Assessing University Research Output Deploying Freely Accessible Bibliometric Tools

Juncal Zabalza-Vivanco [1], Rosa Rio-Belver [1], Ernesto Cilleruelo [2], Gaizka Garechana [3], Javier Gavilanes [1]

Increasingly, there is a need to evaluate the research productivity of universities. This need stems from the fact that new laws concerning research activity have appeared in the last few years. In addition, there are other reasons that can be considered as grounds of this necessity, above all, those related to recruitment requirements and tenure. For that purpose, bibliometric resources such as bibliographic data are important methods to quantify research activity in terms of that what is published as peer-reviewed journal articles. Commercial bibliometric products are commonly used to capture this information, and it is widely used by university rankings and research assessment (European Commission 2006). Another main aspect that should be taken into account is the fact that in last decade the great advances in the field of bibliometrics have resulted in developing new resources, some of them freely available, which bring about the access of mass audience to bibliometrics.

So the main objective of this work is to describe the key features of some free bibliometric resources that can be deployed to assess the research productivity in a university, higlighting among them Google Scholar (GS), Publish or Perish (PoP) and Sci2. Firstly, Google Scholar is a source of bibliometric data. Secondly, Publish or Perish is a software program that retrieves and analyzes data from Google Scholar. Finally, Sci2 is a science mapping software tool that builds bilbiometric maps to describe research production.

Methodology

Data Acquisition: data source and datasets. The two datasets, downloaded from ISI WoS (136 records) and GS (329 records), consist of all publications produced by authors who work at university. The search in both databases is performed according to the author's name. In the case of GS dataset the data were obtainded using PoP. Afterwards Sci2 tool supports the loading and pre-processing of both publication datasets.

Data Preparation *and* Preprocessing. Each dataset is sliced into six intervals of five years. The total period goes from 1986 to 2011. This analysis will be deployed later with Co-authorship network to show the evolution of authors over time.

Types of Analysis. A well designed and executed network analysis satisfies a specific insight need, it provides an answer to a concrete question. Five general types of question exist: when (temporal analysis), where (geospatial analysis), what (topical analysis), with whom (network analysis) and why (modeling) (Börner 2011).

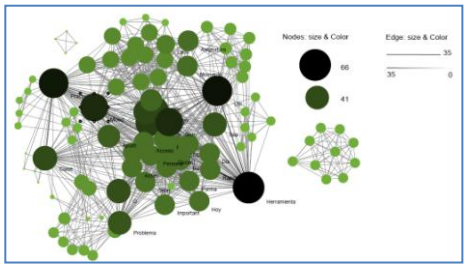
Temporal Analysis (When)aims to show the conceptual, intellectual, or social evolution of the research field, discovering patterns, trends, seasonality, and outliers. We particularly concentrate on **Author Burst detection analysis** that discovers authors who have high intensity over limied durations of time intervals.

[1] University of the Basque Country, SP-01006 Vitoria-Gasteiz, Nieves Cano 12, e-mail: rosamaria.rio@ehu.es, Javier.gavilanes@ehu.es, juncal.zabalza@ehu.es.
[2] University of the Basque Country, SP-48030 Bilbao, Alameda Urquijo s/n, e-mail: ernesto.cilleruelo@ehu.es
[3] University of the Basque Country, SP-48008 Bilbao, Elkano 21, e-mail: gaizka.garechana@ehu.es

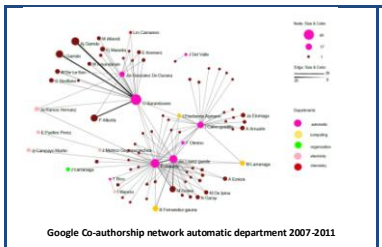
Topical Analysis (what) extracts the set of unique words or word profiles and their frequency from a text corpus. The aim of this map is to visualize the results of a survey. For that purpose, we want to identify the words that most frequently appear in the answers and how often these words appear together with the rest. The main benefit related to this network is that we can manage a clear, quick view of survey´s results without reading all responses

Network Analysis (with whom). In a network, authors, words, papers, journals, patents, and funding are represented as nodes and their interrelations as edges. We will only concentrate on the following examples: **co-authorship network**, **bipartite network** and **bibliographic coupling network.**

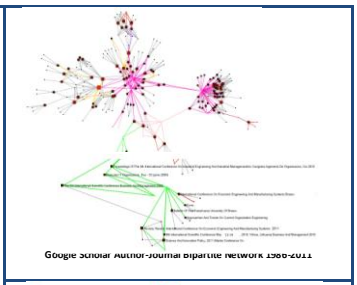
ISI author Burst Detection 1987-2011



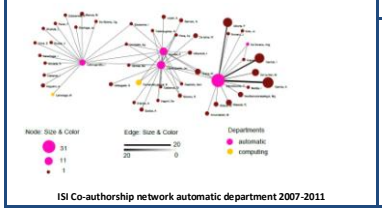
Survey Word Co-occurrence



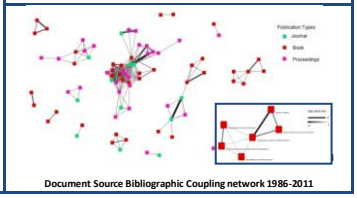
Google Co-authorship network automatic department 2007-2011



Google Scholar Author-Journal Bipartite Network 1986-2011



ISI Co-authorship network automatic department 2007-2011



Document Source Bibliographic Coupling network 1986-2011

Conclusions

With regard to bibliometric databases, for one thing, Google Scholar does not provide its sources and contains a less quality-controlled collection of publications. For another, ISI Web of Science covers scientific research journals with the most considerable impact and basically covers North American, Western European and English-language publications. Consequently, taking into consideration the different database features and the fact that most works comes from proceedings, college materials, book chapters or dissertations and many of them are written in Spanish, the data retrieval from ISI Web of Science in comparison to that of the Google Scholar is not the same, the ISI Web of Science sample is much smaller than the Google Scholar one.

On the question of data cleaning, this could be considered the most laborious step. One of the most relevant deductions that we can draw from this point is the fact that authors should always write their name and affiliation in the same way. What´s more, in the case of Spanish authors, they should use their second surname.

In summary, there is a wide range of analyses and visualization maps obtained through the deployment of freely available bibliometric tools. Moreover, they enable us to evaluate the scientific production over time, as well as, its current situation. This assessment helps us to find its major weak points and new opportunities concerning research output.

References

Börner, K. 2011. Network Science: Theory, Tools and Practice. In *William Sims Bainbridge, Ed. Leadership in Science and Technology: A Reference Handbook*. SAGE Publications, Inc.

European Commission. (2006). Assessing Europe's University-Based Research - Expert Group on Assessment of University-Based Research, doi: 10.2777/80193. 2010.