Journal of Informetrics 12 (2018) 42–58

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

# Reference density trends in the major disciplines

Susana Sánchez-Gil[a,b,c,*], Juan Gorraiz[d], David Melero-Fuentes[e]

[a] Donostia International Physics Center (DIPC), San Sebastián, Spain
[b] Euskampus Fundazioa, Leioa, Spain
[c] Ikerbasque, Bilbao, Spain
[d] University of Vienna, Vienna University Library, Bibliometrics and Publication Strategies, Boltzmanngasse 5, 1090 Vienna, Austria
[e] Institute of Documentation and Information Technologies (INDOTEI), Catholic University of Valencia, Spain

ARTICLE INFO

*Article history:*
Received 26 April 2017
Received in revised form
15 November 2017
Accepted 20 November 2017

*Keywords:*
Reference density
Citation frequency
Publication habits
Open access
Scimago Journal & Country Rank

ABSTRACT

The aim of this study was to determine whether different areas of knowledge presented different behaviour with regard to the number of references cited per journal document or if, conversely, they shared the same reference density practices. Bibliometric and bibliographic data were collected from 27,141 journals (indexed between 2001 and 2015 in the SCImago Journal & Country Rank (SJR)) and the growth rates in reference density and number of documents and journals in each category were calculated at different levels of aggregation.

Our analysis identified that (a) mean reference density values in some Social Sciences and Arts and Humanities categories were equal to or higher than those in the "hard sciences"; (b) reference density growth rates in these disciplines were not as high as those in the hard sciences and, in general, did not correspond with growth rates in the number of documents produced; (c) this can be considered an indication that citation-based evaluation practices affect publication habits; and (d) no significant differences were found in mean values or growth rates between Gold Open Access and Non Gold Open Access journals.



## 1. Introduction

The advancement of science relies on the effective communication of the knowledge generated by members of the scientific community. As part of this process, the rapid exchange of knowledge and experience is a significant factor in the production of new research and, therefore, development of the different areas of knowledge. Publication of research results is the main output of research activity (Gross, Harmon, & Reidy, 2002). There are several formal channels through which to communicate research results, such as oral presentations at conferences, monographs, patents, research reports and scientific journals. The latter predominate in the scientific system (Pacheco-Mendoza & Milanés, 2009) because they publish the latest advances in science, constitute the most rapid means of disseminating research results and, in the majority of cases, are indexed in the major scientific databases. It would be useful here to indicate the differences between areas of knowledge and subject categories. Traditionally, areas such as the Social Sciences and Humanities have tended to publish a higher number of monographs and book chapters, whereas in Engineering it is more usual to communicate knowledge through conference proceedings and scientific articles (which predominate in the hard sciences) (Gumpenberger, Sorz, Wieland, & Gorraiz, 2016; Moed, 2005; Nederhof, 2006; Rovira, 2007; REF, 2014; Sanz-Casado et al., 2002).

* Corresponding author at: Donostia International Physics Center (DIPC), San Sebastián, Spain.
*E-mail address:* susanagil@euskampus.eu (S. Sánchez-Gil).

https://doi.org/10.1016/j.joi.2017.11.003
1751-1577/

The publication of research results enhances a researcher's reputation and brings greater recognition and esteem within his or her scientific community (Sanz-Menéndez, 2003) to the extent that, today, it has become a case of 'publish or perish'. This enhancement of a researcher's reputation, recognition and esteem is closely linked to the evaluation of the research activity. Scientific production *per se* is not an end in itself; impact and visibility are also very important aspects (Gorraiz, Wieland, & Gumpenberger, 2012), so it is of high interest to develop approaches that can be correctly measured and quantified. Consequently, it is necessary to construct and apply indicators that are tailored to evaluate the different processes or aspects (Costas-Comesaña, 2008). These should be designed to be simple to use while still being sensitive to the complexity of the research, standardised within the knowledge area and specific to the subject categories (Gorraiz & Gumpenberger, 2015; Gorraiz, Wieland, & Gumpenberger, 2016; Hicks, Wounters, Waltman, de Rijcke, & Rafols, 2015).

Recent decades have not only witnessed an upsurge in the number of scientific publications (OECD, 2015) but also in the density of citations and references that these receive and contain. According to Moed (2005), and in line with the theories proposed by Merton, references serve two functions. The first is related to content, since they can be used as descriptors of the document they support, and the second is a symbolic payment of intellectual debts or merit recognition (Small, 2004). Citations are also used as a proxy for impact assessment and are commonly used in research evaluation processes.

Citations are generated in the form of literature references and they are usually listed at the end of the publication. Undoubtedly, citations and/or references shape the structure of scientific communication and modern papers contain an abundance of knowledge and information, compiled in their References

Evaluation practices that are largely based on quantitative methods, unquestionably influence the publication habits and practices of researchers. However, it is still a matter of debate as to whether they do so in a manner that is positive or negative for scientific communication. Naturally, scientists tend to follow the rules imposed by the evaluation system in order to increase the worth of their publications, consolidate their prestige and achieve greater recognition (Ellison, 2002; Lange, 1985; Seglen, 1997; Sánchez-Gil, 2014).

Our starting point is the hypothesis that citation habits in different areas of knowledge, and their component categories, can exert a significant influence on the size of a document's reference lists and, therefore, on the mean reference density values.

In this study, we have analysed trends over time regarding the density of references that are cited in articles published in journals. In this context, the term 'references' should be understood to mean the publications listed in the reference list (as well as named literature in some journals or books). In this paper, we use the term *reference density* to refer to the number of references cited per document, which could equally be termed 'citation density' (Garfield, 2007; Van Raan, 2008). However, we have employed this terminology to distinguish between cited and citing papers.

This paper considers whether evaluation systems that are based primarily on citations, lead to an increase in the number of references cited by authors in order to trigger a cumulative effect; thus increasing or inflating the number of citations of each publication, particularly of their own work.

Several earlier studies have analysed references from different perspectives, including the age of the references, the type of document referenced and reference trends in disciplines and journals. Albarrán and Ruiz-Castillo (2011) attempted to determine the distribution of references in an article and the relationship with citations received by articles in different areas of knowledge. Later, Ucar et al. (2014) observed an increase in references in the field of Engineering and related this to greater access to information as a result of the Internet. Finally, Lin and Huang (2012) studied the relationship between co-authorship, self-citing and references; analysing whether a greater number of self-citations is related to lower co-authorship or whether researchers tend to cite their own previous work before that of others.

Behaviour has changed in recent years regarding the number of papers referenced per document, both in different areas of knowledge and in specific subject categories. This change in behaviour has given rise to variability in impact indicators such as the Scimago Journal Rank and the impact factor (IF) (Althouse, West, Bergstrom, & Bergstrom, 2009; Bornmann and Pudovkin, 2017). Consequently, techniques have been introduced to correct these differences, such as the method proposed by Zitt and Small (2008) which formed the basis of the new indicator for measuring journal impact introduced by Moed (2010, 2011), the Source Normalised Impact per Paper (SNIP) which is calculated using data from Scopus (2004) (see also Waltman, van Eck, van Leeuwen, & Visser, 2013).

The aim of this study was to determine whether there were differences in behaviour regarding the density of references generated in different areas of knowledge and at different levels of aggregation, in addition to analysing whether these hypothetical differences were related to an increase in the number of documents available on databases or not.

Given these objectives, we posed the following research questions:

1) How has reference density increased in the 15 years between 2001 and 2015?
2) Are there any differences in the volume and rate of growth between scientific categories and between "hard and soft" sciences?
3) Are reference density growth rates related to an increase in the number of citable documents (i.e. do some categories or fields present a degree of saturation)?
4) Are there differences between Gold Open Access journals and those with more traditional access policies governed by the subscription model?