

ESTUDO BIBLIOMÉTRICO DO SUG

Javier Tarrío-Saavedra¹, Elena Orois² e Salvador Naya¹

¹Escola Politécnica Superior. Grupo MODES. Dpto. de Matemáticas. Universidade da Coruña.

²Servicio de Biblioteca Universidade da Coruña.

RESUMO

Neste traballo preséntase un estudo bibliométrico das universidades galegas usando a información sobre a produción científica que se obtén a partir da base de datos SCOPUS, mediante a aplicación da librería CITAN do Software R. Levouse a cabo a súa comparación mediante a análise descritiva dos datos, a aplicación dos modelos de Price e Lotka e o cálculo e a análise dos diferentes indicadores de produción científica das tres universidades: os índices h e g, e máis aqueles non tan comúns como os $rp1$, $l1$, $linf$, $Ge1$, $Ge5$ e $Slp1$.

Palabras e frases chave: Bibliometría, cienciometría, webmetría, índice h, índice g.

1. INTRODUCCIÓN

A necesidade de coñecer e avaliar de forma precisa a actividade investigadora das universidades galegas e un problema que ocupa e preocupa non só as institucións involucradas senón tamén ao resto da sociedade, por canto se trata de institucións públicas. Neste traballo propónse o emprego do R para analizar a produción investigadora das tres universidades galegas. En primeiro lugar farase unha revisión dos conceptos empregados nas ciencias documentais. A continuación analizaranse distintos indicadores do Sistema Universitario Galego (SUG). Finalmente presentaranse as conclusións e referencias.

2. BIBLIOMETRÍA, CIENCIOMETRÍA, INFORMETRÍA

Os expertos das ciencias da documentación diferencian entre tres tipos de estudos: Bibliometría, Cienciometría e Informetría. Estas tres ferramentas do tratamento da información documental presentan características comúns e certo solapamento en relación ao fluxo de información e aos modelos que utilizan. En moitos estudos sobre a literatura científica a Bibliometría, a Cienciometría e a Informetría aparecen situadas ao mesmo nivel, e diferentes autores utilízanas como sinónimas dificultando a identificación dos seus límites e alcance respectivos. Aínda así é posible atopar algunhas definicións que permiten situar certas fronteiras entres estes conceptos.

O primeiro en falar de Bibliometría foi Pritchard que a entende como “a aplicación dos métodos matemáticos e estatísticos aos libros e outros medios de comunicación”. Por outra parte Brookes, quen realizou importantes aportacións sobre a obsolescencia da literatura científica, entende a Bibliometría como “unha disciplina limitada á actividade bibliotecaria que debe enriquecerse mediante as relacións interdisciplinarias coa estatística para refinar as súas técnicas”. Neste sentido a Bibliometría tería por obxecto estudar libros e revistas científicas e, como obxectivo principal, o de comprender as actividades de comunicación da información. Quizais a mellor definición sexa a que aporta Alcaín (1990) quen di que a Bibliometría é a disciplina que “comprende estudos dirixidos a coñecer o rendemento dos fondos de publicacións científicas, así como a selección e consumo por parte dos usuarios”.

Por outra parte, as primeiras mencións á Cienciometría atopámolas nun estudo de Hulme, de 1923, sobre unha análise estatística da historia das ciencias, porén a popularidade desta disciplina xurdiu a partir do ano 1977 co nacemento da revista *Scientometrics*. A Cienciometría acadou gran desenvolvemento cos traballos de Price e E. Garfield. A partir da obra de Price, *Little Science, Big Science* en 1963,

desenvólvese un novo campo: a análise estatística e sociométrica da literatura científica, que se denominará Cienciometría “cando os estudos se realizan sobre a produción e produtividade dos autores científicos, organismos de investigación”. Unha definición de Cienciometría podería ser “Ciencia da Ciencia”.

O termo Informetría utilízase a partir dos anos 80. Spinak dice que a Informetría basease nas investigacións da Bibliometría e a Cienciometría, e comprende asuntos como o desenvolvemento de modelos teóricos para a medida da información co fin de atopar regularidades nos datos asociados coa produción e o uso da información rexistrada. Ademais da medición da información tamén se ocuparía do seu almacenamento e recuperación, polo tanto podería dicirse que é unha disciplina instrumental das ciencias da documentación que se apoia no emprego das ferramentas informáticas. Tamén está acuñado o termo webmetría ou cibernetría que sería a disciplina cuxo obxectivo é analizar o comportamento dos documentos publicados na web, medir recursos, acceso e utilidade da información publicada en Internet, así como valorar o impacto desa información mediante unha metodoloxía baseada en programas intelixentes que recompilan datos e uso intensivo da análise estatística de grandes mostras (Arroyo et al. 2005). Por último, ao marxe dos indicadores tradicionais, no ano 2010 apareceu o concepto de almetrics, coma a creación e o estudio de novas métricas baseadas na web social para analizar e avaliar aos investigadores (J. Priem et al. 2010).

Entre os indicadores máis utilizados actualmente no campo da Bibliometría están os dous nos que centraremos este estudo, o índice h e o índice g. O índice h foi proposto por Jorge E. Hirsch no ano 2005. Defínese como un número H que coincide co número de publicacións dese autor que recibiron H citas. Un índice $H = X$ significa que hai X artigos con X ou máis citas, pero non $X+1$ que teñan $X+1$ ou máis citas (Arencibia e Carvajal, 2008). Ao igual que se fai para o cálculo de índice h, para obter o índice g lístanse os artigos dun autor en orde descendente de acordo ao número de citas recibidas. O maior número de orde na listaxe onde a suma das citas recibidas polo autor sexa maior ou igual ao cadrado do número de orde será considerado o Índice G do dito autor (Arencibia e Carvajal, 2008).

3. APLICACIÓNS AO SUG

Mediante o uso do paquete CITAN do R, analizouse a produción científica das tres universidades galegas no 2000-2011. Concretamente para o caso da Universidade de Santiago de Compostela (USC) os artigos atopados na base de datos SCOPUS foron 10095 cun total de 19492 autores (o que daría un ratio de 0.19 doc/autor). Para a Universidade da Coruña (UDC) os documentos analizados foron 4080 cun total de 6206 autores (0.66 doc/autor). E no caso da Universidade de Vigo (UVigo) atopáronse 8230 documentos, con 10817 autores (0.76 doc/autor). Hai que salientar o problema das diferentes afiliacións dun mesmo autor, especialmente no caso da Universidade da Coruña. As dificultades na procura son grandes neste caso particular por mor do uso ou non do “ñ”, do uso da voz inglesa, francesa, castelá ou galega ou mesmo por só especificar o campus (Ferrol, por exemplo). En definitiva, o uso do nome estandarizado da afiliación debería terse en conta para a correcta avaliación da produción científica das universidades, evitando a grande perda de tempo recursos debidos o filtrado e tratamento previo dos datos.

Primeiramente próbase que no intervalo 2006-2011 cúmprese a lei de crecemento exponencial da ciencia ou lei de Price (Figura 1). Seguindo o modelo de Price, a UDC e UVigo son universidades novas, con maior capacidade ou potencial de crecemento que a USC, por iso o tempo necesario de duplicación da cadansúa produción científica (D) é menor que o correspondente á USC (Figura 1).

A Figura 2 mostra o cumprimento da lei de Lotka no intervalo 2000-2011. O número de autores da UVigo e da UDC decrece máis lentamente co número de artigos que publican que no caso da USC. En proporción hai máis autores cun maior número de artigos. Por outra banda, a USC presenta un ratio de citas por documento maior que no caso da UVigo e UDC: 10.1 citas/doc > 9.5 citas/doc > 6.4 citas/doc.

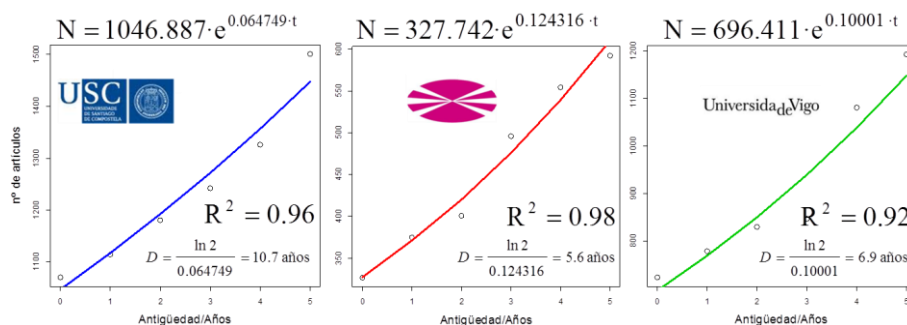


Figura 1. Axuste do modelo de crecemento exponencial de Price. Cálculo do parámetro D.

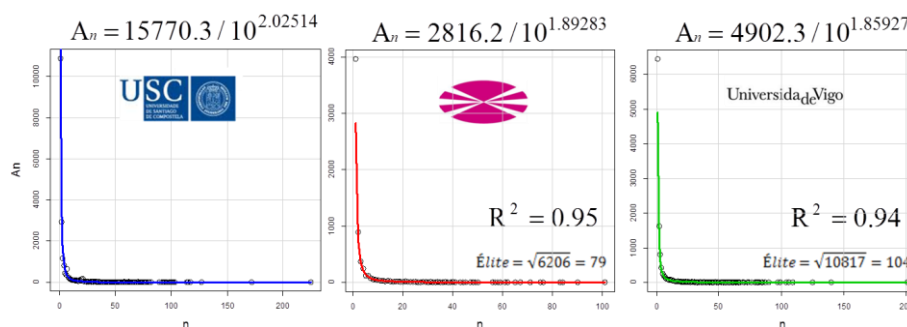


Figura 2. Resultados da aplicación da lei de Lotka.

Na Táboa 1 amósanse os diferentes índices bibliométricos obtidos para cada unha das universidades. A USC presenta, polo xeral índices máis altos que as demais universidades, especialmente no índice h (Hirsch), aínda que para o índice g é a universidade de Vigo a que lle corresponde un maior valor. Isto último, e atendendo á definición do índice g, está relacionado coa produtividade dos autores da UVigo e co, en comparación, elevado número de citas que presentan os seus artigos.

Índice	USC	UDC	UVigo
Nº de documentos	10095	4080	8230
Max nº citas	653	360	1246
Suma de citas	101628	26015	78253
h	93	49	86
g	136	71	145
rp1	93	49	86
l1	332.6	175.5	282.4
linf	172.8	92.09	147.5
Ge1	7694	2673	5923
Ge5	1997	619	1496
SLp1	16057.1	5050.3	12067.1
N autores con $h \geq N$	21	15	19

Táboa 1. Indicadores bibliométricos obtidos para o período 2000-2011 do SUG.

Cada autor ten un índice correspondente, polo tanto, os índices bibliométricos calculados poden ser entendidos como variables aleatorias cunha distribución de probabilidade. Deste xeito, non só se avalía cada universidade, senón tamén a produción dos seus empregados e as súas redes externas de investigación (colexios invisibles). Cabería facer agora a pregunta de se podemos considerar que a distribución de probabilidade do índice h dos autores é diferente dependendo da universidade. Para verificalo emprégase o test de Wilcoxon para a igualdade de medianas. O resultado das comparación múltiples pódense ver na Táboa 2.

Comparacións	Estatístico	p-valor	Resultado
h (UDC) - h (USC)	50440118	$2.2 \cdot 10^{-16}$	Significativo
h (UDC) - h (UVigo)	30671459	$2.2 \cdot 10^{-16}$	Significativo
h (UVigo) - h (USC)	95935123	$2.2 \cdot 10^{-16}$	Significativo

Táboa 2. Resultado do contraste de Wilcoxon aplicado ás distribución do índice h.

Chégase á conclusión de que as diferenzas entre as medias son estatisticamente significativas, así que, por lo xeral, os autores da UDC teñen un menor índice h que os autores de la UVigo, e os autores da UVigo adoitan ter un índice h menor que os da USC: $\bar{h}(\text{autores da USC}) > \bar{h}(\text{autores da UVigo}) > \bar{h}(\text{autores da UDC})$.

4. CONCLUSIONES

Presentouse unha comparativa da produción científica das tres universidades usando a librería CITAN que calcula os índices h e g, e extraendo os datos da base SCOPUS. Se ben existe un problema de definición das autorías, especialmente patente no caso da Universidade da Coruña, é posible establecer comparativas entre institucións e analizar indicadores bibliométricos e cientímetricos para o análise da investigación. En calquera caso, tanto as universidades como os produtores de bases de datos traballan para normalizar a forma de introducir as autorías e a depuración das afiliacións para reducir o máis posible o impacto desde problema, mediante o uso de ferramentas de comunicación en liña cos autores e as plataformas de rexistro coma ORCID. De cara a futuros traballos hai que destacar tamén que dende o pasado mes de xuño Scopus permite descargar a información bibliográfica en formato .csv en ficheiros de 20.000 referencias, cando no momento de realizar este traballo o máximo eran ficheiros de só 2.000 referencias.

A USC (cos seus colexios invisibles) presenta índices bibliométricos superiores ós da UDC e UVigo. De feito a USC é unha universidade antiga, con grupos de investigación consolidados, que se caracteriza por ter un maior número de autores e documentos, un maior número de autores con alto número de citas e a súa vez un maior número de citas por artigo que as demais universidades galegas. Pola súa banda, a UVigo presenta índices similares ós da USC, en parte debido a un tamén semellante número autores e, comparativamente, á alta produtividade destes (de feito, á UVigo lle corresponde o índice g máis alto).

A produtividade autor, medida en número de elementos dividido polo número de autores, é maior na UVigo e na UDC. Asemade, a UDC e a UVigo caracterízanse por ter un menor tempo de duplicación da produción científica a USC (estudando o período 2006-11). Estes resultados están relacionados co feito de que tanto a UVigo como a UDC son dúas universidades relativamente novas, aínda cun amplo potencial de crecemento.

En xeral, os autores da UDC teñen tendencia a ter un índice h máis baixo que os correspondentes da UVigo, mentres que os autores da UVigo adoitan ter un índice h menor que os da USC: $h(\text{USC}) > h(\text{UVigo}) > h(\text{UDC})$. A consolidación dos grupos de investigación da USC, o seu maior número de autores e produción, por unha banda, e por outra a alta produtividade que presenta a UVigo, son posibles causas desta situación. Agás no caso do índice g, os demais índices presentan a mesma tendencia.

O investigador que presenta os índices h, g, rp1, l1, linf, Ge1, Ge5 e Slp1 máis altos no período estudado é Liz-Marzán (UVigo); mentres que Carracedo A. e Posada D. (UDC) son os autores con maior número de documentos e con un maior número de citas a un só documento, respectivamente.

REFERENCIAS

Alcaín, M. D. (1991). Aspectos métricos de la Información Científica, Ciencias de la Información, 160, 32-36.

Alonso, S.; Cabrerizo, F.J.; Herrera-Viedma, E.; Herrera F. (2011). h-Index: A review focused in its variants, computation and standardization for different scientific fields, *Journal of Informetrics* 3 273–289.

Amézquita, A.; Martínez, D.C.; Martínez, J.C.; Maza, F. (2011). *Bibliometría, Infometría e Cienciometría*, Ediciones Unicartagena (Ciencia, Tecnología, Sociedad e Innovación), 2011.

Arencibia Jorge, R., Carvajal Espino, R. (2008) Los índices H, G y R: su uso para identificar autores líderes en el área de la Comunicación durante el período 2001-2006. *ACIMED* 17(4).

Arroyo, N.; Ortega, J.L.; Pareja, V.; Prieto, J.A.; Aguillo, I. (2005). Cibermetría. Estado de la cuestión. En XIX Jornadas Españolas de Documentación, FESABID.

Garfield, E. (1955) Citation indexes to science: a new dimension in documentation thorough the association of ideas, *Science*. 122, 108-111. <http://garfield.library.upenn.edu/essays/v6p468y1983.pdf>.

Gągolewski, M. (2011). Bibliometric Impact Assessment With R and the CITAN Package, *Journal of Informetrics* 5(4) 678–692.

Gągolewski, M.; Grzegorzewski, P. (2010). S-Statistics and Their Basic Properties, In: Borgelt, C. et al (Eds.), *Combining Soft Computing and Statistical Methods in Data Analysis*, Springer-Verlag 281-288.

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output, *Proceedings of the National Academy of Sciences of the United States of America* 102(46) 16569-16572.

Priem, J.; Taraborelli, D.; Groth, P.; Neylon, C. (2010). Altmetrics: a manifesto – altmetrics.org. from <http://altmetrics.org/manifesto/>