

EIXO TEMÁTICO

INDICADORES BIBLIOMÉTRICOS/ CIENTOMÉTRICOS

MODALIDADE: COMUNICAÇÃO ORAL



COMPARANDO O ÍNDICE DE HIRSCH (H) A PARTIR DAS BASES: SCOPUS, WEB OF SCIENCE (WOS) E GOOGLE ACADÊMICO (GA)

Elismar Vicente dos Reis

Doutorando em EMI, Programa de Pós-Graduação em Ciência da Informação
Universidade Estadual Paulista (Unesp)
E-mail: elismarreis@bol.com.br

Ely F. Tannuri de Oliveira

Livre Docente em EMI, Programa de Pós-Graduação em Ciência da Informação
Universidade Estadual Paulista (Unesp)
E-mail: etannuri@gmail.com

INTRODUÇÃO

Com o crescimento da ciência e tecnologia tornou-se necessária a elaboração e adoção de parâmetros que meçam a produção científica nas diversas áreas do conhecimento. Indicadores bibliométricos servem para avaliar o rendimento de comunidades científicas, a eficiência de programas de pós-graduação, a influência da pesquisa dentro da área ou o progresso de setores da nação, dentre os quais podemos destacar os indicadores de produção, indicadores de citação e cocitação, indicadores de ligação e indicadores de influência como o Fator de Impacto (FI) e o índice h.

O índice h foi proposto por Hirsch (2005) que entendia que o registro das atividades de um pesquisador possui informações importantes, tais como, o número de artigos publicados ao longo dos anos, o número de citações por artigo, os periódicos em que publicou, seu FI e outros indicadores. Essa grande quantidade de informações é medida com diferentes critérios por diferentes pessoas, que deverão valer-se desses resultados em diferentes situações, como seleção de pessoal, verificação de desempenho docente, na concessão de bolsas de pesquisas em universidades, e em toda e qualquer situação que exija avaliação. Nesse contexto, e a fim de explicitar a produção científica de um autor por meio de um índice, é que Hirsch propôs o novo indicador.

No Brasil, pesquisadores com Bolsa de Produtividade em Pesquisa (PQ) vinculados ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) da área de Ciência da Informação (CI), notadamente os de nível 1, são os mais profícuos e representam a elite – os mais produtivos – desse domínio de conhecimento. Possuem significativa

produção científica, atuam em diversas universidades como docentes, pesquisadores e orientadores nos programas de pós-graduação, e constroem o conhecimento novo na área de CI.

Diante das proposições acima, o objetivo deste trabalho é verificar o índice h dos bolsistas PQ1 (A, B, C e D) na área de CI presentes na lista do CNPq em janeiro de 2020, em três bases de dados, a Scopus, a Web of Science (WoS) e o Google Acadêmico (GA), para identificar e analisar as variações do referido índice nas diferentes bases.

Esta pesquisa se justifica por permitir comparar os índices em três diferentes bases e relacionar o número de pesquisas e citações, de forma a compreender melhor a relação entre estas duas variáveis – número de pesquisas e citações. O problema-foco de pesquisa, assim se coloca: o índice h de um grupo de pesquisadores de elite da CI se altera em diferentes bases de dados?

BOLSISTAS PRODUTIVIDADE PQ1 EM CI E ÍNDICE H

Para Wainer e Vieira (2013), o CNPq é um dos mais relevantes órgãos de fomento a pesquisa científica no Brasil e a bolsa de produtividade em pesquisa é uma das modalidades de financiamento adotada por este órgão, disponível para pesquisadores advindos de diversas áreas, desde que satisfaçam as condições exigidas pelo órgão financiador. O diferencial dessa bolsa está justamente por qualificar o pesquisador e não apenas a possibilidade e pertinência da submissão de um projeto aos objetivos da unidade financeira. Por esta razão, no país, a bolsa PQ também é considerada significativa, porque indica o desempenho do autor em uma comunidade científica.

As atribuições das bolsas PQ são feitas por classes distintas e em ordem crescente: 2, 1D, 1C, 1B, 1A, e os quatro níveis finais são conhecidos como “níveis 1”, neste caso são chamados “PQ1”. Os PQ1 normalmente possuem alguns privilégios em relação aos “PQ2”, além de poderem participar dos Comitês Assessores (CAs) das áreas. Vários editais de financiamento determinam como requisito que o responsável seja um pesquisador nível 1. Para o CNPq, um pesquisador só pode solicitar uma bolsa de nível 2, depois de três anos de conclusão do doutorado, e, para o nível 1, são necessários ao menos oito anos com o título de doutor. (SILVA, 2018; WAINER; VIEIRA, 2013)

Quanto ao índice h, objetivando facilitar a qualificação de desempenho dos cientistas, Hirsch, em 2005, propôs um novo e único indicador cientométrico, o “índice h”. Para Mugnaini, Packer e Meneghini (2008), o novo índice evitaria alguns problemas encontrados no cálculo da quantidade total de citações, utilizados como referência para avaliar o desempenho da comunidade acadêmica. Para Schreiber (2008, p. 187) “o índice h é definido como o maior número de artigos de um cientista que recebeu ‘h’ ou mais citações. Assim, incorpora a quantidade da publicação e a qualidade da citação em um único número [...], é discutida como uma medida conveniente nas avaliações”. É considerado

um indicativo de mérito do pesquisador ou do impacto de sua pesquisa, retratando sua relevância na área.

De maneira simplificada, demonstra-se na Tabela 1, que o índice h de um autor que tem em seu rol de publicações sete artigos, são dispostos na tabela por ordem decrescente em relação ao número de citações. Busca-se a posição do artigo cuja posição de ordem é igual ao número de citações: o índice h do autor é 5.

Tabela 1 - Como obter o índice h de um pesquisador

Posição do artigo	No de Citações
1	15
2	12
3	10
4	7
5	5
6	4
7	4

Fonte: elaborado pelos autores.

Também demonstra que os artigos de ordem maior que h (posição 6 e 7) não possuem mais que 5 citações, ou pelo menos h citações cada. (SILVA, 2018) Define-se que “um cientista tem o índice h, se h de seus trabalhos têm pelo menos h citações cada, e os demais têm menos de h citações cada”. (ROUSSEAU, 2006, p. 1) Ainda segundo o autor, a Tabela 1 aponta que os artigos 1, 2, 3, 4 e 5 são considerados o núcleo de Hirsch, os quais receberam número de citações iguais ou maiores que o índice h, e são consideradas as publicações científicas de maior impacto daquele pesquisador, de forma cumulativa.

METODOLOGIA

Tomou-se como referência o resultado divulgado pelo CNPq no dia 14 de janeiro de 2020, que continha a listagem atualizada dos bolsistas PQ em CI. Essa listagem foi transferida para um arquivo de planilha eletrônica, separando-se apenas os PQ1 e eliminando os nomes repetidos, já que a tabela inicial apresentava também as bolsas que se encerrariam em fevereiro de 2020. Em seguida, realizou-se o acesso às bases de dados Scopus, WoS e GA para encontrar os índices h dos pesquisadores PQ1 da referida lista. Inicialmente a busca foi feita no GA e, algumas vezes, além do nome completo, procurava-se a instituição de afiliação para se evitar a possibilidade de encontrar nomes iguais e instituições diferentes. Tanto o GA quanto as outras bases não fazem distinção entre letras maiúsculas ou minúsculas e foram considerados todos os documentos encontrados nas plataformas, não havendo filtros para “tipo de documentos” em nenhuma delas.

Na base WoS fez-se as buscas pelas “pesquisas avançadas”, modalidade de busca em que se pode definir os “rótulos” de pesquisa, e, se necessário, o uso de “operadores booleanos”. Na WoS os nomes dos autores devem ser redigidos iniciando-se pelo último nome e as iniciais dos primeiros nomes. No campo de busca era necessário digitar o rótulo “au”(autor), o sinal de igual (=) e o nome, conforme o exemplo: au=otlet p (busca para Paul Otlet). Uma nova página era aberta mostrando o link “resultados” com a relação das publicações, onde é possível conferir os artigos do autor de interesse. Ao selecionar as publicações deve-se adicioná-las na “lista marcada”, e após carregar este link é preciso acessar a opção “criar relatório de citações”, na qual se apresenta o índice h do referido autor.

A Scopus foi visitada por último, na qual era preciso marcar a opção “Authors”. Nessa base a busca se dá pelo último nome do autor, primeira(s) letra(s) do(s) nome(s) inicial(is) com ponto, e, se preciso, a sua afiliação, que ajuda a refinar a pesquisa. Ao pressionar o botão “pesquisar” é apresentado o índice h do nome procurado. Os dados gerados coletados nas bases foram armazenados na planilha que continha somente os respectivos bolsistas PQ1.

RESULTADOS

A Tabela 2 mostra os pesquisadores PQ1 divulgados pelo CNPq em janeiro de 2020, a instituição de afiliação e os índices h obtidos nas bases GA, WoS e Scopus. Pode-se observar certa discrepância entre os índices do GA em relação às outras duas bases, e na grande maioria uma paridade entre WoS e Scopus, corroborando os trabalhos de Bar-Ilan (2008) e de Silva e Grácio (2017), que já apontavam essa diferença.

No GA, 12 pesquisadores tiveram seus perfis e índices localizados e 6 não foram encontrados. Na base WoS, 15 pesquisadores tiveram o índice h apresentado e 3 nomes não foram encontrados na base, sendo que, dos 15 encontrados, 2 deles tem índice 0 (zero), significando que o nome do pesquisador foi encontrado, mas o índice é zero; na base Scopus, todos os pesquisadores tiveram o nome encontrado e todos possuem índice h na plataforma, entretanto, desses 18, 2 apresentam o (zero) como índice, o que significa que o pesquisador tem registro de produção na base, porém não tem citações suficientes para o índice.

Apresenta-se a Tabela 2 com os índices h dos pesquisadores PQ1 em CI nas três bases. Os pesquisadores Maria de Fátima Gonçalves Moreira Tálamo e Marilda Lopes Ginez de Lara não foram encontradas (NE) nas bases do GA e da WoS. Emir José Suaiden, Gilda Olinto de Oliveira, José Maria Jardim e Nair Yumiko Kobashi não foram encontrados (NE) no GA. A pesquisadora Maria Nelida Gonzalez de Gomez não foi encontrada (NE) na WoS. O maior índice h encontrado no GA foi 26 das pesquisadoras Marta Lígia

Pomim Valentim (UNESP) e Sarita Albagli (IBICT), e o menor 6, uma variação de 20 pontos, já as variações na WoS e Scopus não passam de 4 pontos (de 0 a 4).

Tabela 2 - Índice h dos pesquisadores PQ1 em CI nas três bases de dados

Ordem	Nome	Nível	Instituição*	GA	WoS	Scopus
1	Carlos Henrique Marcondes de Almeida	PQ-1D	UFF	18	4	5
2	Emir José Suaiden	PQ-1B	UnB	NE	2	2
3	Gilda Olinto de Oliveira	PQ-1C	IBICT	NE	1	2
4	Guilherme Ataíde Dias	PQ-1D	UFPB	11	1	2
5	Icleia Thiesen	PQ-1D	UNIRIO	6	0	0
6	José Augusto Chaves Guimarães	PQ-1A	UNESP	22	4	4
7	José Maria Jardim	PQ-1D	UNIRIO	NE	1	1
8	Leilah Santiago Bufrem	PQ-1D	UFPE	14	3	2
9	Lena Vania Ribeiro Pinheiro	PQ-1B	IBICT	19	1	1
10	Maria de Fátima Gonçalves Moreira Tálamo	PQ-1D	USP	NE	NE	0
11	Maria Nelida Gonzalez de Gomez	PQ-1A	UFF	22	NE	2
12	Mariângela Spotti Lopes Fujita	PQ-1B	UNESP	20	3	3
13	Marilda Lopes Ginez de Lara	PQ-1B	USP	NE	NE	2
14	Marta Lígia Pomim Valentim	PQ-1D	UNESP	26	3	3
15	Nair Yumiko Kobashi	PQ-1D	USP	NE	1	1
16	Raimundo Nonato Macedo dos Santos	PQ-1D	UFPE	24	0	3
17	Regina Maria Marteleto	PQ-1A	IBICT	21	3	3
18	Sarita Albagli	PQ-1C	IBICT	26	2	2

* Universidade Federal Fluminense (UFF) – Rio de Janeiro-RJ; Universidade de Brasília (UnB); Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) – Rio de Janeiro-RJ; Universidade Federal da Paraíba (UFPB) – João Pessoa-PB; Universidade Federal do Estado do Rio de Janeiro (UNIRIO); Universidade Estadual Paulista Júlio de Mesquita Filho (UNESP) – Marília-SP; Universidade Federal de Pernambuco (UFPE) – Recife-PE; Universidade de São Paulo (USP).

NE – “não encontrou” resultado na base para o pesquisador.

Fonte: elaborado pelos autores.

Nas bases WoS e Scopus, na maioria dos casos, quando os índices h dos pesquisadores não são iguais, estão muito próximos. Infere-se aqui que Maria de Fátima Gonçalves Moreira Tálamo não foi encontrada na WoS e obteve índice 0 na Scopus, e que os pesquisadores que foram encontrados nas duas bases e apresentam disparidade entre os índices são apenas 4. A partir do exposto, pode-se destacar certa proximidade entre os índices da Scopus e da WoS na maioria dos casos, possivelmente pela indexação dos mesmos periódicos ou quase todos.

Os resultados obtidos na Tabela 2 demonstram certa consonância com um trabalho mais abrangente em relação ao tema aqui discutido. Na pesquisa realizada por Martín-

-Martín e demais autores (2018) constatou-se que no geral, o GA encontra 94% de todas as citações de um autor, enquanto o WoS encontra 52% e Scopus 60%. Ainda descobriram que o GA encontrou 95% das citações que o WoS recuperou, e 92% das citações obtidas pela Scopus, que por sua vez encontrou 93% das citações presentes na WoS. No presente estudo, de encontro ao que foi demonstrado por Martín-Martín e demais autores (2018), observou-se que a maior parte das citações encontradas somente pelo GA não são de periódicos, os mais frequentes são teses e dissertações, mas incluem livros, documentos de conferências e até materiais ainda não publicados. Essa diversidade de materiais utilizados pelo GA para calcular o índice h são indícios da superioridade dos valores encontrados nessa base em relação à WoS e Scopus.

CONSIDERAÇÕES FINAIS

Como conclusões, destaca-se a significativa diferença do índice h encontrada nas três bases, com bastante alinhamento entre os valores da WoS e Scopus, enfatizando que o índice h reflete toda amplitude da vida produtiva e acumulada de um pesquisador. Autores com maior tempo de pesquisa e igual intensidade de produtividade, terão possibilidade de apresentar maior índice h que outros em condições diferentes. Salienta-se que a maioria das citações encontradas apenas pelo GA, vem de fontes não pertencentes a periódicos.

Há que se ter certo cuidado na superioridade exclusivamente numérica do GA e até da ausência do pesquisador na base, pois nesta plataforma é necessário que o pesquisador crie seu perfil para que se encontre o índice h. No processo de criação do perfil são apresentadas todas as publicações que a plataforma identifica com alguma relação com o dono do perfil naquele momento, criando grupos das publicações os quais deverão ser selecionados, sendo necessário ainda filtrar apenas os trabalhos que sejam verdadeiramente seus. O índice h do GA depende muito da atividade séria e ética do próprio pesquisador em relação às publicações, é verdadeiramente um trabalho manual, pois novas notificações virão para serem confirmadas, porém, os PQ1 são considerados pesquisadores de alto nível intelectual e bastante qualificados, o que já sugere tal conduta.

Desse modo, as análises dos resultados pedem prudência no uso apenas do GA para verificação do índice h, já que a plataforma precisa estar com o perfil do pesquisador criado e conferido, visto que poderá entrar no cômputo até materiais não publicados ou de outros autores, ou mesmo atividades acadêmicas diversas, além dos artigos publicados.

Destaque-se aqui, uma das dificuldades encontradas em relação a busca dos pesquisadores, nas três bases: a falta de normalização dos nomes. Em um primeiro momento, com a metodologia utilizada nesta pesquisa, deu-se como NE autores de renomada projeto. Em uma busca mais cuidadosa, verificou-se a ausência de utilização do último nome nas suas publicações, ficando assim incompleto e diferente da listagem apresentada pelo

CNPq, casos de Carlos Henrique Marcondes de Almeida (não utiliza “de Almeida”) e Gilda Olinto de Oliveira (não utiliza “de Oliveira”). Indicadores bibliométricos são importantes ferramentas de avaliação de desempenho e impacto acadêmico, tal qual o índice h, mas dada as diferenças encontradas na pesquisa em relação às três bases, não há como eleger a melhor base, ficando esta questão dependente dos objetivos, contexto da pesquisa e publicações que deverão ser levadas em consideração.

AGRADECIMENTOS

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

REFERÊNCIAS

- BAR-ILAN, J. Which h-index?: a comparison of wos, scopus and google scholar. *Scientometrics*, Dordrecht, v. 74, n. 2, p. 257-271, 2008.
- HIRSCH, J. E. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, D. C., v. 102, n. 46, p. 16569-16572, 2005.
- MARTÍN-MARTÍN, A.; ORDUNA-MALEA, E.; THELWALL, M. et al. Google Scholar, Web of Science, and Scopus: a systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, Amsterdam, v. 12, n. 4, p. 1160-1177, 2018. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S1751157718303249>. Acesso em: 16 jan. 2020.
- MUGNAINI, R.; PACKER, A. L.; MENEGHINI, R. Comparison of scientists of the Brazilian Academy of Sciences and of the National Academy of Sciences of the USA on the basis of the h-index. *Brazilian Journal of Medical and Biological Research*, Ribeirão Preto, v. 41, n. 4, p. 258-262, 2008.
- ROUSSEAU, R. New developments related to the Hirsch index. *e-prints in library & information science*, [s. l.], p. 1-5, 2006. Disponível em: http://eprints.rclis.org/7616/1/Hirsch_new_developments.pdf. Acesso em: 17 jan. 2020.
- SCHREIBER, M. The influence of self-citation corrections on Egghe's g index. *Scientometrics*, Dordrecht, v. 76, n. 1, p. 187-200, 2008.
- SILVA, D. D.; GRÁCIO, M. C. C. Índice h de Hirsch: análise comparativa entre as bases de dados scopus, web of science e google acadêmico. *Em Questão*, Porto Alegre, v. 23, p. 196-212, 2017. Disponível em: <https://www.redalyc.org/pdf/4656/465650499011.pdf>. Acesso em: 17 jan. 2020.
- SILVA, D. D. *Medida de dispersão para o índice h: proposta de um indicador do tipo h de Hirsch*. 2018. Tese (Doutorado em Ciência da Informação) - Programa de Pós Graduação em Ciência da Informação, Universidade Estadual Paulista, Marília, 2018.
- WAINER, J.; VIEIRA, P. Avaliação de bolsas de produtividade em pesquisa do CNPq e medidas bibliométricas: correlações para todas as grandes áreas. *Perspectivas em Ciência da Informação*, Belo Horizonte, v. 18, n. 2, p. 60-78, 2013.

