

Development of a General Methodology for the Analysis of the “Durability” of Scientific Documents

Rodrigo Costas¹, Thed van Leeuwen² and Anthony F. J. van Raan³

¹costascomesanar@cwts.leidenuniv.nl, ²leeuwen@cwts.leidenuniv.nl, ³vanraan@cwts.leidenuniv.nl
Centre for Science and Technology Studies (CWTS-Leiden University), PO Box 905, 2300 AX Leiden (The Netherlands)

Introduction

The study of the citation histories and ageing of documents are topics that have been addressed from several perspectives (Aversa, 1985, Moed et al, 1998, Aksnes, 2003), especially in the analysis of documents with “delayed recognition” (Glanzel et al, 2003) or “Sleeping beauties” (van Raan, 2004). However, there are no general methodologies for the analysis of the durability of documents that can be extensively applied for different time periods and/or research fields. In this paper, a tentative general methodology of classification of documents according to general patterns in their citation histories is described.

Objectives

To develop a methodology for the classification of research publications, according to the “durability” of their citations; and also to study the main characteristics of documents according to it.

Methodology

A broad methodology for the analysis of durability of the scientific papers has been developed considering three types of general durability:

- *Normal type*: documents with a ‘normal’ distribution in their citations over time.
- *Flash in the pans*: documents that receive citations immediately after their publication but they are not cited in the long term (van Dalen & Henkens, 2005).
- *Delayed type*: documents that receive their citations later than normal documents.

The methodology of classification of documents is based on the distribution of the percentage of the citations that documents receive each year (citation history), according to the following steps:

1. For each document with citations, the chronological evolution of its citations (self-citations excluded) from the year of publication until the last year considered in the analysis (2008) is provided.

2. The percentage of citations received each year has been calculated, as well as the cumulative value.
3. The year after publication when documents achieve the 50% (or more) of their citations is calculated (“Year 50%”) assigning an ordinal number to the document (for example, for documents published in 1994 that received 50% of their citations in 1994 is 1, is 2 when the 50% of citations come in 1995, 3 if is in 1996, 4 if is in 1997, etc).
4. For the whole population of documents in the same scientific field, percentiles 25 (P25) and 75 (P75) of this new value (“Year 50%”) are calculated for each publication year.
5. The general criteria for classification of documents are: a) *Flash in the pans* < P25; b) *Delayed* documents > P75; c) *Normal* documents >= P25 and <= P75.

An analysis based on the publications of the JCR Subject Category of “Entomology” has been performed. All documents published during the period 1980-2008 in this field were considered (125781). Finally, only those with external citations were included in the analysis (87545, 70%) being classified according to the methodology described.

Results

The evolution of the percentage of citations and the Citation per publication rate (CPP) for Entomology are presented in Figures 1 and 2 (40962 documents with at least 5 citations and until 2003).

The three types of documents present different patterns in the evolution of the percentage of citations and in their CPP. *Flash in the pans* present higher levels of citations during the most immediate years after publication but with a rapid decrease. *Normal* documents present a stable citation pattern, having their peak around the 4th-5th year after publication. *Delayed* documents present an increasing pattern in their citations lingering more time.

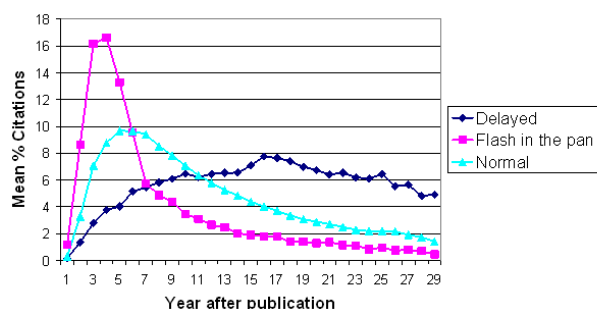


Figure 1. Evolution of percentage of Citations and CPP (included documents ≥ 5 citations and 1980-2003).

According to the classification of documents, it is possible to detect some general properties of the three types. In this sense, *Delayed* documents present a proportional higher rate of papers written in languages other than English, they are also the longer publications (more pages per paper) with the lower rate of authors per document and the documents with the highest percentage of External Citations (EC) and the lowest Self-citations (SC), they also present more citations than *Flash in the pans* but less than *Normal* documents. On the other hand, *Flash in the pans* present the higher percentage of “Other type” documents (including Notes, Letters, etc. –but not reviews–), they are also the documents with the lowest impact (CPP) and the highest percentage of self-citations, while they are the shortest documents as well as the documents with the highest rate of authors involved.

Discussion and Conclusions

The methodology suggested presents several advantages: firstly, it can be applied to all fields and periods of analysis (considering a minimum of 5 years citation history); secondly, it provides a nice picture of the different durability patterns and allows the study of the main characteristics of these documents. This classification and further analyses based on it can contribute to a better determination of the most adequate citation windows for bibliometric analysis, for example taking into consideration the research field and its characteristics.

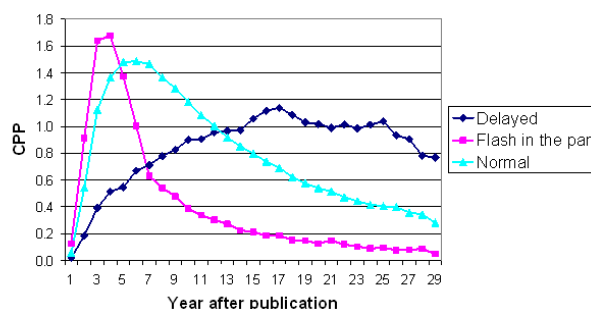


Figure 2. Evolution of CPP (included documents ≥ 5 citations and 1980-2003)

Future research includes the analysis of these different types of documents, how they affect bibliometric indicators, etc. Other research fields will be analyzed and more results will be presented in the final version of the paper.

References

- Aksnes, D.W. (2003). Characteristics of highly cited papers. *Research Evaluation*, 12(3), 159-170.
- Aversa, E.S. (1985). Citation patterns of Highly Cited Papers and their relationship to literature aging. *Scientometrics*, 7(3-6), 383-389.
- Glanzel, W., Schlemmer, B. & Thijs, B. (2003). Better late than never? On the chance to become highly cited only beyond the standard bibliometric time horizon. *Scientometrics*, 58(3), 571-586.
- Moed, H.F., van Leeuwen, T.N & Reedijs, J. (1998). A new classification system to describe the ageing of scientific journals and their impact factors. *Journal of Documentation*, 54(4):387-419.
- Van Dalen, H.P. & Henkens, K. (2005). Signals in science –on the importance of signaling in gaining attention in science. *Scientometrics*, 64(2), 209-233.
- Van Raan, A.F.J. (2004). Sleeping beauties in science. *Scientometrics*, 59(3), 467-472.