

# **Correlações entre a contagem de citações de pesquisadores brasileiros, usando o Web of Science, Scopus e Scholar**

**Paula Vanessa Medeiros Vieira**

**Estudante de Mestrado em Ciência da Computação, Universidade Estadual de Campinas (UNICAMP)**

**Jacques Wainer**

**Professor titular do Instituto de Computação, Universidade Estadual de Campinas (UNICAMP)**

*Este trabalho estuda a correlação entre medidas bibliométricas (índice h e quantidade total de citações), obtidas através de diferentes serviços de contagem de citações (Web of Science, Scholar e Scopus). Calculamos essas medidas para diversos pesquisadores do CNPq. Então, obtivemos a correlação dos resultados obtidos para cada medida, entre cada par de serviços. As correlações foram calculadas para cada uma das subáreas divididas, nas seguintes áreas: Ciências Biológicas, Ciências Exatas, Ciências Humanas, Ciências da Saúde, Ciências Sociais Aplicadas, Engenharia e Letras, Linguística e Artes . Discutimos o significado das correlações obtidas.*

**Palavras-chave:** *Bibliometria; Contagem de citações; Índice h; Análise de correlação.*

## **Correlations between citations countings of brazilian researchers using Web of Science, Scopus and Scholar**

*This work analysis the correlation between bibliometric measures (H index and total amount of citations) obtained using different citation count services (Web of Science, Scholar, Scopus ). We computed these bibliometric measures for different researchers from CNPq. So we obtained the correlation between the results*

*obtained metrics for each pair of services. The correlations were calculated for each subarea divided on the following areas: Agriculture Sciences, Biology Sciences, Exact Sciences, Human Sciences, Health Sciences, Social Sciences, Engineering and Arts. We discussed the meaning of the correlations that were obtained.*

**Keywords:** *Bibliometrics; Citations counting; H index; Correlation analysis.*

Recebido em 14.03.2013 Aceito em 05.06.2013

## 1 Introdução

Há uma crescente tendência de se avaliar pesquisadores por indicadores bibliométricos. A quantidade de citações (uma métrica de impacto) e outras métricas, como índice h (que combina produtividade e impacto), são alguns desses indicadores.

Quando se toma a decisão de avaliar cientistas por esses indicadores, é importante levar em consideração qual serviço de contagem de citações será usado para obter os dados sobre citações obtidas pelos pesquisadores. Há três serviços mais conhecidos, o *Web of Science* (da Thomson Reuters, antigo ISI), o *Scopus* (da Elsevier) e o *Scholar* (da Google). Estes serviços de contagem de citações têm diferentes características. O mais tradicional é o *Web of Science* (*WoS*), que define um conjunto pequeno de revistas científicas que são monitoradas e conta as citações feitas por artigos publicados nestas revistas, para outros artigos também publicados nestas revistas. Assim, apenas artigos publicados nestas revistas são monitorados para citações e as citações são feitas por outros artigos nestas revistas. Como os critérios para que uma revista seja incluída na indexação do *WoS* são suficientemente rigorosos, o próprio fato de uma revista ser indexada já é um aval da sua qualidade.

O *Scopus* é um serviço de contagem de citações recente, patrocinado pela Elsevier, que inclui na indexação não só várias revistas, como, também, conferências científicas, principalmente as das áreas de Engenharia e Ciências da Computação.

Finalmente, o *Scholar* busca referências em qualquer documento que esteja disponível na *Web* (incluindo versões eletrônicas de revistas e conferências). Assim, qualquer documento que é citado por algum outro documento, que está disponível na *Web*, passa a ter suas citações monitoradas. No entanto, o *Scholar* não conta citações feitas por livros (embora ele conte as citações recebidas por livros).

São conhecidas as limitações destes serviços para algumas das áreas da ciência. Por exemplo, áreas como as Engenharias e a Ciência da Computação, que têm uma alta produção científica em conferências (e não em revistas), têm tanto a produção como citações sub-representadas

no *WoS*. Áreas das Ciências Humanas e Sociais, nas quais a produção científica é feita através de monografias, livros e capítulos de livros, na maioria das vezes escritas em português, são muito sub-representadas não só no *WoS* e no *Scopus*, como, também, no *Scholar*, que não conta as citações feitas por livros.

Além das diferenças de cobertura, há diferenças no custo de assinar um ou outro serviço de contagem de citações. Em particular, o *Google Scholar* é gratuito, portanto, há um grande incentivo em utilizá-lo, em vez dos outros dois concorrentes (não sabemos detalhes dos custos de utilizar os outros dois serviços). Por outro lado, muito da pesquisa tradicional em bibliometria é feita usando o *WoS* e recentemente o *Scopus*. Portanto, é importante saber quanto os dados de citações obtidos usando o *WoS* correspondem àqueles obtidos usando o *Scopus* e o *Google Scholar*.

Esta pesquisa tem como objetivo medir a correlação de dados de citações obtidos do *WoS*, do *Scopus* e do *Scholar* para todas as diferentes áreas da Ciência, usando dados de pesquisadores brasileiros. Em particular, usando dados de 10285 pesquisadores divididos em 70 diferentes áreas da Ciência, calcularemos as correlações entre o total de citações recebidas e o índice *h*, usando dados do *WoS*, *Scopus* e *Scholar*.

Nesta pesquisa, calcularemos duas correlações, a mais tradicional, correlação linear de Pearson, e o *ro* de Spearman. A correlação linear mostra quanto uma das medidas, por exemplo, o número total de citações obtidos pelo *WoS*, depende de forma linear de outra, por exemplo, o número total de citações obtidos pelo *Scholar*. Se a correlação é 1, então os dois valores são sempre proporcionais e o valor da proporção permite corrigir um dos valores pelo outro. Assim, se a correlação linear é próxima de 1 e o valor da proporção é 0.3, então, os valores do total de citações obtidos usando o *WoS* são "quase sempre" 30% dos valores obtidos usando o *Scholar*. Baixas correlações indicam que a relação de uma e outra medida varia entre um pesquisador para outro. Uma correlação de 0 indica que os valores são basicamente independentes e saber o número total de citações no *WoS* não nos dá nenhuma informação sobre o número total de citações pelo *Scholar*.

Outra medida de correlação importante neste contexto é o *ro* de Spearman, que mede se há alguma correlação monotônica entre as variáveis. Mesmo que as medidas usando um ou outro serviço bibliométrico tenham baixa correlação linear, pode ser útil saber se as medidas mantêm a ordem relativa, isto é, se uma medida é maior para um cientista usando um serviço bibliométrico, então ela será (provavelmente) maior, também, se usarmos outro serviço. Um valor próximo de 1 para o *ro* indica que as ordens entre as medidas se mantêm. Saber a ordem das medidas pode ser importante em situações nas quais um cientista é avaliado contra outros, por exemplo, em concursos, na atribuição de Bolsa de Produtividade em Pesquisa do CNPq, etc.

Finalmente, neste trabalho, mediremos as correlações de duas medidas, o número total de citações recebidas pelo cientista e o índice *h* do cientista. O número total de citações recebidas é a medida mais tradicional de impacto do cientista na sua área de pesquisa. Outras

medidas derivadas do número total de citações provavelmente terão correlações iguais ou parecidas. Medidas como citações por ano têm a mesma correlação entre si que as medidas de citações totais. Já a medida de citações por artigo, terá correlações similares, mas não necessariamente iguais as de citações totais, já que os diferentes serviços indexam diferentes fontes e, portanto, o número de publicações de um cientista não será igual nos diferentes serviços. As diferenças nas correlações não devem ser grandes. Mas isto não será verdade para o índice *h*, que não depende do total de citações, mas do número de citações recebidas por cada artigo de um subconjunto pequeno das publicações do cientista. Assim, calcularemos também as correlações para o índice *h*.

Assim, este artigo computará as correlações de Pearson (linear) e de Spearman para as medidas de número total de citações e de índice *h*, para dados coletados usando o *Web of Science*, *Scopus* e *Scholar*, para cientistas em 70 diferentes áreas do conhecimento. Também, calcularemos os coeficientes de proporcionalidade para a correlação linear.

## 2 Pesquisas relacionadas

Meho e Yang (2007) fizeram um estudo com base em dados coletados no *Web of Science*, *Google Scholar* e *Scopus* para citações obtidas por publicações de 15 membros do corpo docente de ciência da informação de *Indiana University - Bloomington*. Foram comparados os dados obtidos através do *Scopus* e *Google Scholar* em relação aos obtidos através do *WoS*. Em comparação ao *WoS*, *Scopus* retornou 13.7% mais citações, na média. Combinando os resultados obtidos pelas duas ferramentas, retirando-se a intersecção, foram encontradas 35% mais citações do que quando usado apenas o *WoS*.

A diferença entre a quantidade de citações encontradas pelo *WoS* e pelo *Scopus* variou entre 4.9% a 98.9%, se levada em consideração a contagem individual para cada pesquisador. A sobreposição de citações entre os dois bancos de dados é relativamente pequena (58.2%). O número de citações encontradas exclusivamente no *Scopus* em relação ao *WoS* é relativamente alto (710 ou 26.0%, em comparação a 432 ou 15.8%, respectivamente).

Os resultados também mostraram que *Scholar* retornou 53% citações do que *WoS* e *Scopus* combinados. A diferença variou de acordo com a área analisada. Apesar dessa diferença, combinando-se os três bancos de dados, o *ranking* relativo dos docentes, em comparação com o obtido com resultados apenas pelo *WoS*, não altera significativamente (ro de Spearman = 0.976).

Bar-Ilan (2008) fez um estudo com base em dados obtidos para o índice *h* de um grupo de 40 pesquisadores israelenses, usando *Google Scholar*, *Scopus* e *Web of Science*. No geral, não houve diferenças significativas entre o índice *h* obtido pelo *Web of Science* e pelo *Scopus*. A diferença mais significante foi encontrada entre os resultados obtidos

usando o *Google Scholar* e os outros dois serviços. Bar-Ilan (2008) dividiu os autores em três grupos:

- a) aqueles cujo índice h do *Google Scholar* são 30% maior que a média obtida pelo *Scopus* e *Web of Science*;
- b) aqueles cujo índice h do *Google Scholar* são 30% menor que a média obtida pelo *Scopus* e *Web of Science*; e
- c) aqueles cujo índice h do *Google Scholar* esta entre 0.7 e 1.3 vezes que a média obtida pelo *Scopus* e *Web of Science*.

O primeiro grupo consiste unicamente em matemáticos e cientistas da computação. O segundo grupo é composto por dois físicos de altas energias do grupo de pesquisadores. Mesmo para pesquisadores com um índice h similar para as três plataformas (o terceiro grupo), foram encontradas diferenças na contagem de citações. Por exemplo, o pesquisador Avram Hersho, com 20% menos citações no *Scholar* em relação às outras plataformas; David Harel possui apenas 188 citações no *WoS*, enquanto no *Scholar* possui 3374 citações. O estudo conclui que é importante definir qual será a ferramenta utilizada para analisar o desempenho dos pesquisadores.

Meho e Rogers (2008) fizeram um estudo sobre as diferenças entre *Scopus* e *WoS* na contagem de citações, *ranking* de citações e o índice h dos 22 melhores pesquisadores em Interface Homem-Computador (IHC) do projeto EQUATOR - um grande projeto britânico de pesquisa colaborativa interdisciplinar. Os resultados indicaram que o *Scopus* provê uma cobertura maior da literatura de IHC que o *WoS*. Os 22 membros foram citados 7439 vezes em documentos publicados entre 1996 e 2007. Dessas citações, *Scopus* cobre 6919 (93%), enquanto *Web of Science* cobre 4011 (54%).

O estudo sugere que essa diferença é devido ao fato que a cobertura de conferências é mais ampla no *Scopus* que no *Web of Science* - 775 documentos em comparação a 340. O índice h computado pelo *Scopus* também é significativamente maior que o computado pelo *WoS* (variando de um acréscimo de 55% a 140%). Apesar disso, o *ranking* produzido pelos dois bancos de dados é bastante similar (Spearman  $r_o = 0.97$ ). A soma das citações dos dois bancos de dados também não altera significativamente o *ranking*.

Kulkarni *et al.* (2009) fizeram um estudo para comparar os indicadores de citações de artigos publicados em periódicos médicos entre dados obtidos pelo *Scopus*, *Google Scholar* e *Web of Science*. Resultados do *Google Scholar* e do *Scopus* retornaram mais citações por artigo, com uma média de 160 e 149, respectivamente, do que o *WoS* (média 122). Comparado ao *WoS*, *Scopus* retornou mais citações de fontes em línguas que não a inglesa e menos em artigos, editoriais e cartas. Apesar do *Google Scholar* ter mais citações que o *WoS* e *Scopus* no geral, nos casos específicos de artigos com autoria coletiva, com financiamento declarado de indústrias ou informando sobre estudos drogas e dispositivos médicos, isso não se observa.

Franceschet (2010) fez uma comparação entre resultados de diversos indicadores bibliométricos de pesquisadores de ciência da computação italianos do Departamento de Matemática e Ciência da Computação da Universidade de Udine, obtidos, utilizando dados retirados do *Google Scholar* e do *Web of Science*. O estudo mostrou que os valores calculados para os indicadores utilizando o Scholar foram maiores que utilizando o WoS. Foram encontrados cinco vezes mais artigos no *Google Scholar* do que no *Web of Science* e quase oito vezes mais citações no *Google Scholar* do que no WoS. O autor calculou os coeficientes de correlação (utilizando Kendall e Spearman) entre os indicadores, encontrando:

a) boa correlação (maior que 0.84) para citações, citações por ano, citações por autor, índice g, índice h individual; e

b) correlação moderada (entre 0,62 e 0,69) para número de artigos, artigos com citações, artigos por ano, citações por artigo e índice h  
correlação fraca (menor que 0.53) para índice h contemporâneo e índice m.

A Tabela 1 resume os resultados das pesquisas relacionadas a correlações entre os diferentes serviços bibliográficos.

Tabela 1 – Síntese dos resultados das pesquisas relacionadas

Referência	Fonte	Áreas	Resultados
Meho, L. I & Yang, K (2007)	<i>Web of Science, Google Scholar e Scopus</i>	Biblioteconomia e Ciência da Informação	Scopus retornou 13.7% mais citações que WoS;  A diferença entre a quantidade de citações encontradas pelo WoS e pelo <i>Scopus</i> , variou entre 4.9% a 98.9%, se levado em consideração a contagem individual para cada pesquisador;  <i>Scholar</i> retornou 53% citações do que WoS e Scopus combinados
Bar-Ilan, J (2008)	<i>Google Scholar, Scopus e Web of Science</i>	Diversas áreas	No geral, não houve diferenças significativas entre o índice h obtido pelo <i>Web of Science</i> e pelo <i>Scopus</i>  Diferença mais significativa foi encontrada entre os resultados obtidos usando o <i>Google Scholar</i> e os outros dois serviços.
Meho, L. I & Rogers, Y. (2008)	Scopus e WoS	IHC (Interface Homem-Computador)	<i>Os resultados indicaram que o Scopus provê uma cobertura maior da literatura de IHC que o WoS.</i>  <i>Scopus cobre 6,919 (93%) das citações obtidas pelos autores, enquanto Web of Science cobre 4 011 (54%).</i>
Kulkarni, A. V et al. (2009)	Scopus, Google Scholar e Web of Science	Medicina	<i>Google Scholar e do Scopus retornaram mais citações por artigo, com uma média de 160 e 149, respectivamente, do que o WoS (média 122)</i>
Franceschet, M (2010)	Google Scholar e Web of Science	Ciência da Computação	<i>Foram encontrados cinco vezes mais artigos no Google Scholar do que no Web of Science; Quase oito vezes mais citações no Google Scholar do que no WoS</i>

Fonte: Dados de pesquisa.

### 3 Dados e método

Usamos dados sobre pesquisadores brasileiros com Bolsa de Produtividade do CNPq, vigente em 2012, de acordo com uma lista fornecida pelo CNPq aos autores. O estudo foi feito separando-se os bolsistas por subárea; as subáreas representadas por menos de dez pesquisadores foram removidas da análise. No total, foram incluídos, nesse estudo, 10285 pesquisadores distribuídos em 70 subáreas.

Nós fizemos um conjunto de programas para coletar dados do *Scopus*, *Web of Science* e *Google Scholar*. Procuramos pelo nome de cada pesquisador e cada artigo retornado foi comparado com a lista de artigos presente no seu Currículo Lattes. Para comparar os artigos, retiramos todos os caracteres não-alfanuméricos e não diferenciamos entre letras maiúsculas e minúsculas.

Em muitos casos, o título do artigo está escrito em português no Currículo Lattes e em inglês no *Web of Science*, ou a ordem das palavras difere. Portanto, quando não é retornado nenhum artigo para um pesquisador, comparamos os resultados da busca através do nome do periódico, ano de publicação e volume.

Tabela 2 - Número de cientistas para os quais nosso algoritmo de coleta de dados não obteve dados de citações, para o *WOS*, *Scopus* e *Scholar*. A coluna Total indica o número de cientistas com bolsa do CNPq (em 2012), para cada uma das áreas do conhecimento

	<i>GS</i>	<i>Scopus</i>	<i>WoS</i>	Total
<b>Engenharias (EN)</b>				
Engenharia Civil	31	67	75	272
Engenharia Naval e Oceânica	0	1	1	12
Engenharia Química	11	37	40	169
Engenharia Sanitária	13	27	75	163
Engenharia de Transportes	15	13	32	52
Engenharia Nuclear	10	19	14	77
Engenharia de Produção	9	20	50	146
Engenharia de Minas	3	3	10	35
Engenharia Biomédica	7	25	17	80
Engenharia Elétrica	40	71	64	302
Engenharia Mecânica	22	58	95	304
Engenharia Aeroespacial	6	11	13	60
<b>Ciências Biológicas (CB)</b>				
Farmacologia	14	33	41	205
Imunologia	15	31	46	164
Parasitologia	13	25	30	143
Biofísica	7	11	17	75
Ecologia	15	47	50	208
Genética	17	51	65	264
Morfologia	5	24	36	120
Zoologia	15	44	67	244
Botânica	10	31	57	226
Microbiologia	8	40	36	185
Bioquímica	14	59	39	273
Fisiologia	12	38	35	189
<b>Ciências Sociais Aplicadas (CSA)</b>				

Serviço Social	43	18	70	73
Economia	28	31	133	200
Planejamento Urbano e Regional	31	23	68	78
Arquitetura e Urbanismo	49	24	88	93
Administração	18	21	126	159
Comunicação	71	39	117	128
Ciência Política	34	34	88	114
Demografia	10	12	26	34
Ciência da Informação	17	12	37	46
<b>Ciências Exatas (CE)</b>				
Astronomia	17	26	28	99
Química	56	171	167	759
Matemática	50	88	97	317
Ciência da Computação	57	96	90	383
Oceanografia	6	23	16	117
Probabilidade e Estatística	8	21	17	69
Geociências	38	17	127	470
Física	126	222	229	929
<b>Ciências Humanas (CH)</b>				
Antropologia	47	33	114	144
Arqueologia	19	16	30	47
História	125	55	217	234
Sociologia	55	50	170	192
Psicologia	38	54	195	313
Geografia	46	22	84	90
Filosofia	71	48	122	137
Educação	122	79	349	378
<b>Ciências da Saúde (CS)</b>				
Fisioterapia e Terapia Ocupacional	2	11	19	61
Nutrição	8	17	22	80
Saúde Coletiva	7	31	56	212
Enfermagem	9	34	71	165
Medicina	27	85	112	553
Odontologia	2	27	45	207
Farmácia	2	27	29	158
Educação Física	4	0	35	84
Fonoaudiologia	2	8	22	50
<b>Letras, Linguística e Artes (LLA)</b>				
Linguística	115	54	185	207
Artes	67	40	100	103
Letras	135	70	225	229
<b>Ciências Agrárias (CAG)</b>				
Medicina Veterinária	16	51	69	313
Agronomia	44	134	232	810
Ciência e Tecnologia de Alimentos	10	23	34	189
Engenharia Agrícola	12	29	79	156
Recursos Florestais e Engenharia Florestal	10	17	73	155
Zootecnia	6	40	104	284
Recursos Pesqueiros e Engenharia de Pesca	3	27	31	128

Fonte: Dados de pesquisa e do CNPq.

## 4 Resultados e discussão

A Tabela 2 lista, para cada uma das áreas do conhecimento, o número de cientistas para os quais nosso programa de coleta de dados, nos diferentes serviços bibliométricos, não obteve dados de citações. A coluna total indica o número de pesquisadores com bolsa de pesquisa do CNPq, no ano de 2012. Tabelas 3 e 4 são os principais resultados deste artigo. Tabela 3 indica a correlação de Pearson para as três comparações (*GS vs Scopus*, *GS vs WoS* e *Scopus vs WoS*). A tabela lista as correlações para o número total de citações recebidas e para o índice H. A coluna de correlação contém um “\*”, se a *GS* correlação **não** é estatisticamente significativa, isto é, se o p-valor do teste se a correlação é diferente de 0 é maior que 0.05. A coluna que segue a medida de correlação é a proporção média entre uma medida e outra. A coluna indicada pelo  $\alpha$  que segue a coluna *GS vs Scopus* indica a proporção entre os valores obtidos usando o *Google Scholar* e o *Scopus*. Assim, para Antropologia, não há uma correlação linear estatisticamente significativa entre o total de citações medidas pelo *Google Scholar* e o *Scopus* (primeira linha da Tabela 3). Para a Arqueologia, a correlação é bastante alta (0.89). Além disto, as medidas obtidas usando o *Google Scholar* são, em média, 1.59 vezes as medidas obtidas usando o *Scopus*. A tabela está organizada por grande área: CH indica Ciências Humanas, CB Ciências Biológicas, CSA Ciências Sociais Aplicadas, CE Ciências Exatas, EN Engenharias, CS Ciências da Saúde, LLA Linguística e Letras e Artes, e CAG Ciências Agrícolas.

A Tabela 4 indica as correlações usando o *rho* de Spearman. Da mesma forma, “\*” indica uma correlação não estatisticamente significativa.

Tabela 3 – Correlações utilizando Pearson e inclinação da reta obtida por regressão linear. “\*” indica que a correlação não é estatisticamente significativa (com 95% de confiança). O valor indicado na coluna  $\alpha$  que segue a coluna de correlação indica a proporção média entre os valores obtidos usando os dois serviços

	Citações						Índice H					
	<i>GS</i> x <i>Scopus</i>	$\alpha$	<i>GS</i> x <i>WoS</i>	$\alpha$	<i>Scopus</i> x <i>WoS</i>	$\alpha$	<i>GS</i> x <i>Scopus</i>	$\alpha$	<i>GS</i> x <i>WoS</i>	$\alpha$	<i>Scopus</i> x <i>WoS</i>	$\alpha$
<b>CH</b>												
Antropologia	*		*		0.35	0.37	*		0.31	1.65	0.49	0.59
Arqueologia	0.89	1.59	0.86	1.85	0.99	1.14	0.81	1.15	0.77	1.19	0.95	1.03
Educação	0.38	4.85	0.54	9.54	0.87	1.23	0.34	1.21	0.38	1.75	0.7	0.9
Filosofia	*		0.31	5.15	0.77	0.64	0.33	0.82	0.45	1.09	0.79	0.78
Geografia	*		*		*		*		0.29	1.26	0.59	0.81
História	*		*		*		0.2	0.63	0.22	1.16	0.18	0.3
Psicologia	0.39	0.92	0.34	1.03	0.84	0.82	0.31	0.59	0.25	0.44	0.79	0.75
Sociologia	*		0.21	27.61	0.31	0.68	*		0.15	0.81	0.49	0.73
<b>CB</b>												
Biofísica	0.41	0.58	0.56	0.81	0.7	0.59	0.38	0.41	0.49	0.51	0.6	0.59

Bioquímica	0.85	1.12	0.92	1.14	0.86	0.77	0.46	0.57	0.51	0.61	0.8	0.78
Botânica	0.72	1.34	0.67	1.54	0.82	0.74	0.33	0.44	0.28	0.34	0.82	0.77
Ecologia	0.86	1.24	0.84	1.3	0.93	0.99	0.34	0.43	0.31	0.4	0.86	0.9
Farmacologia	0.54	1.07	0.68	1.33	0.72	0.68	0.36	0.47	0.53	0.71	0.77	0.78
Fisiologia	0.88	1.11	0.93	1.09	0.92	0.85	0.46	0.58	0.52	0.64	0.79	0.79
Genética	0.74	1.02	0.76	0.97	0.81	0.76	0.52	0.62	0.51	0.64	0.7	0.73
Imunologia	0.76	1.2	0.74	0.92	0.89	0.73	0.53	0.69	0.52	0.68	0.73	0.72
Microbiologia	0.52	0.76	0.64	1.01	0.78	0.8	0.26	0.32	0.32	0.45	0.78	0.87
Morfologia	0.82	1.16	0.91	1.29	0.85	0.98	0.28	0.32	0.37	0.43	0.74	0.76
Parasitologia	0.71	0.97	0.7	0.93	0.92	0.89	0.41	0.47	0.47	0.58	0.83	0.9
Zoologia	0.64	1.19	0.54	1.36	0.87	0.95	0.46	0.65	0.34	0.49	0.8	0.82
<b>CSA</b>												
Administração	0.62	2.1	0.59	2.71	0.98	1.29	0.24	0.54	0.2	0.47	0.86	0.92
Arquitetura e Urbanismo	0.94	1.06	0.71	5.8	0.67	4.05	*		0.19	0.38	0.83	1.07
Ciência da Informação	*		*		0.66	0.76	*		0.12	0.75	0.76	0.93
Ciência Política	0.58	13.2 2	0.38	14.9	0.68	0.68	0.39	1.69	0.29	1.54	0.65	0.81
Comunicação	*		0.37	104.6	*		0.2	1.01	0.2	1.68	0.35	0.59
Demografia	*		*		0.92	1.05	*		0.41	0.95	0.91	1.09
Economia	0.59	1.62	0.52	1.33	0.79	0.71	0.49	0.85	0.46	0.77	0.87	0.86
Planejamento Urbano e Regional	*		*		0.63	0.66	*		*		0.57	0.77
Serviço social	*		*		*		0.38	1.66	0.4	1.76	0.68	0.68
<b>CE</b>												
Astronomia	0.29	0.23	0.37	0.21	0.78	0.58	0.4	0.41	0.47	0.43	0.71	0.63
Ciência da Computação	0.51	1.56	0.62	1.49	0.72	0.63	0.38	0.58	0.5	0.79	0.68	0.71
Física	0.76	1.13	0.78	0.91	0.91	0.74	0.51	0.55	0.54	0.56	0.85	0.82
Geociências	0.82	1.31	0.73	1.14	0.82	0.79	0.78	0.87	0.74	0.87	0.83	0.87
Matemática	0.58	1.53	0.53	1.19	0.86	0.8	0.41	0.54	0.38	0.48	0.8	0.78
Oceanografia	0.8	1.17	0.8	1.12	0.87	0.83	0.4	0.56	0.37	0.5	0.74	0.73
Probabilidade e Estatística	0.66	1.18	0.84	1.52	0.88	0.91	0.52	0.77	0.62	0.98	0.85	0.9
Química	0.7	0.99	0.71	1.06	0.82	0.82	0.44	0.49	0.46	0.54	0.77	0.8
<b>EN</b>												
Engenharia Aeroespacial	0.84	1.02	0.9	1.71	0.8	1.23	0.7	0.8	0.7	0.96	0.76	0.9
Engenharia Biomédica	0.93	1.13	0.91	1.22	0.93	1.01	0.45	0.53	0.33	0.43	0.74	0.81
Engenharia Civil	0.87	1.2	0.92	1.41	0.91	1.02	0.56	0.6	0.58	0.69	0.87	0.96
Engenharia de Materiais e Metalúrgica	0.85	1.18	0.93	1.02	0.82	0.55	0.62	0.69	0.57	0.64	0.86	0.86
Engenharia de Minas	0.77	1.11	0.96	1.37	0.86	0.86	0.78	0.82	0.88	0.99	0.91	0.97
Engenharia	0.83	1.75	0.78	2.21	0.92	1.26	0.53	0.76	0.45	0.75	0.84	0.97

de Produção													
Engenharia de Transportes	0.83	2.32	0.5	2.33	0.6	0.83	0.43	0.79	0.38	0.79	0.63	0.71	
Engenharia Elétrica	0.84	1.32	0.85	1.64	0.9	1.12	0.51	0.59	0.52	0.7	0.79	0.92	
Engenharia Mecânica	0.75	1.1	0.76	1.22	0.9	0.97	0.47	0.57	0.44	0.54	0.78	0.81	
Engenharia Naval e Oceânica	0.98	0.87	0.99	1.09	0.98	1.22	0.75	0.75	0.78	0.78	0.85	0.85	
Engenharia Nuclear	0.84	1.22	0.72	0.95	0.81	0.67	0.57	0.72	0.53	0.68	0.89	0.91	
Engenharia Química	0.87	0.99	0.85	1.07	0.86	0.97	0.47	0.53	0.53	0.64	0.8	0.86	
Engenharia Sanitária	0.69	0.85	0.9	1.68	0.79	1.18	0.59	0.69	0.74	0.98	0.84	0.97	
<b>CS</b>													
Educação Física	*		*		0.94	1.81	0.33	0.32	0.17	0.21	0.8	1.01	
Enfermagem	0.64	2.15	0.42	3.11	0.82	1.19	0.34	0.7	0.26	0.74	0.73	1.01	
Farmácia	0.83	0.9	0.86	1.14	0.82	0.98	0.47	0.54	0.46	0.58	0.75	0.83	
Fisioterapia e Terapia Ocupacional	*		0.53	0.89	0.52	0.41	0.47	0.65	0.43	0.63	0.71	0.76	
Fonoaudiologia	*		*		*		*		*		0.7	0.81	
Medicina	0.74	0.94	0.77	1.04	0.84	0.87	0.41	0.48	0.41	0.54	0.76	0.85	
Nutrição	0.8	2.5	0.86	2.95	0.86	0.96	0.52	0.88	0.45	0.86	0.83	0.93	
Odontologia	0.82	0.89	0.82	1.3	0.93	1.34	0.5	0.61	0.52	0.76	0.83	1	
Saúde Coletiva	0.95	1.43	0.91	1.3	0.97	0.93	0.68	0.93	0.68	0.92	0.88	0.87	
<b>LLA</b>													
Artes	*		*		*		*		0.1	0.26	*		
Letras	*		*		0.45	24.58	0.34	0.98	0.15	2.1	0.39	1.86	
Linguística	*		0.41	13.92	0.57	1	0.21	0.86	0.31	1.82	0.61	0.89	
<b>CAG</b>													
Agronomia	0.81	1.5	0.67	1.37	0.84	0.9	0.49	0.74	0.43	0.69	0.76	0.82	
Ciência e Tecnologia de Alimentos	0.9	1.2	0.91	1.36	0.96	1.07	0.58	0.69	0.59	0.72	0.9	0.92	
Engenharia Agrícola	0.68	2.23	0.4	2.2	0.67	1.14	0.28	0.49	0.15	0.37	0.54	0.75	
Medicina Veterinária	0.84	1.19	0.81	1.03	0.85	0.81	0.52	0.64	0.55	0.74	0.85	0.93	
Recursos Florestais e Engenharia Florestal	0.89	1.54	0.92	1.69	0.98	1	0.39	0.48	0.44	0.49	0.88	0.82	
Recursos Pesqueiros e Engenharia de Pesca	0.4	0.66	0.51	0.95	0.77	0.85	0.4	0.59	0.35	0.53	0.73	0.76	

Zootecnia 0.56 1.8 0.33 0.91 0.5 0.48 0.39 0.75 0.28 0.62 0.45 0.52

Fonte: Dados de pesquisa.

Tabela 4 – Resultados obtidos utilizando-se correlação de Spearman

	Citações			Índice H		
	Scopus X GS	GS X WoS	Scopus x WoS	Scopus X GS	GS X WoS	Scopus x WoS
<b>CH</b>						
Antropologia	*	0.31	0.48	0.27	0.19	0.53
Arqueologia	0.64	0.61	0.97	0.5	0.41	0.89
Educação	0.24	0.31	0.42	0.28	0.23	0.38
Filosofia	*	*	0.51	0.23	0.27	0.54
Geografia	*	0.34	*	*	0.18	0.32
História	*	*	*	0.18	0.07	0.27
Psicologia	0.37	0.32	0.68	0.27	0.19	0.67
Sociologia	0.26	0.2	0.42	*	0.13	0.43
<b>CB</b>						
Biofísica	0.63	0.6	0.59	0.48	0.55	0.66
Bioquímica	0.54	0.76	0.75	0.34	0.42	0.76
Botânica	0.6	0.63	0.76	0.34	0.36	0.77
Ecologia	0.59	0.59	0.74	0.23	0.18	0.8
Farmacologia	0.55	0.71	0.75	0.29	0.4	0.78
Fisiologia	0.56	0.71	0.7	0.33	0.35	0.75
Genética	0.42	0.51	0.61	0.39	0.4	0.64
Imunologia	0.54	0.6	0.65	0.35	0.39	0.66
Microbiologia	0.51	0.67	0.75	0.22	0.26	0.78
Morfologia	0.58	0.73	0.75	*	0.19	0.71
Parasitologia	0.5	0.63	0.76	0.38	0.47	0.79
Zoologia	0.54	0.52	0.72	0.45	0.36	0.76
<b>CSA</b>						
Administração	0.5	0.46	0.77	*	*	0.68
Arquitetura e Urbanismo	0.64	*	0.47	*	0.09	0.38
Ciência da Informação	*	*	0.72	*	*	0.78
Ciência Política	0.44	*	0.59	0.31	0.22	0.56
Comunicação	*	*	*	0.2	0.12	0.42
Demografia	*	0.67	0.74	*	0.15	0.65
Economia	0.46	0.45	0.69	0.42	0.33	0.68
Planejamento Urbano e Regional	*	*	0.46	*	*	0.52
Serviço social	*	0.4	*	0.41	0.41	0.68
<b>CE</b>						
Astronomia	0.36	0.5	0.72	0.34	0.39	0.64
Ciência da Computação	0.62	0.64	0.69	0.36	0.45	0.69
Física	0.62	0.69	0.83	0.45	0.47	0.83
Geociências	0.7	0.67	0.84	0.69	0.69	0.79
Matemática	0.55	0.58	0.77	0.33	0.32	0.81
Oceanografia	0.5	0.62	0.68	0.38	0.34	0.7
Probabilidade e Estatística	0.55	0.74	0.77	0.42	0.48	0.79
Química	0.67	0.69	0.73	0.41	0.44	0.77
<b>EN</b>						
Engenharia	0.64	0.89	0.68	0.55	0.59	0.74

Aeroespacial						
Engenharia Biomédica	0.48	0.39	0.71	0.34	0.23	0.71
Engenharia Civil	0.74	0.81	0.81	0.42	0.44	0.79
Engenharia de Materiais e Metalúrgica	0.79	0.76	0.84	0.54	0.49	0.84
Engenharia de Minas	0.52	0.59	0.84	0.51	0.67	0.82
Engenharia de Produção	0.54	0.55	0.77	0.34	0.23	0.78
Engenharia de Transportes	0.69	0.5	0.47	0.46	0.42	0.57
Engenharia Elétrica	0.57	0.65	0.75	0.36	0.38	0.69
Engenharia Mecânica	0.57	0.57	0.75	0.42	0.34	0.74
Engenharia Naval e Oceânica	0.93	0.97	0.88	0.84	0.85	0.86
Engenharia Nuclear	0.76	0.65	0.82	0.52	0.48	0.89
Engenharia Química	0.75	0.74	0.76	0.41	0.47	0.77
Engenharia Sanitária	0.64	0.6	0.82	0.31	0.38	0.79
<b>CS</b>						
Educação Física	0.64	0.54	0.73	0.4	0.29	0.8
Enfermagem	0.65	0.5	0.84	0.6	0.36	0.72
Farmácia	0.67	0.56	0.65	0.36	0.37	0.65
Fisioterapia e Terapia Ocupacional	0.59	0.72	0.71	0.42	0.42	0.7
Fonoaudiologia	0.45	0.56	0.77	0.42	0.42	0.75
Medicina	*	*	0.66	*	*	0.68
Nutrição	0.61	0.64	0.71	0.38	0.38	0.72
Odontologia	0.69	0.71	0.82	0.46	0.49	0.8
Saúde Coletiva	0.55	0.51	0.72	0.38	0.34	0.71
<b>LLA</b>						
Artes	*	*	*	*	0.1	*
Letras	*	*	*	0.22	*	0.16
Linguística	*	0.33	0.65	0.23	0.21	0.62
<b>CAG</b>						
Agronomia	0.62	0.54	0.71	0.36	0.33	0.67
Ciência e Tecnologia de Alimentos	0.67	0.75	0.87	0.46	0.48	0.87
Engenharia Agrícola	0.48	0.36	0.58	0.21	0.16	0.5
Medicina Veterinária	0.67	0.72	0.81	0.4	0.44	0.81
Recursos Florestais e Engenharia Florestal	0.45	0.34	0.57	0.28	0.3	0.58
Recursos Pesqueiros e Engenharia de Pesca	0.46	0.44	0.66	0.37	0.25	0.67
Zootecnia	0.54	0.32	0.53	0.35	0.25	0.47

Fonte: Dados de pesquisa.

Subáreas de LLA possuem baixa correlação ou correlação não significativa, com exceção de Linguística, que possui uma correlação alta

entre citações e índice h obtidos pelo *WoS* e pelo *Scopus* (0,65 e 0,62 com *Spearman* e 0,57 e 0,61 com *Pearson*, para citações e índice h, respectivamente).

Subáreas de CS possuem, em geral, alta correlação entre resultados do *Scopus* e do *WoS*, entre citações obtidas pelo *Scholar* e pelo *Scopus* e correlações médias para os outros resultados.

Subáreas de Engenharia possuem, em geral, correlações altas com algumas exceções. Sendo, também, maior a correlação entre dados obtidos pelo *Scopus* e pelo *WoS*. Há um destaque para Engenharia Naval e Oceânica, que possui correlações muito altas para todos os dados.

Subáreas de CAG possuem correlações médias no geral, também com correlações menores para o índice h do que para quantidade de citações.

Subáreas de CE possuem, no geral, correlações altas para quantidade de citações. A correlação entre índice h obtido com dados do *WoS* e *Scopus* que se destaca, obtendo uma correlação alta em todas as subáreas.

Subáreas de CH possuem, no geral, correlações baixas, com destaque para Arqueologia, que possui correlações altas no geral e para Psicologia, que possui correlações altas entre dados obtidos pelo *WoS* e pelo *Scopus*.

Franceschet (2010) encontrou uma boa correlação para pesquisadores de Ciência da Computação entre o *Scholar* e *WOS*. Para citações por autor,  $r$  é de 0,87 e tau de Kendal 0,68 ; para o índice h,  $r=0,84$  e tau= 0,70. Nós encontramos 0,62 (citações) e 0,5 (índice h) por Pearson e 0,64 (citações) e 0,45 (índice h) por Pearson, que são valores mais baixos. A quantidade de citações tem uma boa correlação, mas o índice h não.

## 5 Conclusão

Esta pesquisa deve ser encarada como uma pesquisa exploratória sobre as correlações entre medidas bibliométricas, usando os três diferentes serviços de contagem de citações mais comuns: *Web of Science*, *Scopus* e *Google Scholar*.

Esperamos que o principal uso para esta pesquisa seja que pesquisadores de cada uma das áreas incluídas verificarão as correlações entre os vários serviços e usarão esta informação para decidir se um particular serviço bibliográfico é o mais apropriado para ser usado nas avaliações da área. Em particular, a tradição bibliométrica tem sido usar o *Web of Science*. Caso as correlações entre os outros serviços e o *WoS* são baixas, pesquisadores da área podem argumentar que o *WoS* não representa bem a área e que a utilização de um dos outros serviços seria mais justificável.

Altas correlações de Pearson indicariam que um serviço pode ser substituído por outro, desde que os valores fossem corrigidos pelos coeficientes de proporção calculados neste artigo. Baixas correlações indicam que usar um serviço é muito diferente de usar outro e, portanto,

pesquisadores da área devem concordar em qual serviço melhor representa a área.

Altas correlações de Spearman indicam que os valores obtidos por dois serviços provavelmente mantêm a ordem relativa entre si e, assim, em avaliações competitivas, nas quais um pesquisador é avaliado em relação a outros, o uso de um ou outro serviço é *provavelmente* equivalente. Obviamente, um pesquisador pode ser injustiçado por um dos serviços (a não ser que a correlação de Spearman seja exatamente 1.0) e, portanto, ainda assim, cuidado em utilizar um ou outro serviço deve ser tomado nestas situações. Um baixo valor de Spearman indicaria que nem a ordem de avaliações é preservada quando se usa outro serviço. Neste caso, como no caso da correlação de *Pearson*, pesquisadores devem decidir qual dos serviços melhor representa a área.

Deve-se acrescentar que esta pesquisa apenas analisou a correlação entre as duas medidas (número total de citações e índice h), entre os vários serviços bibliométricos. Uma questão diferente e muito relevante para a avaliação de cientistas é a cobertura de cada um dos serviços para cada uma das subáreas discutidas, no que diz respeito a publicações e citações. Quanto à cobertura no número de publicações, tanto *WoS* quanto *Scopus*, explicitamente, decidiram não incluir alguns/vários veículos de publicação científica. A política de inclusão da *Scopus* é menos clara, embora a empresa diga que só inclua, na indexação, veículos de publicação que passaram por algum critério de qualidade. A política de inclusão da *WoS* é mais clara e conhecida. A empresa monitora vários veículos (quantos não é claro), mas apenas aqueles que receberam um número mínimo de citações, nos últimos dois anos (o famoso índice de impacto), são incluídos no *Web of Science*. Finalmente, *Scholar* indexa todos os documentos disponíveis na *Web*, inclusive documentos não publicados em veículos científicos, como monografias, teses e dissertações, relatórios, etc. Para algumas áreas científicas, estas limitações na cobertura das publicações pode ser o maior problema em usar estes serviços para a avaliação de cientistas.

Finalmente, se um serviço tem sérias limitações na cobertura dos veículos de publicações de uma área é muito provável que ele também tenha limitações quanto à cobertura das citações que o cientista recebe – se a maioria das publicações de uma área não está contemplada em um serviço bibliométrico, então, provavelmente, a maioria das citações que um cientista recebe não estará listada naquele serviço.

As limitações de cobertura de publicações e citações não necessariamente invalidam o uso do serviço bibliométrico na avaliação de cientistas – algumas áreas podem ter um fluxo de produção científica que inclui publicações em veículos não indexados, mas as “melhores” pesquisas ou os resultados “finais” da pesquisa acabam sendo publicadas em veículos indexados. Nestes casos, os serviços bibliométricos acabam indexando o “melhor” da pesquisa do cientista e estas medidas podem fazer sentido para avaliá-lo. No entanto, para áreas nas quais nem as “melhores” pesquisas são indexadas nestes serviços, seja porque elas são publicadas em veículos não indexados (por exemplo, revistas em

português ou conferências) ou porque elas não são publicadas em revistas, mas, sim, como livros, o uso destes serviços na avaliação do pesquisador pode fazer pouco sentido.

## Referências

BAR-ILAN, J. Which h-index? A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, v. 74, n. 2, p. 257-271, 2008.

FRANCESCHET, M. A comparison of bibliometric indicators for computer science scholars and journals on Web of Science and Google Scholar. *Scientometrics*, v. 83, n.1, p. 243-258, 2010.

KULKARNI, A. V. *et al.* Comparisons of citations in Web of Science, Scopus, and Google Scholar for articles. *JAMA*, v. 302, n. 10, p. 1092-1096, 2009.

MEHO, L. I.; YANG, K. Impact of Data sources on citation counts and rankings of LIS Faculty: Web of Science Versus Scopus and Google Scholar. *Journal of the American Society for Information Science and Technology*, v. 58, n. 13, p. 2105-2125, 2007.

MEHO, L. I.; ROGERS, Y. Citation counting, citation ranking, and h-index of

human-computer interaction researchers: a comparison of Scopus and Web of Science. *Journal of the American Society for Information Science and Technology*, v. 59, n. 11, p. 1711-1726, 2008.