Sleeping Beauties in Mathematical Research

Samuel Hansen[1]

[1] *hansensm@umich.edu*
University of Michigan, Shapiro Science Library, Mathematics & Statistics Librarian, 919 S University Ave, Ann Arbor, MI 48109

Introduction

Mathematics is often called the language of science, but there is much which separate the two. For example there is the very nature of mathematical knowledge. Unlike scientific results, once a mathematician proves a theorem it is true for the rest of mathematics. Then there is the axiomatic non-empirical disposition of mathematics and its focus on describing a realm of pure abstraction versus science's focus on describing the real world.

Given these clear differences it is surprising that during scientometric study mathematics is often lumped together with science rather than being treated as its own entity. This poster paper represents a set of initial research results from a comprehensive bibliometrics study trying to fix this oversight. Specifically this poster paper will present an initial analysis of mathematical Sleeping Beauties, or research receives a spike of citations after years of relatively few.

Research Aging and Sleeping Beauties

Research aging, primarily through the study of references and citations ages, is an active area of bibliometrics research and has been for many decades (Anker, 1979; Glänzel & Schoepflin, 1995). Thanks to large-scale literature databases and citation indexes scientometric researchers have recently been able to do comprehensive aging studies of both references and citations (Zhang & Glänzel, 2017a; Zhang & Glänzel, 2017b). These large-scale databases have also allowed researchers to develop and formalize the idea of Sleeping Beauties (SBs).

The idea of SBs was first put forth by van Raan (2004) who identified a collection of papers which received fewer than two citations for years and then began to receive large numbers of citations. This work was expanded by Redner (2005) who conducted the first in-depth search of SBs, in the area of Physics. Both of these works were limited by arbitrary definitions of SBs, which drove Ke, Ferrara, Radicchi, & Flammini (2015a) to develop the "Beauty Coefficient" to provide a measure for how deeply research slept. They calculated the Beauty Coefficient for papers in Clarivate's Web of Science database and defined SBs as those papers with Beauty Coefficients in the top 0.1% of all research. This provided them with SB Beauty Coefficient thresholds in different disciplines, for mathematics it was 90.62(Ke et. al., 2015b, p. 19).

Methodology

The data used to conduct this study is from Clarivate's Web of Science citation database, 1900-2017. The Big Ten Academic Alliance has an agreement with Clarivate where they have provided the contents of their database in the form of XML documents with a license allowing for academic research by members of Big Ten academic institutions. These documents were parsed into a PostgreSQL database based off of a data model and python scripts created by the University of Indiana (Indiana University Network Science Institute, n.d.; Light, Halsey, & Herr, n.d.). The scripts used to extract, process, and calculate the Beauty Coefficient for the mathematical citation data are available via the University of Michigan Gitlabs instance (Hansen, n.d.).

Mathematical Sleeping Beauties

There are three mathematics subjects used by Web of Science to classify its contents, Mathematics, Mathematics, Applied, and Mathematics, Interdisciplinary Applications. These subjects are not applied in a mutually exclusive manner, and can overlap. These subjects are applied to 1,343,970 entries of all document types in the Web of Science database. Over half were considered Mathematics, with around 45% Applied and 15% Interdisciplinary Applications. There were 3847 cases of SBs, e.g. Beauty Coefficients higher than 90.62. Mathematics was most likely to generate a SB, with a rate around three times the other subjects.

Table 1. Counts of Mathematics Research and Sleeping Beauties in Web of Science by subject.

Table	Total	SBs
Mathematics	742541	3044
Applied	611160	743
Interdisciplinary Applications	199652	324
Total	1343970	3847

The ratio stays nearly the same when only mathematical research which has received more than 100 citations is considered. Interestingly this is true even though research classified as Mathematics is less likely than the other subjects to reach 100 citations, less than half as likely as Interdisciplinary

Applications. Since the number of citations, specifically the peak, plays a major role in the Beauty Coefficient this implies that the less applied or interdisciplinary a highly cited mathematics paper is the more likely it is go without citation for an extended period.

Table 1. Counts of Mathematics Research with citation counts of at least 100 and Sleeping Beauties in Web of Science by subject.

Table	Total	SBs
Mathematics	6485	938
Applied	6635	342
Interdisciplinary Applications	3995	174
Total	15745	1354

This analysis also identified a new SB (citation peak of 2017) with one of the highest known Beauty Coefficients, Clive Granger's "Investigating Causal Relations by Econometric Models and Cross-spectral Methods" (1969). With a beauty coefficient of just over 6737 it is only behind two of the SBs found by Ke et. al. (2015a, p. 7429). Related to their findings of a relationship between SBs and interdisciplinarity this work was assigned Web of Science subjects of Economics, Mathematics, Interdisciplinary Applications, Social Sciences, Mathematical Methods, and Statistics & Probability.

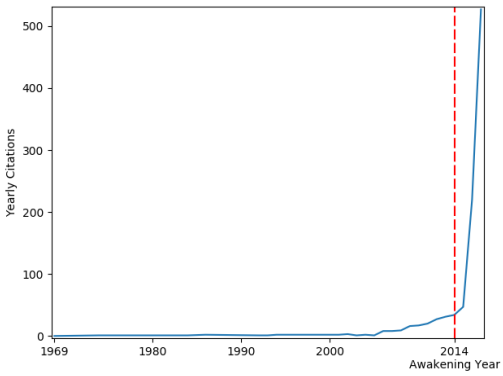

Figure 1. Citation History for Investigating Causal Relations by Econometric Models and Cross-spectral Methods with a dotted line indicating its awakening year.

Continuing Work

This analysis only represents the first step in ongoing research into the bibliometrics of mathematics research. This research will include thorough analyses of mathematical reference and citation aging, bibliographic coupling, and sub-discipline networks.

This work has also indicated there is still work to be done with regard to identifying different forms of SBs. The Beauty Coefficient is a useful measure but it cannot identify research which has an initial burst of citations, and then falls asleep, also known as all-elements-sleeping-beauties (Li, 2014), or papers which awake and fall asleep multiple times. A measure which could identify such SBs, perhaps related to peak analysis, could open up a new range of potential analyses, not only in mathematics, but for scientometrics as a whole.

References

Anker, A. L., Servi, P. N., Griffith, B. C., & Carl Drott, M. (1979). The aging of scientific literature: a citation analysis. Journal of Documentation, 35(3), 179–196.

Glänzel, W., & Schoepflin, U. (1995). A bibliometric study on ageing and reception processes of scientific literature. Journal of Information Science, 21(1), 37–53.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-spectral Methods. Econometrica, 37(3), 424–438.

Hansen, S. (n.d.). Citation Aging Gitlabs Repository. Retrieved April 5, 2019, from GitLab website: https://gitlab.eecs.umich.edu/hansensm/citationaging

Indiana University Network Science Institute. (n.d.). Web of Science (WoS). Retrieved April 5, 2019, from Indiana University Network Science Institute website: http://iuni.iu.edu/resources/web-of-science

Light, R., Halsey, D., & Herr, B. (n.d.). Generic Parser Github Repository. Retrieved April 5, 2019, from https://github.com/cns-iu/generic_-parser

Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015a). Defining and identifying Sleeping Beauties in science. Proceedings of the National Academy of Sciences, 112(24), 7426–7431.

Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015b). Defining and identifying Sleeping Beauties in science. Proceedings of the National Academy of Sciences, 112(24), 7426–7431.

Li, J. (2014). Citation curves of "all-elements-sleeping-beauties": "flash in the pan" first and then "delayed recognition." Scientometrics, 100(2), 595–601.

Redner, S. (2005). Citation Statistics from 110 Years of Physical Review. Physics Today, 58(6), 49–54.

van Raan, A. F. J. (2004). Sleeping Beauties in science. Scientometrics, 59(3), 467–472.

Zhang, L., & Glänzel, W. (2017a). A citation-based cross-disciplinary study on literature ageing: part II—diachronous aspects. Scientometrics, 111(3), 1559–1572.

Zhang, L., & Glänzel, W. (2017b). A citation-based cross-disciplinary study on literature aging: part I—the synchronous approach. Scientometrics, 111(3), 1573–1589.