J2J-GR: Journal-to-Journal references by Greek researchers

Leonidas Pispiringas[1] [0000−0002−7111−094X], Dimitris A. Dervos[2] [0000−0002−5441−0086], and Georgios Evangelidis[1] [0000−0003−1639−2152]

[1] University of Macedonia, 54636 Thessaloniki, Greece
lpispir@gmail.com, gevan@uom.gr
[2] International Hellenic University, 57400 Sindos, Greece
dimitris.dervos@gmail.com

Abstract. The Hellenic Academic Libraries Link[3] operates since 1998. Its members include all the academic and/or research institutions operating under the auspices of the Hellenic Ministry of Education, plus the Academy of Athens, the National Library, and the Institute of Educational Policy. The present paper reports on the research effort undertaken in order to facilitate the decision-making process and the development of a HEAL-Link strategy plan. The focus is on scientific journals the Greek academic/research community publish their works with, and on the journals they reference. It is assumed that when a researcher makes a reference to an article in a scientific journal, the latter is considered to comprise a valuable source of information. The more references made to articles published with a given journal, the higher the value of the latter as a source of scientific information to the HEAL-Link user community. In order to exploit the aforementioned research goal, bibliographic metadata from nearly 63,000 research publications have been collected and pre-processed. The publications involve at least one (co-)author affiliated to a Greek academic/research institution. They span over a period of nine years (2010-2018), and nearly 10,000 journals. The bibliographic data include metadata on subject (discipline) area(s). The findings are made public via a Web application[4]. The latter utilizes interactive graphs that facilitate the interpretation of the relevant bibliography data, and it is seen to comprise a springboard for conducting further data analytic and mining tasks.

Keywords: bibliography metadata · journal-to-journal references · data visualization · journal evaluation

1 Introduction

The HEAL-Link Consortium actively participates in projects, initiatives, and relevant developments that shape the scientific publications landscape in Greece.

[3] HEAL-Link, https://www.heal-link.gr/
[4] http://j2j.heal-link.gr/

This is done in a way that is beneficial to its members and to the scientific community, in general. Its main goals are:

- to establish common policies on journals subscriptions, and promote the rational growth of its journals collection,
- to combine financial savings with access to a large number of electronic resources that meet the educational and research needs of its user community, and
- to negotiate the signing of joint subscription agreements with publishers, and provide/manage remote access to electronic resources and information services (the electronic journals included) to its members.

The HEAL-Link Consortium receives central funding from the Hellenic Ministry of Education to provide access to electronic scientific subscriptions to its members. In 2019, the latter represent 661,647 users, 13,536 of which are academic faculty members. During the 1998-2018 period, the HEAL-Link users have conducted nearly 85 million article downloads from a total of 17,100 scientific journal titles[5]. Access to the latter is made possible via license agreements made with twenty-three (23) e-journal publishers. To effectively negotiate with the latter, HEAL-Link elaborate on a strategy plan involving the evaluation of journals on the basis of their usage as reference (source) material by the Greek academic/research community.

In order for the HEAL-Link consortium to exploit its financial resources and meet the needs of its academic community when it comes to accessing scientific publications, a set of criteria have been depicted that are taken into consideration and shape HEAL-Link's actions in relation with the evaluation of the value a scientific journal represents to the Consortium's members and user community. Criteria that relate to the drafting of a strategic plan for HEAL-Link when it comes to negotiating with the publishers who provide access to electronic subscriptions. The scientific journals evaluation process emphasizes on the usage of each journal as reference (source) material by the Greek academic/research community. The criteria used include:

- What journals do the Greek researchers publish with, and what (other) journals do they reference in the works? Once this question is addressed, its outcome could very well comprise a means for "quantifying" the value a journal represents to the Greek research community. Having done so, it will then become possible to quantify the effect/impact the discontinuation of a subscription will have to HEAL-Link members.
- The number of full-text downloads per scientific journal, conducted by HEAL-Link users over given time periods. This information is provided by the corresponding publishers.
- The journal's interdisciplinary value.

[5] HEAL-Link, https://www.heal-link.gr/

A journal can be rated to be important for HEAL-Link by also considering its scientific impact at the international level. It represents added value for HEAL-Link when Greek researchers reference its articles in their works, and, to a lesser extent, when they publish their works with it. The present paper focuses on scientific journals the Greek academic/research community publish their works with, and on the journals they reference in their works. The terms “citation” and “reference” are often used interchangeably. Some researchers define them as two complementary actions: “reference” as “acknowledgement to” and “citation” as “acknowledgment from” [7]. It is assumed that when a researcher makes a reference to a scientific article in his work, the journal the latter is published with deserves the credit for being a valuable source of scientific information. In this respect, the more references made by researchers to articles published with a given journal, the higher is the value the latter as a source of scientific information to the HEAL-Link user community.

2 Related work

Data visualization 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 results. Data visualization is the interpretation of information in visual terms by forming a mental picture based on data and the application of suitable methods to put data into visible form [5]. Additionally, [1] defines data visualization as “the use of computer supported, interactive, visual representations of data to amplify cognition”, while [8] point out that “data visualization is for exploration, for uncovering information, as well as for presenting information. It is certainly a goal of data visualization to present any information in the data, but another goal is to display the raw data themselves, revealing their inherent variability and uncertainty.”

Most related research work in this field and especially in bibliometrics has been done on science mapping. Science mapping aims to build science / bibliometric maps that describe how specific disciplines, scientific domains, or research fields are conceptually, intellectually, and socially structured [3]. It comprises an important research topic in the field of bibliometrics [4]. Science mapping involves a number of discrete steps: data retrieval, preparation and preprocessing, network extraction, mapping, analysis and visualization [3]. The preprocessing step is the most critical one. 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 [4]. Many software tools on science mapping have been developed, reviewed, analyzed and evaluated [3]. The ones most widely used in the literature can be found in [2] and [3].

Visualization tools have always been key elements not only in science and research, but also in any industry relating to the production of data that can be visualized [6]. The aim is to assist anyone who is involved in the production of knowledge, or in the decision making process. Visualization reveals information that cannot be found as such in raw data. The ability to see data clearly creates a capacity for building intuition that is unsurpassed by summary statistics [6].

The two most popular subscription-based bibliographic databases are (a) the Web of Science[6], an online subscription-based scientific citation indexing service maintained by Clarivate Analytics, and (b) Scopus[7], 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 Analytics[8] 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 communicate effectively the collected information, helping the subscribers make informed decisions, and assisting strategic initiatives. In higher education, an institution may specify various criteria in order to determine its peer and aspirational peer institutions. Peer institutions are other institutions that are at a similar institutional level and they have similar institutional characteristics. Aspirational peer institutions are institutions with similar institutional characteristics, yet they have better key performance indicators, such as significantly higher graduation rates and higher level performance. They are a want to be state of the institutions that set them as asparational. Peer and aspirational peer institutions can be defined for the overall institution, but they can also be defined for different schools or departments. Peer and aspirational peer institutions are appropriate for benchmarking purposes.

SCImago Journal & Country Rank from Elsevier[9] 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 maps 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.

3 Data collection, preparation and unification

In the direction of exploiting the aforementioned research goal, bibliographic metadata from nearly 63,000 research publications have been collected and processed. The collection consisted of research publications involving at least one (co-)author affiliation to a Greek HEAL-Link member academic/research institution. The Scopus database was used to retrieve research publications spanning

[6] https://clarivate.com/products/web-of-science
[7] https://www.scopus.com
[8] https://clarivate.com/products/incites
[9] https://www.scimagojr.com/

a period of nine years (2010-2018) and involving nearly 10,000 journals. The PostgreSQL RDBMS was used to organize the bibliographic data collection.

A number of models exist that classify scientific journals (Library of Congress classification, Web of Science, Scopus ASJC schema, etc.). We adopted the Scopus ASJC (All Science Journal Classification System) schema[10]. Our decision was based on the fact that the ASJC schema involves a small and manageable number of subject areas facilitating the handling of important cognitive areas/categories. ASJC 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 categories. Bibliographic data are coupled with metadata on subject (discipline) area(s). This information will be used to identify journals of interdisciplinary value during the next stages of our research.

To harvest the publications made during the 2010-2018 period by Greek researchers as well as the references they make to other journals, two API services were used: the Crossref REST API[11], and the Scopus APIs[12].

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 (co-)author who is affiliated to a Greek academic/research institution.

Table 1 presents the number of publications retrieved, grouped by publication type.

Table 1. Information about the publications

Publication Type	**Count**
Article	47035
Article in Press	1320
Book	341
Chapter	3437
Conference Paper	15271
Editorial	1644
Erratum	242
Letter	813
Note	467
Review	4094
Short Survey	96
Total	74760

The book and chapter type publications, plus any other publication not assigned ISSN number were removed from the dataset. ISSN is a unique 8-digit

[10] https://service.elsevier.com/app/answers/detail/a_id/15181/supporthub/scopus/
[11] https://api.crossref.org
[12] https://dev.elsevier.com/sc_apis.html

identifier assigned to periodical publications, irrespective of their medium (print or electronic)[13]. For the publications data used, ISSN primarily identifies scientific journals.

The publications harvesting stage was followed by the harvesting of metadata relating to the references made by the published works. The references harvesting process was split into two sub-processes. One for works published with Elsevier journals, and one for all other publications. This became necessary due to Elsevier's not making available to the Crossref REST API the references data of their published works. A separate data harvesting stage was conducted using the Scopus API in order to harvest the Elsevier journals references data. Most of the publications data records retrieved via the Crossref REST API include the corresponding references lists. The latter are of a diverse format, one that lacks a specific pattern even for works published with the same publisher. As an example, the digital object identifier (DOI) [14], which is the most important field as it is explained below, is either not present in the Crossref record, or it exists in a field of its own, or it is incorporated alongside with other information in other fields.

Today, nearly all published works are assigned a unique DOI. The latter is a unique alphanumeric string assigned by a registration agency (the International DOI Foundation[15]). It is used to identify content and provide a persistent link to its location on the Internet. As such, the DOI identifier facilitates the retrieval of the corresponding scientific journal metadata. In this respect, for the Crossref records that would not conform to a pre-specified pattern, the DOI identifiers were used to track down and retrieve the corresponding (referenced) journal metadata. For the Crossref records that lacked a DOI identifier, a separate data harvesting stage was conducted in order to retrieve the ISSN number of the journal each (referenced) work was published with.

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 identifiers for each one member of the latter (DOI, or Scopus identifier[16]). 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 of a number of records as possible and by doing so remain within the Scopus imposed quotas.

The next stage involved the harvesting of journal ISSNs, for journals the referenced works are published with. For references with DOIs, the ISSN identifiers were harvested from the Crossref REST API. For references lacking their DOI identifiers, data harvesting was conducted by using the the Scopus APIs, plus the Arxiv electronic preprints repository[17]. Additional checking was conducted

[13] https://www.issn.org/understanding-the-issn/what-is-an-issn/
[14] https://en.wikipedia.org/wiki/Digital_object_identifier
[15] https://www.doi.org
[16] https://www.elsevier.com/solutions/scopus/how-scopus-works/high-quality-data
[17] https://arxiv.org

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 fields, e.g. the journal title. When the number of records to be checked was reduced to a small value, even manual checking was conducted in order to identify the target journal.

4 The J2J-GR service

In addition to conducting the relevant data (pre-)processing and unification tasks, a Web application has been developed and implemented utilizing interactive graphs to facilitate the interpretation of the findings[18]. Aiming for the development of a powerful and informative Journal-to-Journal (J2J) associations service, the current version comprises one first step. It involves a dynamically generated set of graph networks from the journal to journal references data, focusing on references originating from publications involving at least one (co-)author who is affiliated to a Greek HEAL-Link member institution (target authors population). In this respect, the current version of the service targets a specific population of authors by considering scientific publications the latter have made during the 2010-2018 years period. The interactive graphs generated present the references made by the published works of the targeted authors population.

The application has been coded in R[19], utilizing the RStudio[20] project development environment. To construct the interactive Web apps straight from R, the Shiny[21] R package and shinydashboard[22] have been adopted.

As it is shown in Figure 1, the user interface involves two sections. The sidebar on the left, where the publication year, the subject area and the journal title are set together with some relevant fine tuning parameters, and the main section where the generated graph is presented.

The sidebar, marked as (1) in Figure 1, includes three selection menus: one for the publication year, one for the subject area, and one for the journal title of interest. In addition, the sidebar includes three filters for narrowing down the scope of the journals referenced by works published with the journal of interest. Filter number one is used to specify a range for the number of references made to other journal publications by the works of the target authors group published with the set journal during the set publication year. Filter number two enables the user to specify whether or not the list of referenced journals includes publications of the target authors group during the set publication year. Filter number three enables the user to include or exclude HEAL-Link subscriptions/journals.

[18] http://j2j.heal-link.gr/
[19] https://www.r-project.org
[20] https://www.rstudio.com
[21] https://shiny.rstudio.com
[22] https://rstudio.github.io/shinydashboard

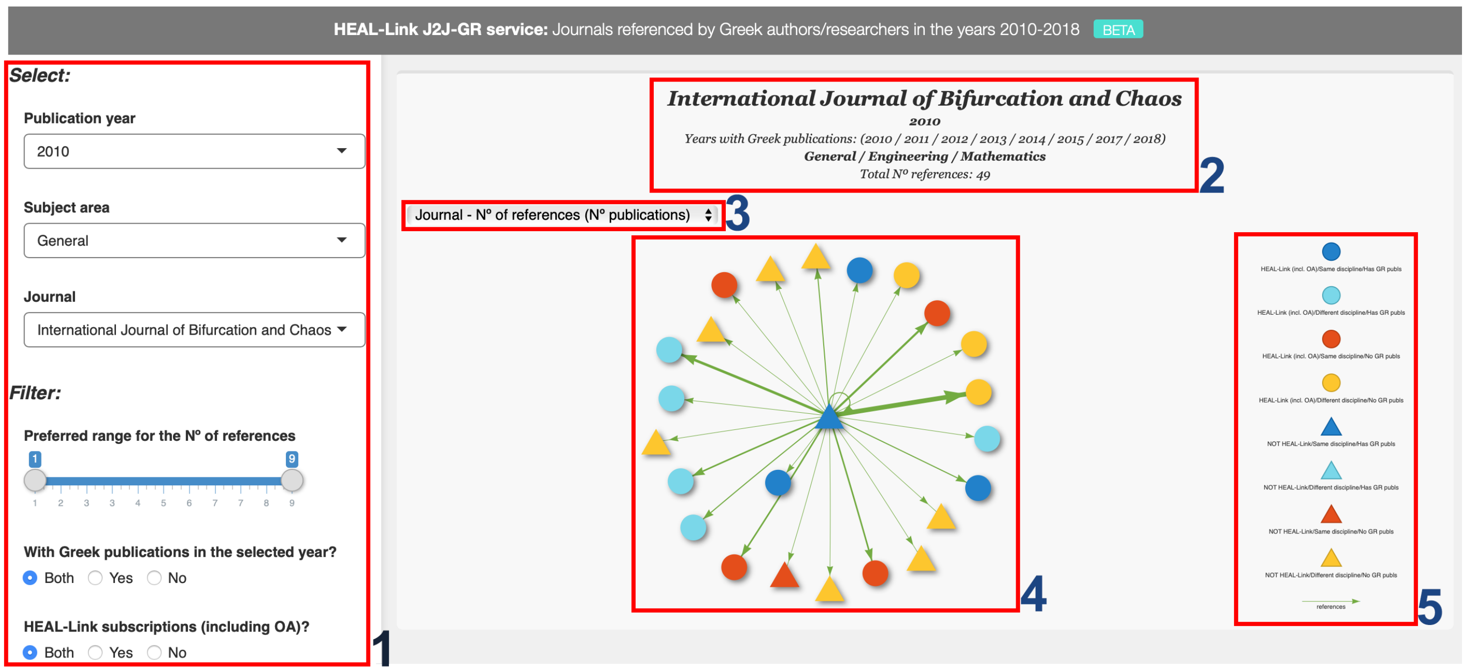

Fig. 1. The J2J-GR Web application user interface screen

The main page displays the graph generated in accordance with the settings made in the sidebar. The graph consists of four parts: the journal's title and information marked as (2), a drop down list of referenced journals marked as (3), the graph itself marked as (4), and its informative legend marked as (5) in Figure 1.

The journal selected in the sidebar becomes the central node in the graph. Its title section lists the year of interest, the years that this journal includes publications made by Greek researchers, the subject area(s) covered, and the total number references made to other journals from works (co-)authored by Greek researchers during the selected year. The graph is re-generated each time a (new) filter is applied on the sidebar.

The thickness of the graph's edges in Figure 1 encodes the number of references made to the corresponding journals. By hovering the mouse over any (target) journal the user can see information, such as its title, the number of works (co-) authored by Greek researchers published with it in the year of interest (if any) as well as the number of such publications during the 2010-2018 period, plus the number of references the central node (reference) journal makes to the journal in question.

The content of the dropdown list marked as number 2 in Figure 1 is displayed in Figure 2. It involves the same information displayed in the interactive graph, this time presented in textual form. When the user selects a (target) list element/journal, the corresponding graph node gets highlighted for the user to click on and have it become the new central node, if so desired. In this respect, the interactive graph supports a fan-out type of functionality. This is useful especially in cases of the user having to navigate himself in a large numbers of referenced (target) journals.

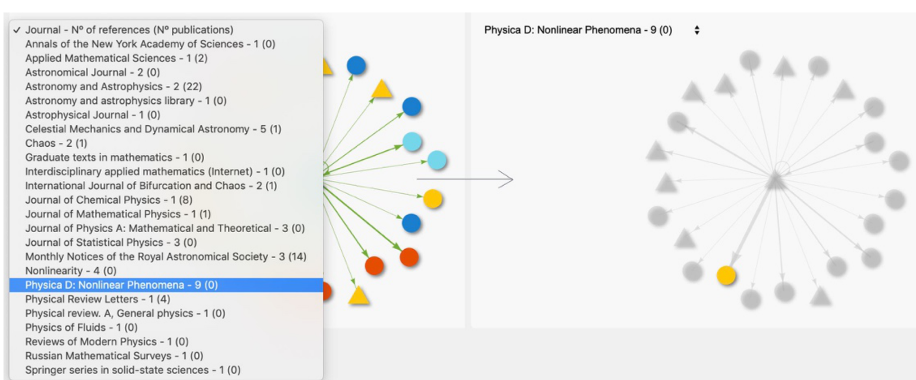

Fig. 2. The dropdown list.

The graph's legend marked as element number 5 on the screen shown in Figure 1 encodes characteristics of the graph's nodes/journals. Journal characteristics are encoded by combining geometrical shapes and colors. Geometrical shapes depict the inclusion or not of the journal in the the HEAL-Link subscriptions list: a circle indicates a journal that HEAL-Link members have access to. This includes HEAL-Link paid subscriptions, and HEAL-Link listed Open Access (OA) journals. On the other hand, a triangle indicates a journal outside the HEAL-Link paid subscriptions and OA list. Colors encode additional journal characteristics as follows: (a) blue indicates a journal that shares a common discipline area with the central node in the graph the Greek researchers have published works with in the year considered/selected, (b) turquoise indicates a journal that shares no common discipline area with the central node, yet one where Greek researchers have published works with in the year considered/selected, (c) red indicates a journal that shares a common discipline area with the central node with no works published by Greek researchers in the year considered/selected, and (d) yellow indicates a journal that shares no common discipline area with the central node with no works published by Greek researchers in the year considered/selected.

All journal nodes in the graph are clickable. When a user clicks on a node, a new graph is generated with the clicked journal as the (new) selected central node/journal. When this happens, the year selected remains the same, and the (new selected) journal title appears in the areas marked (1) and (2) in Figure 1. If the new central node shares the same subject (discipline) area considered so far then the name of the latter remains selected in the area marked (1) in Figure 1, otherwise a new subject area name (of the new central node) is selected automatically. Lastly, the three filter settings in the area marked (1) in Figure 1 they are all reset. Figures 3, and 4 demonstrate the stated fan-out type of functionality of the interactive graph shown in Figure 1.

5 Future work

Our next goal is to conduct exploratory analysis and data mining operations on the HEAL-Link bibliographic dataset. It will be interesting, for example, to consider the journals Greek researchers reference to next to the references made by of all the authors who publish in the same journals with the former. The exploratory analysis outcome could then be used to proceed and conduct predictive analytics in order to identify new potential additions to the HEAL-Link journals subscriptions list. Equally well, the same type of processing could facilitate the HEAL-Link decision making process for cancelling subscriptions dictated by budget cutbacks. Last but not least, the aforementioned analytical processing is expected to reveal journals of an interdisciplinary value to the HEAL-Link user community, i.e. journals used as a scientific information source material by researchers representing diverse scientific discipline areas. Such journals usually represent a high reference value to the HEAL-Link user community.

Fig. 3. Graph generation after clicking on a referenced journal Greek researchers have published with during the selected year.

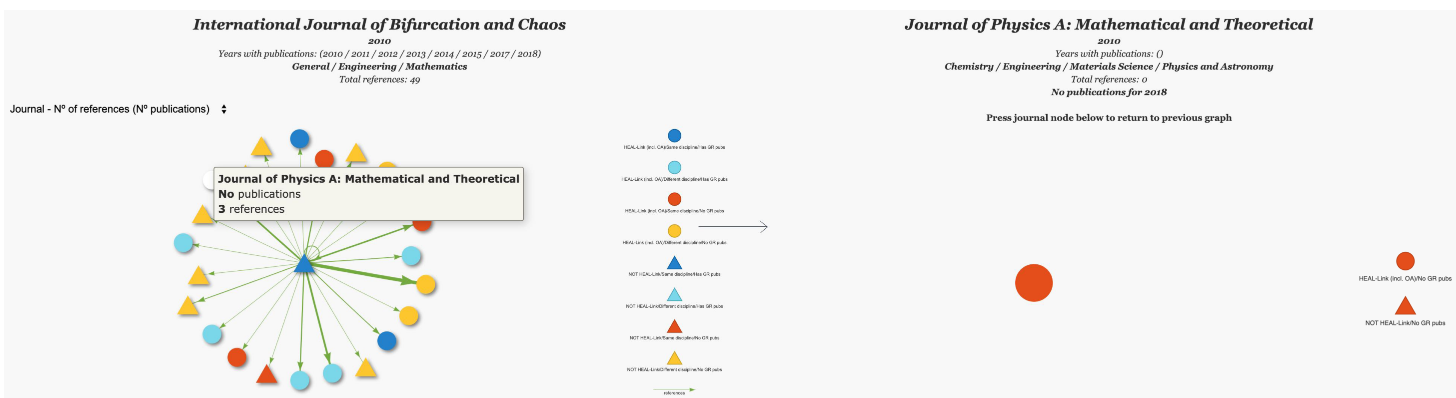

Fig. 4. Graph generation after clicking on a referenced journal Greek researchers have not published with during the selected year.

6 Conclusion

We report on the current of a research undertaken in order to exploit journal-to-journal references by considering publications made by a specific target authors group, namely researchers affiliated with academic institutions and research centers in Greece. The usage value each scientific journal represents is quantified by considering the references made to it by published works involving at least one (co-)author affiliation to a Greek academic and/or research institution. It should be borne in mind that a specific journal may represent a different usage value when considered in relation with the references received by works published by Greek researchers, as compared to when the same journal is considered in relation with the references received by all works published in journals the Greek researchers publish with. This possibility is worth exploiting further by conducting exploratory and predictive analytics type of processing to bibliographic datasets the Hellenic Academic Libraries Link (HEAL-Link) Consortium has access to.

The work reported herewith comprises a first step towards the research goal outlined in the previous paragraph. Having collected and pre-processed a first collection of releveant bibliographic data, an attempt has been made to realize the inherent journal-to-journal associations with respect to the references they make to one another. A first (beta) version of the J2J-GR Web application has been implemented and it is publicly available on the Internet. The focus is on works published in the 2010-2018 years period, and the references made by Greek researchers in works they have published with research journals during the stated period. It is anticipated that the presentation of the relevant information in the form of interactive graphs as in [23] will evolve into a bibliographic data exploratory tool useful to both the HEAL-Link Consortium and its user community.

References

1. Arnheim, R.: Visual Thinking. University of California Press, Berkeley, CA (1969)
2. Borner, K., Huang, W., Linnemeier, M., J. Duhon, R., Phillips, P., Ma, N., Zoss, A., Guo, H., A. Price, M.: Rete-netzwerk-red: Analyzing and visualizing scholarly networks using the network workbench tool. Scientometrics **83**, 863–876 (06 2010). https://doi.org/10.1007/s11192-009-0149-0
3. Cobo, M., LopezHerrera, A., HerreraViedma, E., Herrera, F.: 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 (2011), https://doi.org/10.1002/asi.21525
4. van Eck, N.J., Waltman, L., Dekker, R., van den Berg, J.: A Comparison of Two Techniques for Bibliometric Mapping: Multidimensional Scaling and VOS. J. Am. Soc. Inf. Sci. Technol. **61**(12), 2405–2416 (2010). https://doi.org/10.1002/asi.21421
5. Hinterberger, H.: Data visualization. In: LIU, L., ÖZSU, M.T. (eds.) Encyclopedia of Database Systems, pp. 652–657. Springer US, Boston, MA (2009).

[23] http://j2j.heal-link.gr/

https://doi.org/10.1007/978-0-387-39940-9_1370, https://doi.org/10.1007/978-0-387-39940-9_1370
6. Moody, J., McFarland, D., BenderdeMoll, S.: Dynamic network visualization. American Journal of Sociology **110**(4), 1206–1241 (2005). https://doi.org/10.1086/421509, https://doi.org/10.1086/421509
7. Narin, F., Carpenter, M.P.: National publication and citation comparisons. Journal of the American Society for Information Science **26**(2), 80–93 (1975). https://doi.org/10.1002/asi.4630260203
8. Unwin, A., Theus, M.: Graphics of a Large Dataset, pp. 227–249. Springer New York, New York, NY (2006). https://doi.org/10.1007/0-387-37977-0_11, https://doi.org/10.1007/0-387-37977-0_11