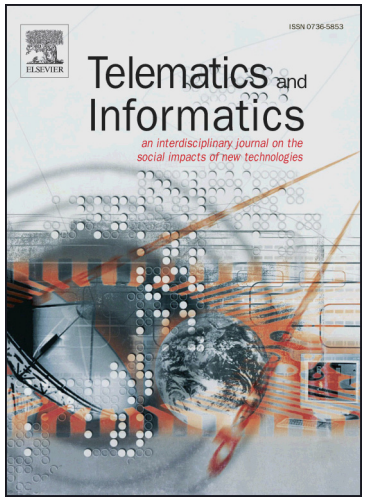

A fast method for identifying worldwide scientific collaborations using the Scopus database

Francisco G. Montoya, Alfredo Alcayde, Raúl Baños, Francisco Manzano-Agugliaro

PII: S0736-5853(17)30541-5
DOI: https://doi.org/10.1016/j.tele.2017.10.010
Reference: TELE 1025

To appear in: *Telematics and Informatics*

Received Date: 18 August 2017
Revised Date: 11 October 2017
Accepted Date: 15 October 2017

Please cite this article as: Montoya, F.G., Alcayde, A., Baños, R., Manzano-Agugliaro, F., A fast method for identifying worldwide scientific collaborations using the Scopus database, *Telematics and Informatics* (2017), doi: https://doi.org/10.1016/j.tele.2017.10.010

A fast method for identifying worldwide scientific collaborations using the Scopus database

Francisco G. Montoya[a], Alfredo Alcayde[a], Raúl Baños[a], Francisco Manzano-Agugliaro[a]

[a] *CeiA3, Department of Engineering, University of Almería, Carretera de Sacramento s/n, 04120 Almería (Spain)*

Abstract

Science is essential for human prosperity because social and technological advances often depend on scientific advances. Science is living a golden era characterized by a rapidly growing number of researchers worldwide exploring different disciplines and research fields. Keeping in mind that funding is limited, many researchers are encouraged to establish new collaborations with individuals or groups of researchers. Furthermore, the funding bodies use increasingly complex criteria to determine the researchers and projects to be supported. In this regard, the analysis of scientific collaboration networks can help to determine the main areas of specialization of universities and research centres, as well as the type of internal and external collaborations of their researchers. This paper presents an advanced method for analysing scientific collaboration networks at universities and research institutions. This method is based on automatically obtaining bibliographic data from scientific publications through the use of the Scopus Database API Interface, which are then analysed using graph visualization software and statistical tools. This model has been validated through the analysis of a real university, and the results show that it is possible to determine in a fast way and with high reliability the main research lines of an institution as well as the structure of the collaboration network. The method opens new perspectives for the study of scientific collaboration networks because it can be applied at different levels of detail, from small research groups to large academic and research centres, and over different time frames.

Keywords: Data mining methods, Database management, Information retrieval, Scientific collaboration, Network Analysis

1. Introduction

Now, as never before, the scientific community has the material and human resources to carry out relevant research activities in multiple disciplines and research fields. According to some studies, world scientific production doubles every nine years Noorden (2014). Thanks to the development of Information and Communication Technologies (ICT), it is now possible to access most of the scientific literature through the Internet (Asadi et al., 2017; Ravishankar, 2013), which allows transferring knowledge effectively. However, this process also leads to increasing competition between researchers (Whitley, 2003) and higher quality standards that increase the demands on research productivity (Guthrie et al., 1993). These changes encourage researchers, especially those in the early stages of their careers, to establish collaborative relationships with other researchers from their own or another institution (Lewis et al., 2012). The dynamism of these collaborations makes the analysis of the scientific collaboration network a topic of great interest (Shafiq et al., 2015; Shen et al., 2010; Aron, 2009), both for researchers and for funding agencies (Laudel and Gläser, 2014).

Research activities are funded by public institutions or private organizations, which invest many millions of dollars to provide technical and human resources to guarantee that the established objectives will be reached (McNutt, 2014). However, evaluating scientific quality is a notoriously difficult problem that has no standard solution (Seglen, 1997). In fact, the evaluation criteria used to assign these budgets are usually based on scientometric indicators retrieved through web search engines such as Scopus, Google Scholar, etc., which include the number and quality of publications by researchers based on journal and author metrics. Nevertheless, these evaluation criteria do not explicitly take into account the relationships between