

tosr: Create the Tree of Science from WoS and Scopus

Sebastian Robledo^{1,*}, Luis Valencia², Martha Zuluaga¹, Oscar Arbelaiz Echeverri³, Jorge W. Arboleda Valencia^{4,5}

¹Dirección Académica, Universidad Nacional de Colombia, Sede de La Paz, La Paz 202017, Cesar, COLOMBIA.

²Centro de Bioinformática y Biología Computacional de Colombia – BIOS, COLOMBIA.

³Core of Science, IRELAND.

⁴FITOBIOLOGÍA, Instituto de Biología, Facultad de Ciencias Exactas y Naturales, Universidad de Antioquia, Cl. 67 # 53-108, Medellín 050010, COLOMBIA.

⁵Centro de Investigaciones en Medio Ambiente y Desarrollo—CIMAD, Universidad de Manizales, Cra. 9a # 19-03, Manizales 170001, COLOMBIA.

ABSTRACT

The R package 'tosr' enables the construction of the Tree of Science (ToS), a metaphorical representation of scientific papers on a specific topic. The ToS's roots symbolize seminal works, the trunk stands for structural works, and the leaves depict the current literature. Traditionally, researchers have had to limit their ToS to data from a single database, such as Scopus or Web of Science (WoS). The 'tosr' package overcomes this limitation by allowing researchers to merge seed files from both Scopus and WoS, thereby facilitating a more comprehensive bibliometric analysis. This paper describes the development and application of the 'tosr' package, demonstrating its unique capabilities in creating a complete and more cohesive ToS and citation network for any scientific topic. By bridging the gap between these two major databases, 'tosr' offers researchers an unprecedented tool for scientometric research.

Keywords: tosr, citation network, scientometrics, Tree of Science, bibliometrics, Web of Science, Scopus, R package.

Correspondence:

Sebastian Robledo

Dirección Académica, Universidad Nacional de Colombia, Sede de La Paz, La Paz 202017, Cesar, COLOMBIA.

Email: srobledog@unal.edu.co

ORCID ID: 0000-0003-4357-4402

Received: 26-07-2023;

Revised: 10-04-2024;

Accepted: 10-05-2024.

INTRODUCTION

The digital age has fueled an exponential surge in academic literature production and accessibility. While reachability was an issue before the 20th century, the current challenge lies in managing the overwhelming volume of new research. In response to this, the Tree of Science (ToS) metaphor has proven instrumental in streamlining the identification of pertinent works within academic literature.^[1] Leveraging graph theory, the ToS algorithm positions papers within a tree structure: classic or seminal works as roots, structural works as the trunk, and the latest research as leaves.^[2,3]

An evolved version of this algorithm, known as the SAP algorithm, has further improved the accuracy of results, especially within the leaves.^[4] In furtherance of this work, the 'tosr' package seeks to automate the SAP algorithm while facilitating the merger of Scopus and Web of Science data.

The SAP algorithm, extensively detailed by Valencia-Hernandez *et al.*,^[4] has expedited the creation of review papers. For instance, Duque *et al.*^[5] and Tabares & Duque^[6] have effectively applied it

to identify significant literature on social economy and school cyberbullying, respectively. The integration of the SAP algorithm with machine learning techniques has enabled the testing of different software.^[7] The ToS has also proved valuable in guiding early-career researchers toward essential academic papers on specific topics.

Notable examples include Rubio *et al.*'s discourse on the governance of tourist destinations^[8] and Uribe *et al.*'s work on blended learning in education.^[9] Additionally, the ToS has enhanced the visibility of academic papers within the scientific community, as evidenced by the increased citation rates of works like Grisales-Aguirre *et al.*,^[10] Ariza-Colpas *et al.*,^[11] and Hernández-Leal.^[12]

Figure 1 elucidates the overall workflow of 'tosr.' To generate a mixed source Tree of Science, users should download two files—one from Scopus (.bib) and another from Web of Science (WoS, .txt)—both containing the references of each paper. Subsequently, an RStudio cloud session should be initiated, activating the 'tosr,' 'biblometrix',^[13] and 'tidyverse'^[14] libraries for data interaction.

The primary function of the package, `tosr()`, accepts the two files as input, transforming the data into a data frame that designates papers as roots, trunks, and leaves. The `tosr_load()` function creates three separate files, each representing a part of the tree and ready for in-depth analysis using different programs.



DOI: 10.5530/jscires.13.2.36

Copyright Information :

Copyright Author (s) 2024 Distributed under
Creative Commons CC-BY 4.0

Publishing Partner : EManuscript Tech. [www.emanuscript.in]

The 'tosr' package complements existing scientometric analysis packages in both R and Python—like 'bibliometrix' with its sophisticated algorithms and aesthetic figures,^[13] 'pybliometrics' for Scopus data access,^[15] 'scientopy' for general scientometric analysis,^[16] and 'litstudy' for network analysis.^[17]

In addition to its stand-alone capabilities, 'tosr' can preprocess data for other scientometric software, such as VOSviewer,^[18] Citespace,^[19] and VantagePoint. It also allows for the export of input files suitable for software like Gephi^[20] for more elaborate scientometric data analysis.

Overview

A stable version of *tosr* is available at CRAN¹ (Comprehensive R Archive Network) and the current development version is on Github.²

"tosr" architecture

The 'tosr' package comprises three core functions: *tosR()*, *tosr_load()*, and *tosSAP()*. Its operation requires two input files—a .txt file from WoS encompassing all records and cited references, and a .bib file from Scopus containing comprehensive data (Figure 1). Utilizing the *bibliometrix* package's *convert2df* function,^[13] 'tosr' converts these files into data frames. The *mergeBD* function^[13] subsequently merges data devoid of references, facilitating the creation of an ID_TOS derived from reference data, specifically, the first author's last name and publication year. Notably, despite the distinct reference formats of WoS and Scopus, their ID_TOS data align.

Figure 1 outlines the 'tosr' package operation. The *tosR()* function employs Scopus and WoS files to generate a ToS pertinent to a specified research topic. There exist two methodologies for creating the ToS—a concise path and a comprehensive one.

In the concise path, the user submits the WoS and Scopus files to the *tosR()* function, which returns a dataframe distinguishing papers into roots, trunks, and leaves. Conversely, the comprehensive path necessitates the use of both *tosr_load()*

and *tosSAP()* functions. The former receives Scopus and WoS files, generating a citation network, a merged dataframe, and a dataframe containing reference names (WoS and Scopus). This process benefits users seeking a refined citation network analysis. The subsequent *tosSAP()* function employs these three files to yield a dataframe that classifies papers into roots, trunks, and leaves.

"tosr" functionalities

The 'tosr' package generates a citation network from WoS and Scopus files. Recognizing the disparate reference formats between the two databases, 'tosr' establishes a common identifier (ID_TOS) drawn from both reference types, enabling their unification and merging. In this citation network, nodes represent papers, while links symbolize references connecting two papers. Thus, if paper A cites paper B, it creates a link—a directed graph that, by definition, is unidirectional as the cited paper cannot reciprocate the citation.

The 'tosr' package ensures the citation network's integrity by extracting the giant component^[21] and eliminating nodes (papers) exhibiting an in-degree of 0 and an out-degree of 1.

The refined network is then ripe for SAP algorithm application. Mirroring a tree's sap process, the SAP algorithm first identifies the most frequently cited papers (those with a high citation count and zero out-degree), papers with the highest number of references to others, and those with significant betweenness. Upon determining these metrics, the SAP algorithm initiates the shortest path identification process among the paper groups. Finally, the algorithm designates papers into roots, trunks, and leaves (For a detailed explanation, refer to Valencia-Hernandez *et al.*).^[4]

1 <https://cran.r-project.org/web/packages/tosr/index.html>

2 <https://github.com/coreofscience/tosr>

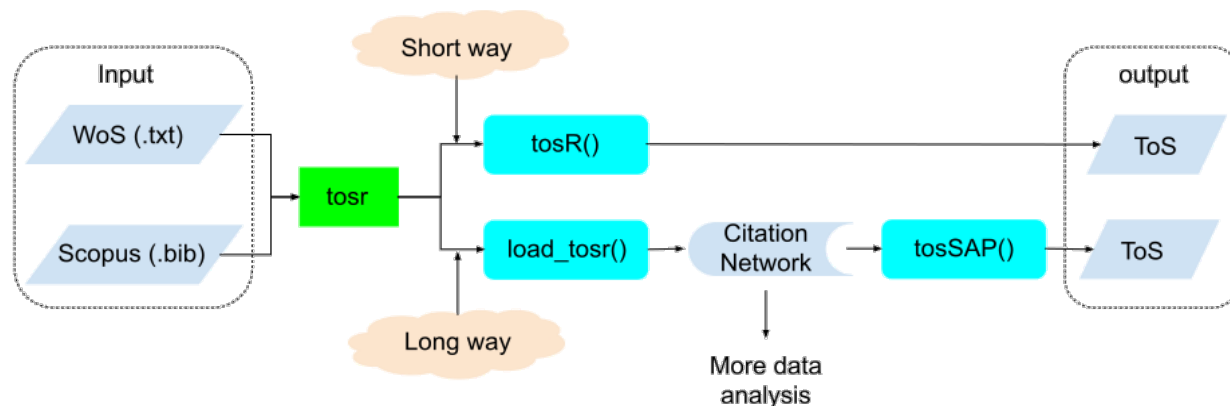


Figure 1: General overview to create the Tree of Science with tosr.

METHODOLOGY

Data Acquisition

This section illustrates the process of constructing the ToS using the 'scientometrics' topic as an example. To start, the user needs to download data from both WoS and Scopus databases, ensuring that references to the papers are included. It's important to note that access to these databases was facilitated through licenses provided by one of the institutions affiliated with the researchers, ensuring that the use of these data complies with all relevant copyright and licensing agreements. Figure 2 provides examples from both platforms. In WoS, the user should opt for a 'plain text file' configured to include 'Full Record and Cited References.' Simultaneously, in Scopus, the user should select a BibTeX file encompassing all the relevant information.

RESULTS

Creating ToS - Concise Method

In the concise method, the user employs the `tosR()` function to construct the ToS for the 'Scientometrics' topic using the acquired files from WoS and Scopus. The analysis necessitates the activation of 'tosr', 'tidyverse', and 'bibliometrix' libraries. Source Code 1 delineates the code required to generate the ToS.

Source Code 1: Example code to create ToS with tosr in a short way.

```
library(tosr)
library(tidyverse)
library(bibliometrix)
# Data loading and creating ToS
ToS_short <- tosr::tosR("scientometrics.txt",
"scientometrics.bib")
```

The outcome of this process is a dataframe classifying papers into roots, trunks, and leaves. Table 1 presents an exemplar ToS, featuring the first three papers from each tree segment. The metaphor unravels the narrative of 'Scientometrics,' commencing with Hirsh's seminal paper introducing the h-index,^[22] followed by the trunk's representation of HistCite software's application in evaluating scientometrics' impact.^[23] The tree concludes with a leaf signifying a contemporary paper on scientometrics' application in photocatalytic degradation.^[24]

Creating ToS - Comprehensive Method

The comprehensive method of generating a ToS from a research topic serves users desiring a citation network for more detailed data analysis (refer to section 3.4). The `tosr_load()` function utilizes WoS and Scopus files as inputs, generating a list comprising three elements: a dataframe merging Scopus and WoS files (`df`), a graph depicting the citation network (`graph`), and a dataframe listing the names of the papers (`nodes`) (Source Code 2).

Source Code 2: Sample code illustrating the comprehensive method of creating ToS using 'tosr'.

```
library(tosr)
library(tidyverse)
library(bibliometrix)
# Creating the merged dataframe, graph, and nodes dataframe
tosr_files <- tosr::tosr_load("scientometrics.txt",
"scientometrics.bib")
ToS_large <- tosr::tosSAP(graph = tosr_files$graph,
df = tosr_files$df,
nodes = tosr_files$nodes)
```

The variable 'tosr_files' is a list comprising the objects 'df', 'graph', and 'nodes'. 'tosr_files\$df' incorporates a dataset of 628 rows and 34 columns. 'tosr_files\$graph' represents an 'igraph' object with 1568 nodes (corresponding to papers) and 3585 links (signifying references). 'tosr_files\$nodes' constitutes a dataframe featuring 13287 paper names. The `tosSAP()` function utilizes these three variables to construct the ToS for the 'Scientometrics' topic. Consequently, 'ToS_large' is generated as a variable with 75 rows and 2 columns.

Extended Scientometric Analysis - Citation Network

The citation network forms the core of the ToS process and can be created using the 'tosr' package. Figure 3 exhibits a scientometrics citation network, segmented into three clusters.

The directed citation network comprises 1,568 nodes (papers) and 3,585 edges (references). Figure 3a categorizes the ten most significant subfields using clustering analysis,^[25] with only three selected for detailed study. Figure 3b illustrates the longitudinal evolution of these clusters; despite being the largest, cluster 1 has displayed diminished production over the past four years. Figure 3c provides a visual representation of the citation network, where cluster 1 is the densest, primarily because it contains seminal papers such as^[26] and^[27] renowned and frequently cited in academic circles.

The initial phase of scientometric analysis involves data acquisition and preparation, given that WoS and Scopus export data in .txt and .bib formats, respectively. Meanwhile, software tools such as R and Python necessitate a dataframe or JSON structure for effective data analysis. Source Code 3 outlines the principal libraries ('tosr', 'tidyverse', 'bibliometrix', and 'tidygraph') required to transform the data. The `tosr_load()` function, sourced from the 'tosr' package, generates three files, one of which is the citation network in an 'igraph' object format. This is subsequently converted into a 'tidygraph' object. The 'tidygraph' package facilitates the addition of attributes and metrics to the network, employing 'tidyverse' syntax for enhanced ease of use.

³ Permanent link to reproducible capsule <https://codeocean.com/capsule/4217921/tree>

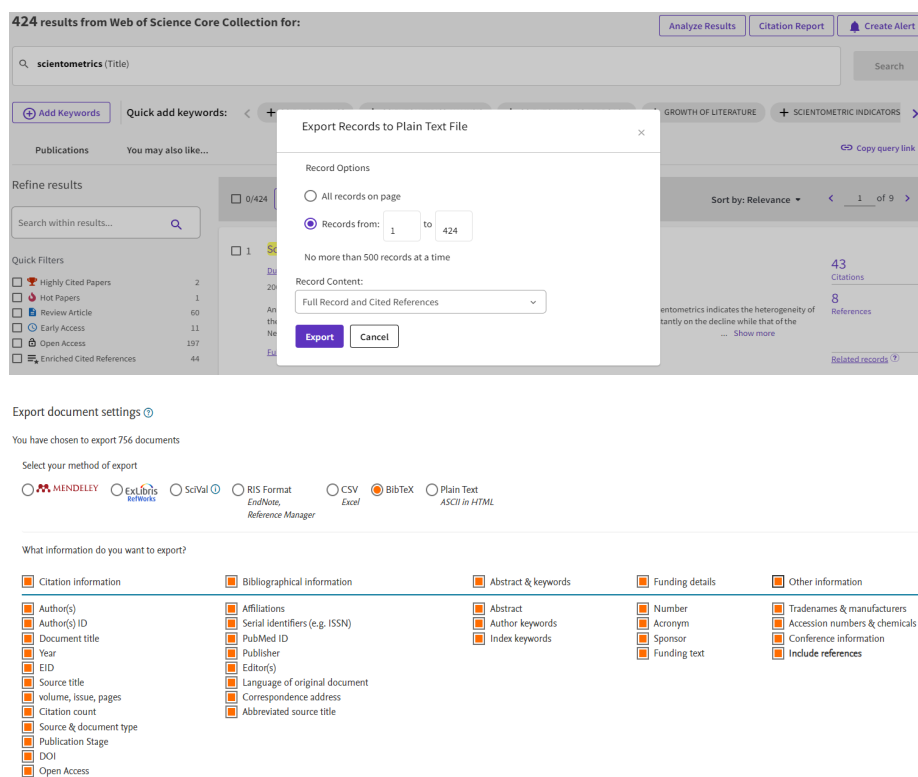


Figure 2: Data features from WoS and Scopus.

Table 1: Key scientometric papers in the Tree of Science hierarchy.

ToS	Paper
Roots	HIRSCH JE, 2005, P NATL ACAD SCI USA, V102, P16569, DOI 10.1073/PNAS.0507655102
Trunk	GARFIELD E, 2009, FROM THE SCIENCE OF SCIENCE TO SCIENTOMETRICS VISUALIZING THE HISTORY OF SCIENCE WITH HISTCITE SOFTWARE, JOURNAL OF INFORMETRICS
Leaf	BRINDHA R; RAJESWARI S; JENNET D J; RAJAGURU P, 2022, EVALUATION OF GLOBAL RESEARCH TRENDS IN PHOTOCATALYTIC DEGRADATION OF DYE EFFLUENTS USING SCIENTOMETRICS ANALYSIS, JOURNAL OF ENVIRONMENTAL MANAGEMENT

Source Code 3: Getting and tidying data.

```
# Creating the environment
library(tosr)
library(tidyverse)
library(bibliometrix)
# Data getting
tosr_files <-
tosr::tosr_load("scientometrics_scopus.bib",
"scientometrics_wos.txt")# Data tidying
tosr_citation_network <-
tosr_files$graph %>% tidygraph::as_tbl_graph() %>%
tidygraph::activate(nodes) %>% dplyr::left_join(tosr_
files$nodes,
by = c("name" = "ID_TOS")) %>% dplyr::mutate(PY =
stringr::str_extract(string = name,
pattern = "[0-9]{4}"), in_degree = tidygraph::centrality_
degree(mode = "in"), out_degree = tidygraph::centrality_
degree(mode = "out"))
```

Source Code 4 presents the code to create Figure 3a. The `tosr_citation_network` created before (Source Code 3) is transformed into a data frame to count the number of papers in each subfield.

Source Code 4: Example code to create the subfields figure.

```
figure_subfields <- tosr_citation_network |>
activate(nodes) |> data.frame() |> dplyr::count(subfield) |>
slice(1:10) |> ggplot(aes(x = reorder(subfield, n), y = n)) +
geom_point(stat = "identity") + geom_line(group = 1) +
labs(title = "Subfields by size", x = "Subfields", y = "papers") +
theme(text = element_text(color = "black",
face = "bold", family = "Times New Roman"), plot.title =
element_text(size = 12), panel.background = element_rect(fill =
"white"),
axis.text.y = element_text(size = 12,
colour = "black"), axis.text.x = element_text(size = 12,
colour = "black"), axis.title.x = element_text(size = 12), axis.
title.y = element_text(size = 12))
```

To execute a longitudinal analysis of a citation network, it is crucial to incorporate zero values using the `'pivot_wider'` function from the `'tidyr'` library, and subsequently select the three target subfields. This approach ensures accurate representation and

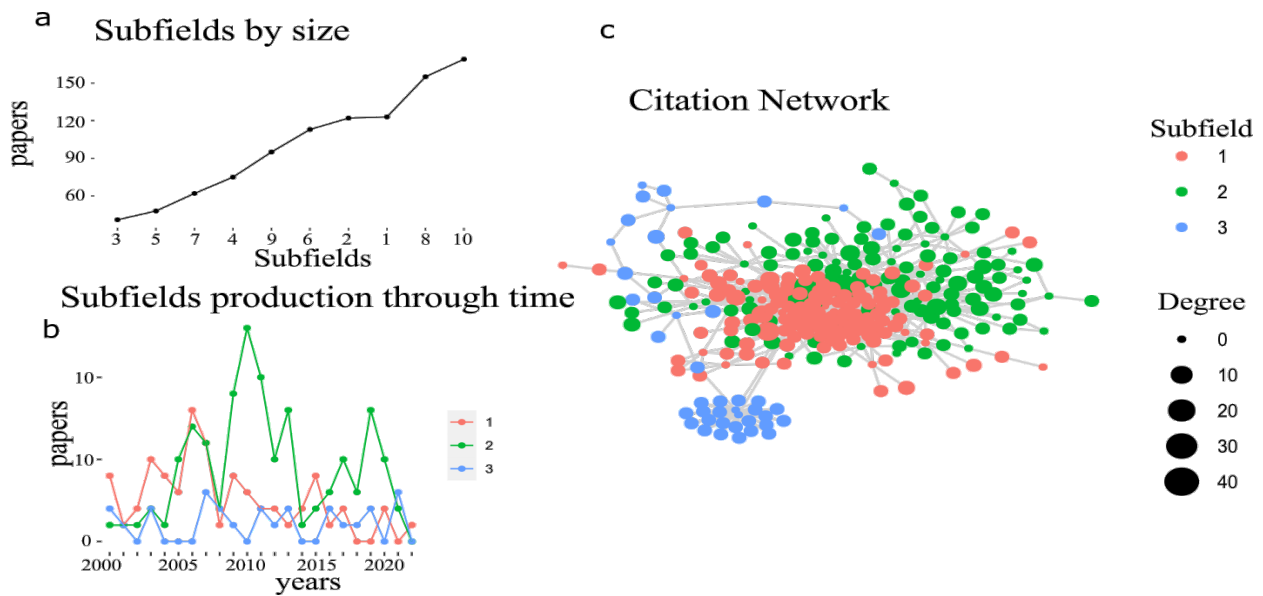


Figure 3: Citation network for scientometrics using tosr. a. Biggest subfields and ordered by size. b. Longitudinal analysis of the three biggest clusters since 2000. c. Citation Network of the three main clusters.

examination of temporal trends in the citation network (Source Code 5).

Source Code 5: Example code to create the longitudinal analysis.

```
tosr_graph <- tosr_citation_network %>%
tidygraph::activate(nodes) %>% tidygraph::as_tibble() %>%
select(PY, subfield) %>% filter(subfield <= 3, PY >= 2000)
%>% tidyr::drop_na() %>% group_by(PY, subfield) %>%
dplyr::count(PY) %>% tidyr::pivot_wider(names_from =
subfield, values_from = n) %>%
replace(is.na(.), 0) %>% tidyr::pivot_longer(cols = !PY,
names_to = "subfield", values_to = "n")
figure_subfields_time <-
tosr_graph %>% ggplot(aes(x = factor(PY),
y = n, group = as.character(subfield), color =
as.character(subfield))) + geom_point() + geom_line(stat =
"identity") +
theme( legend.position = "right",
text = element_text(color = "black",
face = "bold",
family = "Times New Roman"), plot.title = element_text(size
= 12), panel.background = element_rect(fill = "white"), axis.
text.y = element_text(size = 12,
colour = "black"), axis.text.x = element_text(#size = 12,
colour = "black", angle = 45,
vjust = 0.5), axis.title.x = element_text(size = 12), axis.title.y =
element_text(size = 12),
# legend.text = element_text(size = "12"), legend.title =
element_blank()
) + labs(title = "Subfields production through time", y =
"Papers", x = "years")
```

The visualization of a citation network is a prevalent component of the scientometric analysis, often included in academic papers for better representation and understanding. The 'ggraph' package is

particularly well-suited for creating such network visualizations. The provided Source Code 6 demonstrates an example of generating a network visualization using the pre-constructed 'tosr_citation_network'.

Source Code 6: Example code to create the citation network.

```
figure_citation_network <- tosr_citation_network %>%
tidygraph::activate(nodes) %>% tidygraph::filter(subfield <=
3) %>% tidygraph::mutate(Subfield = as.character(subfield))
%>% ggraph(layout = "graphopt") + geom_edge_link(
colour = "lightgray") + scale_edge_width(name = "Link
strength") + geom_node_point(aes(color = Subfield,
size = in_degree)) +
# geom_node_text(aes(label = name), repel = TRUE) +
scale_size(name = "Degree") +
# scale_color_binned(name = "Communities") + theme_
graph() +
labs(title = "Citation Network") # Plotting the citation
networklibrary(patchwork)fig_abc <- (figure_subfields /
figure_subfields_time) | figure_citation_network
fig_abx + plot_annotation(tag_leves = "a")
```

Influence and Implications

The 'tosr' package facilitates researchers in extracting the Tree of Science (ToS) from a research topic, leveraging datasets from Web of Science (WoS) and Scopus. This serves as a potent tool to address exploratory research queries concerning their respective research topics, for instance, delineating significant contributions from inception to contemporary literature.

One of the salient features of 'tosr' is its ability to construct a citation network utilizing the two most widely employed datasets, WoS and Scopus. This citation network enables researchers to

conduct more complex analyses, such as clustering and topic modeling.

Constructed in R, an open-source language accessible via CRAN, 'tosr' simply necessitates an understanding of R code from its users. It aligns with the Tidyverse ethos, a collection of R packages for data analysis that adhere to a unified philosophy and grammar. Thus, researchers can leverage the advantages of the R language and its dynamic community to conduct data analysis.

The 'tosr' package allows researchers to gain an understanding of a research topic via ToS by amalgamating WoS and Scopus data, offering limited restrictions on the volume of data they wish to process. The package can be employed either on a personal computer or in the cloud (via rstudio.cloud). Users have the ability to download up to 100,000 records per day from WoS and 2,000 from Scopus, integrate them via 'tosr', and subsequently perform statistical analysis. Moreover, users have the option to export the data in an Excel format for uploading to 'biblioshiny' (a Shiny application of bibliometrix), facilitating more intricate data analyses.

The 'tosr' package is a component of the Core of Science suite of products, employed in courses at various universities and institutions. The ToS concept was initiated with a web application exclusively for WoS data,^[2] and subsequently expanded to Scopus data.^[3] The 'tosr' package paves the way for the development of new web applications that can allow users to amalgamate both datasets, fostering the prospect of new courses with in-depth scientometric analyses. The impact of the ToS concept is evident, with over one hundred citations for the inaugural paper.^[28]

CONCLUSION

In this scholarly discourse, we have succinctly outlined the salient functionalities of the 'tosr' package, demonstrated its application in a scientometrics investigation, and elucidated its profound implications on academic research. The primary merits of the 'tosr' package lie in its capacity to amalgamate Scopus and WoS datasets, thereby enabling more sophisticated data analysis, and its ability to construct the Tree of Science for a research topic.

Researchers stand to benefit immensely from this package as it simplifies the process of tracing the evolutionary trajectory of a specific research topic. Furthermore, this package can be complemented with additional tools and software to augment scientometric analysis, including topic modeling and burst analysis.

Future scholarly pursuits can concentrate on establishing a tidyverse environment conducive to data preprocessing and conceptualizing a web application capable of handling both WoS and Scopus. This innovative development in scientometrics significantly extends the scope of data analysis, thus enriching the process of knowledge discovery in various research fields.

ACKNOWLEDGEMENT

This work was also made possible by the financial support of Minciencias, Ministry of Science, Technology and Innovation of Colombia. Also, the Core of Science corporation supported this research.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Eggers F, Risselada H, Niemand T, Robledo S. Referral campaigns for software startups: The impact of network characteristics on product adoption. *J Bus Res* 2022;145:309-24. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S0148296322002351>
2. Zuluaga M, Robledo S, Arbelaez-Echeverri O, Osorio-Zuluaga GA, Duque-Méndez N. Tree of Science - ToS: A Web-based Tool for Scientific Literature Recommendation. *Search Less, Research More! Issues in Science and Technology Librarianship* 2022;100. DOI: 10.29173/istl2696
3. Robledo S, Zuluaga M, Valencia LA, Arbelaez-Echeverri O, Duque P, Alzate-Cardona JD. Tree of Science with Scopus: A Shiny Application. *Issues in Science and Technology Librarianship* 2022;100. DOI: 10.29173/istl2698
4. Valencia-Hernandez DS, Robledo S, Pinilla R, Duque-Méndez ND, Olivar-Tost G. SAP Algorithm for Citation Analysis: An improvement to Tree of Science. *Ingeniería e Investigación* 2020;40(1). DOI: 10.15446/ing.investig.v40n1.77718
5. Duque P, Meza OE, Giraldo D, Barreto K. Economía Social y Economía Solidaria: un análisis bibliométrico y revisión de literatura. *REVESCO Revista de Estudios Cooperativos* 2021;138:e75566-e75566. Available from: <https://revistas.ucm.es/index.php/REVE/article/view/75566>
6. Tabares ASG, Duque MCC. La asociación entre acoso y ciberacoso escolar y el efecto predictor de la desconexión moral: una revisión bibliométrica basada en la teoría de grafos. *EducXXI* 2022;25(1):273-308. Available from: <https://revistas.uned.es/index.php/educacionXXI/article/view/29995>
7. Robledo S, Grisales-Aguirre AM, Hughes M, Eggers F. Hasta la vista, baby - will machine learning terminate human literature reviews in entrepreneurship? *J Small Bus Manage* 2021;
8. Rubio AE, Yepes GYF, Marín LAV. Gobernanza para el desarrollo y la sostenibilidad de los destinos turísticos: una revisión de la literatura con ToS. *Interfaces* 2022;5(1). Available from: <https://revistas.unilivre.edu.co/index.php/interfaces/article/view/9459>
9. Uribe JCV, Rocha ACS, Rodríguez OV, Tuberquia AO. Blended Learning: una revisión bibliométrica. *Interfaces* 2022;5(1). Available from: <https://revistas.unilivre.edu.co/index.php/interfaces/article/view/9458>
10. Grisales-Aguirre AM, Robledo S, Zuluaga M. Topic Modeling: Perspectives From a Literature Review. *IEEE Access* 2023;11:4066-78. Available from: <https://doi.org/10.1109/ACCESS.2022.3232939>
11. Ariza-Colpas PP, Piñeres-Melo MA, Morales-Ortega RC, Rodríguez-Bonilla AF, Butt-Aziz S, NazS, Contreras-Chinchilla LdC, Romero-Mestre M, Vacca Ascanio RA. Sustainability in Hybrid Technologies for Heritage Preservation: A Scientometric Study. *Sustainability* 2024;16:1991. <https://doi.org/10.3390/su16051991>
12. Hernández-Leal EJ, Duque-Méndez ND, Moreno-Cadavid J. Big Data: una exploración de investigaciones, tecnologías y casos de aplicación. *Tecnologica* 2017;20(39):17-24. Available from: http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0123-77992017000200002
13. Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. *J Informetr* 2017;11(4):959-75.
14. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the tidyverse. *J Open Source Softw* 2019;4(43):1686. Available from: <https://joss.theo.org/papers/10.21105/joss.01686>
15. Rose ME, Kitchin JR. pybliometrics: Scriptable bibliometrics using a Python interface to Scopus. *SoftwareX* 2019;10:100263.
16. Ruiz-Rosero J, Ramirez-Gonzalez G, Viveros-Delgado J. Software survey: SciToPy, a scientometric tool for topics trend analysis in scientific publications. *Scientometrics* 2019;121(2):1165-88. Available from: <https://doi.org/10.1007/s11192-019-03213-w>
17. Heldens S, Sclocco A, Dreuning H, van Werkhoven B, Hijma P, Maassen J, et al. litstudy: A Python package for literature reviews. *SoftwareX* 2022;20(101207):101207.
18. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 2010;84(2):523-38. DOI: 10.1007/s11192-009-0146-3
19. Chen C. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *J Am Soc Inf Sci Technol* 2006;57(3):359-77. DOI: 10.1002/asi.20317

20. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. In: Third international AAAI conference on weblogs and social media. 2009. Available from: <https://www.aaai.org/ocs/index.php/ICWSM/09/paper/viewPaper/154>
21. Robledo S, Grisales Aguirre AM, Hughes M, Eggers F. "Hasta la vista, baby" – will machine learning terminate human literature reviews in entrepreneurship? *J Small Bus Manage* 2021;1-30. DOI: 10.1080/00472778.2021.1955125
22. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A* 2005;102(46):16569-72. DOI: 10.1073/pnas.0507655102
23. Garfield E. From the science of science to Scientometrics visualizing the history of science with HistCite software. *J Informetr* 2009;3(3):173-9.
24. Brindha R, Rajeswari S, Jennet Debora J, Rajaguru P. Evaluation of global research trends in photocatalytic degradation of dye effluents using scientometrics analysis. *J Environ Manage* 2022;318:115600. DOI: 10.1016/j.jenvman.2022.115600
25. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech: Theory Exp* 2008; 2008(10):P10008. DOI: 10.1088/1742-5468/2008/10/p10008
26. Bornmann L, Mutz R, Daniel HD. Are there better indices for evaluation purposes than the index? A comparison of nine different variants of the index using data from biomedicine. *J Am Soc Inf Sci Technol* 2008;59(5):830-7. DOI: 10.1002/asi.20806
27. Leydesdorff L, Opthof T. Scopus's source normalized impact per paper (SNIP) versus a journal impact factor based on fractional counting of citations. *J Am Soc Inf Sci Technol*. 2010;61(11):2365-9. DOI: 10.1002/asi.21371
28. Robledo S, Osorio-Zuluaga GA, Lopez-Espinosa C. Networking en pequeña empresa: una revisión bibliográfica utilizando la teoría de grafos. *Revista Vinculos* 2014;11(2):6-16.

Cite this article: Robledo S, Valencia L, Zuluaga M, Echverri OA, Arboleda JW. tosr: Create the Tree of Science from WoS and Scopus. *J Scientometric Res.* 2024;13(2):459-65.