

Enviesamento segundo a origem nacional da produção científica na edição de dados históricos da Web of Knowledge

Poster

António Tavares Lopes
Universidade de Coimbra, Faculdade de Letras (Portugal)
atlopes@fl.uc.pt

Palavras-chave: Web of Knowledge; indicadores nacionais; produção científica; enviesamento

Keywords: Web of Knowledge; national indicators; scientific production; bias

INTRODUÇÃO

Este texto apresenta de forma condensada evidência empírica de tratamento diferenciado na edição das bases de dados que compõem a Web of Knowledge (WoK), tendo por eixo de avaliação a origem nacional das referências que integram o catálogo disponível ao público. Os efeitos deste tipo de distorção estão para além do propósito deste trabalho, mas podem ser revistos na literatura, em particular no contexto geral da crítica à aplicação e opacidade na determinação do Journal Impact Factor (Moed, 2002; Archambault & Larivière, 2009; van Leuween *et al.*; 2001; Rossner, van Epps & Hill, 2007).

OBJETIVOS E METODOLOGIA

Durante um ano, de novembro de 2011 a outubro de 2012, e com periodicidade mensal, mediram-se as contribuições nacionais de diversas origens registadas na Web of Knowledge, através de pesquisas globais sobre o conjunto de dados incidindo em cada um dos 20 anos do período de 1991 a 2010. O levantamento ocorreu de forma a procurar diminuir efeitos de edição concorrente à recolha de dados, centrando-o em dias que são, na tradição ocidental, de descanso semanal.

Na impraticabilidade de cobrir um número mais alargado de países, foram selecionados os Estados Unidos da América, Portugal, Brasil e Espanha (por interesse de proximidade), a França (como país não anglófono de produção científica relevante, em termos de volume) e a Irlanda (como país anglófono de níveis de produção mais periféricos), a que se juntaram a partir de março de 2012 a Inglaterra e a China (como os países que, segundo a WoK, apresentam em 2010 um maior volume de colaboração científica com os Estados Unidos).

RESULTADOS E DISCUSSÃO

Para todas as origens nacionais que foram objeto de controlo, o pico de crescimento anual registado na WoK no período de 1991-2010 ocorre em 1998. O elevado crescimento observado de 1997 a 1998 não serve de avaliação da produção científica real, mesmo que sob uma sua visão particular, e pode explicar-se apenas com fatores relacionados com edição da base de dados. No primeiro ponto de observação (novembro de 2011 ou março de 2012), esse pico vai de 53,6%, para a França, até aos 120,92%, para os Estados Unidos (contra um registo do crescimento médio anual, de 1991 a 1997, de 9,26 e 17,88%, respetivamente).

No segundo segmento da série de observação, ou seja, no período de 1998 a 2010, as taxas de crescimento anual descem significativamente, fruto provável de uma base inicial mais próxima da cobertura máxima que a gestão editorial da base de dados prevê, e que corresponde ao interesse instrumental destes dados na formação do JIF atual: nessa segunda fase, os países centrais crescem pouco (Estados Unidos: 1,73% de crescimento médio anual; Inglaterra: 1,84%; França: 2,39%; com a exceção da China, no seu movimento ascendente para estabelecimento como potência científica: 18,65%) e os países mais periféricos reforçam a sua presença com um crescimento mais acentuado (Espanha: 6,62%; Brasil: 10,45%; Irlanda: 6,51%; Portugal: 11,65%).

O ano de 1997 marca, portanto, uma fronteira que separa dois estados das bases de dados que suportam a WoK: essa data distingue os *dados correntes*, com grande proximidade entre cobertura potencial e cobertura realizada, dos *dados históricos*, em que se verifica maior distância entre essas coberturas. No entanto, o acompanha-

mento da evolução da frequência das diversas origens nacionais, realizado ao longo do período de observação, permite identificar um comportamento editorial do catálogo que não ignora os dados históricos.

De novembro de 2011 a outubro de 2012, os dados históricos das bases de dados da Web of Knowledge foram sendo continuamente editados, resultando num crescimento das referências associadas à localização (*address*) dos Estados Unidos em cada um dos meses de observação que não encontra paralelo nas restantes origens nacionais, centrais ou periféricas, anglófonas ou de outros domínios linguísticos, como se pode ver na tabela 1.

Tabela 1: Crescimento médio de referências relativas aos anos de 1991 a 1997, reportado pela Web of Knowledge, segundo a origem nacional, observado em cada mês de novembro de 2011 a outubro de 2012 (em percentagem)

Origem	1991	1992	1993	1994	1995	1996	1997
Estados Unidos	2,02	2,01	1,88	2,11	1,41	1,32	0,91
Inglaterra *	0,00	0,00	0,00	0,01	0,05	0,06	0,06
China *	0,01	0,00	0,00	0,01	0,01	0,04	0,02
França	0,00	0,01	0,00	0,01	0,02	0,05	0,04
Espanha	0,00	0,00	0,00	0,01	0,15	0,19	0,13
Brasil	0,00	0,00	0,00	0,01	0,04	0,05	0,04
Irlanda	0,01	0,01	0,01	0,02	0,03	0,03	0,05
Portugal	0,03	0,01	0,02	0,02	0,00	0,01	0,55

* início da observação em março de 2012

Este esforço de edição dos dados históricos associados aos Estados Unidos é realizado de forma contínua e também sistemática: verifica-se em todos os meses de observação e, embora nem sempre, em valores de frequência absoluta normalmente próximos: de novembro a dezembro de 2011, cada um dos anos deste período cresce de 5000 a 6200 referências; no mês seguinte, cerca de 2000; atingindo o máximo de abril a maio de 2012, com crescimentos de 24753 (relativo ao ano de 1993) a 31676 (1995), contrastando com as poucas centenas que se verificam de agosto a setembro.

Durante a observação, o número de referências reportadas pelo catálogo da WoK, relativas aos anos de 1991 a 1997, cresceu 282174. Para se entender a importância deste esforço, a cobertura relativa aos anos de 1998 a 2010 cresceu apenas 123720 (ou seja, menos de metade para o dobro dos anos). Como resultado, a fronteira entre os dois períodos, estabelecida no crescimento de 1997 para 1998, diminuiu de 120,92%, observado em novembro de 2011, para 99,99%, observado em outubro de 2012.

As outras origens nacionais não beneficiaram do mesmo esforço de edição dos dados históricos, mesmo entre os países com quem os Estados Unidos têm relações científicas mais profundas, razão pela qual o esbatimento da distinção entre dados históricos e dados correntes é muito menos acentuado (como se resume na tabela 2).

Tabela 2: Crescimento de 1997 a 1998, em referências reportadas pela Web of Knowledge, no início e fim da observação, relativa a diversas origens nacionais (em percentagem)

Origem	Novembro / 2011	Outubro / 2012
Estados Unidos	120,92	99,99
Inglaterra *	60,08	59,49
China *	74,11	73,90
França	53,60	52,98
Espanha	61,57	59,27
Brasil	69,10	68,50
Irlanda	60,15	59,24
Portugal	62,26	52,86

* início da observação em março de 2012

CONCLUSÕES

A origem nacional da produção científica é um critério marcadamente determinante na edição retrospectiva das bases de dados da Web of Knowledge. Em termos observáveis, este enviesamento reforça ainda mais a centralidade dos Estados Unidos e acentua a distorção, nessa direção, que está na própria origem desse serviço comercial.

REFERÊNCIAS

- Archambault, É., & Larivière, V. (2009). History of the journal impact factor: Contingencies and consequences. *Scientometrics*, 79(3), 639-653.
- van Leuween, T. N., Moed, H. F., Tijssen, R. W., Visser, M. S., & van Raan, A. F. J. (2001). Language biases in the coverage of the Science Citation Index and its consequences for international comparisons of national research performance. *Scientometrics*, 51(1), 335-346
- Moed, H. F. (2002). The impact-factors debate: the ISI's uses and limits. *Nature*, 415(6873), 731-732.
- Rossner, M., Van Epps, H. & Hill, E. (2007). Show me the data [editorial]. *The Journal of Cell Biology*, 179(6), 1091-1092.