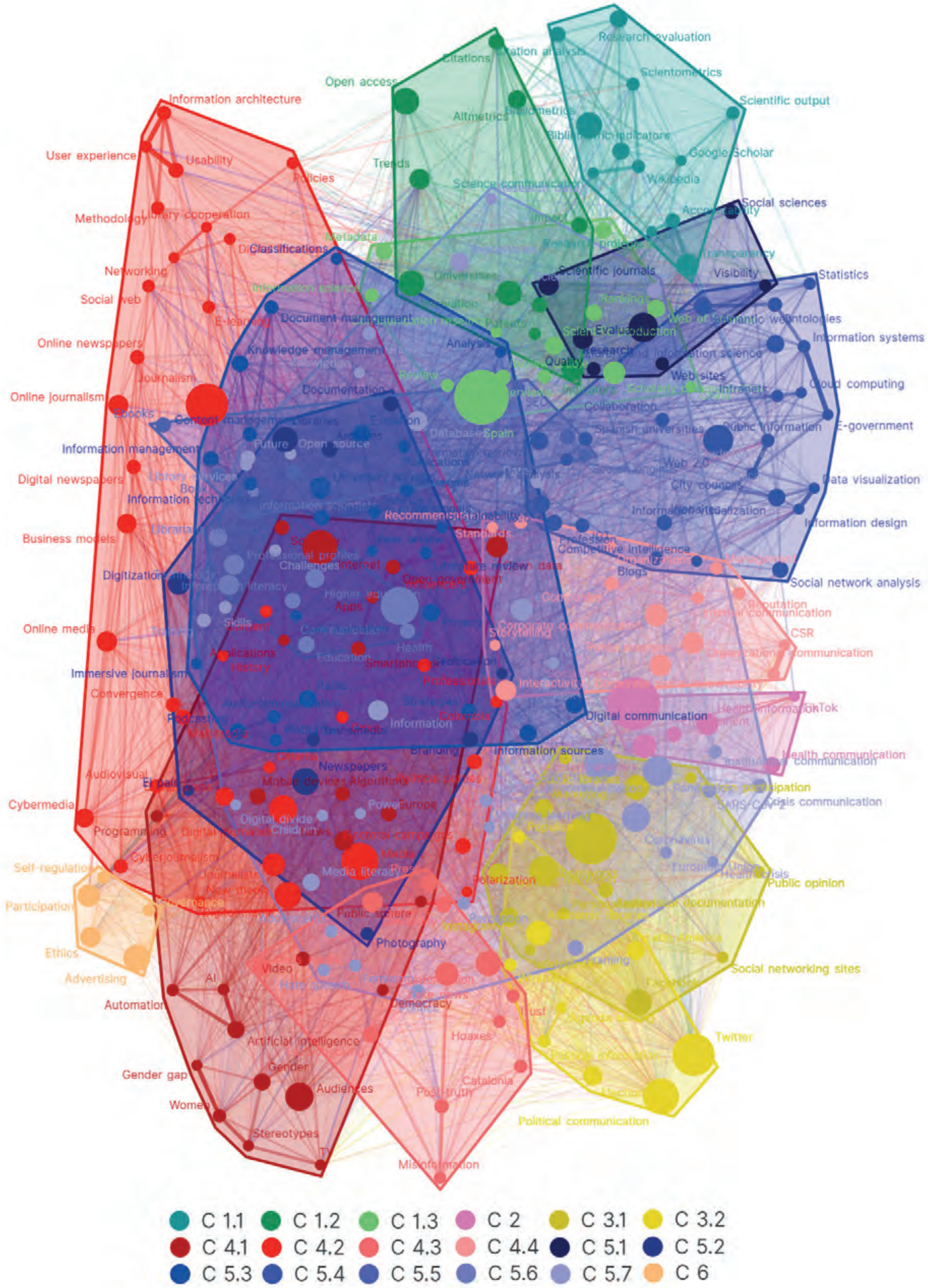


Figure 3. General view of the co-word map with shading for the zone corresponding to each community.

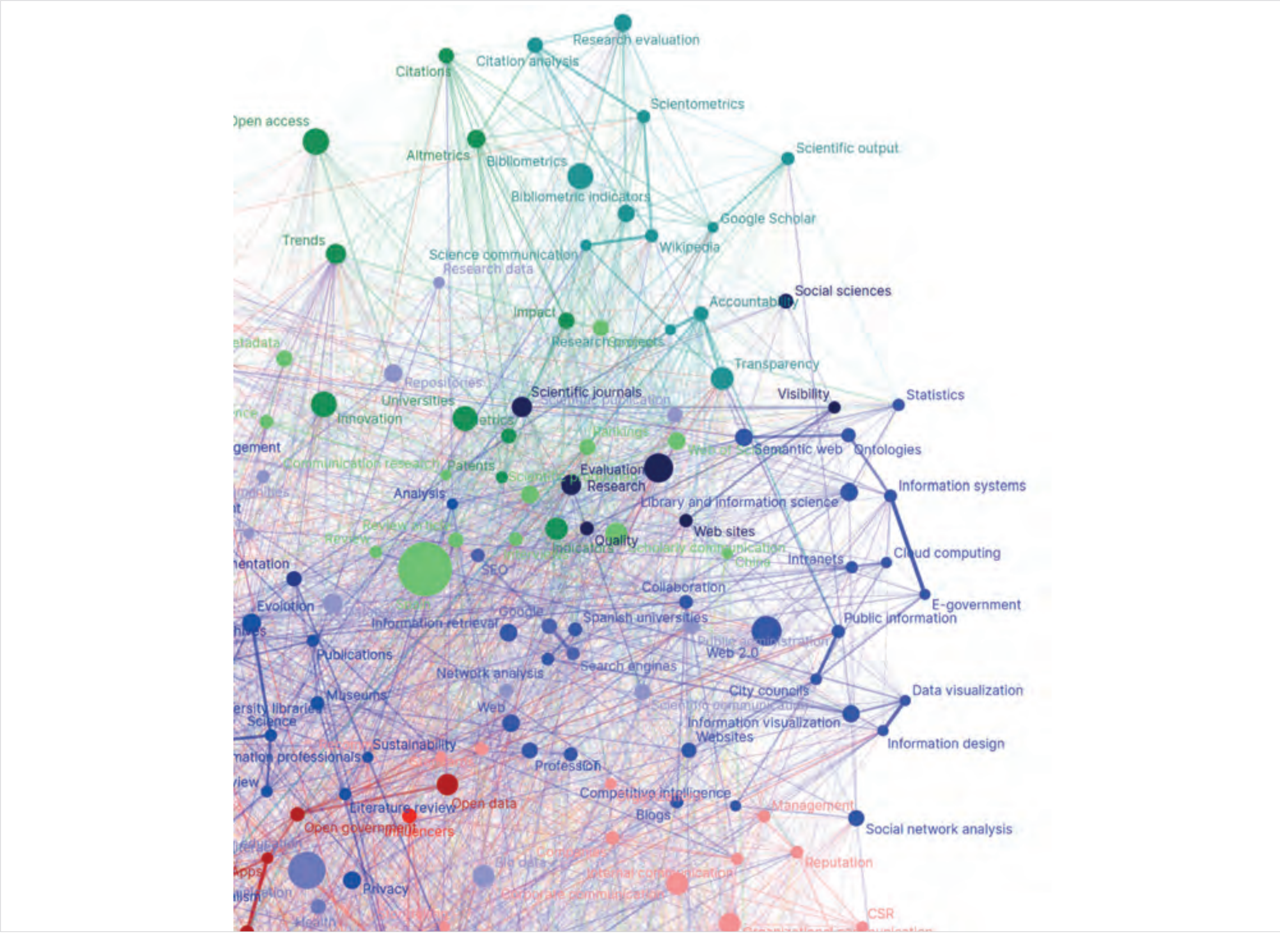


Figure 4. Zoom-in on the upper right corner of the co-word map.

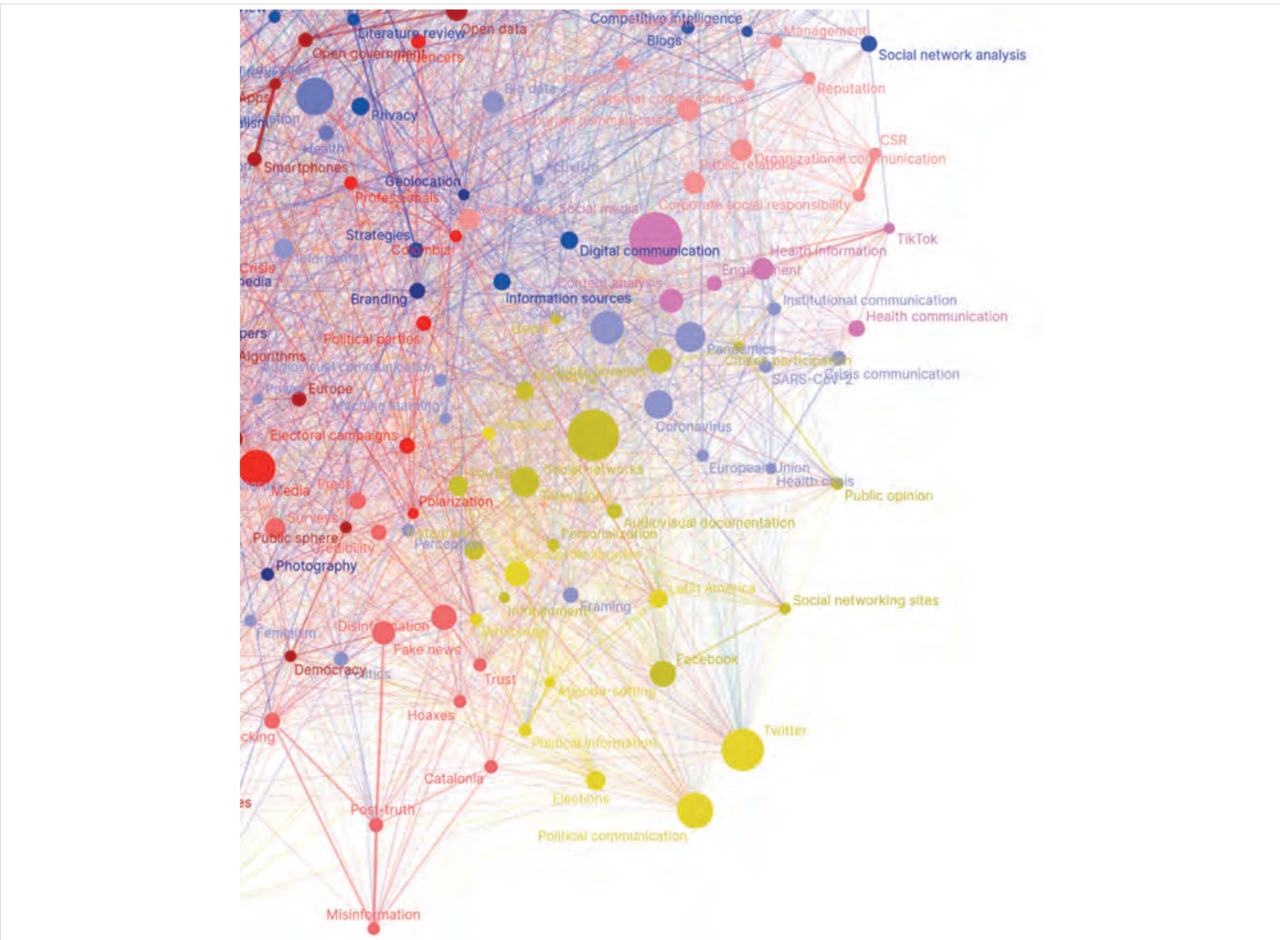


Figure 5. Zoom-in on the lower right corner of the co-word map.

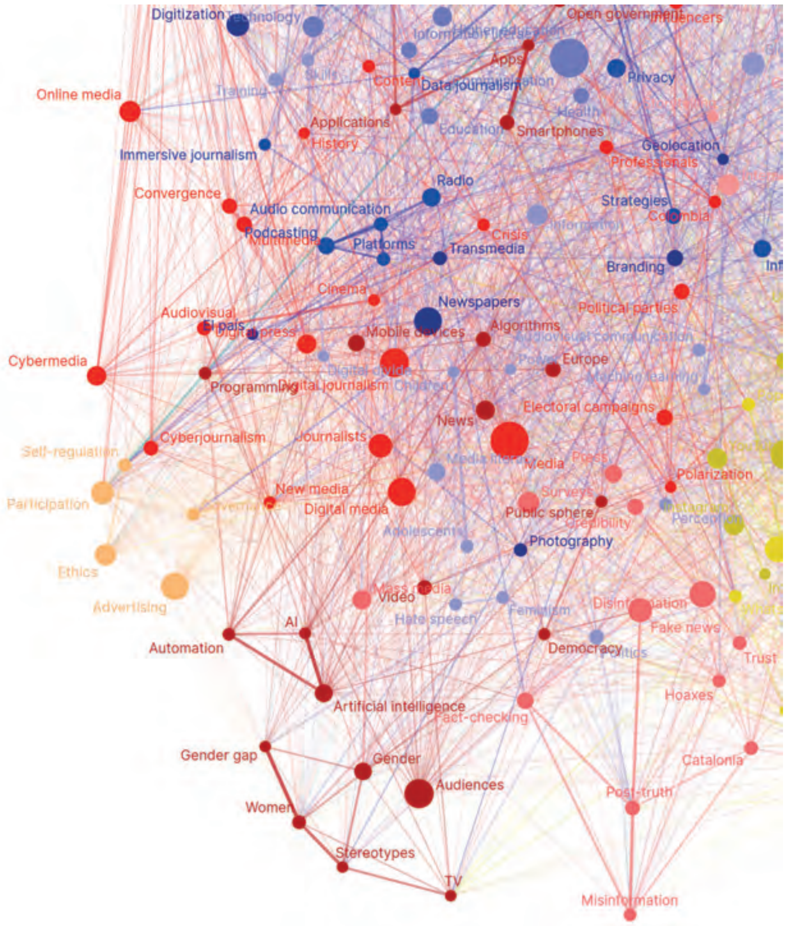


Figure 6. Zoom-in on the lower left corner of the co-word map.

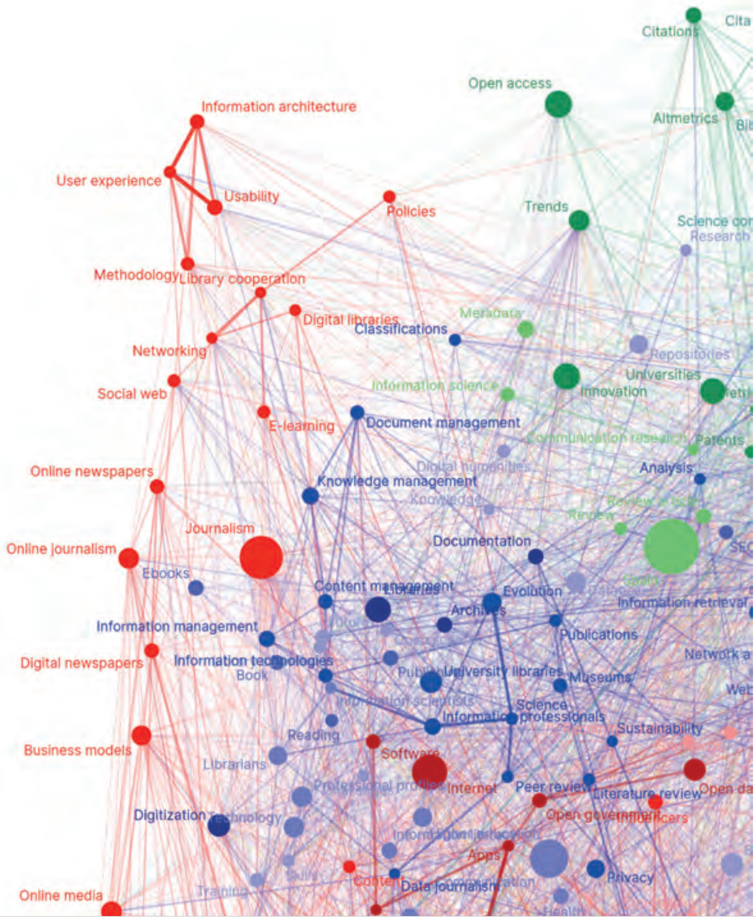


Figure 7. Zoom-in on the upper left corner of the co-word map.

Close to 10% of the studies involve collaborations, and there was considerable growth in this parameter during the period –around 8% per annum. The co-authorship index is 2.20, and also grew during the period, although more discreetly.

Table 2. Scientometric Indicators of the journal and communities, with number of documents, normalized impact, percentage of international collaboration, percentage of excellence, percentage of excellence 01, number of authors, and their respective annual rates of change.

	Ndoc	%ARC	NI	%ARC	%Int	%ARC	% Exc.	%ARC	%Exc1	%ARC	Authors	%ARC
Profesional de la Información	**1809**	**4.51**	**1.23**	**13.83**	**9.12**	**8.27**	**29.43**	**14.11**	**3.25**	**18.65**	**2.20**	**2.08**
C 1 Informetrics	**593**	**9.91**	**1.70**	**11.06**	**12.31**	**10.01**	**37.11**	**10.08**	**4.39**	**16.97**	**2.33**	**1.79**
C 1.1 Bibliometrics-Scientometrics	162	10.32	2.64	11.18	14.20	8.81	39.31	4.76	6.90	16.34	2.40	-0.96
C 1.2 Altmetrics	252	9.12	1.30	8.97	13.89	15.46	37.89	10.79	4.41	12.27	2.32	2.87
C 1.3 Scholarly Communication	358	11.36	1.62	11.68	12.85	8.32	40.66	11.96	4.52	22.90	2.41	1.45
C 2 Health Social media	**269**	**20.24**	**2.47**	**15.14**	**10.41**	**9.22**	**56.20**	**11.19**	**8.26**	**18.74**	**2.22**	**2.07**
C 3 Social Networks	**505**	**12.55**	**1.69**	**13.51**	**10.30**	**2.92**	**44.28**	**16.91**	**4.87**	**17.15**	**2.25**	**0.87**
C 3.1 Entertaiment networks	375	11.88	1.39	12.09	8.27	0.44	42.53	16.38	3.16	18.60	2.17	0.57
C 3.2 Political networks	264	14.89	2.21	12.27	12.88	-0.84	55.14	13.91	8.64	13.10	2.32	-1.14
C 4 Communication	**865**	**11.35**	**1.66**	**14.47**	**8.79**	**4.44**	**39.45**	**13.23**	**4.34**	**17.95**	**2.20**	**1.86**
C 4.1 Data & automation	332	13.81	2.03	15.00	8.43	9.34	42.16	10.14	3.59	19.20	2.19	3.24
C 4.2 Journalism	511	10.31	1.50	13.47	8.22	7.06	39.62	13.77	3.77	15.43	2.17	1.92
C 4.3 Disinformation	160	18.22	2.88	20.14	8.13	-12.65	54.29	19.24	10.71	23.24	2.21	1.62
C 4.4 Corporate communication	151	13.71	1.25	14.79	9.27	7.47	37.93	16.39	4.14	28.04	2.19	-0.23
C 5 Information	**1143**	**6.36**	**1.38**	**13.78**	**8.75**	**11.45**	**30.46**	**14.58**	**3.52**	**20.49**	**2.24**	**1.02**
C 5.1 Information Research	150	6.32	0.96	8.56	12.67	16.51	27.66	8.72	0.71	29.17	2.57	1.07
C 5.2Information Stores	191	5.41	1.66	19.05	6.28	16.04	21.86	15.33	2.73	20.43	1.92	2.23
C 5.3 Information Management	326	7.03	1.22	12.15	7.98	13.91	27.97	17.43	3.54	17.53	2.14	2.26
C 5.4 Information Systems	330	-0.04	1.03	9.00	9.39	12.85	26.56	13.63	1.56	10.56	2.34	0.86
C 5.5 Search Engines & Books	83	3.54	0.67	9.16	7.23	8.90	25.00	14.46	0.00	0.00	2.28	-0.78
C 5.6 Information Professionals	228	11.35	1.25	12.75	10.53	9.37	35.68	16.14	3.29	22.57	2.20	2.16
C 5.7 Information Subjects	444	12.40	1.95	13.49	7.88	11.10	40.69	14.81	7.20	20.92	2.17	2.10
C 6 Advertising	**124**	**14.57**	**1.09**	**11.76**	**8.87**	**18.49**	**38.84**	**17.83**	**2.48**	25.77	2.37	3.22

Observing the evolution of the first-level communities, one sees that they are all growing, although those that are growing most are the two smallest, especially *C 2 Health Social media* which has grown by about 20%. Of the second-level communities, those of *C 4.3 Disinformation* (18%) and *C 3.2 Political networks* (15%) are also growing notably.

In terms of impact, the *C 2 Health Social media* community stands out with more than twice the average impact. There also stand out *C 1 Informetrics* (1.7%), *C 3 Social Networks* (1.69%), and *C 4 Communication* (1.66%). In terms of evolution, *C 2 Health Social media* and *C 4 Communication* are the most noteworthy, with growth of more than 14%. Of the second-level communities, *C 4.3 Disinformation* stands out with an impact of 2.88% and an annual growth rate of more than 20%, and *C 1.1 Bibliometrics-Scientometrics* with an impact of 2.64%. Only *C 5.1 Information Research* (0.96%) and *C 5.5 Search Engines & Books* (0.67%) are below the mean.

In percentage of excellence, *C 2 Health Social media* and *C 3 Social Networks* stand out, having more than 40% of articles within the top 10% most cited. In the evolution of this indicator, *C 6 Advertising* and *C 3 Social Networks* stand out with increases of around 17% per annum. In the second level, there stand out *C 3.2 Political networks* (55%) and *C 4.3 Disinformation* (54%).

In the percentage of excellence 01, *C 2 Health Social media* clearly stands out, with more than 8% of its articles in the top 1% of the discipline. Among those of the second level, *C 4.3 Disinformation* (10.7%) and *C 3.2 Political networks* (8.64%) stand out.

Regarding international collaboration, *C 1 Informetrics* stands out with more than 12%, but all reach more than 8%. Of the increase in this period, *C 6 Advertising* stands out with an annual increase of

Table 3. Keywords from articles with a greater average normalized impact.

Id	Keyword	Ndoc	Ac. Weight	NI	C l2
172	Scientometrics	13	1.47	13.49	C 1.1
148	Software	15	1.44	10.27	C 4.1
262	Health crisis	9	0.66	9.16	C 5.7
214	Misinformation	11	1.00	7.98	C 4.3
235	Democracy	10	0.88	7.15	C 4.1
247	Analysis	9	0.74	6.37	C 5.3
238	Machine learning	10	1.05	6.10	C 5.7
31	Fake news	40	3.50	5.94	C 4.3
15	Coronavirus	60	4.93	5.73	C 5.7
10	Covid-19	81	6.44	4.97	C 5.7
22	Bibliometrics	50	6.17	4.78	C 1.1
223	Statistics	11	1.06	4.70	C 5.4
96	Fact-checking	19	1.64	4.63	C 4.3
107	Credibility	18	1.88	4.62	C 4.3

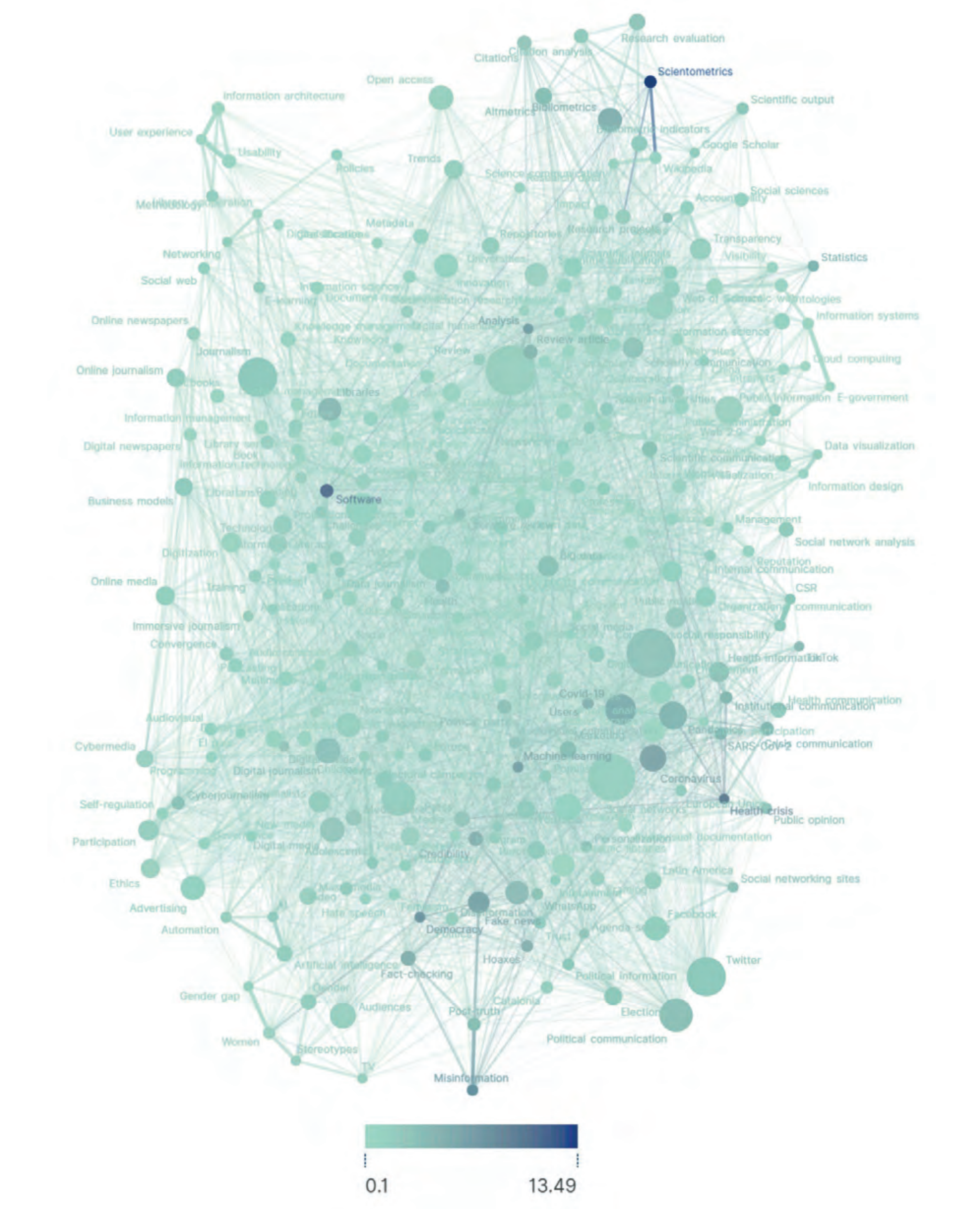


Figure 8. Map of co-words coloured on the basis of the normalized citation of the articles that include the corresponding keywords.

more than 18%. Of the second level, *C 1.1 Bibliometrics-Scientometrics* (14.2%) and *C 1.2 Altmetrics* (13.89%) stand out.

In the authorship index, there stand out *C 6 Advertising* (2.37%) and *C 1 Informetrics* (2.33%) with more than 2.30%. The former grew the most during the period. Of the second level, *C 5.1 Information Research* stands out with a co-authorship index of 2.57%.

Figure 8 is the same as the previous co-word map, but this time each node is coloured based on the impact of the documents in which the corresponding word occurs. Table 3 lists the keywords from articles with a greater average normalized impact. The number of documents (Ndoc), the weight they accumulate (Ac. Weight), the second level community (C l2), and the order by size are given.

The keyword from articles with a greater average normalized impact is Scientometrics, with a normalized impact greater than 13. It is followed by Software (10.27), Health crisis (9.16), Misinformation (7.98), and Democracy (7.14). All of these keywords occur in a small number of articles (fewer than 15). The keywords with more than 40 documents and a high impact are Fake news (5.94), Coronavirus (5.73), Covid-19 (4.97), and Bibliometrics (4.78).

Table 4 lists the most notable keyword burst periods. Included are all those that exceed a strength of 5. They are ordered by community, with which the distribution can be seen not to be balanced but instead very uneven. Most of the bursty periods noted correspond to keywords included in *C 5 Information*.

Table 4. Keyword bursty periods, ordered by level-2 community.

Word	Length	Strength	Start	End	C l2
Indicators	1	10.53	2018	2018	C 1.2
Open access	1	6.42	2012	2012	C 1.2
Metadata	9	5.40	2006	2014	C 1.3
Health information	2	6.11	2019	2020	C 2
Public libraries	11	6.23	2006	2016	C 3.1
Audiovisual documentation	6	5.67	2009	2014	C 3.1
Political communication	4	14.03	2017	2020	C 3.2
Academic libraries	5	6.44	2012	2016	C 3.2
Internet	7	7.58	2006	2012	C 4.1
Audiences	2	5.29	2015	2016	C 4.1
Artificial intelligence	4	5.29	2021		C 4.1
Online journalism	8	7.00	2010	2017	C 4.2
Information architecture	8	6.07	2007	2014	C 4.2
Social web	8	5.08	2008	2015	C 4.2
Disinformation	6	7.87	2019		C 4.3
Surveys	5	5.83	2014	2018	C 4.3
Post-truth	2	5.75	2018	2019	C 4.3
Fake news	1	5.01	2019	2019	C 4.3
Organizational communication	1	8.31	2019	2019	C 4.4
Public relations	2	7.51	2019	2020	C 4.4
Corporate communication	2	6.28	2019	2020	C 4.4
Libraries	7	8.27	2009	2015	C 5.2
Digitization	8	6.07	2007	2014	C 5.2
Documentation	4	5.37	2010	2013	C 5.2
Audio communication	3	7.74	2022		C 5.3
Information management	9	7.29	2006	2014	C 5.3
Podcasting	3	7.15	2022		C 5.3
Knowledge management	11	7.09	2006	2016	C 5.3
Document management	7	5.74	2006	2012	C 5.3
Content management	9	5.26	2006	2014	C 5.3
Museums	4	5.02	2011	2014	C 5.3
Web 2.0	6	23.95	2007	2012	C 5.4
Semantic web	9	7.94	2007	2015	C 5.4
Information retrieval	8	6.21	2007	2014	C 5.4
Ontologies	5	5.54	2007	2011	C 5.4
Intranets	6	5.32	2006	2011	C 5.4
Information design	2	5.05	2017	2018	C 5.4
Covid-19	5	23.75	2020		C 5.7
Pandemics	5	20.62	2020		C 5.7
Coronavirus	2	20.37	2020	2021	C 5.7
Repositories	7	8.78	2007	2013	C 5.7
Big data	3	7.59	2016	2018	C 5.7
Databases	11	6.81	2006	2016	C 5.7
Crisis communication	1	5.76	2020	2020	C 5.7
Ethics	1	6.47	2017	2017	C 6

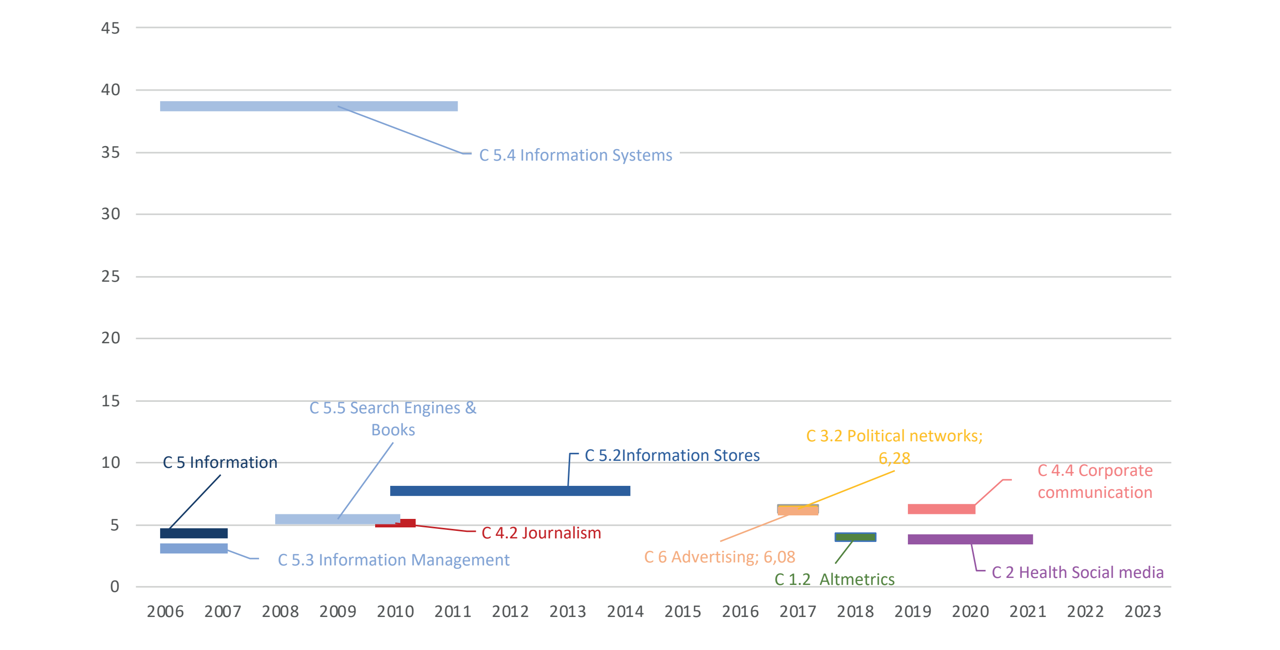


Figure 9. The communities' bursty periods.

The most notable burst period is that corresponding to the keyword *Web 2.0*, with a strength of almost 24 and a period that began in 2007 and ended in 2012.

The next three keywords, *Covid-19*, *Pandemics*, and *Coronavirus*, could well have been unified, all of them having a strength greater than 20 and having started in 2020. The periods relating to the first two have yet to be taken as concluded.

There are only two keywords with strength between 10 and 20. With 14.03, there is that corresponding to *Political communication* which lasted for the 4 years from 2017 to 2020, and in second place, with 10.53, that corresponding to *Indicators* which lasted only one year, 2018.

Table 5. The communities' bursty periods.

Community	Length	Strength	Start	End
C 1.2 Altmetrics	1	3.97	2018	2018
C 2 Health Social media	3	3.85	2019	2021
C 3.2 Political networks	1	6.28	2017	2017
C 4.2 Journalism	1	5.14	2010	2010
C 4.4 Corporate communication	2	6.23	2019	2020
C 5 Information	2	4.34	2006	2007
C 5 Information	1	3.15	2011	2011
C 5.2 Information Stores	5	7.69	2010	2014
C 5.3 Information Management	2	3.13	2006	2007
C 5.3 Information Management	3	5.28	2011	2013
C 5.4 Information Systems	6	38.68	2006	2011
C 5.5 Search Engines & Books	3	5.42	2008	2010
C 6 Advertising	1	6.08	2017	2017

Figure 9 and Table 5 present the bursty periods of the different communities, both first level and second level. The period of the *C 5.4 Information Systems* community stands out above all for its strength, which is more than four-fold that of the next. This period occurred between 2006 and 2011, and one can see from Table 2 that the said community does not increase during the production period.

The only first-level community (with second-level communities) that has separate bursts is *C 5 Information* which had two brief bursts of moderate intensity in 2006 and 2011.

There also stands out the digitalization fostered burst from 2010 to 2014 of the community denominated *C 5.2 Information Stores*. Also recognizable are the burst of the community denominated *C 3.2 Political networks* with the end of bipartisanship in Spain, that of *C 2 Health Social media* with the pandemic, and that of *C 1.2 Altmetrics*.

4. Conclusions

Since 2006, the journal under study has had its articles indexed in the major scientific literature databases. During this period, the journal has progressed considerably in both the quantity and the quality of what it has published. The number of published articles has increased, as have its international collaboration and average normalized impact.

The journal's content can be represented by the authors' keywords since more than 98% of the works contain keywords. Indeed, the 279 most used keywords are sufficient to represent its content since they are present in more than 92% of the keyword-containing works.

Co-word analysis identified 6 top-level thematic areas in the journal. Four of these first-level communities subdivide into 16 more communities.

The journal can be said to combine content of Library and Information Science with content of high technology which is where it comes from, together with other content of Audiovisual Communication. This gives rise to one front of Communication, another of Information, and two smaller ones of Informetrics and Social Networks.

While the Library and Information Science content occupies the central part of the map, serving as mortar that holds the rest of the areas together, it is the type with the slowest growth. Specifically, the area labeled *C 5.7 Information Subjects* is spread across much of the map. The other major central theme is *C 4 Communication* which is distributed over the left part of the map, also touching most of the other areas. It is smaller than the previous area because it was incorporated later, but it has a rapid growth rate.

The *C 2 Health Social media* and *C 3 Social Networks* areas are very close to each other and strongly related. The former obtains the greater impacts, although the latter's impact values are also good.

C 1 Informetrics is one of the most specialized areas and another of those which obtain the greatest average impacts.

C 6 Advertising is seen to be related to communication, although it is very specialized.

Different burst periods are observed, notable being that of *C 5.4 Information Systems* which can be regarded as the beginnings of the journal. The burst periods of the first part of the period correspond to areas included within *C 5 Information*. A burst of another area is not seen until 2010, specifically in *C 4.2 Journalism*. Also recognizable is the burst of the community denominated *C 3.2 Political networks* with the end of bipartisanship in Spain, or that of *C 2 Health Social media* with the pandemic, or of *C 1.2 Altmetrics*.

Periods of a keyword's boiling (when it has a sudden particular strength) are more recognizable and easier to detect. The most intense was that of *Web 2.0* from 2007 to 2012. Those corresponding to *Covid-19* keywords are also very intense. From 2017 to 2020 there is one of *Political communication* and in 2018 another of *Indicators*.

In 2006, *EPI* began to be indexed by international scientific literature databases, and entered a virtuous cycle that has led it to successfully expand its thematic scope, thus pushing it to levels of impact and excellence superior to those it originally had.

5. References

Blázquez-Ruiz, J.; **Guerrero-Bote, V. P.**; **De-Moya-Anegón, F.** (2016). New Scientometric-Based Knowledge Map of Food Science Research (2003 to 2014). *Comprehensive Reviews in Food Science and Food Safety*, 15(6), 1040-1055.
https://doi.org/10.1111/1541-4337.12223

Blázquez-Ruiz, J.; **Guerrero-Bote, V. P.**; **De-Moya-Anegón, F.** (2017). Nuevo mapa conceptual de la investigación científica en ciencias de los alimentos basado en técnicas cienciométricas en España (2003-2014). *Ibersid: revista de sistemas de información y documentación*, 11(1), 13-31.
https://doi.org/10.54886/ibersid.v11i1.4403

Bornmann, L. (2011). Scientific peer review. *Annual Review of Information Science and Technology*, 45(1), 197-245.
https://doi.org/10.1002/aris.2011.1440450112

Bornmann, Lutz; **De-Moya-Anegón, Félix**; **Leydesdorff, Loet** (2012). "The new excellence indicator in the World Report of the SCImago Institutions Rankings 2011". *Journal of Informetrics*, v. 6, n. 2, pp. 333-335.
https://doi.org/10.1016/j.joi.2011.11.006

Callon, M.; **Courtial, J.**; **Laville, F.** (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics*, 22, 155-205.
https://doi.org/10.1007/BF02019280

Callon, M.; **Rip, A.**; **Law, J.** (eds.). (1986). *Mapping the dynamics of science and technology: Sociology of science in the real world*. Springer.
https://doi.org/10.1007/978-1-349-07408-2

Clauset, A.; **Newman, M.**; **Moore, C.** (2004). Finding community structure in very large networks. *Physical Review E*, 70(6).
https://doi.org/10.1103/PhysRevE.70.066111

Hassan-Montero, Y.; **De-Moya-Anegón, F.**; **Guerrero-Bote, V. P.** (2022). SCIMago Graphica: a new tool for exploring and visually communicating data. *Profesional de la Información*, 31(5), e310502.
https://doi.org/10.3145/epi.2022.sep.02

Herrera-Viedma, E.; **López-Robles, J.-R.**; **Guallar, J.**; **Cobo, M.-J.** (2020). Global trends in coronavirus research at the time of Covid-19: A general bibliometric approach and content analysis using SciMAT. *Profesional de la Información*, 29(3).
https://doi.org/10.3145/epi.2020.may.22

Irvine, J.; **Martin, B.**; **Peacock, T.**; **Turner, R.** (1985). Charting the decline of British science. *Nature*, 316, 587-590.
https://doi.org/10.1038/316587a0

Kleinberg, J. (2003). Bursty and Hierarchical Structure in Streams. *Data Mining and Knowledge Discovery*, 7(4), 373-397.
https://doi.org/10.1023/A:1024940629314

Lechuga-Sancho, M.-P.; **Martínez-Fierro, S.**; **Ramos-Rodríguez, A. R.** (2023). Thirty years of research on high-growth entrepreneurship: bibliometric overview of its H-Classics. *Profesional de la Información*, 32(3).
https://doi.org/10.3145/epi.2023.may.20

López-Robles, J. R.; **Guallar, J.**; **Otegi-Olaso, J. R.**; **Gamboa-Rosales, N. K.** (2019). "El Profesional de la Información (EPI): bibliometric and thematic analysis (2006-2017)". *El Profesional de la Información*, v. 28, n. 4, e280417.
https://doi.org/10.3145/epi.2019.jul.17

Noack, A. (2007). Energy models for graph clustering. *Journal of Graph Algorithms and Applications*, 11(2), 453-480.
https://doi.org/10.7155/jgaa.00154

Noack, A. (2009). Modularity clustering is force-directed layout. *Physical Review E*, 79(2).
https://doi.org/10.1103/PhysRevE.79.026102

Olmeda-Gómez, C.; **Ovalle-Perandones, M. A.**; **Perianes-Rodríguez, A.** (2017). Co-word analysis and thematic landscapes in Spanish information science literature, 1985-2014. *Scientometrics*, 113(1), 195-217.
https://doi.org/10.1007/s11192-017-2486-8

Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, 14(3), 130-137.
https://doi.org/10.1108/eb046814

Rehn, Catharina; **Kronman, Ulf** (2008). *Bibliometric handbook for Karolinska Institutet*. Karolinska Institutet University Library. Version 1.05.
https://doi.org/10.13140/RG.2.1.1480.9447

Romo-Fernández, L. M.; **Guerrero-Bote, V. P.**; **Moya-Anegón, F.** (2013). Co-word based thematic analysis of renewable energy (1990-2010). *Scientometrics*, 97(3), 743-765.
https://doi.org/10.1007/s11192-013-1009-5

Segado-Boj, F.; **Gómez-García, S.**; **Díaz-Campo, J.** (2022). Intellectual and thematic structure of Communication research in Scopus (1980-2020). A comparative perspective among Spain, Europe, and Latin America. *Profesional de la Información*, 31(1).
https://doi.org/10.3145/epi.2022.ene.10

Velasco-López, J. E.; **Carrasco, R.-A.**, **Cobo, M. J.**; **Fernández-Avilés, G.** (2023). Data-driven scientific research based on public statistics: a bibliometric perspective. *Profesional de la Información*, 32(3).
https://doi.org/10.3145/epi.2023.may.14

***Profesional de la información* (*EPI*) journal no longer accepts subscriptions.**

Since January 1, 2023, EPI is a 100% Open Access journal, that is, all published content is available free.

If you wish to receive a summary when a new issue is published, please register here:

https://revista.profesionaldelainformacion.com/index.php/EPI/user/register

RULES FOR AUTHORS

Contributions must be original and unpublished, they cannot have been previously published on paper or electronically.

The text must be sent in Word or rtf. The tables must be pasted in the same document. All graphic materials (diagrams, photographs, screenshots, etc.) must be pasted into Word and also sent in separate files (in xls, jpg, pdf, etc. formats). Jpg images must have a resolution of at least 300 pp (about 200 KB each).

The text must be completely plane, without Word autoformats or automatisms (subsections, bullets, linked citations, footers, indentations, tabs, colors, etc.), but must follow the EPI style in terms of bold (names of cited authors), italics (institutions, journal titles, brands) and capital letters. The urls must be without hyperlink.

Bibliographic citations in the text will be made in the form: (Surname, year) or (**SurnameAuthor1**; **SurnameAuthor2**, year).

The wording must be concise and precise, avoiding rhetoric.

Papers must include: a) title, b) abstract of 250-300 words, c) 14-15 keywords

In addition, the name of the authors, their place of work and address (postal and electronic), their passport type photo greater than 100 KB in jpg, and their orcid will be included.

PEER REVIEW

Papers are reviewed according to the traditional double-blind system by at least two experts on the subject, from the journal's Advisory Board and/or external experts. The journal undertakes to inform the authors of the result.

BIBLIOGRAPHIC REFERENCES

Sorted alphabetically by author, they will be limited to the works cited in the article.

No filler bibliography or "for more information" is accepted.

Journal article:

Lastname name1, First name1; **Lastname2, First name2** (year). "Article title". *Journal name*, v., n., pp. xx-yy.
Url address starting in a new line, without period at the end
https://doi.org/10.xxxx/doi

Conference paper:

Lastname name1, First name1; **Lastname2, First name2** (year). "Title of paper". In: Name of the congress, pp. xx-yy.
Url address starting in a new line, without period at the end
https://doi.org/10.xxxx/doi

Books and reports:

Lastname name1, First name1; **Lastname2, First name2** (year). *Work title*. Place of publication: Publisher. ISBN: spaces, without hyphens and without period at the end
Url address starting in a new line, without period at the end
https://doi.org/10.xxxx/doi

Book chapter:

Lastname name1, First name1; **Lastname2, First name2** (year). "Chapter title". In: Lastname name1, First name1; Lastname2, First name2. *Title of the book*. Place of publication: Publisher, pp. xx-yy. ISBN: spaces, without hyphens or end period
Url address starting in a new line, without period at the end
https://doi.org/10.xxxx/doi

Online resource:

Lastname name1, First name1; **Lastname2, First name2** (year). *Title of the resource (without date of consultation)*.
Url address starting in a new line, without period at the end
https://doi.org/10.xxxx/doi

All contributions have to be sent to *EPI*'s *OJS* platform
https://revista.profesionaldelainformacion.com/index.php/EPI

Previously, authors must register at:
https://revista.profesionaldelainformacion.com/index.php/EPI/user/register