

Title:

Citation Based Journal-to-Journal Associations in the microcosm of an Academic Libraries Consortium.

Authors:

- Leonidas Pispiringas (Corresponding author), University of Macedonia, 54636 Thessaloniki, Greece, lpisp@uom.edu.gr; lpisp@heal-link.gr, <http://orcid.org/0000-0002-7111-094X>
- Dimitris A. Dervos, International Hellenic University, 57400 Sindos, Greece, dad@ihu.gr, <http://orcid.org/0000-0002-5441-0086>
- Georgios Evangelidis, University of Macedonia, 54636 Thessaloniki, Greece, gevan@uom.gr, <http://orcid.org/0000-0003-1639-2152>

Declarations of interest: none

Title:

Citation Based Journal-to-Journal Associations in the Microcosm of an Academic Libraries Consortium.

Abstract:

Scholarly journals are usually rated on the basis of the number of citations received which directly translates to impact in the advancement of research. In this paper, the emphasis is on considering citations as a means for unveiling journal-to-journal associations. An Academic Libraries Consortium could use such kind of information in the process of planning for the collection of journal titles (subscriptions) that best serves the needs of its research user community. The innovative contribution is summarized as follows: (a) journal research impact measures are considered in the context of the well-defined user community (Target Authors Group, TAG) of the Hellenic Academic Libraries Link (HEAL-Link) Consortium, and (b) research impact measures are used to reveal interesting journal-to-journal associations, as well as journals of an interdisciplinary value, specifically for the microcosm of the given TAG community. The research outcomes are publicized by means of a number of interactive graphs via the J2J-GR web service (<https://j2j.heal-link.gr>). The latter is expected to be of use not just for the HEAL-Link office staff, but also for all the interested researchers, in Greece or abroad.

Keywords:

Academic Consortia; Citation Network; Citation Analysis; Data Visualization; Data Analyse Approach; Journal Evaluation

1. Introduction

The scholarly journal is a means for communicating ideas, knowledge and know-how to researchers within a specific or across many discipline areas. In this respect, it effectively promotes the evaluation and the advancement of scholarly research. Scholarly journals are usually rated on the basis of the impact they represent in the advancement of research, namely the number of citations received from articles published with other scholarly journals. The terms “citation” and “reference” are often used interchangeably. They are seen to imply two complementary actions: “reference” as “acknowledgment” and “citation” as “acknowledgment from” (Narin & Carpenter, 1975). It is assumed that when a researcher makes a reference to an article in her work, the journal the latter is published with deserves the credit to comprise a valuable source of information. The larger the number of references made to articles published with a given journal, the higher is the value of the latter as a source of information for the academic and research community.

The HEAL-Link consortium is funded by the Hellenic Ministry of Education to provide access to subscriptions of electronic journals to the Greek academic and research community. In 2019, the consortium involved 720,655 users, 12,276 of which were academic faculty members¹. During the 1998–2018 period its users conducted nearly 85 million article downloads from a total of 17,100 scholarly journal titles².

¹ Quality Assurance Unit of Academic Libraries: Usage reports for Academic Libraries for 2019

http://mopab.seab.gr/sites/default/files/mopab_report_2019_0.pdf

² HEAL-Link, <https://www.heal-link.gr/>

Access to the latter is made possible via license agreements HEAL-Link makes with publishers / providers of electronic content (scholarly journals, e-books, bibliographic databases, etc.). The consortium participates in projects and initiatives aiming to elevate and contribute added value to the academic publications landscape in Greece. Its main goals are to:

- establish common policies on electronic academic subscriptions, and promote the rational growth of its collection,
- combine financial savings with access to a large number of electronic academic content that meet the educational and research needs of its research community,
- negotiate the signing of joint subscription agreements with electronic academic content publishers / providers,
- facilitate and manage the remote access to its electronic academic content, and,
- develop and maintain information services for the management and the distribution of its electronic academic content to its members.

In the present research, a new viewpoint is established whereby the value a scholarly journal represents is not considered in the context of the global academic/research community. Instead, it is considered in the context of the microcosm of a libraries consortium, HEAL-Link in particular. Greek academics and researchers who (co-)author scholarly publications comprise the Target Authors Group (TAG) for HEAL-Link.

Considering the above, the value each scholarly journal represents to the consortium's TAG is used for justifying the investment made for its acquisition and inclusion into the HEAL-Link collection. In order to best exploit its financial resources and in order to meet the needs of its members when it comes to accessing scholarly publications, the consortium uses a set of criteria like the following:

- The scholarly journals the TAG members publish with, and the scholarly journals they reference in their works.
- The number of full-text downloads per scholarly journal over given time periods.
- The (possible) interdisciplinary value of each one scholarly journal.

Thus, on the basis of effectively negotiating with the publishers / providers, HEAL-Link elaborates on a strategic plan, part of which has to do with the evaluation of scholarly journals on the basis of their usage as reference (source) material by the HEAL-Link TAG. A scholarly journal, in addition to its impact at the international level, represents an asset for HEAL-Link when the TAG members reference its articles in their works, and, to a lesser extent, when they publish their works with it. In this respect, it makes sense to process bibliographic data in order to address questions like *“To what extent is a scholarly journal cited by the consortium's TAG members?”*, *“Considering the TAG articles published with a specific journal, which journals host articles included in the “References” sections of the former?”*, or, equivalently: *“To what extent is scholarly journal A associated with scholarly journal B, on the basis of the number of references made by (TAG) articles in A that direct the reader to (any, not just TAG) articles in B?”*.

Consequently, the ultimate goal is to quantify the value each one scholarly journal represents for the HEAL-Link TAG members. Once calculated, such a value

could then be used to realize the impact the acquisition or the discontinuation of the journal title in question will have to the consortium's TAG members.

Motivated by what is outlined above, the present research focuses on utilizing HEAL-Link TAG related bibliographic data, their processing, cleansing, transformation and analytical processing in order to reveal and visualize journal-to-journal associations along the lines of the stated objectives. Namely, to pinpoint scholarly journals proven to be of value to the TAG in question, thus being worthy to either retain in, or to acquire/include into the HEAL-Link collection.

Section 2 "Related work" comprises a brief literature review on the relevant research in bibliometrics/citation analysis and visualization. In Section 3 "Data collection, preparation, cleansing and unification", we present the dataset used for our exploratory analysis, and the data preparation, cleansing, and transformation tasks conducted. Section 4 "J2J-GR service" focuses on the extended, improved and enriched version of the J2J-GR service. In Section 5 "Discussion", we discuss the results and interpret the findings by considering examples of use of the J2J-GR interactive graphs. Finally, in Section 6 "Future work", we comment on the novelty of our plans to conduct data mining tasks in order to reveal pairwise causal associations between scholarly journals in the microcosm of the HEAL-Link users community.

2. Related work

Bibliometric analysis focuses mainly on extracting useful information about the impact of the research outputs. Bibliometric analysis involves the use of qualitative and quantitative statistical methods to analyze publication contents, references, citations and (co)authorships (Appio, Cesaroni, & Di Minin, 2014). Through these methods bibliometric analysis provides credible and judicial results. It is also complex because it entails several steps that employ numerous and diverse analyses and mapping software tools, which are frequently available only under commercial licenses (Guler, Waaijer, & Palmblad, 2016). Bibliometric techniques introduce objective measures of evaluation of scholarly publications that contrast the potential bias embedded in subjective evaluations (Fonseca & Borges-Tiago, 2021). Mas-Tur et al. (2021) review studies where bibliometric analysis has been applied on various research discipline areas.

One of the most commonly used bibliometric techniques is citation analysis, through which graphical representations of the citations among documents are formulated. Data visualization through these graphical representations is important because it gives the data a specific form of expression so that they are broadly readable and understandable, and it helps to interpret and analyze the results.

Most of the related research work on data visualization in conjunction with bibliometric techniques is based on science mapping. Science mapping is a generic process of domain analysis and visualization (Chen, 2017). Science mapping aims to build science/bibliometric maps that describe how specific disciplines, scientific domains, or research fields are conceptually, intellectually, and socially structured (Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011). Different types of maps have been studied, some revealing relations among authors, documents, journals, or keywords, usually constructed by utilizing citations, co-citations, or bibliography coupling based on co-occurrences of keywords in documents (van Eck, Waltman, Dekker, & van den Berg, 2010). Science mapping comprises an important research topic in the field of bibliometrics (van Eck et al., 2010). It involves a number of discrete

steps/stages with a few differences amongst scholars, but with the same workflow principals. We mention two representative approaches:

1. Aria & Cuccurullo (2017) propose a standard workflow on science mapping, which consists of the following 5 stages: (1) Study design, where scholars define the research questions and choose the appropriate bibliometric methods to answer these questions; (2) Data collection, where the bibliometric data are collected to and exported from a selected database; (3) Data analysis where one or more bibliometric or statistical software tools are employed - alternatively, scholars can write their own source code to meet their requirements; (4) Data visualization, where scholars determine the most suitable visualization method to be used in connection with the results of the third step; (5) Interpretation where scholars interpret and describe their findings.
2. According to Cobo et al. (2011) the steps involving science mapping are (1) data retrieval, (2) preparation and preprocessing, (3) network extraction, (4) mapping, and, (5) analysis and visualization with the preprocessing step being the most critical one.

Many software tools on science mapping have been developed, reviewed, analyzed and evaluated (Cobo et al., 2011). Additional analyses and reviews of the ones most widely used in the literature can be found in Börner et al. (2010) and Aria & Cuccurullo (2017). Chen (2017) elaborated a systematic literature review concerning major aspects of science mapping to demonstrate the use of a science mapping approach and to identify major areas of research activities concerning science mapping.

Citation networks (or citation graphs) are one representative form of a graphical representation of the citations among documents. "A citation graph is a graph created from the metadata provided for each paper and is constructed based on the links that exist between papers and are defined via the references section of each paper" (Fragkiadaki & Evangelidis, 2014). There are many studies that use citation networks broadly. Most of them go beyond the aspects of visualization and in conjunction with complex methods and clustering or ranking algorithms extract the corresponding desired outcome. Leicht, Clarkson, Shedden, & Newman (2007) present three methods of analyzing citation networks with each method based on different algorithms. Fragkiadaki & Evangelidis (2014) introduce and construct three types of citations graphs and they further discuss a number of indirect indicators for papers such as PageRank, EigenFactor and indirect h-index. Šubelj, van Eck, & Waltman (2016) focus on systematically comparing the performance of a large number of clustering methods when applied to the problem of clustering scholarly publications in a citation network based on citation relations. Knezevic & Savic (2020) list a number of studies that use and analyze citation networks on the basis of extracting information on different aspects of the scientific environment. Hou, Yang, & Chen (2020) applied structural variation analysis (SVA) to measure the potential scholarly impact of researchers in the IS field and they utilize co-author networks to reveal the position and structural features of different types of researchers with potential scholarly impact.

When the focus of the citation analysis is on journal-to-journal citations then the formulated citation networks are called journal citation networks. Various approaches on journal citation networks have been conducted by many scholars. Leydesdorff (2007) by using Pajek³, a software mapping tool for analyzing large networks,

³ <http://mrvar.fdv.uni-lj.si/pajek/>

generates journal networks (cited and citation impact environments) from citation matrices comprising of a number of journals. Leydesdorff, Wagner, & Bornmann (2018) focus on measuring interdisciplinarity. They generate journal citation networks by using Visualization of Similarities Viewer (VOSViewer)⁴ (van Eck & Waltman, 2010), another software tool for constructing and visualizing citation networks and the journals listed in the Science and Social Sciences Citation Index in the 2015 Journal Citation Reports⁵. Knezevic & Savic (2020) study and rank the relations between scientific fields by analyzing an empirically observed journal citation network constructed from bibliographic data from Web of science. Wang & Bowers (2016) study and uncover how knowledge is exchanged and disseminated in the educational administration research literature through the journal citation network.

The two most popular subscription-based bibliographic databases that provide products and services based on bibliometrics and citation analysis are (a) the Web of Science⁶, an online subscription-based scientific citation indexing service maintained by Clarivate Analytics, and (b) Scopus⁷, Elsevier's abstracts and citations database. Both vendors have built online applications for data visualization on various aspects of the scientific environment and on their publications.

InCites from Clarivate Analytics⁸ is a citation-based evaluation tool for academic and government administrators to analyze institutional productivity and benchmark output against peers and aspirational peers in the national or international context. It provides the means to gather and analyze data with multiple visualization types that effectively communicate the collected information, helping the subscribers make informed decisions, and assisting strategic initiatives.

SCImago Journal & Country Rank from Elsevier⁹ is a publicly available portal that includes journal and national scientific indicators developed from the information contained in the Scopus database. These indicators can be used to assess and analyze scientific domains from journals to country rankings. The SCImago Journal & Country Rank network visualization tools map multiple networks based on each journal selection. The information visualization projects *Shape of Science*, *Subject Bubble Chart* and *World Report* aim to reveal the structure of science. They involve the construction of real time maps and bubble charts offering detailed information for the analysis of the world as whole, plus for each one of eight large geographic regions.

It is not simple to define interdisciplinarity since there are no strict methods or a unique international standard to delimit or dis sever the disciplines. There are different models/schemas of disciplines across platforms and across the most popular bibliographic databases. Interdisciplinarity can be attributed to departments, journals, oeuvres, emerging disciplines, etc. (Leydesdorff et al., 2018). Additionally, many scholars use various clustering algorithms to dis sever disciplines areas based on their specific datasets. Leydesdorff et al. (2018) focus on ranking journals based on their interdisciplinarity. Further analysis for interdisciplinarity goes beyond the scope of this research paper. Our approach on selecting the disciplines model is described in Section 3 (*Data collection, preparation, cleansing and unification*).

In the present paper, we focus on visualizing journal-to-journal associations and utilize, among other representational data visualization objects, simple journal-to-

⁴ <https://www.vosviewer.com/>

⁵ <https://clarivate.com/webofsciencegroup/solutions/journal-citation-reports/>

⁶ <https://clarivate.com/products/web-of-science>

⁷ <https://www.scopus.com>

⁸ <https://clarivate.com/products/incites>

⁹ <https://www.scimagojr.com/>

journal citation networks for each individual journal the TAG has published their works. The goal is to reveal relations and interdisciplinarity among journals. The main groups of interest for our research are the academic libraries' consortia on the aspect of providing an additional criterion on the journals' subscriptions evaluation as we have already described it in the "Introduction" Section. Our methodology has been applied to the HEAL-Link consortium and its corresponding TAG. We acquire a similar but novel approach on creating the J2J citation networks inspired by the data visualization model used by (Leydesdorff et al., 2018). As stated in the previous, only citations that originate from HEAL-Link TAG (co-)authored works are considered. For any one journal, all incoming (fan-in) journal and outgoing (fan-out) journal citations are considered. A J2J citations mapping network is thus constructed. By means of an interactive graphical interface, the J2J-GR service facilitates user navigation in the citations mapping network in search for journal-to-journal associations of interest to the HEAL-Link leadership and members. The service and its interactive visual interface are presented in Section 4 (The J2J-GR Service) below.

3. Data collection, preparation, cleansing and unification

In order to pursue the aforementioned research goal, bibliographic metadata from nearly 93,000 research publications have been collected and processed. The collection consists of research publications involving at least one (co-)author affiliated to a HEAL-Link member academic/research institution (Target Authors Group, TAG). The Scopus database was used to retrieve research publications that span a period of eleven years (2010–2020) and involve nearly 10,000 journals. The PostgreSQL RDBMS¹⁰ was used to organize the bibliographic data collection.

A number of models exist that classify scholarly journals (Library of Congress classification, Web of Science subject classification, Scopus ASJC scheme, etc.). We have adopted the Scopus ASJC (All Science Journal Classification System) scheme¹¹. Our decision was based on the fact that the ASJC scheme involves a small and manageable number of Subject areas facilitating the handling of important cognitive areas/categories. The ASJC scheme is a formal model used by many researchers. It involves 3 levels of classification. We have decided to use the second level which comprises of 27 Subject category Classifications. Bibliographic data are coupled with metadata on subject categories. This information will be used to identify journals of interdisciplinary value.

To harvest the TAG co-authored publications of the 2010–2020 period as well as the journals they make references to, two API services were used: the Crossref REST API¹², and the Scopus APIs¹³.

In order to make the dataset consistent and to further improve the accuracy of our research results, a set of data preparation/cleansing tasks were conducted. During the data harvesting stage, the Scopus APIs service was used to harvest all publications involving at least one TAG (co-)author. Table 1 presents the harvested publications grouped by publication type.

¹⁰

¹¹ https://service.elsevier.com/app/answers/detail/a_id/12007/supporthub/scopus/

¹² <https://api.crossref.org>

¹³ https://dev.elsevier.com/sc_apis.html

Publication type	Count
Article	70829
Article in Press	1257
Book	454
Business Article	1
Chapter	5034
Conference Paper	21868
Editorial	2634
Erratum	495
Letter	1776
Note	933
Review	7445
Short Survey	150

Table 1: Information about the publications

The book- and chapter- type, as well as other publications not assigned an ISSN number were removed from the dataset. ISSN is a unique 8-digit identifier assigned to periodical publications, irrespective of their medium (print or electronic)¹⁴. For the publication data used, ISSN is taken to identify scholarly journals.

The publications harvesting stage was followed by the harvesting of metadata relating to the references made by the published works. The ISSN numbers of the latter comprised the means for achieving this. Most of the publications data records retrieved via the Crossref REST API include the corresponding references lists. As this has not been the case for the references relating metadata of the Elsevier publications, a separate data harvesting stage had to be conducted using the Scopus API in order to harvest the corresponding references lists.

An additional data preprocessing stage was carried out in order to homogenize the Crossref REST API originating references metadata. The latter are of a diverse format, one that lacks a specific pattern even for works of the same publisher. As an example, the digital object identifier (DOI)¹⁵ which, as it is explained below, comprises the most important metadata item is either not present in the reference record, or it is listed as a field of its own, or it is listed alongside with other metadata in various fields.

Today, nearly all published works are assigned a unique DOI (Digital Object Identifier). The latter is a unique alphanumeric string assigned by a registration agency in cooperation with the International DOI Foundation¹⁶. It is used to identify content and provide a persistent link to its location on the Internet. As such, the DOI facilitates the retrieval of the corresponding scholarly journal metadata. In this respect, for the Crossref records that included DOIs the latter were used to identify and retrieve the corresponding journal metadata. For Crossref records lacking a DOI, a separate data harvesting stage had to be conducted in order to locate the missing DOI values which were then used to retrieve the corresponding journal ISSN numbers.

The Scopus API is more coherent than the Crossref REST API. Almost all of Scopus API retrieved works include the corresponding references list, including unique

¹⁴ <https://www.issn.org/understanding-the-issn/what-is-an-issn/>

¹⁵ https://en.wikipedia.org/wiki/Digital_object_identifier

¹⁶ <https://www.doi.org>

identifiers for each one member of the latter (DOI, or Scopus identifier¹⁷). Unfortunately, it imposes quotas that restrict API-based data retrieval operations for an extended number of records. This is the reason why the Scopus APIs were used for all the references that could not be harvested from the Crossref REST API. An effort was made to use the Scopus APIs for as small a number of records as possible and by doing so remain within the Scopus imposed quotas.

For references with DOIs, the ISSN identifiers of the corresponding journals were harvested from the Crossref REST API. For references lacking their DOI identifiers, data harvesting was conducted by using the Scopus APIs, plus the ArXiv electronic preprints repository¹⁸. Additional checking was conducted to ensure that each sought for bibliographic record existed in the harvested dataset. As a follow-up, additional techniques were applied during the ISSN harvesting stage. They included similarity checking on other metadata items, e.g., the journal title. When the number of records to be checked was small enough and manageable, manual checking was conducted in order to identify the target journal. Additional checking for errors was conducted to ensure the integrity of the dataset. For example, when more than two ISSNs were found to correspond to the same title. At this stage it is important to note that references found not to correspond to an ISSN number (books, book chapters, proceedings, conferences, handbooks, etc.) were removed from the dataset. The proportion of the discoverable references found to correspond to was 83.29%. Figure 1 is a diagrammatic representation of the data preprocessing phase.

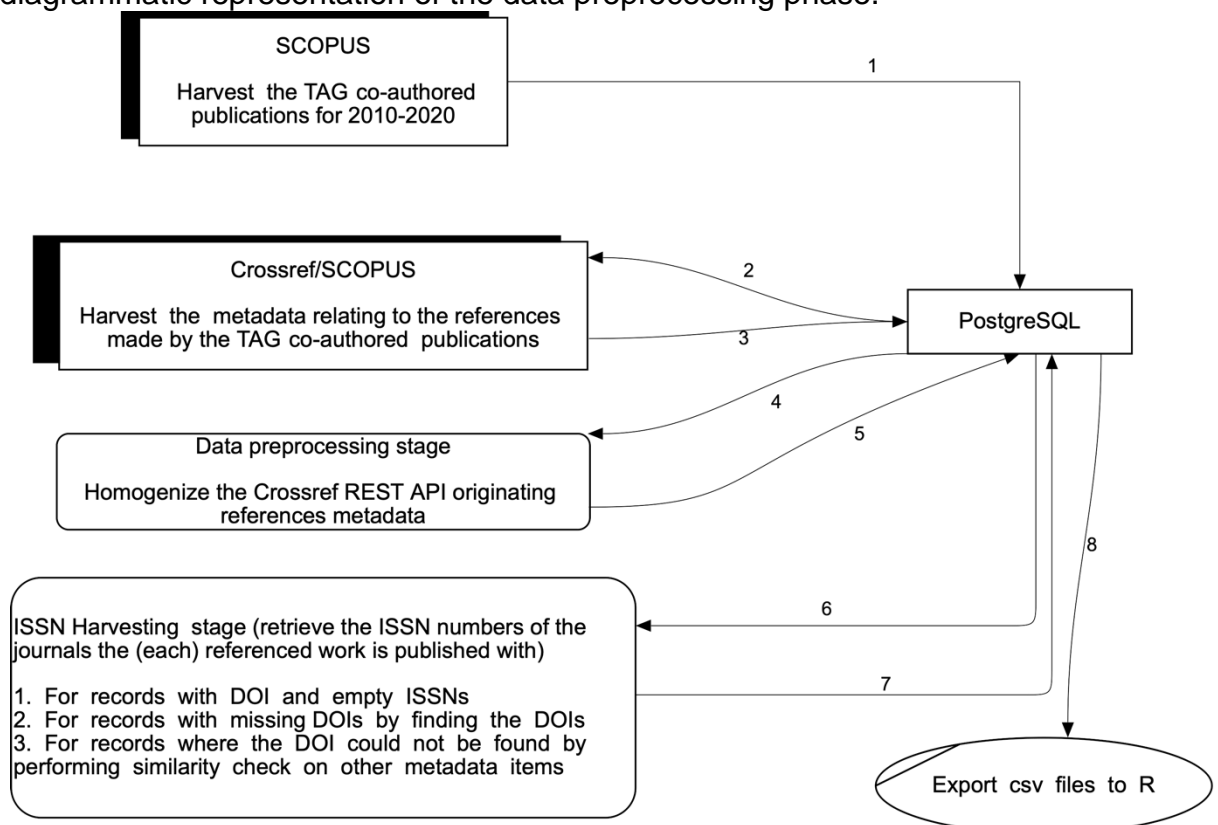


Figure 1. Data collection, preparation, cleansing and unification process

Following the aforementioned data collection, preparation, cleansing and unification stage, specific views were created to facilitate the visualization, exploratory

¹⁷ <https://www.elsevier.com/solutions/scopus/how-scopus-works/high-quality-data>

¹⁸ <https://arxiv.org>

analysis, and interpretation of the information present in the bibliographic dataset. The latter are outlined in the following Section.

4. The J2J-GR Service

The J2J-GR service is a web application that utilizes interactive graphs in order to assist the HEAL-Link user interpret and realize existing HEAL-Link specific associations between scholarly journals. More specifically, journal-to-journal associations based on the references made from works published by the HEAL-Link TAG community. As reported in Pispiringas et al. (2019), the service provides the interactive references network graph shown in Figure 2. The details of the user interface pertaining to the network graph in Figure 2 are considered and they are commented upon next.

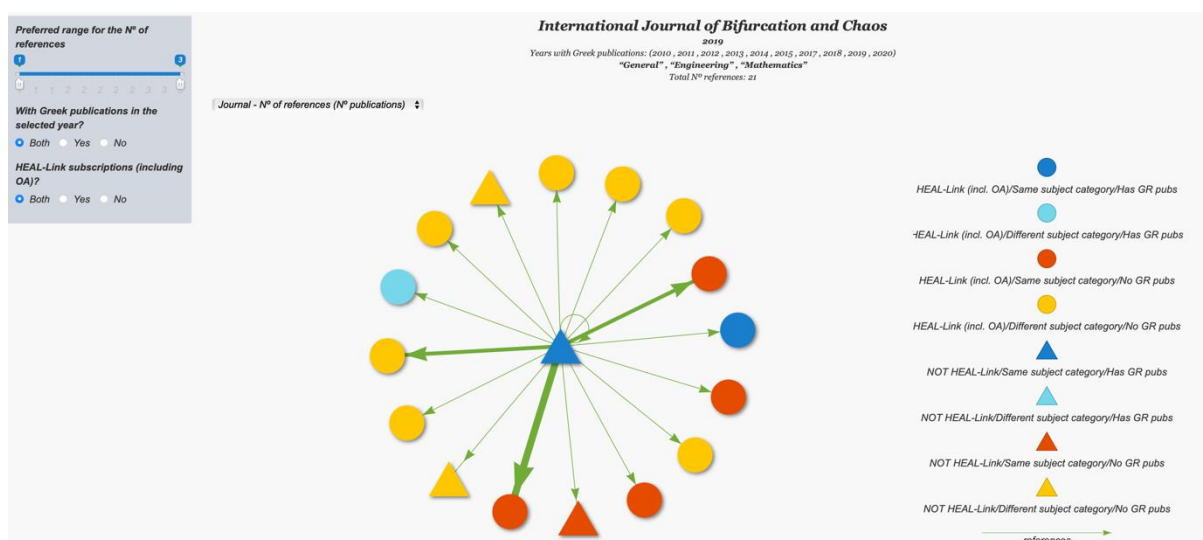


Figure 2. J2J-GR interactive references network graph for the International Journal of Bifurcation and Chaos

Since Pispiringas et al. (2019), the service has been extended to provide interactive citation network graphs, plus two types of bar charts: “cited-by” and “references”, grouped by subject category. It also provides co-cited journals lollipop type charts. The new additions are considered in the next sections. It is noted that all graphs and charts are synchronized to each other, and they all refer to the one central journal title the user wishes to focus on, listed in the top of Figure 2.

The application has been coded in the R¹⁹/RStudio²⁰ project development environment. The interactive graphs are constructed with the visNetwork package²¹; an R package for network visualization based on the vis.js JavaScript library²². The bar charts are constructed with the R ggplot2 package for declaratively creating graphics²³, which is part of the tidyverse²⁴ collection of R packages for data science. To implement the web application and its interactive user interface straight from R, the

¹⁹ <https://www.r-project.org>

²⁰ <https://www.rstudio.com>

²¹ <https://datastom-open.github.io/visNetwork/>, <https://github.com/datastom-open/visNetwork>

²² <http://visjs.org/>

²³ <https://ggplot2.tidyverse.org/>

²⁴ <https://www.tidyverse.org/>

Shiny²⁵, and shinydashboard²⁶ packages have been adopted. Last but not least, the R parallel library²⁷ is used in order to reduce data processing and response times.

Next follows a detailed presentation of each one type of interactive network graph and chart supported in the current version of the J2J-GR service, namely: (a) the interactive references network graph, (b) the interactive citations network graph, (c) the cited by/references subject categories bar charts, and (d) the co-cited journals lollipop charts.

4.1 Interactive references network graph

Upon initiating the service, the user is presented with the dialogue screen shown in Figure 3.



Figure 3. J2J-GR service: graphs type selection, and journal/publication year settings

In the left part of the screen in Figure 3, the user alternates between (a) network graphs and bar charts, and (b) co-cited journals lollipop charts. Beginning with “Network graphs and bar charts”, the target publication year, subject category, and journal title values are set by means of the three dropdown menus shown in Figure 3. This translates into the user wishing to have the synchronized graphs and bar charts display information relating to the journal title in question (“International Journal of Bifurcation and Chaos”), under the subject category selected (“General”), and for the year set in “Publication year” (2019).

Once the “Go” button is pressed (cf. Figure 3), the set journal title becomes the central node for the network of journal associations. The uppermost title part of the “Interactive references network graph” displays information relating to the journal title set to comprise the central network node (Figure 4). It lists the year that this journal is considered with its TAG publications, the subject categories covered, and the total number of references its TAG publications make to other journals during year in question.



Figure 4. Interactive references network graph title part: journal information

²⁵ <https://shiny.rstudio.com>

²⁶ <https://rstudio.github.io/shinydashboard>

²⁷ <https://stat.ethz.ch/R-manual/R-devel/library/parallel/doc/parallel.pdf>

The central network node is surrounded by nodes/journal titles, each one node corresponding to a journal cited by TAG (co-)authored articles published with the central node during the year of interest. By hovering the pointer over the geometrical shape representing any one of the central or peripheral displayed nodes, the user is presented with the corresponding journal information. In Figure 5, the peripheral journal titled “Nature” is registered to include nine (9) TAG authored articles, and it has been cited by one (1) TAG article published with the central node in the year 2019. The user is also informed that a total of fifty four (54) TAG (co-)authored articles are registered to have been published with the cited journal titled “Nature” in the entire collection (years 2010-20).

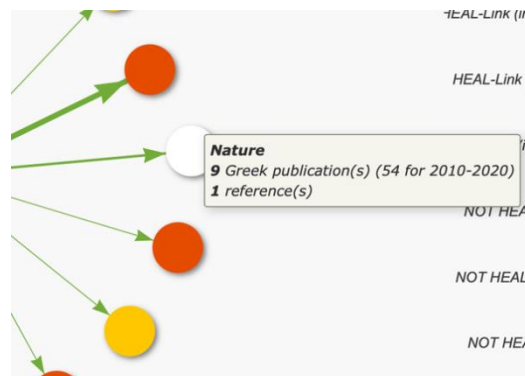


Figure 5. Interactive references network graph: cited journal title info

The interactive references network graph’s user interface consists of four sections, marked by red encircled numbers 1,2,3, and 4 in Figure 6.

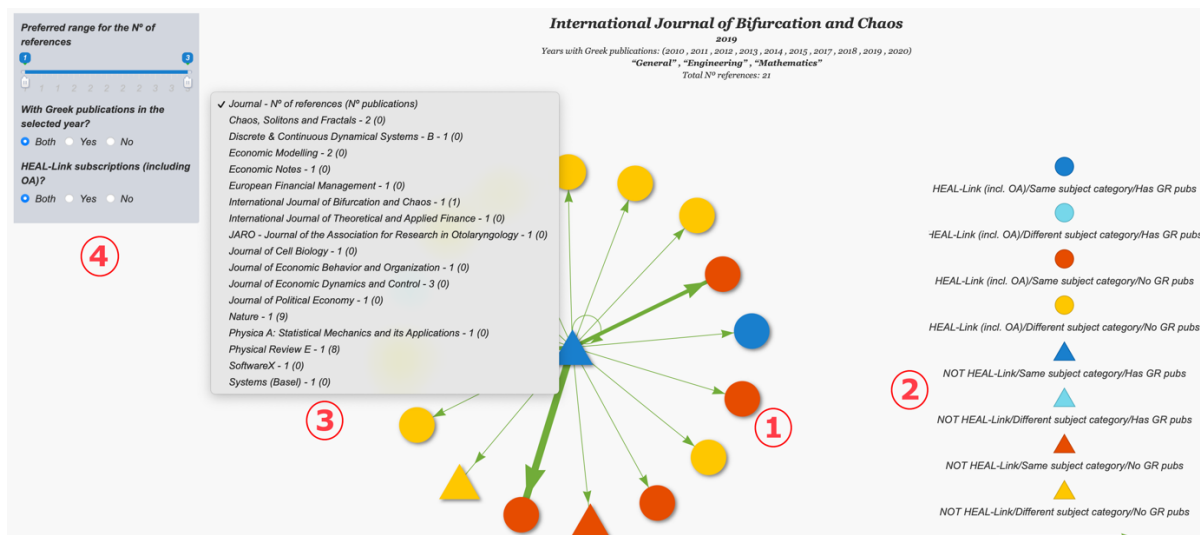


Figure 6. J2J-GR service: Interactive references network graph

The subnetwork graph with the selected journal as its central node and its cited journals in the perimeter is marked “1” in Figure 6. The thickness of the arrows (edges) from the central to each one perimetric node varies in analogy with the number of references registered to have been made from the former to the latter. Detailed information on the number of references made in each one case is displayed by the “Journal - N° of references (N° of publications)” drop-down marked “3” in Figure 6. The drop-down menu displays in textual form the same information displayed in the

interactive network graph. When the user selects a target list journal, the corresponding graph node is highlighted for the user to click on and have it become the new central node. In this respect, the interactive network graph supports a *fan-out* type of functionality. This is useful especially in cases when the users have to navigate themselves in a network of citing and cited journals.

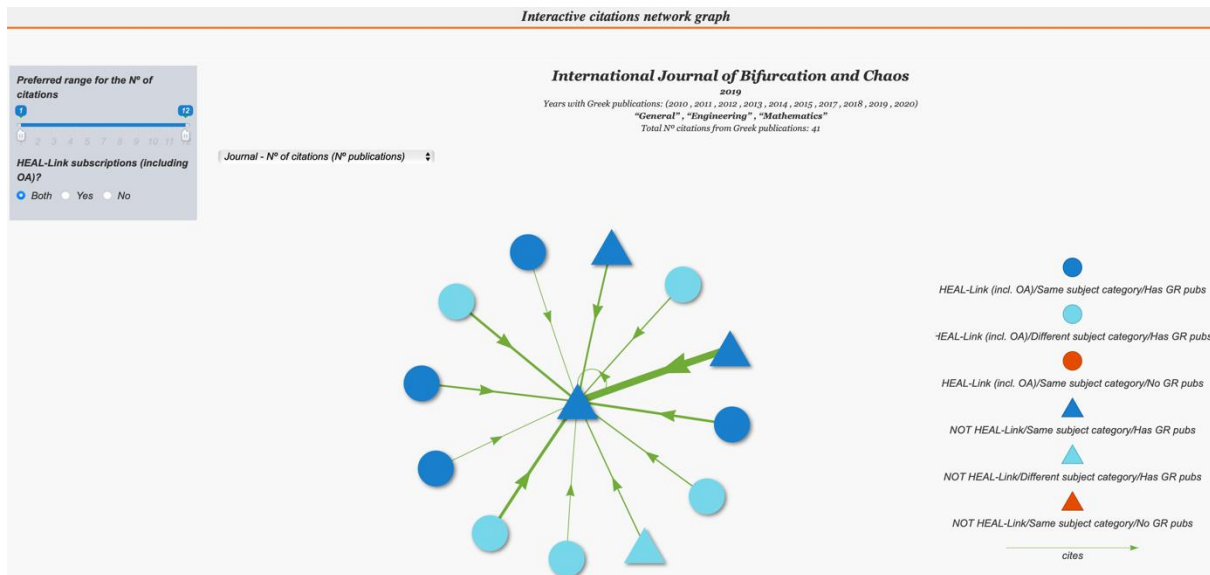
The colors and shapes encoding scheme outlined in the legend marked “2” in Figure 6 relates to information of interest to HEAL-Link. Geometrical shapes depict the inclusion or not of the journal in the HEAL-Link subscriptions list. A circle indicates a journal HEAL-Link members have access to. This includes HEAL-Link paid subscriptions and HEAL-Link listed Open Access (OA) journals. A triangle indicates a journal that HEAL-Link members do not have access to. Additional color encoded journal characteristics have as follows: (a) blue indicates a journal that shares a common discipline area with the journal of the central node, plus one that includes TAG (co-)authored works in the year considered/selected, (b) turquoise indicates a journal that shares no common discipline area with the central node, yet one that includes TAG (co-)authored works in the year considered/selected, (c) red indicates a journal that shares a common discipline area with the central node with no TAG (co-)authored works in the selected year, and (d) yellow indicates a journal that shares no common discipline area with the central node, plus one with no TAG (co-)authored works in the selected year.

The sidebar labeled “4” in Figure 6 provides three filters for narrowing down the scope of the journals cited by the central node-journal. Filter number one is used to specify a range for the number of references made to other journal publications by the TAG works published with the central node during the set publication year. Filter number two enables the user to specify whether or not the list of cited journals includes TAG publications during the set publication year. Filter number three enables the user to include or exclude HEAL-Link subscriptions/journals.

It is worth noting that all four graphs under “*Network graphs, bar charts*” remain synchronized with one another, i.e. they are automatically updated to always reflect the current journal/year information set and displayed in the top part of the web page shown in Figure 3.

4.2 Interactive citations network graph

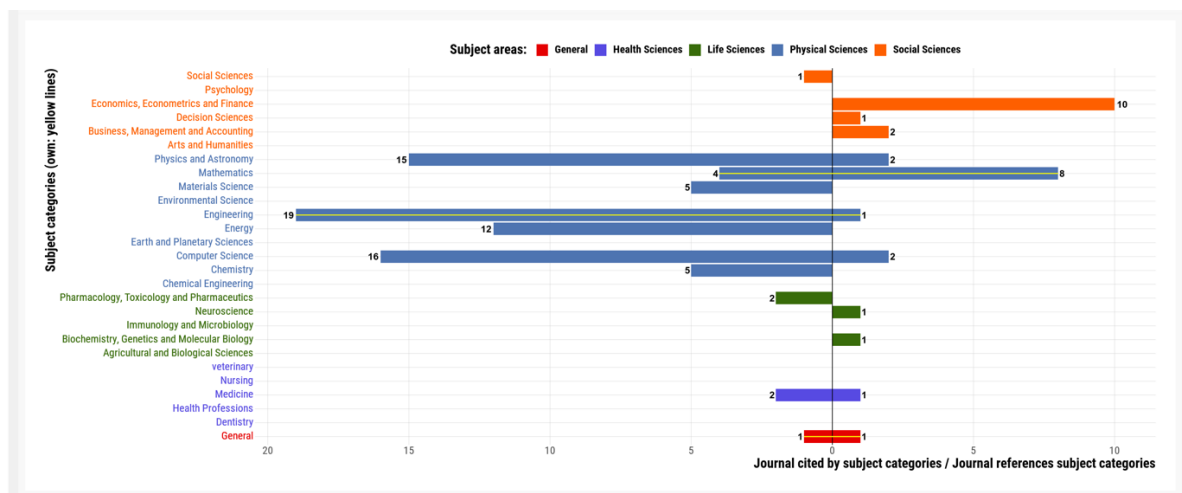
Contrary to the interactive references network graph, which is of a fan-out type, the interactive citations network graph is one of a fan-in type (Figure 7). Other than this, its user interface and interactivity are analogous to those of the interactive references network graph.



As with the references graph, the central network node is the one set in Figure 3. This time however, perimetric nodes represent journals that cite the central journal node. By construction, the database used by J2J-GR involves only references made by TAG (co-)authored publications. In accordance with the journal types encoding scheme (cf. the legend labeled “2” in Figure 6), nodes in the perimeter of the citations graph may be colored blue or turquoise (Figure 7). The central node may also be colored red when it does not have any TAG (co-)authored publications.

4.3 Cited by/references subject categories bar charts

Each one of the two bar charts under “*Networks graphs, bar charts*” is of the 2-in-1 type (Figures 8 and 9, below). The “categorical” y-axis notches depict ASJC subject category labels. In Figure 8, the horizontal bars to the left of the y-axis represent the numbers of citations received, and the ones to the right represent the numbers of references made by TAG (co-)authored articles. Citations and references are grouped by the ASJC subject category name. They all relate to the journal title and publication year specified in Figure 3.



As an example, let “2019” be specified to comprise the year of interest (cf. Figure 3). In Figure 8 the journal “*International Journal of Bifurcation and Chaos*” is seen to have been cited fifteen (15) times by articles published with journals covering the “*Physics and Astronomy*” subject category, nineteen (19) times by articles published with journals in the “*Engineering*” subject category, etc. Likewise, TAG (co-)authored articles published in 2019 with the journal in question are seen to have made ten (10) references to journals covering the “*Economics Econometrics and Finance*” subject category, etc.

Complementary to the information revealed by Figure 8 is the information present in Figure 9. This time, the horizontal bars to the left and to the right of the y-axis represent numbers of distinct articles that make/receive the corresponding citations listed in Figure 8.

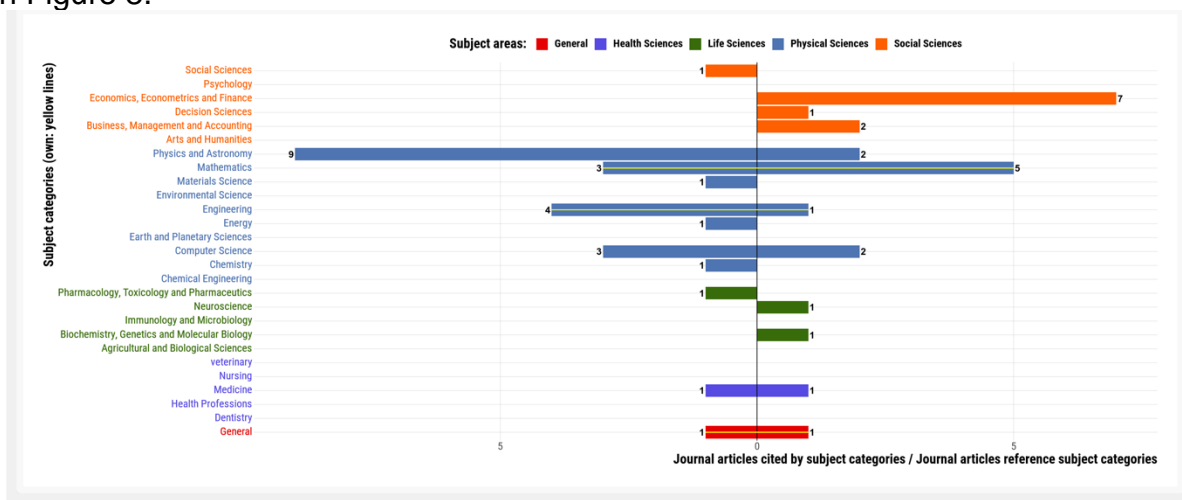


Figure 9. Journal articles citations / Journal articles references by ASJC subject category

For the “*International Journal of Bifurcation and Chaos, 2019*” pair in question and subject category “*Physics and Astronomy*”, the bar to the left of the y-axis in Figure 9 reads as follows: nine (9) articles with one or more journals covering the subject category “*Physics and Astronomy*” are found to have cited TAG (co-)authored articles published in 2019 with “*International Journal of Bifurcation and Chaos*”. The bar to the right of the y-axis for the same subject category reads as follows: two (2) TAG (co-)authored works published with “*International Journal of Bifurcation and Chaos*” in 2019 are found to have cited articles published with journal(s) covering the “*Physics and Astronomy*” subject category.

A scholarly journal may cover one or more subject categories. The latter are indicated by drawing (a) yellow spinal strip(s) along the corresponding bar(s) in Figures 8 and 9, above. By combining the information revealed by the two graphs, one may obtain a first indication on the selected journal’s degree of interdisciplinarity.

For example, considering the “*Physics and Astronomy*” subject category: (a) it is not covered by the selected journal, (b) the fifteen (15) citations made to the latter (cf. Figure 8) originate from nine (9) articles published with one or more “*Physics and Astronomy*” journal titles (cf. Figure 9); an average of one point seven (1.7) citations per article. Analogously, the “*Computer Science*” subject category also lies beyond the scope of the central node journal. This time, the sixteen (16) citations from “*Computer Science*” (cf. Figure 8) are made by just three (3) articles published with one or more “*Computer Science*” journals; five point three (5.3) citations per article, on the average. Thus, as a source of information, the “*International Journal of Bifurcation and Chaos*” is seen to have a broader influence on Physics and Astronomy researchers as

compared to the influence it has on researchers who publish with Computer Science journals. This is so, despite the fact that in the year 2019 the number of citations received from the former is smaller than that of the latter. This is due to having sixteen (16) citations originate from just three (3) Computer Science articles, possibly of a literature review type. On the other hand, fifteen (15) citations originate Physics and Astronomy articles (9, cf. Figure 9); three times as many when compared to the case of Computer Science articles.

4.4 Co-cited journals lollipop charts

Co-cited journals lollipop charts generation is based on the entire dataset collection (currently covering the 2010-2020 period); in this respect, there is no option for specifying a *Publication year* value/filter in the corresponding dialogue screen (Figure 10). For the example considered, the focus remains on the “*International Journal of Bifurcation and Chaos*” central node journal. In the first lollipop chart, citations originate from TAG (co-)authored articles published in all journals covering the “*General*” subject category (Figure 11). In the second lollipop chart, citations originate from TAG (co-)authored articles published in all journals covering the entire dataset, regardless of subject categories (Figure 12).

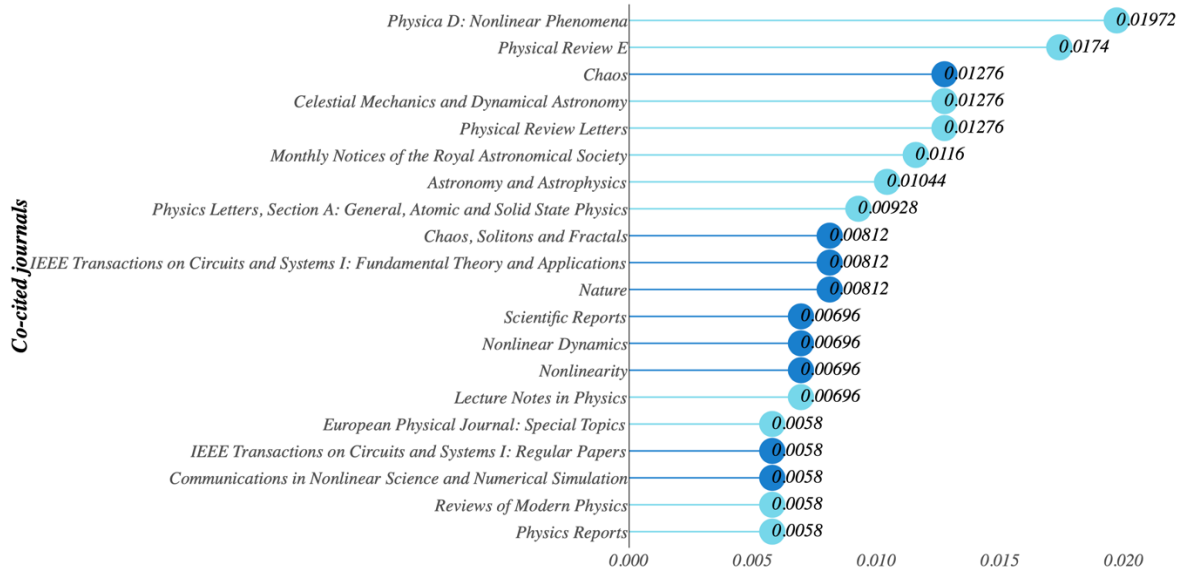
The screenshot shows a web interface for generating co-cited journals lollipop charts. On the left, under 'Journal graphs', the option 'Co-cited journals lollipop charts' is selected. The main panel contains three dropdown menus: 'Publication year' is set to 'All years', 'Subject category' is set to 'General', and 'Journal' is set to 'International Journal of Bifurcation and Chaos'. A 'Go' button is located at the bottom of the main panel.

Figure 10. Scope of citing journals in “Co-cited journals lollipop charts”

For the first lollipop chart listed along the vertical axis of Figure 11 are the journal titles found to represent a measurable frequency/probability of occurrence as co-cited journals, alongside with the “*International Journal of Bifurcation and Chaos*”. Co-citation is defined as the frequency with which two documents are cited together by other documents (Small, 1973). Consequently, co-cited journals are two journals, which are cited together in the TAG (co-) authored publications. The corresponding probability values are displayed for each one of the lollipop tops in Figure 11. The latter are colored blue when they cover at least one of the selected journal categories (“*General*”, “*Engineering*”, and “*Mathematics*”), otherwise they are colored turquoise.

Publications in subject category "General" (using frequent itemsets and support)

■ In subject categories: "General", "Engineering", "Mathematics"
 ■ Other subject categories



Probability (862 articles in subject category "General")

Figure 11. Co-cited journals lollipop chart in Subject category "General"

In Figure 11, each one of the twenty (20) lollipops displayed corresponds to a journal title that tends to frequently be co-cited with the journal title considered. The chart is limited to consider TAG (co-)authored articles published with journals in the "General" subject category, a total of 862 for the database considered.

Of the twenty lollipops, nine (9) are colored blue and the remaining eleven (11) they are colored turquoise. The former are journals that have at least one (covered) subject category in common with the three subject categories covered by the journal title considered ("General", "Engineering", and "Mathematics"); for example, "Chaos" (cf. Figure 12). Lollipops colored turquoise depict journal titles that have no subject category in common with the "International Journal of Bifurcation and Chaos"; for example, "Physica D: Nonlinear Phenomena" (cf. Figure 13).

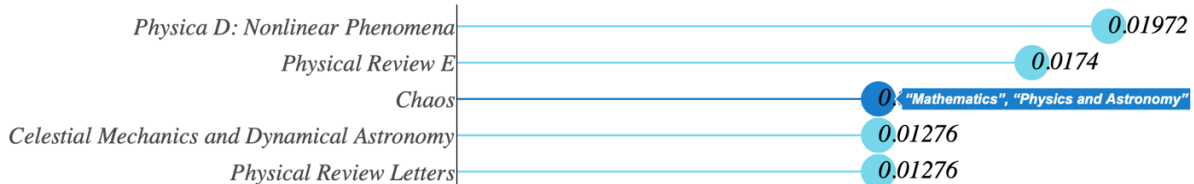


Figure 12. The "Chaos" (co-cited) journal: subject categories covered

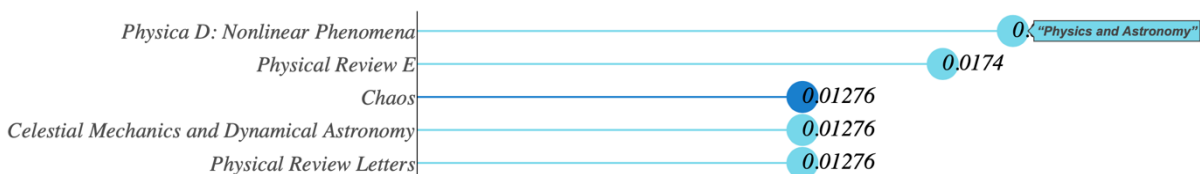


Figure 13. The "Physica D: Nonlinear Phenomena" (co-cited) journal: subject categories covered

Analogous to the lollipop chart in Figure 11 is the one in Figure 14, only this time the TAG (co-)authored articles are not limited to have been published with journals of any given subject category. Considering all of the 71,063 TAG (co-)authored articles in the dataset, around 124 (= $0.00174 * 71,063$) of them co-cite the “*Physical Review E*”-“*International Journal of Bifurcation and Chaos*” pair, 98 (= $0.00138138 * 71,063$) co-cite the “*Physica D: Nonlinear Phenomena*”-“*International Journal of Bifurcation and Chaos*” pair, etc.

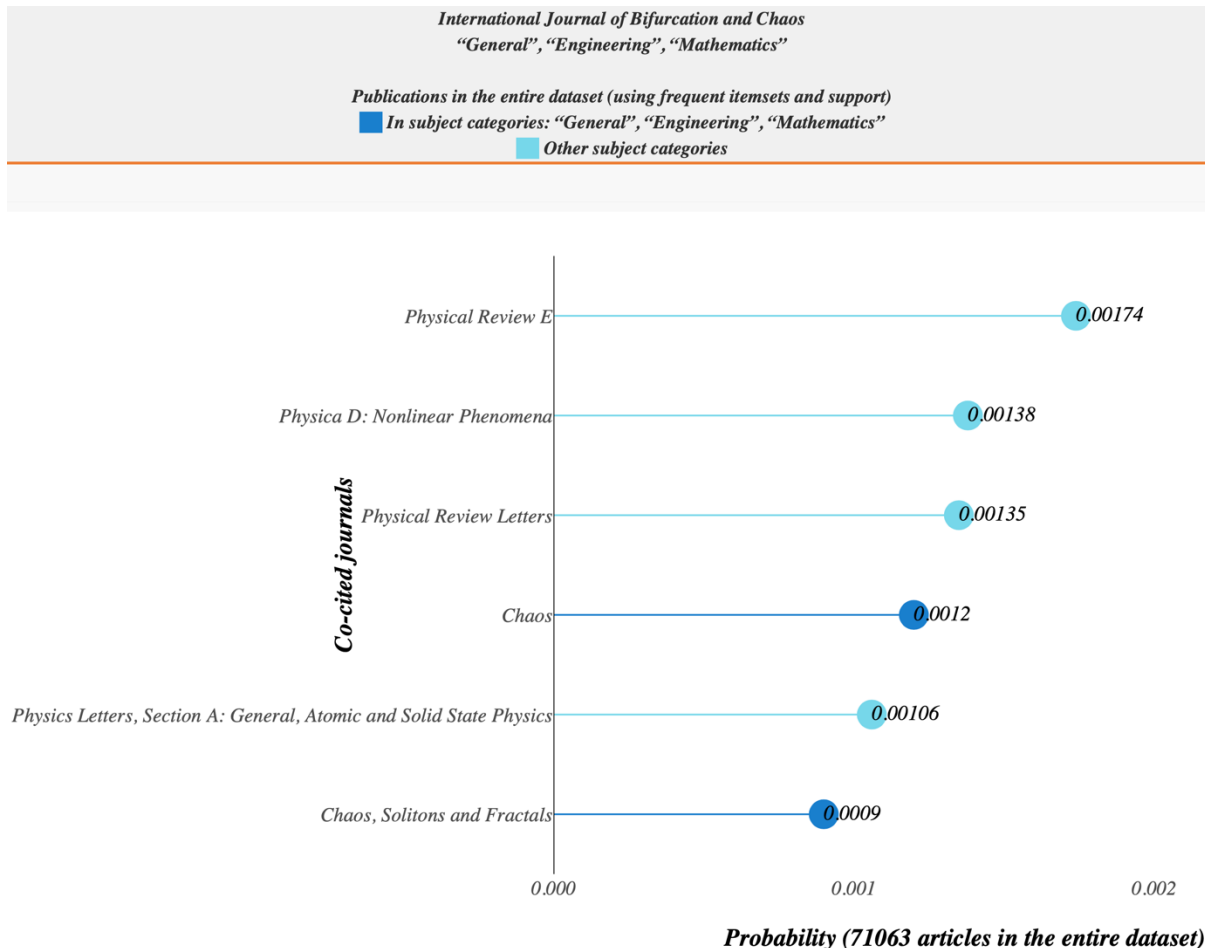


Figure 14. Co-cited journals lollipop chart: articles in the entire dataset

It is worth noting that by observing the lollipop charts in Figures 11 and 14 it is evident that the “*International Journal of Bifurcation and Chaos*” journal tends to be co-cited mainly with journal titles of subject categories that lie outside the scope of its own subject categories set (“*General*”, “*Engineering*”, and “*Mathematics*”). This is indicative of the journal’s interdisciplinary value, an issue to be given special consideration in Section 5 (Discussion) that follows next.

5. Discussion

The J2J-GR service has been designed to provide information relating to the value a journal title represents as a source of information specifically for the TAG community of the HEAL-Link consortium. The information is expected to be of use primarily for the consortium itself, and partly for its member institutions and user community.

Considering a journal as an asset in the HEAL-Link collection, it is important to know:

- a. the journals cited by TAG researchers who publish works with the journal in question,
- b. the degree to which the journal itself is used as a reference in TAG authored works,
- c. the journals TAG authors tend to co-cite alongside with the journal in question
- d. whether the journal in question comprises a useful source of information just for TAG authors who publish with journals in the given journal's subject categories or is it of use to broader/interdisciplinary TAG audiences.

The graph in Figure 2 presents the distribution of references made by TAG works published with the journal in the middle. Colors and shapes encode the intensity (i.e. relative weights) of the references made as well as the types and subject category proximities of the cited journals. Journal titles and numbers of references made can be displayed by means of the drop-down menu marked “3” in Figure 6. In this respect, the graph shown in Figure 2 is used to address requirement (a) in the above, namely to identify the journals that tend to be used as reference material by TAG researchers who publish their works with a given journal.

Analogously, the “fan-in” type graph in Figure 7 comprises a visual representation of information relating to journals that cite the journal title in the middle. This time, citing journals are color coded only in light blue or purple. This is due to J2J-GR’s input data: only citations originating from HEAL-Link TAG works are considered, as dictated by requirement (b) above.

Subject to the same input data limitations (i.e. involving journals color coded in light blue or purple, only) are the lollipop journal depictions in Figures 11 through to 14. This time, the aforementioned user requirement (c) is addressed, namely: journals that TAG authors tend to co-cite alongside with the journal in question. The calculated probability values quantify the intensity of co-citation for each (J_c, J_i) pair, J_c being the journal appearing in the header section of Figures 11 and 14, and J_i (with i ranging from 1 up to i_{max}) being the journals listed along the vertical axis of the lollipop chart (cf. Figures 11 through to 14). Equivalently, one may say that co-citation probability values comprise a measure of relativeness for journal pairs that tend to be co-cited in TAG authored works.

With regard to user requirement (d) in the above, it has already been explained how lollipop charts can be used to exploit interdisciplinarity: journals that are highly probable to be co-cited with journals of different subject categories deserve further investigation with regard to their interdisciplinary characteristics (cf. Figures 11 and 13). In addition, the bar charts in Figures 8 and 9 can be used to exploit journal interdisciplinarity. This is best demonstrated by means of an example. Figures 15 and 16 represent information relating to the way TAG authored works cite “The Lancet” journal. The latter is stated to cover just one subject category of the Scopus ASJC scheme: “Medicine”. Color coded purple, “Medicine” is seen to be one of the five (5) subject categories of the “Health Sciences” subject area. It is noted that references made to “The Lancet” include within-journal references made by works published with “The Lancet” itself. The difference between the two bar charts is that the one in Figure 15 depicts numbers of references “The Lancet” articles make to other works and numbers of citations “The Lancet” articles receive from other works, whereas the one in Figure 16 depicts numbers of distinct articles published with “The Lancet” that reference and receive citations from other works. By combining information from the

two bar charts and for each one subject category, one calculates the average number of citations made to the Lancet journal by a typical TAG authored article known to have used the journal as a source of information. Also, one calculates the average number of references a typical TAG authored article published with “The Lancet” makes to journals covering the subject category in question.

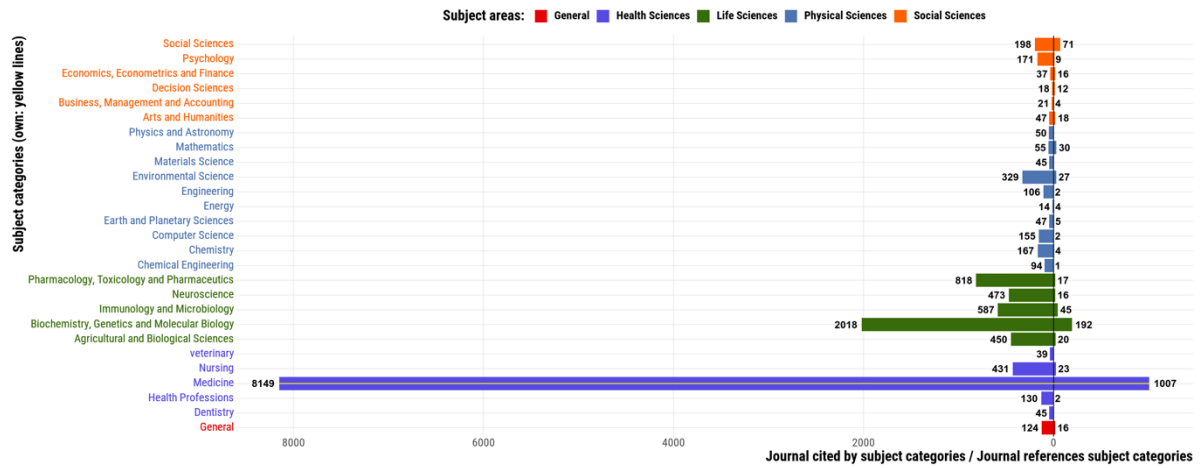


Figure 15. "The Lancet" Cited by/References bar chart

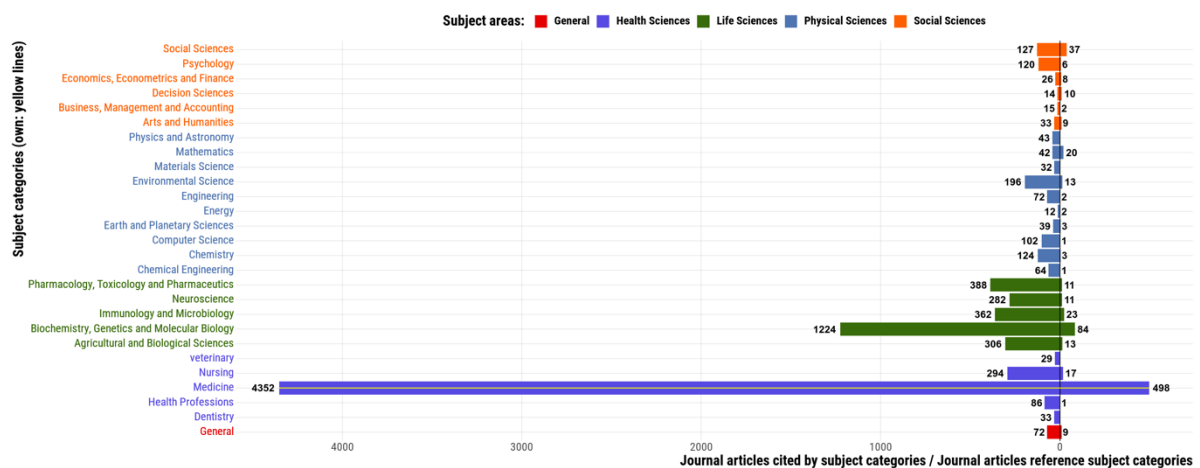


Figure 16. "The Lancet" Cited by Articles/References Articles bar chart

Why is it important to know the average number of distinct articles the references originate from? Because, for journal interdisciplinarity it is different to have many references collectively made by a small number of articles published with journals from within a wide range of subject areas, as opposed to having a large number of references collectively made by a large number of articles from the same wide range of subject areas.

Back to the Lancet journal example: having excluded the “General” subject area, the journal is seen to receive a total of 6545 citations from subject categories other than its own (“Medicine”). This is 44.5% of the total number of citations received within the time period considered (2010-2020). Moreover, the 6545 citations are seen to have originated from a total of 4065 distinct journal articles (cf. Figure 16). The value is calculated by summing up all the non-“Medicine” article numbers. Thus, citations from non-“Medicine” publications are seen to occur at a rate of $6545/4065=1.61$ citations per article. For within-“Medicine” publications the corresponding rate is

8149/4352=1,87 citations per article (cf. Figures 15 and 16). In both cases, citations are made at a most satisfactory/expected average rate of a little less than two references per citing article.

To further exploit interdisciplinarity in the case of the example considered, one may sum up citations in the broader subject area level “Health Sciences”, instead of the “Medicine” subject category level. By conducting analogous calculations, 40.2% of the total citations are found to originate from non-“Health Sciences” journals, occurring at a rate of 1.63 citations per article, as compared to 1.83 citations per article from within-“Health Sciences” journals. It thus becomes possible to quantify a journal’s degree of interdisciplinarity. For the example considered, the Lancet journal is clearly seen to represent a significant degree of interdisciplinarity for the HEAL-Link TAG community.

Admittedly, the Lancet journal example is not representative of the general case. Journals are often indexed to cover more than one ASIJC subject category in the same or even different subject areas. The Lancet journal happens to be indexed to cover only the “Medicine” subject category in the “Health Sciences” subject area. This makes it an ideal test case to consider in order to visualize and quantify its degree of interdisciplinarity by means of the bar charts in Figures 15 and 16. One may argue that a journal indexed to cover a number of subject categories spanning across more than one subject areas is interdisciplinary, by definition. Even then, the degree of its interdisciplinarity may be calculated by means of the bar charts in Figures 15, and 16, in a way analogous to the one presented for the Lancet journal.

Considering the aforementioned user requirements (a), (b), (c), and (d), the J2J-GR service can be used as an easy-to-access exploratory and explanatory analysis tool both by the HEAL-Link leadership/management, as well as by interested HEAL-Link TAG member institutions and individual researchers. The latter, for example, may use the service to identify new potential sources of information in their research areas, and journals that they may consider publishing their works with. They can also acquire insight on journals in new subject areas/categories that they can turn to in order to look for knowledge and methodologies that they can use in their research.

6. Conclusion and Future Work

We report on the current stage of our research effort to exploit journal-to-journal associations as they are shaped by the works of the Greek academics and researchers who (co-)author scholarly publications. The latter comprise the Target Authors Group (TAG) for the Hellenic Academic Libraries Link (HEAL-Link) consortium. Having collected, organized, and pre-processed the corresponding dataset, together with the associated bibliographic metadata, an attempt has been made to further realize the inherent journal-to-journal associations.

Currently, the focus is on conducting exploratory data analysis and presenting the findings by exploiting the potential of the R/RStudio suite of libraries for explanatory data analytics. In this respect, the J2J-GR service has been developed and it is continuously enriched and extended to support new insights to the underlying bibliographic data. Its “Interactive references graph” and “Interactive citations graph” user interface can be used to quantify the value journals represent as sources of information specifically for the HEAL-Link TAG community. This is done by considering (1) the references to other journals made by TAG (co-)authored articles published with a given journal, and (2) the citations originating from TAG (co-)authored articles that target TAG (co-)authored articles published with a given journal.

In addition to the above, the J2J-GR graphs and charts may also be used to obtain a first indication of a journal's degree of interdisciplinarity. More specifically, they can be used to identify journals that tend to cite and be cited by works published with journals of a noticeable degree of diffusion across ASJC subject categories, not registered to be covered by the journal considered. Last but not least, the J2J-GR "Co-cited journals lollipop charts" reveal journals frequently co-cited with the journal considered. For each one pair of co-cited journals, the corresponding frequency/probability values are displayed. The latter are calculated with respect to (a) articles published with journals in a selected ASJC subject category, and (b) articles from the entire dataset.

The current version of the J2J-GR service is publicly available on the Internet²⁸. The service has evolved to comprise a comprehensive bibliographic data exploratory tool not just for the HEAL-Link leadership, but also for its user community.

Our next goal is to elaborate further and conduct data mining operations on the HEAL-Link TAG bibliographic dataset. By doing so we expect to extract additional and more intuitive information regarding journal-to-journal associations. More specifically, we intend to identify and exploit causal associations between scholarly journals. The outcome of such an analytical data processing could be used to conduct predictive analytics in order to facilitate the decision making process of HEAL-Link, e.g., for cancelling existing, or for identifying new potential subscriptions. This type of analytical processing is expected to reveal co-cited journals with even stronger and causal pairwise relations. The latter will be utilized as additional useful information to the HEAL-Link user community, e.g., the readability of their published works as well as journal recommendations on new potentially useful resources that they can benefit from in their research.

7. Acknowledgements

The authors wish to thank the anonymous referees and the editor for their constructive comments and advice.

8. Bibliography provide researchers with a venue

- Appio, F. P., Cesaroni, F., & Di Minin, A. (2014). Visualizing the structure and bridges of the intellectual property management and strategy literature: a document co-citation analysis. *Scientometrics*, *101*(1), 623–661. <https://doi.org/10.1007/s11192-014-1329-0>
- Aria, M., & Cuccurullo, C. (2017). bibliometrix : An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, *11*(4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Börner, K., Huang, W., Linnemeier, M., Duhon, R. J., Phillips, P., Ma, N., ... Price, M. A. (2010). Rete-netzwerk-red: Analyzing and visualizing scholarly networks using the Network Workbench Tool. *Scientometrics*, *83*(3), 863–876. <https://doi.org/10.1007/s11192-009-0149-0>
- Chen, C. (2017). Science Mapping: A Systematic Review of the Literature. *Journal of Data and Information Science*, *2*(2), 1–40. <https://doi.org/10.1515/jdis-2017-0006>
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011).

²⁸ <https://j2j.heal-link.gr>

- Science mapping software tools: Review, analysis, and cooperative study among tools. *Journal of the American Society for Information Science and Technology*, 62(7), 1382–1402. <https://doi.org/10.1002/asi.21525>
- Fonseca, J. M. R. da, & Borges-Tiago, M. T. (2021). Cyberbullying From a Research Viewpoint (pp. 182–200). <https://doi.org/10.4018/978-1-7998-5728-0.ch010>
- Fragkiadaki, E., & Evangelidis, G. (2014). Review of the indirect citations paradigm: Theory and practice of the assessment of papers, authors and journals. *Scientometrics*, 99(2), 261–288. <https://doi.org/10.1007/s11192-013-1175-5>
- Guler, A. T., Waaijer, C. J. F., & Palmblad, M. (2016). Scientific workflows for bibliometrics. *Scientometrics*, 107(2), 385–398. <https://doi.org/10.1007/s11192-016-1885-6>
- Hou, J., Yang, X., & Chen, C. (2020). Measuring researchers' potential scholarly impact with structural variations: Four types of researchers in information science (1979-2018). *PLoS ONE*, 15(6 June), 1–26. <https://doi.org/10.1371/journal.pone.0234347>
- Knezevic, D., & Savic, M. (2020). Analysis of scientific fields using journal citation networks: An empirical study. *INISTA 2020 - 2020 International Conference on INnovations in Intelligent SysTems and Applications, Proceedings*. <https://doi.org/10.1109/INISTA49547.2020.9194640>
- Leicht, E. A., Clarkson, G., Shedden, K., & Newman, M. E. J. (2007). Large-scale structure of time evolving citation networks. *The European Physical Journal B*, 59(1), 75–83. <https://doi.org/10.1140/epjb/e2007-00271-7>
- Leydesdorff, L. (2007). Visualization of the citation impact environments of scientific journals: An online mapping exercise. *Journal of the American Society for Information Science and Technology*, 58(1), 25–38. <https://doi.org/10.1002/asi.20406>
- Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2018). Betweenness and diversity in journal citation networks as measures of interdisciplinarity—A tribute to Eugene Garfield. *Scientometrics*, 114(2), 567–592. <https://doi.org/10.1007/s11192-017-2528-2>
- Mas-Tur, A., Roig-Tierno, N., Sarin, S., Haon, C., Segó, T., Belkhouja, M., ... Merigó, J. M. (2021). Co-citation, bibliographic coupling and leading authors, institutions and countries in the 50 years of Technological Forecasting and Social Change. *Technological Forecasting and Social Change*, 165(November 2020). <https://doi.org/10.1016/j.techfore.2020.120487>
- Narin, F., & Carpenter, M. P. (1975). National publication and citation comparisons. *Journal of the American Society for Information Science*, 26(2), 80–93. <https://doi.org/10.1002/asi.4630260203>
- Pispiringas, L., Dervos, D. A., & Evangelidis, G. (2019). J2J-GR: Journal-to-Journal References by Greek Researchers. In K.-D. Schewe & N. K. Singh (Eds.), *Model and Data Engineering* (Vol. 11815 LNCS, pp. 83–95). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-32065-2_6
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265–269. <https://doi.org/10.1002/asi.4630240406>
- Šubelj, L., van Eck, N. J., & Waltman, L. (2016). Clustering Scientific Publications Based on Citation Relations: A Systematic Comparison of Different Methods. *PLOS ONE*, 11(4), e0154404. <https://doi.org/10.1371/journal.pone.0154404>
- van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538.

<https://doi.org/10.1007/s11192-009-0146-3>

van Eck, N. J., Waltman, L., Dekker, R., & van den Berg, J. (2010). A Comparison of Two Techniques for Bibliometric Mapping: Multidimensional Scaling and VOS. *Journal of the American Society for Information Science and Technology*, 61(12), 2405–2416. <https://doi.org/10.1002/asi.21421>

Wang, Y., & Bowers, A. J. (2016). Mapping the field of educational administration research: a journal citation network analysis. *Journal of Educational Administration*, 54(3). <https://doi.org/10.1108/JEA-02-2015-0013>