Groups of Highly Cited Publications: Stability in Content with Citation Window Length

Nadine Rons

Vrije Universiteit Brussel, Research Coordination Unit and Centre for R&D Monitoring (ECOOM), Pleinlaan 2, B-1050 Brussels (Belgium)

Introduction
The growing focus in research policy worldwide on top scientists makes it increasingly important to define adequate supporting measures to help identify excellent scientists. Highly cited publications have since long been associated to research excellence. At the same time, the analysis of the high-end of citation distributions still is a challenging topic in evaluative bibliometrics. Evaluations typically require indicators that generate sufficiently stable results when applied to recent publication records of limited size. Highly cited publications have been identified using two techniques in particular: pre-set percentiles, and the parameter free Characteristic Scores and Scales (CSS) (Glänzel & Schubert, 1988). The stability required in assessments of relatively small publication records, concerns size as well as content of groups of highly cited publications. Influencing factors include domain delineation and citation window length. Stability in size is evident for the pre-set percentiles, and has been demonstrated for the CSS-methodology beyond an initial citation period of about three years (Glänzel, 2007). Stability in content is less straightforward, considering for instance that more highly cited publications can have a later citation peak, as observed by Abt (1981) for astronomical papers. This paper investigates the stability in content of groups of highly cited publications, i.e. the extent to which individual publications enter and leave the group as the citation window is enlarged.

Data and methodology

Database, document type and time frame
The bibliometric data were obtained from the online Web of Science. This study focuses on articles as primary vehicles for new research results (reviews, less frequent and typically more highly cited, are in itself related to high esteem). Results were calculated for articles published in 2004. Citations were collected up to 7 years after publication, until 2011.

Aggregation level
The group of top-cited publications is highly dependent on the level of aggregation at which these are determined (Zitt, Ramanana-Rahary & Bassecoulard, 2005). In this paper, the reference sets within which highly cited publications are identified are the partition cells formed by the structure of overlapping subject categories (Rons, 2012). These partition cells are an aggregation level of intermediate size, situated between journals and entire subject categories, proposed particularly for usage at the level of individual scientists.

Domains
Data were collected from a domain with fast citation characteristics (a sub-domain of physics), and from a domain with slow citation characteristics (mathematics). In both domains, three adjacent partition cells are observed, containing all journals assigned to a particular combination of subject categories:
- 'Astronomy Astrophysics' only (A), 'Physics Particles Fields' only (P), and both 'Astronomy Astrophysics' and 'Physics Particles Fields' (A&P), with 8047, 1977 and 2440 articles respectively.
- 'Mathematics' only (M), 'Mathematics Applied' only (MA), and both 'Mathematics' and 'Mathematics Applied' (M&MA), with 10022, 3938 and 3286 articles respectively.

Groups of highly cited publications
Groups of highly cited publications were identified using both the technique of pre-set percentiles (5% and 1%), and the CSS-methodology (at least 'remarkably' and 'outstandingly' cited publications, stabilizing in size towards 8 to 11%, and 2 to 3% respectively in the largest citation windows).

Results and discussion
The citation distributions vary with domain and highly cited level. Peak citation years tend to fall later for more highly cited publications. Differences depend on the domain, with in the physics cells 0 to 3 years between the top-1% or outstandingly cited and the less cited publications, and in the mathematics cells 1 to 2 years between the top-5% or at least remarkably cited and the less cited publications. In particular for publications below the top-1% and outstandingly cited, peak citation years in the slow mathematics domain fall later than in the fast physics domain (3 to 6, and 1 to 2 years after the publication year respectively). Figure 1
shows how groups of highly cited publications converge in content with increasing citation window length, in a similar way for both methodologies used. A yearly fluctuation in content remains between the larger consecutive citation windows.

Conclusions
The results of the present study show how, in stable size groups of highly cited publications, the content of individual publications converges with an extending citation window. To the author's knowledge, no similar investigations of the extent to which individual publications enter and leave the group of highly cited publications as the citation window is enlarged, figure in scientific literature. The process evolves towards a limited remaining yearly fluctuation and depends on domain and highly cited level. Additional factors outside the scope of this paper may also exert an influence, such as the structure within which the highly cited publications are identified, including subject category as well as publication based structures. Stability in content of groups of highly cited publications is relevant in particular in a context of publication records of limited size, such as those of individual scientists. The studied domains with different bibliometric characteristics indicate that a same stability in content can require citation windows of different lengths depending on the domain. Whether a sufficient stability in content can be attained for successful general and comparative applications at the micro-level, with citation windows of acceptable size in an evaluative context, requires further investigation.

Acknowledgments
This paper is related to research on individual excellence carried out at the Flemish Centre for R&D Monitoring (ECOOM).

References
Abt, H.A. (1981). Long-term citation histories of astronomical papers. Publications of the Astronomical Society of the Pacific, 93(552), 207-210.
Glänzel, W. (2007). Characteristic scores and scales: A bibliometric analysis of subject characteristics based on long-term citation observation. Journal of Informetrics, 1(1), 92-102.
Glänzel, W. & Schubert, A. (1988). Characteristic scores and scales in assessing citation impact. Journal of Information Science, 14(2), 123-127.
Price, D. de S. (1976). A general theory of bibliometric and other cumulative advantage processes. Journal of the American Society for Information Science, 27(5-6), 292-306.
Rons, N. (2012). Partition-based field normalization: An approach to highly specialized publication records. Journal of Informetrics, 6(1), 1-10.
Zitt, M., Ramanana-Rahary, S. & Bassecoulard, E. (2005). Relativity of citation performance and excellence measures: from cross-field to cross-scale effects of field-normalization. Scientometrics, 63(2), 373-401.