

Trends in science mapping: The co-use of scientific literature as evidence of the interests of researchers

Daniel Torres-Salinas¹, Evaristo Jiménez-Contreras² & Nicolás Robinson-García²

¹ EC3 Research Group & EC3Metrics

Universidad de Navarra, Pamplona Spain, torressalinas@gmail.com

² EC3 Research Group, Departamento de Información y Comunicación,
Universidad de Granada, Granada, Spain, elrobin@ugr.es

Abstract: This paper explores the possibility of constructing science maps based on the co-use of the staff of an academic institution made of scientific literature. For this, we define co-use as the co-occurrence of scientific information requests by a given user in platform of scientific journals. We use request data from the University of Navarre to the *ScienceDirect* platform in 2012 in order to analyze the potential of such methodological approach. We conclude by emphasizing the viability of such methodology when exploring the research interests of an academic institution along with the relations between different disciplines.

Keywords: science maps; information visualization; University of Navarre; co-use; consumption of scientific literature; co-downloading

Preprint version: 02-March-2014

Introduction

The shift to the electronic format has changed the way researchers consume scientific literature. In 2005 a fifth of the scientific papers read by researchers was in electronic format, currently, they read more than half of the publications this way (Tenopir, et al., in press). Also, their reading behaviour has changed: they read a larger number of papers, they do not read them thoroughly, spending less time per paper, and they tend to pay more attention to those which are shorter (van Noorden, 2014).

At the same time, new indicators have emerged as an alternative to the traditional bibliometric indicators (Cabezas-Clavijo & Torres-Salinas, 2010; Torres-Salinas & Cabezas-Clavijo, 2013), such as social bookmarking (Haustein & Siebenlist, 2011), requests to digital

libraries (Bollen et al., 2005), hyperlinks (Aguillo et al., 2006) or mentions in the social media (Torres, Cabezas & Jiménez, 2013). Hence, the range of alternatives to analyze the use made of scientific information now goes far beyond mere citation counts, extrapolating bibliometric methodologies to the world of the so-called *altmetrics*.

In this methodological extrapolation, one of the aspects to which less attention has been paid is the visualization and mapping of science, which could have its equivalent in the new metrics world. This maps would offer an interesting perspective from the traditional maps based on relational indicators such as co-citation (Small, 1973) or coauthorship (Luukkonen, Persson & Sivertsen, 1992). Here, the relationship would be established by the co-use of

papers, that is, two papers requested or used by a single user or in a single session.

This paper takes such approach exploring how to transfer the mapping techniques used for co-citation to online platforms of scientific journals, aiming at generating science maps based on co-use. With this perspective in mind, the goal is to explore the possibility of creating science maps of subject categories based on the co-use of information. We analyze the use of these maps for identifying the scientific interests of a particular research community, in this case, a university.

Material and methods

Here we have selected the University of Navarre as a case study in order to develop a science map based on the online requests - mainly PDF downloads and full-text HTML views - of the staff of the university in 2012. The usage data is based on the requests recorded by *ScienceDirect* and has been given as part of the *Elsevier Bibliometric Research Program*¹. The *ScienceDirect* platform includes more than 2500 journals covering the main areas of scientific knowledge. For each request, we obtained the IP number from which the query was made, an ID of the session, the date of the request and various identification fields referred to the journal. In order to assign a subject category to each request, we crossed the journal data with that from the *Scimago Journal Rank* (SJR). **Table 1** offers a basic overview of the dataset.

Then, we developed a matrix relating subject categories according to co-requests or co-use. The underlying idea is that each session in *ScienceDirect* responds to a

Table 1. General description of the requests made from the University of Navarre to ScienceDirect in 2012

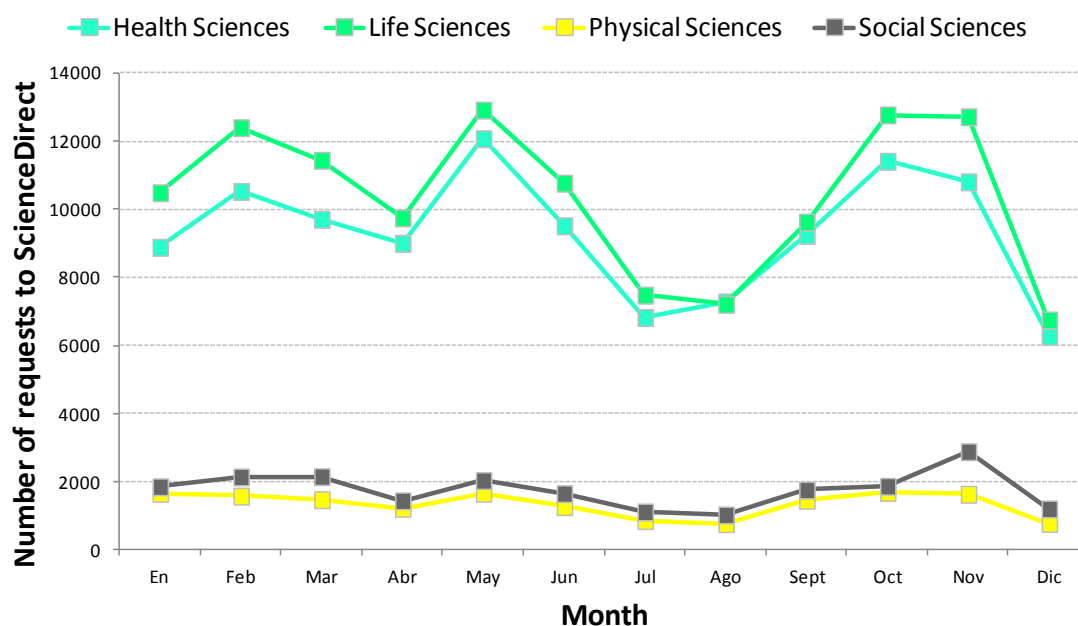
Indicator	Results
Requests	259547
Sessions	79693
IP Numbers	1845
Requests per session	3.26
Sessions per IP Number	43.21

single information demand, therefore, there is a thematic or content link between the papers requested in a single session, similarly to what happens with co-citation, where to papers referenced in a third paper are presumed to be related. Therefore, the term of co-use refers to the relations established between documents requested by a single user in the same session. Hence, if to papers belonging to journals of different categories are co-requested, co-downloaded or co-viewed in a single session, they establish a relation between those two categories. Parting from this premise, we developed a normalized distance matrix between SJR categories based on co-use based on the Jaccard coefficient (Leydesdorff, 2008). The final representation of the matrix was elaborated with Pajek, including just the most significant relations and representing the number of requests through the size of the nodes.

Results: Co-use science maps

75.98% of the requests recorded from the University of Navarre in 2012 were directed to papers from the fields of Life Sciences and Health Sciences. As observed

Figure 1. Chronological distribution by areas of the requests made from the University of Navarre to ScienceDirect in 2012



in **Figure 1** the most requested papers belonged to the Life Sciences followed by the Health Sciences. On the other hand, Social Sciences and Humanities represent barely more than 8% of the total requests, while Physical Sciences reach approximately 6% of the total share.

The overrepresentation of Life and Health Sciences is confirmed when observing the co-use map in **Figure 2**. Here we see a map characterized by the predominance of categories belonging to these areas, as observed on the upper part of the figure. The upper left is formed by categories belonging to the Life Sciences and connects through Pharmacy and Biochemistry with Medicine (in the upper right). The lower right represents categories related with Economics, Business and Social Sciences, and connect with Medicine through

Nursing. Engineering and Computer Science connect with the Environmental Sciences (lower left) and the Physical Sciences are situated in the middle of the map connecting Physics with Chemistry and this with Biochemistry, hence closing the cycle as this field connects again with the Health Sciences.

If we focus in each area we will see in more detail the disciplines that form each of these areas. This is what we do in **Figure 3**, where we focus on Medicine. Here we observe that there are three branches which are of greater interest for the staff of the University of Navarre. All of them converging through Surgery. The first branch is that related with Cancer Research, Oncology, etc. It is located on the upper right side of the figure. The Second branch is that formed by disciplines

Figure 2. General co-use map of the University of Navarre based on requests made to ScienceDirect in 2012

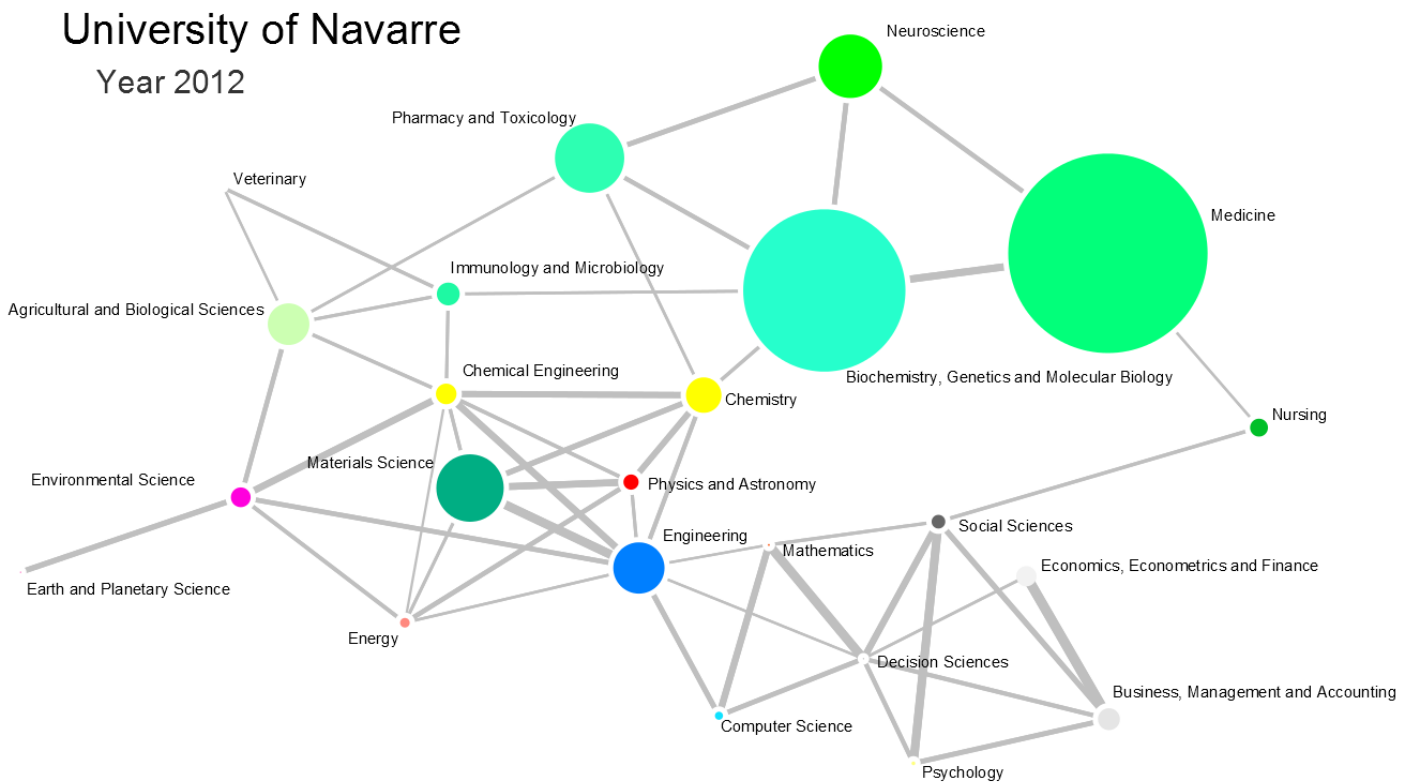
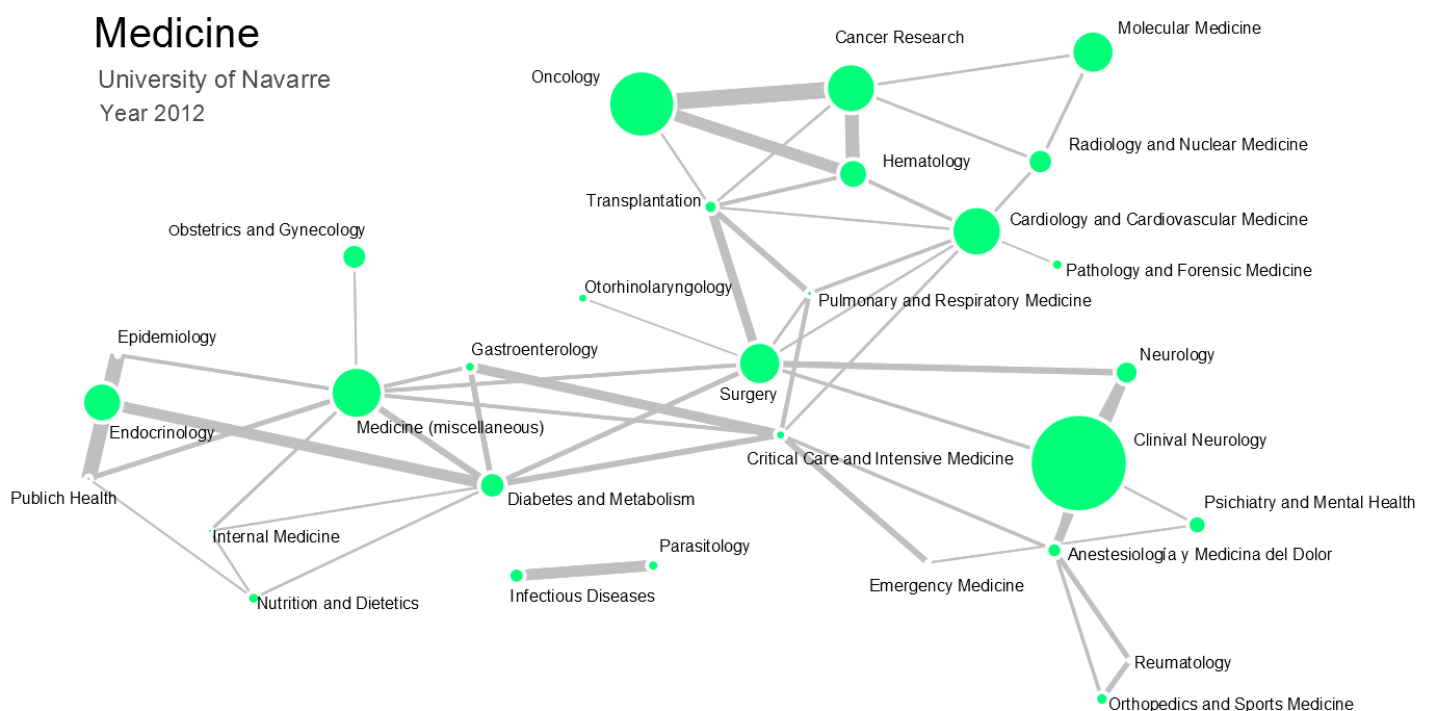


Figure 3. Co-use map of Medicine based on the requests made by the University of Navarre to ScienceDirect in 2012



related with Clinical Neurology as the main focus. A subgroup falls from this branch connected with Anesthesiology and Pain Medicine. This subgroup is formed by Rheumatology and Orthopedics and Sport Medicine. Finally, the third branch is the one related with Endocrinology, Diabetes and Metabolism. However, in this case there is a miscellaneous category that may blur the picture. We must highlight a forth cluster formed by Parasitology and Infectious Diseases separate from the rest of the network.

These maps seem to reflect the research structure of the University of Navarre, as we observe when comparing it with its scientific output. This university shows high rates of disciplinary specialization when compared with the rest of the Spanish universities, in the fields of Medicine (2,13) and Economics (2,02) (**Torres-Salinas; Delgado López-Cózar; Moreno-Torres; Herrera**, 2011). These two areas are well-represented in **Figure 2**. Also the size of the nodes according to the online requests seems to be well-represented as Medicine and Biochemistry are in fact the most productive disciplines of the university in 2012, according to the information provided by the Scopus database. Only these two areas represent 46.1% of the whole production of the university.

Concluding remarks

In this paper we explore the possibility of developing maps that illustrate the co-use of publications as an alternative to classical techniques based on co-citation and bibliographic coupling. This approach was already suggested by others (**Bollen et al.**, 2005; **Brody; Harnad; Carr**, 2006), however this is the first study in which we apply such methodology at an institutional level in order to explore the research interests

of a given community of users. Co-use maps enrich the analysis as they adopt a much more inclusive perspective, encompassing students, practitioners and teachers.

Having said that and despite showing promising results, there is still much research to be done deepening on the significance of this type of maps and their potential use as a complement to the traditional science maps. Their broader approach, contemplating other uses of the scientific information rather than research itself makes us consider their utility for two possible users:

- 1) Library boards and committees. These maps would allow librarians to easily visualize the use their patrons make of their subscriptions to journal platforms such as *ScienceDirect* as well as the relation that exists between their subscriptions and the interests of their community.
- 2) Research policy managers. Monitoring the real-time use that researchers make of scientific literature may be very useful for decision making.

Finally, this visualization technique based on co-use could be easily extrapolated to other platforms and contexts from the academic social web. Most of these social media platforms allow us to calculate altmetric indicators which could be used for developing such maps. An example of such extrapolation could be Mendeley, where we could construct similarity matrices based on articles that co-occur within a single user's library. The outreach of social media along with researchers' shift towards the digital format offer a new landscape in which we can redefine the concept of use when developing information visualization techniques. The

emergence of new indicators offer new ways for tracking the relationship between disciplines and researchers' interests.

Acknowledgements

The authors would like to thank Gali Halevi and Henk Moed for granting usage data from *ScienceDirect* and Víctor Sanz Santacruz, Head of the Library of the University of Navarre for allowing the access to such dataset. Nicolás Robinson-García is currently supported by a FPU grant from the Spanish Ministerio de Economía y Competitividad.

Notes

Information regarding the *Elsevier Bibliometric Research Program* is accessible at: <http://ebrp.elsevier.com>. This study is part of the on-going project *Viability of co-downloading data analysis form mapping interdisciplinary research at institutional level*. For more details of the project, please visit the aforementioned website..

References

Aguillo, I.F.; Granadino, B.; Ortega, J.L.; Prieto, J.A. (2006). Scientific research activity measured with cybermetrics indicators. *Journal of the American Society for Information Science and Technology*, vol. 57, n. 10, pp. 1296-1306. <http://dx.doi.org/10.1002/asi.20433>

Bollen, J.; Van de Sompel, H.; Smith, J.A.; Luce, R. (2005). Toward alternative metrics of journal impact: A comparison of download and citation data. *Information Processing & Management*, vol. 41, n. 6, pp. 1419-1440. <http://dx.doi.org/10.1016/j.ipm.2005.03.024>

Brody, T.; Harnad, S.; Carr, L. (2006). Earlier web usage statistics as predictors of later citation impact. *Journal of the American Society for Information Science and Technology*, vol. 57, n. 8, pp. 1060-1072. <http://dx.doi.org/10.1002/asi.20373>

Cabezas-Clavijo, Á.; Torres-Salinas, D. (2010). Indicadores de uso y participación en las revistas científicas 2.0: El caso de PLoS One. *El profesional de*

la información, vol. 19, n. 4, pp. 431-434. <http://dx.doi.org/10.3145/epi.2010.jul.14>

Haustein, S.; Liebenlist, T. (2011). Applying social bookmarking data to evaluate journal usage. *Journal of Informetrics*, vol. 5, n. 3, pp. 446-457. <http://dx.doi.org/10.1016/j.joi.2011.04.002>

Leydesdorff, L. (2008). On the normalization and visualization of author co-citation data: Salton's Cosine versus de Jaccard Index. *Journal of the American Society for Information Science and Technology*, vol. 59, n. 1, pp. 77-85. <http://dx.doi.org/10.1002/asi.20732>

Luukkonen, T.; Persson, O.; Sivertsen, G. (1992). Understanding patterns of international scientific collaboration. *Science, Technology & Human Values*, vol. 17, n. 1, pp. 101-126. <http://dx.doi.org/10.1177/016224399201700106>

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 and Technology*, vol. 24, n. 4, pp. 265-269.

Tenopir, C.; King, D.W.; Christian, L.; Volentine, R. (en prensa). Scholarly article seeking , reading and use: A continuing evolution from print to electronic. *Learned Publishing*. <http://libvalue.cci.utk.edu/sites/default/files/TenopirChanges.2.3.14.pdf>

Torres-Salinas, D.; Delgado López-Cózar, E.; Moreno-Torres, J.G.; Herrera, F. (2011). Rankings de las universidades españolas según campos científicos: Descripción y resultados. *El profesional de la información*, vol. 20, n.1, pp. 111-118.

Torres-Salinas, D.; Cabezas-Clavijo, Á. (2013). Altmetrics: no todo lo que se puede contar, cuenta. *Anuario ThinkEPI*, vol. 7, pp. 114-117.

Torres, D.; Cabezas, Á.; Jiménez, E. (2013). Altmetrics: Nuevos indicadores para la comunicación científica en la Web 2.0. *Comunicar*, vol. 41, pp. 53-60. <http://dx.doi.org/10.3916/C41-2013-05>

Van Noorden, R. (2014). Scientists may be reaching a peak in reading habits. *Nature*. <http://dx.doi.org/10.1038/nature.2014.14658>