

INDICADORES BIBLIOMÉTRICOS PARA A AVALIAÇÃO DA REPRESENTATIVIDADE DO ÍNDICE h COMO DESCRITOR DO DESEMPENHO CIENTÍFICO: UM ESTUDO NA ÁREA DA MATEMÁTICA NO BRASIL

Deise Deolindo Silva

Docente da Fatec-Garça e Fatec-Pompéia
Centro Estadual de Educação Tecnológica Paula Souza
deise.deolindo@fatec.sp.gov.br

Maria Cláudia Cabrini Grácio

Docente do Programa de Pós-Graduação em Ciência da Informação
Universidade Estadual Paulista, Marília, SP, Brasil
cabrini@marilia.unesp.br

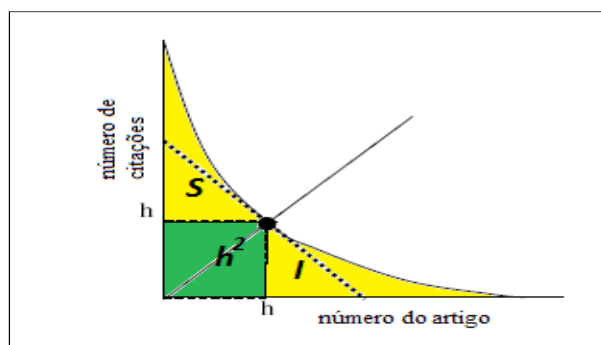
INTRODUÇÃO

Ao longo dos anos, a análise de citação vem se consolidando como um procedimento metodológico para mensurar e caracterizar o impacto das publicações científicas e sendo usada, constantemente, como aporte para a avaliação científica e tomadas de decisão em agências de fomento à pesquisa. Hirsch (2005) propôs o índice h que, desde então, tem sido muito utilizado na mensuração do desempenho e impacto científico dos pesquisadores. É formulado da seguinte forma: se h dos N artigos de um pesquisador têm pelo menos h citações cada e os outros $(N - h)$ artigos não têm mais que h citações cada, então o pesquisador tem índice h com valor igual a h . (HIRSCH, 2005)

O índice h identifica o núcleo mais significativo das publicações científicas de um pesquisador em termos daqueles trabalhos que receberam mais citações, denominado núcleo Hirsch ou h -core, termo inserido por Rousseau (2006). A Figura 1 ilustra a representatividade do índice h de um pesquisador em relação ao total de citações recebidas por seus artigos.

O índice h representa h^2 do Impacto Total (IT) da obra do pesquisador (Figura 1), em que $IT = S + h^2 + I$. Assim, o índice h não inclui o total de citações nas caudas da distribuição de citações: cauda superior (S), composta por artigos mais citados, e cauda inferior (I) pelas publicações menos citadas.

Figura 1 - Representação gráfica do índice h



Fonte: Silva (2018) adaptado de Hirsch (2005).

Portanto, o índice h representa tão melhor o desempenho científico de um pesquisador quanto menor forem as áreas S e I . Beneficia os pesquisadores que têm as áreas (“caudas”) S e I menores, uma vez que à medida que estas aumentam, a área h^2 representada por esse índice é menos significativa em relação à área total (impacto total do conjunto de artigos), subestimando o desempenho real do pesquisador. Por outro lado, prejudica aqueles com uma proporção das caudas S/I elevada. Desse modo, considerando que o índice h não leva em consideração o peso das caudas, pode acontecer de dois pesquisadores terem o mesmo valor para h , um com $S > I$ e outro com $S < I$. (DORTA-GONZALEZ; DORTA-GONZALEZ, 2010)

Em função dessa limitação, diversos estudiosos vêm se dedicando a analisar a significância do índice h como indicador para a representação do impacto científico dos pesquisadores. (BORNMAN; MUTZ; DANIEL et al., 2008; EGGHE, 2010) As fraquezas identificadas no índice h não diminuíram sua popularidade, ao contrário, o uso e a literatura nessa temática estão em plena expansão. (LEYDESDORFF; WOUTERS; BORNMAN, 2016)

Nesse sentido, Silva (2018) propôs um par de indicadores destinados a avaliar a adequação do índice h como representação consistente da descrição do desempenho científico de um pesquisador, contribuindo para o aperfeiçoamento dos processos avaliativos da ciência.

Silva (2018) propôs dois indicadores, denominados dc_i (dispersão das citações no interior do h -core) e dc_o (dispersão das citações fora do h -core), a fim de mensurar e avaliar as caudas superior e inferior da distribuição das citações de um pesquisador. O índice dc_i é definido pela soma das distâncias entre o número de citações recebidas pelos artigos do núcleo Hirsch e o índice h , dividida por h . Desse modo, dc_i representa a distância média entre as citações dos artigos do h -core e h . O índice dc_o é definido como a soma das distâncias entre o número de citações recebidas pelos artigos que estão fora do núcleo Hirsch

e o índice h , dividida por $(N - h)$. Logo, dc_o determina a distância média de citações dos artigos menos citados em relação ao valor de h . Pode, assim, ser considerado um índice que mede a expectativa de crescimento do índice h . (SILVA, 2018)

Esses dois novos indicadores buscam identificar e caracterizar a representatividade do índice h , ao evidenciar a intensidade do impacto não representada pelo valor de h . Contribuem, assim, para se discernir, de forma mais precisa, as diferenças entre o impacto científico acumulado de pesquisadores de um mesmo domínio científico. (SILVA, 2018)

Em função do exposto, esta pesquisa tem por objetivo avaliar a contribuição dos indicadores dc_i e dc_o para a avaliação do impacto científico de pesquisadores com bolsa produtividade em pesquisa nível 1 (PQ1) do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) da área da Matemática, na complementação e aprimoramento da descrição dos seus desempenhos científicos, ao possibilitar distinguir as diferenças entre as distribuições de citação.

Esta pesquisa busca contribuir para o aprofundamento da análise do índice h , assim como para a sua compreensão como descritor da trajetória acadêmica de um pesquisador. Além disso, contribui para o refinamento da metodologia dos indicadores bibliométricos e propicia aportes que tornam possíveis uma avaliação mais precisa, a fim de subsidiar as tomadas de decisões de forma mais consistente.

PROCEDIMENTOS METODOLÓGICOS

Optou-se por analisar o conjunto de pesquisadores PQ1 em Matemática pela consolidação desta área em âmbito nacional, com publicações e pesquisadores de referência e impacto científico nacional e internacional. Inicialmente, em janeiro de 2018, por meio da busca no portal do CNPq, identificaram-se os pesquisadores PQ1, nos níveis A, B, C e D, ativos da área da Matemática no Brasil, correspondendo a um total de 116 pesquisadores. A seguir, para cada um dos pesquisadores, recuperaram-se, na base de dados Scopus, seus seguintes indicadores: total de artigos publicados, índice h e, para cada artigo, ano de publicação e total de citações recebidas. Os dados foram organizados e analisados no *software* Excel. Na sequência, para cada pesquisador, calcularam-se os índices dc_i e dc_o .

O índice dc_i foi calculado pela fórmula $dc_i = \frac{\sum_{j=1}^h (cit_j - h)}{h}$, em que, cit_j : número de citações do j -ésimo artigo, para $1 \leq j \leq h$; h : índice h . Um valor baixo para dc_i indica que, em média, há poucas citações excedendo h . Neste caso, o índice h representa adequadamente o desempenho científico do pesquisador. Por outro lado, um valor alto para dc_i significa que o pesquisador tem artigos influentes que receberam muitas citações, e, portanto, o índice h está subestimando o seu desempenho científico. O índice dc_o foi calculado pela

fórmula $dc_o = \frac{\sum_{j=h+1}^{Np} (h - cit_j)}{N - h}$, onde cit_j : número de citações do j -ésimo artigo, para $h + 1 \leq j \leq N$; N : número de publicações, h : índice h . Um valor baixo para dc_o significa que, em média, os artigos que não compõem o h -core têm um número de citações próximo ao valor de h e, por consequência, indica o potencial de se incrementar o índice h . (SILVA, 2018)

ANÁLISE E DISCUSSÃO DOS RESULTADOS

A Tabela 1 apresenta as estatísticas descritivas dos indicadores, em que se observa que, em média, os pesquisadores têm índice h de aproximadamente 11, ou seja, têm pelo menos 11 artigos com pelo menos 11 citações cada. O alto valor para o CV (51%) evidencia uma significativa dispersão do valor do índice h entre os pesquisadores PQ1 analisados.

Tabela 1 - Estatísticas descritivas relativas ao total (#) de artigos, tempo desde a primeira publicação na Scopus (t), índice h , dc_i e dc_o dos 116 pesquisadores PQ1 da Matemática

Estatísticas	# artigos	t	Índice h	dc_i	dc_o
Média	47,8	24,7	11,1	17,0	7,7
Mediana	37,5	24	9,5	11,1	6,8
Mínimo	8	9	4	0,4	2,1
Máximo	149	53	32	240,1	24,8
CV	62%	36%	51%	140%	53%

Fonte: elaborado pelo autor.

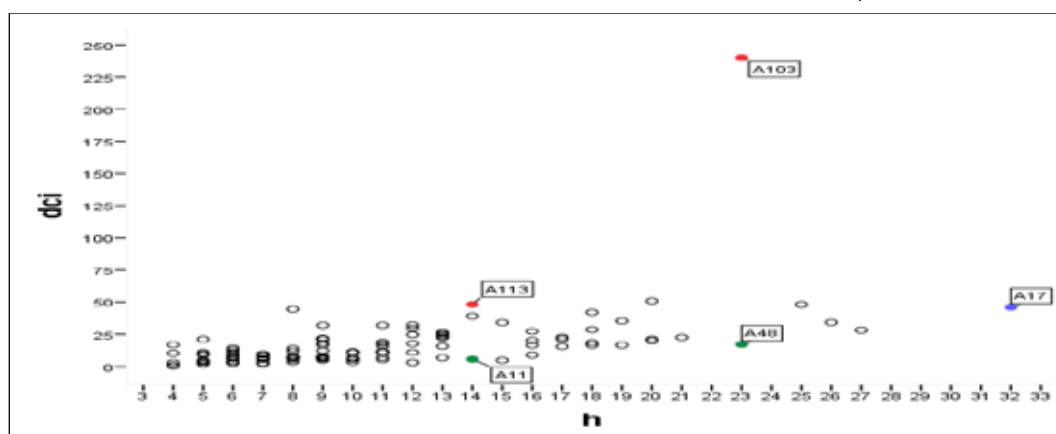
Em média, os pesquisadores têm dc_i baixo (17,0), evidenciando que, a tendência geral entre esses pesquisadores é de haver poucas citações excedendo h . O valor mediano para este indicador é significativamente menor (35%) que o valor médio, sugerindo uma assimetria na distribuição do indicador, ou seja, há alguns pesquisadores *outliers* (entre eles, aquele com $dc_i=240$) em relação à tendência central de desempenho do dc_i . Ademais, corroborando este fato, observa-se um alto valor para o CV do indicador. Têm-se, assim, que em geral, o índice h representa adequadamente o desempenho científico do pesquisador, mas há pesquisadores *outliers* a esta tendência, ou seja, para os quais o índice h está subestimando o seu desempenho científico, por terem artigos influentes que receberam muitas citações.

Em relação ao indicador dc_o , observou-se uma média baixa (7,7), indicando uma baixa variabilidade na cauda inferior da distribuição de citações e que, em geral, para os pesquisadores analisados, os artigos que não compõem o h -core têm um número de citações próxima ao valor de h e, por consequência, potencial para incrementar seus respectivos índices h . Esta tendência pode ser considerada ratificada pelo fato de o valor mediano do indicador ser próximo do valor médio, embora CV represente uma dispersão significativa.

Para os 116 pesquisadores PQ1 analisados da área da Matemática, as Figuras 2 e 3 propiciam uma melhor visualização e compreensão da contribuição dos indicadores dc_i e dc_o , respectivamente, para a descrição mais precisa do desempenho e impacto científico dos pesquisadores. Alguns autores foram destacados a fim de exemplificar essa contribuição.

Observa-se (Figura 2) que o autor A17 possui o maior valor de índice h , mas não possui o maior dc_i ($=39,1$), evidenciando que para A17 o índice h representa mais adequadamente seu impacto que a representatividade de h para o pesquisador A103, que possui o maior valor de dc_i ($=240,1$). Para A103, o índice h subestima seu impacto científico.

Figura 2 - Dispersão entre o índice h (eixo horizontal) e o indicador dc_i (eixo vertical)



Fonte: elaborado pelo autor.

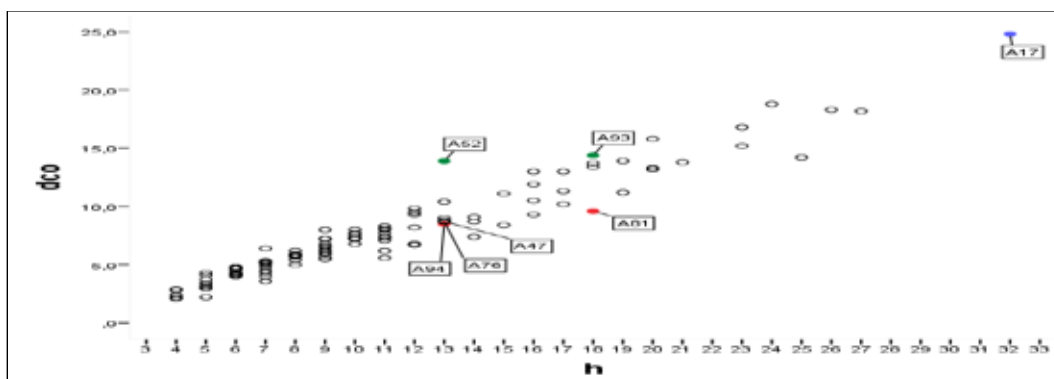
Ainda, os autores A11 e A113 apresentam índice h igual (14). No entanto, A11 apresentou $dc_i = 5,6$ e A113 apresentou $dc_i = 48,2$, evidenciando que o valor de h ($=14$) é mais representativo do impacto científico de A11 do que de A113. O pesquisador A113 tem mais artigos altamente citados não representados pelo índice h .

Similarmente, os pesquisadores A48 e A103 têm índice $h = 23$, diferindo, todavia, em relação ao dc_i , ou seja, em relação à distribuição das citações dos artigos que compõem o h -core, com $dc_i = 17,1$ e $240,1$, respectivamente, evidenciando que o valor de h é muito mais representativo do impacto científico de A48 do que de A103. Para A103, houve muitos artigos altamente citados que levaram a uma dispersão altíssima no h -core e à principal sub-representação do impacto científico pelo índice h no conjunto de pesquisadores analisados.

A partir da Figura 3, é possível observar que há uma tendência de quanto maior o índice h de um pesquisador maior tende a ser seu valor de dc_o . Considerando que dc_o mede a dispersão na cauda inferior da distribuição das citações do pesquisador e representa, então, a potencialidade de se incrementar o valor de h , essa tendência de associação

evidencia que quanto maior o índice h , mais difícil o pesquisador aumentar seu valor. Assim, dc_o é sensível à questão apresentada por Grupo Scimago (2006), que afirma que quanto maior o índice h de um pesquisador, maior dificuldade este terá para aumentá-lo, exigindo um maior esforço (de característica não linear). O índice dc_o pode, então, ser considerado um indicador que representa de forma objetiva e quantificável esta característica do índice h .

Figura 3 - Dispersão entre índice h (eixo horizontal) e o indicador dc_o (eixo vertical)



Fonte: elaborado pelo autor.

Observa-se (Figura 3) que o pesquisador A17 (maior índice h) apresenta o maior valor para dc_o ($=24,8$), o que sugere que encontrará mais dificuldade de incrementar o valor de h que os demais pesquisadores analisados. Comparativamente, os pesquisadores A47, A52, A76 e A94 têm índices $h = 13$. No entanto, A52 tem dc_o (13,9) bem maior que o de A47, A76 e A94, com valores igual a 8,7; 8,7; 8,5, respectivamente, o que indica que estes últimos têm mais chance de incrementar seus índices h que A52. Similarmente, os autores A81 e A93 têm índices $h = 18$ e dc_o igual a 2,9 e 9,6, respectivamente, indicando que para A81, o potencial de incrementar o índice h é pelo menos três vezes maior que para A93.

CONSIDERAÇÕES FINAIS

Por meio dos resultados, identificou-se a contribuição dos dois novos índices, dc_i e dc_o , complementares ao índice h , na mensuração da dispersão do h -core e na cauda inferior da distribuição das citações de pesquisadores PQ1 da área da Matemática, para o refinamento da descrição dos seus desempenhos científicos. Os indicadores dc_i e dc_o mostraram-se capazes de caracterizar adequadamente a dispersão do desempenho dos pesquisadores em torno do seu índice h . Nesse sentido, os índices de dispersão permitiram distinguir os pesquisadores que, embora com valores iguais para os índices h , apresentavam desempenhos distintos.

Conclui-se que os indicadores dc_i e dc_o podem contribuir para avaliar a representatividade do índice h como indicador do impacto científico acumulado de um pesquisador. Sugere-se, desse modo, se descrever o impacto científico de um autor por meio da terna (h, dc_i, dc_o) . Desse modo, quando necessárias avaliações do desempenho científico, os indicadores dc_i e dc_o podem contribuir para as tomadas de decisão, em especial, como critério de desempate em situações de similaridade quanto ao índice h dos avaliados.

REFERÊNCIAS

- BORNMANN, L.; MUTZ, R.; DANIEL, H.-D. Are there better indices for evaluation purposes than the h index? A comparison of nine different variants of the h index using data from biomedicine. *Journal of the American Society for Information Science and Technology*, v. 59, n. 5, p. 830-837, 2008.
- DORTA-GONZÁLEZ, P.; DORTA-GONZÁLEZ, M. I. Indicador bibliométrico basado en el índice h . *Revista Española de Documentación Científica*, p. 225-245, 2010.
- EGGHE, L. The Hirsch index and related impact measures. *Annual review of information science and technology*, v. 44, p. 65-114, 2010.
- GRUPO SCIMAGO. El índice h de Hirsch: aportaciones a un debate. *El profesional de la información*, v. 15, n. 4, 2006.
- 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*, v. 102, n. 46, p. 16569-16572, 2005.
- LEYDESDORFF, L.; WOUTERS, P.; BORNMANN, L. Professional and citizen bibliometrics: complementarities and ambivalences in the development and use of indicators – a state-of-the-art report. *Scientometrics*, v. 109, p. 2129-2150, 2016.
- ROUSSEAU, R. New developments related to the Hirsch index. *e-prints in library & information science*. 2006.
- SILVA, D. D. *Medida de dispersão para o índice h* : proposta de um indicador do tipo h de Hirsch. Tese (Doutorado) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Marília, 2018.

