Scholar Metrics: the impact of journals according to Google, just an amusement or a valid scientific tool?

Álvaro Cabezas-Clavijo y Emilio Delgado-López-Cózar

acabezasclavijo@gmail.com; edelgado@ugr.es

EC3: Evaluación de la Ciencia y de la Comunicación Científica, **Universidad de Granada**

Cabezas-Clavijo, **Álvaro**; **Delgado-López-Cózar, Emilio (2012).** "Scholar Metrics: the impact of journals according to Google, just an amusement or a valid scientific tool?" *E-LIS. E-prints in Library and Information Science.*

INTRODUCTION

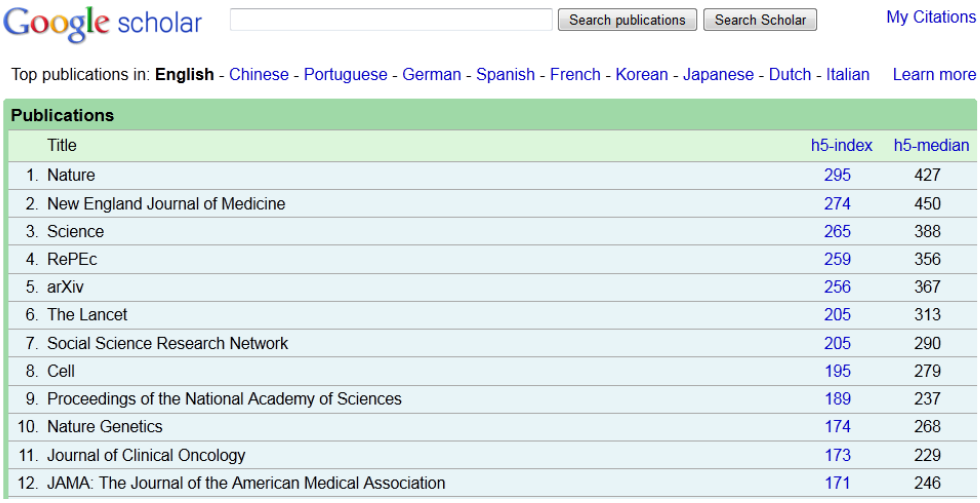
Last April 1, Google launched Google Scholar Metrics, a new tool that provides impact metrics for scientific journals, measured from citation counts. The research community was caught on guard and many wondered whether it was a real product or just a joke (April 1 is Fools' Day). However, to the delight of bibliometricians and uneasiness of some major scientific publishers, the Google product was not a joke but a natural step for Google in the expansion of the Scholar family.

Even if this move was expected, it is nevertheless relevant. After some time calibrating the utility of Google Scholar as a tool for scientific evaluation (Torres-Salinas, Ruiz-Perez and Delgado-Lopez-Cózar, 2009), Google is now entering the very heart of bibliometrics: calculating journals' impact. The launch of Google Scholar Citations a few months ago (Cabezas-Clavijo and Torres-Salinas, 2012), - a tool that measures the impact of researchers - sent a warning message of Google's future plans in the scientific information assessment world. The launching of the classification of journals according to their h-index is in direct competition with the different products for calculating journals' impact on the market, and particularly with the Journal Citation Reports (JCR), the traditional index for measuring the impact of journals. In this note we review the most significant features of Google Scholar Metrics, pointing out their strengths and weaknesses (always bearing in mind that Google products are progressively incorporating new functionalities), and discussing the possibilities of the adoption of these tools by bibliometricians and policy makers.

DESCRIPTION

Scholar Metrics provides in its first version a table of the first one hundred journals worldwide by language of publication. These lists are ranked on the basis of the h-index of journals, calculated from the articles published in the last five years (2007-2011). That is, a journal with

an h index of 12 (eg, El Profesional de la Información) means that it has published 12 papers with at least 12 citations each. As additional data, the median number of citations obtained by the articles that contribute to the index h is also provided. Obviously, the median value can never be less than the value of the journal's h index. Scholar Metrics uses this indicator (h5-median) to rank journals with the same h-index value, which is necessary since, by taking discrete values, this indicator has little discriminatory power. Also items contributing to the h-index of each journal are listed.


Google scholar Search publications Search Scholar My Citations

Top publications in: **English** - Chinese - Portuguese - German - Spanish - French - Korean - Japanese - Dutch - Italian Learn more

Publications

	Title	h5-index	h5-median
1.	Nature	295	427
2.	New England Journal of Medicine	274	450
3.	Science	265	388
4.	RePEc	259	356
5.	arXiv	256	367
6.	The Lancet	205	313
7.	Social Science Research Network	205	290
8.	Cell	195	279
9.	Proceedings of the National Academy of Sciences	189	237
10.	Nature Genetics	174	268
11.	Journal of Clinical Oncology	173	229
12.	JAMA: The Journal of the American Medical Association	171	246

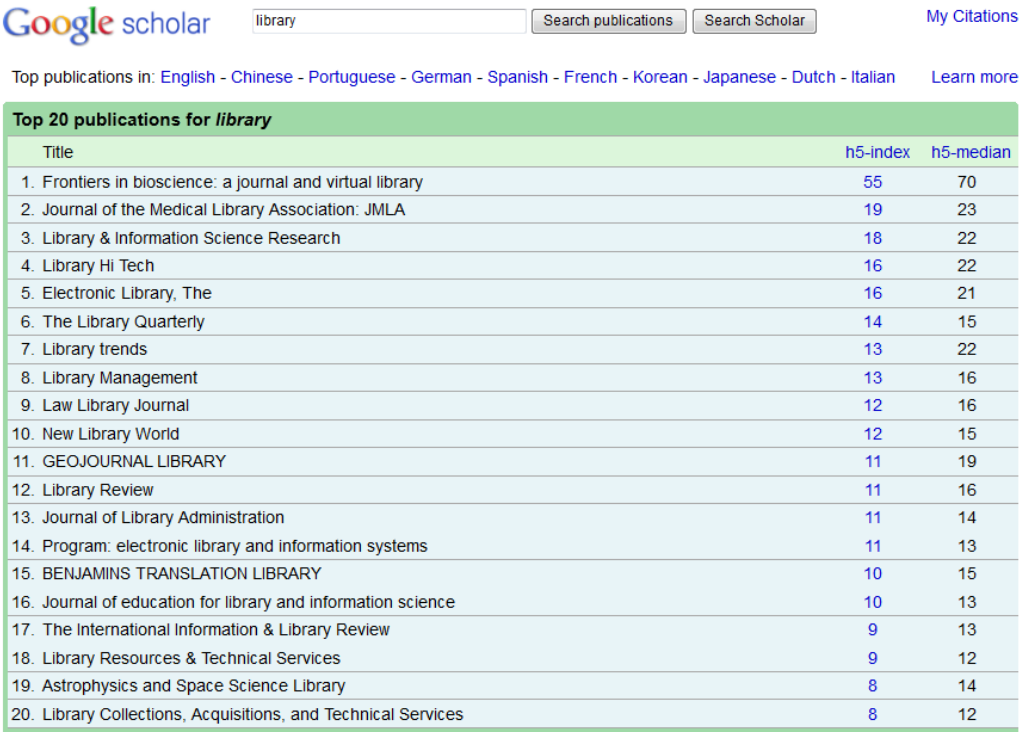
Figure 1: Homepage of Google Scholar Metrics listing the publications with the highest h-index

The Scholar Metrics can be explored in two ways:

• Accessing the lists by language (currently ten: English, Chinese, Portuguese, German, Spanish, French, Korean, Japanese, Dutch and Italian). For every language a ranking of the 100 journals with the highest h-index is shown.

• Using the search engine to look up for words in the titles of journals. In this case, only 20 results are shown.

In this sense, it is noted that the journals included in this product are not all the journals indexed in Google Scholar. A selection has been made, based on two criteria: journals that have published at least 100 articles in the period 2007-2011 and which have received at least one citation (ie, excluding journals with h-index=0). Data shown from the journals is fixed to the date of launching, April 1, 2012. Therefore, it is not a dynamic information system as it is not updated live as journals receive more citations. Although not announced, it is expected that Google will periodically update these lists.



Google scholar library Search publications Search Scholar My Citations

Top publications in: English - Chinese - Portuguese - German - Spanish - French - Korean - Japanese - Dutch - Italian Learn more

Top 20 publications for *library*

	Title	h5-index	h5-median
1.	Frontiers in bioscience: a journal and virtual library	55	70
2.	Journal of the Medical Library Association: JMLA	19	23
3.	Library & Information Science Research	18	22
4.	Library Hi Tech	16	22
5.	Electronic Library, The	16	21
6.	The Library Quarterly	14	15
7.	Library trends	13	22
8.	Library Management	13	16
9.	Law Library Journal	12	16
10.	New Library World	12	15
11.	GEOJOURNAL LIBRARY	11	19
12.	Library Review	11	16
13.	Journal of Library Administration	11	14
14.	Program: electronic library and information systems	11	13
15.	BENJAMINS TRANSLATION LIBRARY	10	15
16.	Journal of education for library and information science	10	13
17.	The International Information & Library Review	9	13
18.	Library Resources & Technical Services	9	12
19.	Astrophysics and Space Science Library	8	14
20.	Library Collections, Acquisitions, and Technical Services	8	12

Figure 2: The 20 journals with the highest h-index with the term "Library" in the title.

LIMITATIONS

A first review of the features of Google Metrics points out some of the major limitations of the product.

• **Coverage**. It is frankly surprising the ambiguous if not misleading definition of which documents are to be measured in Google Metrics. Although in the brief methodological note that accompanies the product, explicit references to scientific journals are constant ("... If you cannot find the journal you're looking for ..." "If you're wondering why your Journal ... "), it is stated that journal articles from websites that meet the inclusion criteria for Google are included as well as "conference articles and preprints from a small number of manually identified sources". We must browse the rankings by language or do some searches to meet the big surprise: the indiscriminate mixture of sources as diverse as journals, repositories (RePEc, Arxiv and Social Science Research Network), databases (Cochrane database of Systematic Reviews), conference proceedings (IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Proceedings of the ESSCIRC, Proceedings of SPIE, AIP Conference Proceedings) and working papers (NBER Working Paper Series). Could it be that those responsible for Google Metrics are unaware of the different nature of these products preventing from any bibliometric comparison? Thus, it is not surprising that of the ten sources with the highest h-index in English, three are repositories (RePEc, Arxiv and Social Science Research Network), probably some of those "hand-selected sources." The obvious question is: why these are selected and not others? Worthy repositories as Dialnet or E-LIS, even if they have different purposes and coverage, may well appear in the listings. In short, this decision is surprising and certainly it is not justified, especially because the product is limited to a selected

number of sources and not to all conference proceedings, databases or repositories of research materials.

• **Source Control**: In his state-of-the-art paper, Peter Jacso (2008) pointed out Google Scholar's main shortcomings. Among other discoveries he learned from prolific authors such as Password, Building, Introduction or View. In the same vein, the not so productive but equally rare journal "Age (years)" can now be found in the list, as it ranks 99th in the Spanish-language listing.

• **Lack of standardization**. Calculating an impact factor for journals requires a laborious pace that includes annoying tasks as the standardization of journal titles. This is the basics bibliometrics. So it surprising that, although Google acknowledges they have found 959 ways to name the journal PNAS (Proceedings of the National Academy of Sciences), it commits such childish mistakes like the one of Revista Española de Cardiología (Spanish Journal of Cardiology) that appears with two distinct forms (one with the spanish character ñ with a h-index of 24 and one without ñ reaching a h-index of 19 and that just appears when searching by the word Cardiologia but not in the Spanish language ranking). More unacceptable is that Google has paid so little attention to the formal presentation of the product that it has not even bothered to show standardized journal titles. This way we can find most of themwith their full name and others with the abbreviation (Nutr Hosp, Rev Argent Cardiol Rev. CEFAC, Rev. bras. Enferm ...), some titles are capitalized (REVISTA DE SALUD PÚBLICA, BOLETÍN GEOLÓGICO Y MINERO, REVISTA BRASILEIRA DE OTORRINOLARINGOLOGIA…) and most are not, etc. Little details that wouldn't have much cost considering they only list 1200 journals.

H-index articles for El Profesional de la Informacion 1-12

Title / Author	Cited by	Year
Conceptos de web 2.0 y biblioteca 2.0: origen, definiciones y retos para las bibliotecas actuales DM Arnal El Profesional de la Informacion 16 (2), 95-106	58	2007
El factor de impacto de las revistas científicas: limitaciones e indicadores alternativos R Aleixandre-Benavent, JC Valderrama-Zurián, G González-Alcaide El Profesional de la Informacion 16 (1), 4-11	53	2007
Las bibliotecas universitarias y Facebook: cómo y por qué estar presentes D Margaix-Arnal El Profesional de la Informacion 17 (6), 589-602	26	2008
Ciencia 2.0: catálogo de herramientas e implicaciones para la actividad investigadora Á Cabezas-Clavijo, D Torres-Salinas, E Delgado-López-Cózar El Profesional de la Informacion 18 (1), 72-80	25	2009
Animating the development of Social networks over time using a dynamic extension of multidimensional scaling L LEYDESDORFF, T SCHANK, A SCHARNHORST, W DE NOOY El Profesional de la información 17 (6), 611-626	24	2008
Repositorios de publicaciones digitales de libre acceso en Europa: análisis y valoración de la accesibilidad, posicionamiento web y calidad del código C Rovira, MC Marcos, L Codina El Profesional de la Informacion 16 (1), 24-38	20	2007

Figure 3: List of some of the items contributing to the h-index of the journal El Profesional de la Información

• **Results browsing**. The main critic is not directed to these aspects but with how Google has arranged the data. The ability to browse only by language is certainly unprecedented in

bibliometrics, and has little practical value. The logical thing would be to provide rankings by areas or scientific disciplines, as any bibliometric indicator, the h-index depends on output and citation patterns that prevent from making any kind of comparison across disciplines. Thus the only way to check the impact of journals in a given area is to search one by one, and with no certainty of whether it is or not included, as no directory with the information sources collected by Scholar Metrics is provided.

A FEW FINAL THOUGHTS

However and after acknowledging these limitations, we have to welcome this step taken Google which will facilitate the access to journals' metrics by many researchers who are not allowed to browse traditional citation databases, and which can stimulate competition between different products for assessing journals' impact. Moreover, the popularity of the h-index to evaluate researchers, and its ease of calculation and comparison may stimulate this measure for the evaluation of journals, especially in the humanities, an area that has few bibliometric indicators of journal impact.

From the research evaluation viewpoint, it is unquestionable that Google Scholar moves into this niche market and is working on products that are in direct competition with those of Elsevier and Thomson Reuters. Its success among scientists and their usefulness for the bibliometrics community may well depend on how they manage the limitations previously referred and how the integration with personal profiles (Google Scholar Citations) and with the Google Scholar search engine is performed.

Nevertheless, we have to be very critical: it is disappointing to see how Google, a lavish economic empire which manages impressive resources, presents a product of such short goals and bad execution from a bibliometric point of view. Google should be aware that products do demand some serious bibliometric effort and some means more than algorithm and spiders that automatically produce results. It seems that for Google, scientific evaluation with bibliometric tools is a field for "playing" at the time, and has not become yet a threat to its competitors with good products which may lead to improve its profits.

Finally, we must have in mind that this product comes at a time of a growing debate among the research community on two different but related fronts. On the one hand, there is a heated debate about the laws of access to scientific information such as the U.S. Research Works Act, supported in principle by major publishers such as Elsevier (which withdrew its support after the boycott of their journals sponsored by a group of scientists, giving new lease of life to the open access movement). On the other hand, within the evaluation community there is also a search for new indicators of impact and visibility of scientific production within the initiatives known as Alt-metrics or alternative metrics. Probably and depending on Google's capability to awake sympathies within both movements, this move will strong its position in the academic community. What will certainly make a difference is the free of charge products developed by the Silicon Valley company compared with the high costs of Thomson Reuters' and Elsevier' databases.

ACKNOWLEDGEMENTS

The authors would like to thank Nicolás Robinson-García for the translation of this text.

REFERENCES

Cabezas-Clavijo, Á. & Torres-Salinas, D. (2012). "Google Scholar Citations y la emergencia de nuevos actores en la evaluación de la investigación". Anuario ThinkEPI, 6. http://ec3noticias.blogspot.com.es/2011/12/thinkepi-google-scholar-citations-y-la.html

Jacso, P. (2008). "The pros and cons of computing the h-index using Google Scholar". Online Information Review 32(3), 437-452. http://www.jacso.info/PDFs/jacso-GS-for-h-index-unabridged.pdf

Torres-Salinas, D., Ruiz-Pérez, R. & Delgado-López-Cózar, E. (2009). "Google Scholar como herramienta para la evaluación científica". El profesional de la Información, 18(5), 501-510. http://ec3.ugr.es/publicaciones/d700h04j123154rr.pdf