

## ANÁLISES COMPARATIVAS ENTRE INDICADORES BIBLIOMÉTRICOS

**Luana Carneiro Brandão**

Pós Graduação em Engenharia de Produção, Universidade Federal Fluminense  
Rua Passo da Pátria 156, São Domingos, 24210-240, Niterói, RJ  
luanabrandao@id.uff.br

**João Carlos Correia Baptista Soares de Mello**

Departamento de Engenharia de Produção – Universidade Federal Fluminense  
Rua Passo da Pátria, 156, São Domingos, 24210-240, Niterói, RJ, Brazil  
jcsmello@producao.uff.br

### RESUMO

Por sua importância e simplicidade, o índice  $h$ , indicador bibliométrico proposto em 2005, é bastante popular na literatura, havendo centenas de estudos publicados com críticas e propostas de novos índices, os quais são conhecidos como índices tipo  $h$ . O presente artigo estuda alguns desses índices propostos para amenizar a insensibilidade do índice  $h$  a publicações muito citadas, a saber, os índices  $g$ ,  $A$  e  $R$ . Primeiramente, são realizadas análises teóricas para comparar os valores desses índices. Além disso, estuda-se o comportamento de tais índices, frente a um aumento do número de citações, bem como a sua estabilidade, mediante variações na distribuição de citações. Nesses estudos, são utilizados exemplos numéricos para fins ilustrativos.

**PALAVRAS-CHAVE.** Índice  $h$ , Índices Bibliométricos, Comparações entre índices

**Área principal:** Outras aplicações em PO

### ABSTRACT

The  $h$ -index, proposed in 2005, has become very popular in the literature, due to its importance and simplicity. In fact, many studies propose new indexes, known as  $h$ -type indexes. This work studies a few of these indexes, which decrease the  $h$ -index's disadvantage of being insensitive to highly cited publications, namely the  $g$ ,  $A$  and  $R$  indexes. Firstly, we compare the different values of these indexes, theoretically. Then, we study their reaction to an increase in the number of citation, and their stability, when the distribution of citations varies. We also use numerical examples to illustrate these studies.

**KEYWORDS.**  $h$ -index, Bibliometric indexes, Comparisons between indexes

**Main area:** Other applications in OR

## 1. Introdução

O índice bibliométrico proposto por Hirsch (2005) para medir impacto e relevância da produção científica, denominado índice  $h$ , é um dos indicadores mais populares na área da ciência da informação, provavelmente por sua importância, simplicidade e fácil compreensão (Franceschini e Maisano, 2010), sendo, não somente, muito estudado pela literatura, mas também, bastante utilizado informalmente, em diversas áreas associadas à pesquisa científica (Egghe, 2010).

Dentre as centenas de estudos relativos ao índice  $h$ , muitos apresentam críticas, e.g. Bornmann e Daniel (2007), Braun *et al.* (2006), Harzing e Van der Wal (2008), Norris e Oppenheim (2010), Panaretos e Malesios (2009), e propõem índices alternativos para superar as desvantagens, e.g., Alonso *et al.* (2010), Anderson *et al.* (2008), Batista *et al.* (2006), Cabrerizo *et al.* (2010), Egghe e Rousseau (2008), Schreiber (2008), Tol (2009).

O presente artigo realiza análises comparativas com o índice  $h$  e com alguns desses índices alternativos, chamados índices tipo  $h$ , quais sejam, os índices  $g$  (Egghe, 2006a),  $A$  (Jin, 2006) e  $R$  (Jin *et al.*, 2007). Esses índices buscam reduzir a insensibilidade do índice  $h$  a publicações muito citadas e não consideram outras variáveis além daquelas do índice  $h$ , quais sejam, o número de publicações e de citações de cada pesquisador.

As análises comparativas do presente estudo são divididas em três categorias, quais sejam, valores, efeitos do crescimento de citações e estabilidade. Na primeira categoria, enquadram-se as análises teóricas que definem as relações de valor entre os índices. Na segunda, enquadram-se as análises de como os índices se comportam frente a um crescimento do número de citações. Na terceira categoria, enquadram-se as comparações da estabilidade dos índices frente a alterações na distribuição de citações.

Na próxima seção, é feita uma revisão dos índices bibliométricos estudados. Na terceira seção, constam as comparações de valor dos índices. Na quarta, são comparadas as reações dos índices frente a um crescimento de citações, com base em um exemplo ilustrativo. Na quinta seção, analisa-se a estabilidade dos índices, sendo também fornecido um exemplo numérico. Por fim, a sexta seção apresenta as conclusões.

## 2. Revisão de alguns índices bibliométricos

### 2.1. Índice $h$

Proposto por Hirsch (2005) para suprir a necessidade de avaliar o desempenho de pesquisadores, o índice  $h$  mede, de maneira simples e intuitiva, o impacto e relevância da produção científica. Por definição, um pesquisador apresenta o índice no valor de  $h$ , se o mesmo é autor de  $h$  publicações com, pelo menos,  $h$  citações cada, enquanto as demais publicações apresentam no máximo  $h$  citações. As  $h$  publicações com pelo menos  $h$  citações compõem o chamado núcleo do índice  $h$ .

Muitos trabalhos apontaram desvantagens desse índice, como (a) ser influenciado por autocitações, o que pode ser significativo, em especial para pesquisadores jovens (Schreiber, 2007); (b) não considerar múltiplas coautorias (Schreiber, 2008); (c) não considerar o tipo de publicação, uma vez que artigos de revisão ou em áreas populares costumam ser mais citados que outros tipos de publicação (Castillo *et al.*, 2007); (d) não permitir comparar pesquisadores de áreas diferentes (Antonakis e Lalive, 2008; Batista *et al.*, 2006; Braun *et al.*, 2006); (e) não considerar a idade das publicações, o que também beneficia pesquisadores antigos (BiHui, 2007; BiHui *et al.* 2007; Sidiropoulos *et al.*, 2007; Franceschini e Maisano, 2010); (f) ser insensível a publicações muito citadas (Egghe, 2006a, 2010; Schreiber, 2010b); e (g) apresentar comportamento discrepante quando aplicado a grupos de cientistas (Waltman e Van Eck, 2012; Rubem *et al.*, 2014), embora o índice não tenha sido originalmente proposto para isso.

Algumas dessas críticas levaram a propostas de novos índices tipo  $h$  que consideram outras variáveis, além do número de publicações e citações, e.g., como a idade da publicação, o tempo de carreira do pesquisador, o número de coautores, etc. Assim, os demais índices, que consideram apenas as variáveis originalmente propostas por Hirsch (2005), são ditos clássicos (Egghe, 2010). No presente estudo, são analisados apenas índices clássicos, e que buscam superar a crítica de insensibilidade às publicações muito citadas.

## 2.2. Índice $g$

Uma das variantes mais populares do índice  $h$  (Franceschini e Maisano, 2010), o índice  $g$  é definido como o maior número natural de publicações que juntas receberam ao menos  $g^2$  citações (Egghe, 2006a), as quais formam o chamado núcleo  $g$ . Em outras palavras, o índice equivale ao número de publicações citadas, em média, pelo menos  $g$  vezes (Schreiber, 2010a). Dessa forma, o índice  $g$  valoriza publicações bastante citadas, sendo sempre maior ou igual do que o índice  $h$ , por definição (Egghe, 2010).

Para ilustrar o cálculo do índice  $g$  e compará-lo ao índice  $h$ , a Figura 1 fornece um exemplo arbitrário de um pesquisador com 13 publicações citadas e 154 citações no total. Nesse caso, o índice  $h$  é 7 porque existem 7 publicações com pelo menos 7 citações cada, sendo que as demais publicações são citadas não mais que 7 vezes; enquanto o índice  $g$  é 12, uma vez que existem 12 publicações citadas que receberam pelo menos 144 citações.

rank da publicação	nº de citações	rank <sup>2</sup>	nº acumulado de citações
1	45	1	45
2	32	4	77
3	19	9	96
4	15	16	111
5	12	25	123
6	9	36	132
7	7	49	139
8	5	64	144
9	3	81	147
10	3	100	150
11	2	121	152
12	1	144	153
13	1	169	154

Figura 1: Cálculo dos índices  $h$  e  $g$

Dentre as críticas do índice  $g$ , destaca-se a questão da precisão de dados (Jin *et al*, 2007). Isso porque o mesmo apresenta usualmente núcleos maiores do que o índice  $h$ , dependendo de um conjunto maior de informações precisas para que seja adequadamente calculado. Assim, foi proposto o índice  $A$  para lidar com a insensibilidade do índice  $h$ , sem acrescentar problemas de precisão.

## 2.3. Índice $A$

O índice  $A$  é definido como o número médio de citações das publicações do núcleo  $h$  (Jin, 2006), podendo ser matematicamente representado como em (1), onde  $h$  é o número de publicações no núcleo  $h$ , e  $c_j$  é o número de citações para cada publicação  $j = 1, \dots, h$ . Observa-se que  $A \geq h$  (Jin *et al*, 2007), uma vez que  $h$  é o número mínimo de citações do núcleo (podendo ser, ainda, menor que esse número mínimo), enquanto  $A$  é a média dessas citações.

$$A = \frac{\sum_{j=1}^h c_j}{h} \quad (1)$$

Em certo sentido, trata-se de índice semelhante ao índice  $g$ , uma vez que ambos são o valor médio de citações dentre um conjunto de publicações. Entretanto, esse conjunto de publicações difere, sendo o núcleo do índice  $h$  para um e o núcleo do próprio índice  $g$  para o outro.

Entretanto, o índice  $A$  não aumenta monotonicamente com o aumento das citações. Em outras palavras, é possível que um pesquisador receba novas citações e tenha o seu índice consequentemente reduzido. Isso porque as novas citações podem levar a um aumento do índice  $h$  e, com isso, o seu novo núcleo pode apresentar uma média de citações menor. Essa característica indesejável é consequência da divisão por  $h$  (Jin *et al.*, 2007), como observado em (1). Para lidar com essa questão, foi proposto o índice  $R$ .

## 2.4. Índice $R$

O índice  $R$  é definido como a raiz do número total de citações das publicações do núcleo  $h$  (Jin *et al.*, 2007), podendo ser matematicamente representado como em (2), onde  $h$  é o número de publicações no núcleo  $h$ , e  $c_j$  é o número de citações para cada publicação  $j = 1, \dots, h$ .

$$R = \sqrt{\sum_{j=1}^h c_j} \quad (2)$$

Observa-se, ainda, que  $R \geq h$ , uma vez que  $c_j \geq h$  (Jin *et al.*, 2007). Acrescenta-se, ainda, que  $R$  apresenta valor mínimo quando todas as publicações do núcleo  $h$  são citadas o número mínimo de vezes, i.e.,  $c_j = h, \forall j = 1, \dots, h$ . Nesse caso,  $\sum_{j=1}^h c_j = h^2$  e  $R = h$ .

Como não se trata de uma divisão por  $h$ , mas sim de uma função que depende somente da soma de citações no núcleo  $h$ , o índice  $R$  cresce monotonicamente com o aumento do número de citações. Assim,  $R$  não apresenta as desvantagens de  $A$ , embora apresente também a característica de ser sensível a publicações muito citadas, mantendo, ainda, o mesmo núcleo  $h$  de publicações, diferentemente do índice  $g$ .

Por outro lado, apesar de Jin *et al.* (2007) afirmarem que o problema de precisão associado aos índices  $A$  e  $R$  é idêntico ao do índice  $h$ , e não aumentado, como para o índice  $g$ , vale notar que existe, sim, um aumento no seu problema de precisão. Isso porque o índice  $h$  permite dúvidas a respeito da quantidade exata de citações para as publicações, ainda que dentro do núcleo, desde que não sejam muito significativas, ao contrário dos índices  $A$  e  $R$ , os quais não permitem tais dúvidas sem ser influenciados.

## 3. Comparações de Valor entre os Índices

Da mesma forma que os índices tipo  $h$  foram comparados ao índice  $h$ , nessa seção, são apresentadas outras comparações de valores. A primeira subseção demonstra quantas citações podem estar associadas ao núcleo  $h$ , para que os índices  $g$  e  $h$  apresentem o mesmo valor. Na subseção seguinte, demonstra-se que existe a relação  $A \geq g$ , o que já foi apresentado em Jin *et al.* (2007). A última subseção é dedicada às relações com o índice  $R$ , sendo demonstrado que  $A \geq R$ , porém que não existe relação fixa entre  $R$  e  $g$ .

### 3.1. Índices $g$ e $h$

Conforme visto anteriormente, existe a relação  $g \geq h$  (Egghe, 2010). Entretanto, vale acrescentar que tais índices apresentam o mesmo valor se cada publicação no núcleo apresentar  $h$  citações, i.e., se houver um total de  $h^2$  citações no núcleo, independente do número de citações das demais publicações, como será mostrado a seguir.

Suponha-se uma série de publicações listadas em ordem decrescente de citações, em que as primeiras  $h$  publicações tenham recebido exatamente  $h$  citações cada, e a publicação seguinte tenha

recebido  $\sigma$  citações, sendo  $\sigma \leq h$ , pela ordenação das publicações. Nesse caso, as primeiras  $h$  publicações formam o núcleo do índice  $h$ , independente do valor de  $\sigma$ , tendo em vista a própria definição do índice. Em relação ao índice  $g$ , sabe-se, também por definição, que o núcleo é composto por, pelo menos, as primeiras  $h$  publicações. Para mostrar que o núcleo do índice  $g$  não poderia ser formado por  $h + 1$  publicações, utiliza-se o método de redução ao absurdo.

Suponha-se o núcleo do índice  $g$  formado por  $h + 1$  publicações, o número de citações nesse núcleo deve ser pelo menos  $(h + 1)^2$ , pela definição do índice. Segue que o total de citações no núcleo, para o caso em tela, é de  $h^2 + \sigma$ . Para esse número de citações ser suficiente, deve-se ter  $h^2 + \sigma \geq (h + 1)^2 = h^2 + 2h + 1$ , ou seja,  $\sigma \geq 2h + 1$ . Por outro lado,  $\sigma \leq h$ , tendo em vista a ordenação das publicações, havendo, portanto, uma incoerência. Assim, pode-se afirmar que o núcleo do índice  $g$  não pode ser formado por  $h + 1$  publicações, independente do valor de  $\sigma$ .

Além disso, os índices  $h$  e  $g$ , por serem números naturais, também apresentam o mesmo valor quando, além das  $h^2$  citações no núcleo, houver algumas citações extras. Esse número de citações pode ser calculado, conforme a seguir.

Seja  $\delta$  o número de citações acrescentadas ao núcleo de publicações, definido anteriormente, que não provoca acréscimo do índice  $g$ , i.e., o núcleo não pode incluir as primeiras  $h + 1$  publicações. Em outras palavras, a soma de citações para essas  $h + 1$  publicações deve ser inferior a  $(h + 1)^2$ , por definição. Como o núcleo original apresentava  $h^2$  citações, sendo acrescentado ao mesmo  $\delta$  citações, a soma de citações para as primeiras  $h + 1$  publicações seria de  $h^2 + \delta + \sigma$ . Dessa forma, deve-se ter  $h^2 + \delta + \sigma < (h + 1)^2$ , ou seja,  $\delta < 2h + 1 - \sigma$  é o número de citações que podem ser acrescentadas ao núcleo mantendo os índices  $h$  e  $g$  equivalentes. Para que  $\delta$  não dependa de  $\sigma$ , vale lembrar que  $\sigma \leq h$  e, assim,  $\delta < h + 1$  ou  $\delta \leq h$ .

### 3.2. Índices $g$ e $A$

Além dos índices  $g$  e  $A$  serem maiores que  $h$ , vale ressaltar que existe a relação  $A \geq g$ , conforme demonstrado em Jin *et al.* (2007). Essa relação pode ser observada pelo fato de o núcleo de  $g$  ser sempre maior ou igual que o núcleo de  $h$ , englobando, assim, publicações menos citadas além daquelas presentes no núcleo  $h$ . Assim, a média de citações das publicações no núcleo  $h$  será sempre maior ou igual à média de citações das publicações no núcleo  $g$ .

Para demonstrar isso, de maneira diferente daquela apresentada em Jin *et al.* (2007), considera-se primeiramente os casos em que  $g$  é maior que  $h$  para depois considerar os casos em que os índices são iguais. Em qualquer caso, ordenam-se as publicações em ordem decrescente de citações, sem perda de generalidade. Assim, suponha que as publicações do núcleo  $h$  receberam  $x_1, \dots, x_h$  citações, respectivamente, e que as publicações do núcleo  $g$  receberam, respectivamente,  $x_1, \dots, x_h, \dots, x_g$  citações. Considerando que as publicações estão em ordem decrescente,  $x_1 > \dots > x_h > \dots > x_g$ . Assim, a média de  $x_1, \dots, x_h$  é sempre maior que a média de  $x_1, x_2, \dots, x_h, \dots, x_g$ . Nesses casos,  $A$  é estritamente maior que  $g$ .

Quando os índices  $h$  e  $g$  são os mesmos, o núcleo é o mesmo, bem como a sua média de citações. Entretanto, como  $g$  é um número natural, enquanto  $A$  é um número inteiro, ainda assim é possível ter  $A > g$  mesmo quando  $h = g$ . Mais especificamente, sempre que houver  $\delta$  citações a mais no núcleo, além das  $h^2$  citações mínimas, conforme calculado na seção 3.1, haverá essa situação.

### 3.3. Comparações com o índice $R$

É possível observar que  $A \geq R$ . Para isso, basta rescrever os índices, i.e.,  $A.h = \sum_{j=1}^h c_j$  e  $R.R = \sum_{j=1}^h c_j$ , ou seja,  $A.h = R.R$ . Como  $h \leq R$ , então, para se manter a equação,  $A \geq R$ .

Em compensação, não é possível definir uma relação entre  $g$  e  $R$ , sendo um índice maior em alguns casos, e o outro maior em outros. Geralmente, tem-se  $g \geq R$ . Isso porque os índices podem



ser reescritos da seguinte forma:  $R \cdot R = \sum_{j=1}^h c_j$  e  $g \cdot g \cong \sum_{j=1}^g c_j$ . Neste último caso, o sinal de igualdade não pôde ser colocado por se tratar de um número natural e não de um real. Como  $\sum_{j=1}^h c_j \leq \sum_{j=1}^g c_j$ , dado que  $g \geq h$ , tem-se, em via de regra,  $g \geq R$ . Na verdade, se os índices apresentassem a mesma natureza numérica, seria garantido que  $g \geq R$ . Entretanto, se os valores das somas de citações nos núcleos  $g$  e  $h$  forem próximos, é possível ter  $g < R$ , justamente porque  $g$  é natural enquanto  $R$  é real.

#### 4. Efeitos do Crescimento das Citações

Apesar de as relações de valores terem sido estabelecidas na seção 3, vale observar como os índices se comportam quando há um crescimento do número de citações para as publicações do núcleo  $h$ . Ressalte-se que essa situação ocorre na prática quando um pesquisador cessa as suas atividades, mas continua tendo as suas publicações citadas, em especial as suas publicações mais populares.

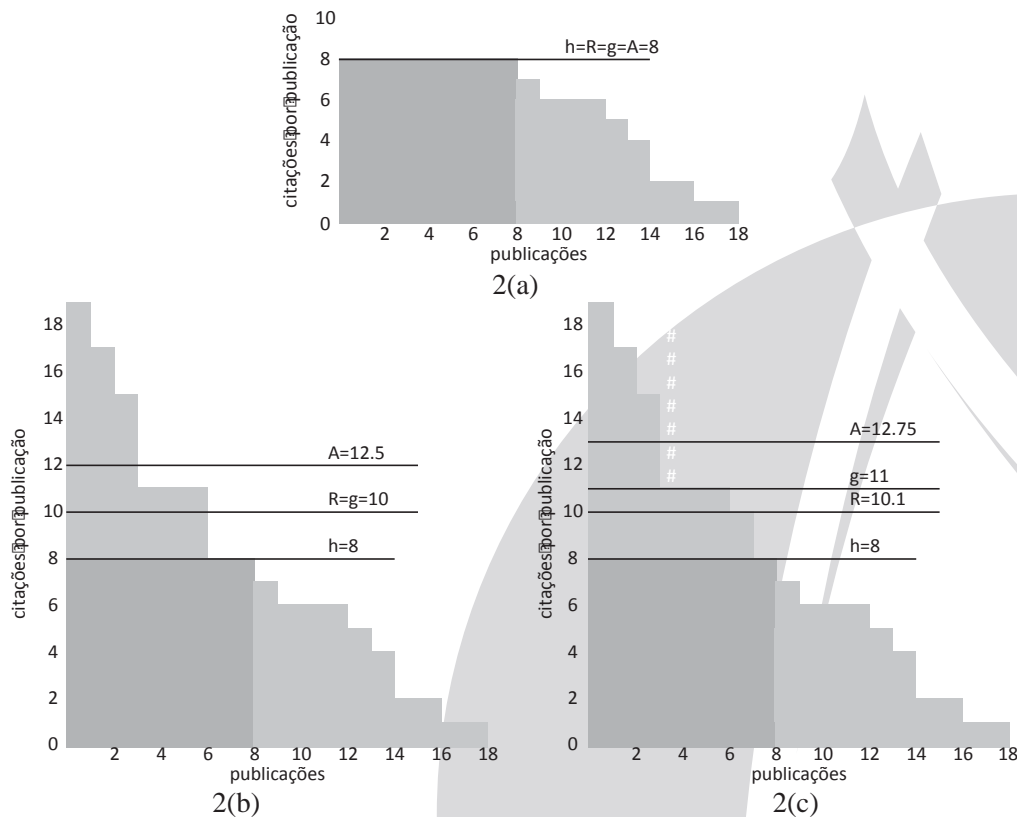


Figura 2: Comportamento dos índices frente a um aumento do número de citações

Combinando as equações 1 e 2, observa-se que  $A$  e  $R$  são iguais somente quando  $\sum_{j=1}^h c_j = h^2$ , ou seja, quando cada publicação do núcleo tiver recebido exatamente  $h$  citações, caso em que  $A = R = h$ , conforme ilustrado em 2(a). Acrescenta-se que, nessa situação, o índice  $g$  também apresenta o mesmo valor, conforme apresentado anteriormente. Um aumento somente do número de citações das publicações do núcleo  $h$  eleva os valores de  $A$  e  $R$ , mantendo o índice  $h$  constante, sendo o crescimento de  $A$  maior que o de  $R$ , como ilustrado em 2(b) e 2(c). Para entender o comportamento de  $g$ , em comparação com esses índices, considera-se a Figura 2(a), em que 64 citações são atribuídas

às 8 publicações do núcleo, e a nona publicação apresenta 7 citações. Quando se acrescentam algumas citações no núcleo,  $A$  e  $R$  crescem, mas  $g$  não cresce até que as 9 publicações juntas apresentem 81 citações. Em outras palavras, enquanto não forem acrescentadas 10 citações, tem-se  $A > R > g$ ; já com 10 citações extras, tem-se  $g = 9$ ;  $R = \sqrt{74} = 8.6$ ; e  $A = 74/8 = 9.25$ , ou seja,  $A > g > R$ .

Mantendo-se o crescimento das citações,  $R$  atinge o valor de  $g = 9$ , quando houver 81 citações no núcleo; e ultrapassa  $g$ , a partir desse valor. Quando o núcleo apresentar 87 citações,  $g$  terá o valor de 10, uma vez que a nona e a décima publicação juntas apresentam 13 citações, enquanto  $R = \sqrt{87} = 9.3$  e  $A = 10.9$ . Com o incremento de novas citações,  $A$  e  $R$  crescem até que, com mais 13 citações, i.e., 100 no núcleo  $h$ , tem-se  $R = g$ , como ilustrado em 2(b). Com duas novas citações no núcleo, tem-se a situação ilustrada em 2(c). Com novos incrementos no número de citações,  $g$  permanecerá sempre maior ou igual a  $R$ , para o exemplo em questão, até o final, i.e., até que  $g$  seja equivalente ao número total de publicações.

## 5. Estabilidade dos Índices

Outra característica importante para índices bibliométricos, é o seu comportamento frente a mudanças na distribuição das citações pelas publicações, sem modificar o seu total. Vale notar que tais mudanças não são possíveis na prática, no âmbito de um mesmo pesquisador, como as mudanças analisadas na seção 3. Assim, trata-se de análise interessante para comparar diferentes pesquisadores.

De maneira geral, os índices estudados permitem certas variações na distribuição das citações dentro dos respectivos núcleos. Entretanto, o índice  $g$  permite uma variação mais significativa do que os demais índices, por dois motivos básicos. Primeiramente, o núcleo  $g$  é mais amplo do que o núcleo  $h$ , permitindo, assim, uma variação da distribuição de citações entre um número maior de publicações do que para os índices  $h$ ,  $A$  e  $R$ . Em segundo lugar, todas as publicações do núcleo  $h$  devem ser citadas minimamente  $h$  vezes, o que não é necessário para o núcleo  $g$ . Assim, a variação da distribuição de citações deve respeitar esse limite para os índices  $h$ ,  $A$  e  $R$ , mas não para o índice  $g$ .

Dessa forma, o índice  $g$  permite que as citações de seu núcleo sejam distribuídas a um número mínimo de publicações, i.e., sejam muito concentradas, desde que mantido o mesmo número de publicações citadas, sem ter o seu valor alterado. Por outro lado, quando as citações são concentradas acima do limite definido anteriormente, o núcleo  $h$  diminui, o que leva a um aumento do índice  $A$ , conforme esclarecido em 2.3. Já o comportamento do índice  $R$  não é fixo, o que será analisado adiante, pela Tabela 1.

Em contrapartida, quando as citações são dissipadas, i.e., distribuídas pelas publicações de maneira mais uniforme, o índice  $h$  tende a aumentar, uma vez que novas publicações podem atingir o número mínimo de citações para entrarem no núcleo, levando a uma diminuição do índice  $A$ , conforme esclarecido em 2.3. Nessa situação, o índice  $g$  pode diminuir, se a dissipação ultrapassar o núcleo  $g$ , de forma que um menor número de citações permaneça no núcleo. Entretanto, vale esclarecer que, mesmo com a possível redução de  $g$  e o possível aumento de  $h$ , frente a uma dissipação das citações, permanece válida a relação  $g \geq h$ . Novamente, o índice  $R$  não apresenta comportamento previamente definido, como ilustrado a seguir.

A Tabela 1 ilustra o comportamento indefinido de  $R$ , frente a concentrações e dissipações das citações pelas publicações, com base em um exemplo fabricado, cujos dados iniciais estão representados na primeira coluna. O núcleo  $h$  é indicado em cinza, sendo apresentados os valores dos índices  $h$  e  $R$ , bem como o total de publicações ( $p$ ) e de citações ( $N_c$ ), ambos mantidos fixos para todos os casos.

A segunda coluna apresenta o caso em que as citações de dentro do núcleo original sofrem uma concentração, levando a uma redução do núcleo  $h$ , bem como uma redução de  $R$ , uma vez que o

total de citações no novo núcleo é menor que o total de citações no núcleo anterior. A terceira coluna indica o caso em que as citações do núcleo são dissipadas, levando a um aumento de  $h$ , bem como um aumento de  $R$ , uma vez que o total de citações no novo núcleo torna-se maior do que no núcleo original.

Ordem da Publicação	Dados Iniciais	Concentração limitada	Dissipação limitada	Concentração Geral	Dissipação Geral
1	10	10	8	11	6
2	8	8	6	10	5
3	5	6	6	8	5
4	4	3	5	2	5
5	3	3	5	1	5
6	3	3	3	1	4
7	1	1	1	1	4
p	7	7	7	7	7
Nc	34	34	34	34	34
h	4	3	5	3	5
R	5.2	4.9	5.5	5.4	5.1
Resultados em comparação com os dados originais		$h$ menor, $R$ menor	$h$ maior, $R$ maior	$h$ menor, $R$ maior	$h$ maior, $R$ menor

Tabela 1: Comportamento do índice  $R$  frente a dissipações e concentrações de citações

As demais colunas apresentam casos em que o índice  $R$  apresenta comportamentos contrários aos apresentados nas colunas 2 e 3. A quarta coluna apresenta uma concentração geral, em que o novo núcleo  $h$  recebe citações de publicações externas a ele, levando a uma redução de  $h$ , mas um aumento de  $R$ . Já a quinta coluna apresenta uma dissipação geral, em que o núcleo perde citações para publicações externas, levando a um aumento do índice  $h$ , mas uma redução do índice  $R$ .

Ordem da Publicação	Egghe	Rousseau	Leydesdorff	White	Martin	Van Raan
1	47	25	79	128	156	108
2	42	18	32	106	74	51
3	37	18	26	103	52	49
4	36	16	24	45	38	41
5	21	16	23	37	35	35
6	18	15	22	28	33	32
7	17	15	19	22	33	31
8	16	14	17	21	30	30
9	16	13	17	20	29	25
10	16	13	16	15	28	25
11	15	13	15	14	24	23
12	13	13	13	14	23	22
13	13	13	13	12	22	22
14	13	12	13	12	20	21
15	13	12	11	12	19	20
16	12	12	11	12	18	19
17	12		11	11	16	19
18	12		10	10	16	19
19	12		10	8	16	19
20	11		9	6	16	18
21				5	14	18
22				5	14	17
23				5	11	17
24				5	9	17
25				4	9	15
26				4	9	14
27					6	14
28					4	14
h	13	13	13	12	16	19
g	19	15	19	25	27	27
A	23.6	15.5	24.3	46.1	39.6	32.2
R	17.5	14.2	17.8	23.5	25.2	24.7

Tabela 2: Base de dados de citações e publicações de seis autores, extraída de Egghe (2006b)



Para ilustrar a estabilidade de todos os índices analisados no presente, utiliza-se o exemplo, extraído da base de dados de Egghe (2006b). A Tabela 2 apresenta o número de citações para cada publicação de seis autores, quais sejam, L. Egghe, L. Leydesdorff, R. Rousseau, H. White, B. Martin e A. Van Raan. As publicações são posicionadas em ordem não crescente de citações. Além do total de citações ( $N_c$ ), a Tabela 2 também apresenta os índices  $h$ ,  $g$ ,  $A$  e  $R$  para cada autor.

Observa-se que a diferença entre  $h$  e  $g$  é pequena para Rousseau e grande para White e Martin. Isso é consequência do fato de as citações de Rousseau estarem distribuídas de maneira relativamente uniforme, enquanto as citações de White e Martin estarem concentradas em poucas publicações. Outra consequência desse mesmo fato é a proximidade de  $A$  e  $h$  para Rousseau, enquanto que para White e Martin,  $A$  é muito superior a  $h$ .

Para analisar a estabilidade dos índices a partir dos dados da Tabela 2, as citações foram primeiramente dissipadas, i.e., distribuídas pelas publicações de maneira mais uniforme, mantendo o mesmo número de citações totais e o mesmo número de publicações citadas. Como existem diversas maneiras de se dissipar as citações, neste caso, maximiza-se o índice  $h$ . Mesmo com a definição desse objetivo, há algumas citações extras, para cada autor, indiferentes para o seu índice  $h$ . Essas citações foram, então, alocadas no núcleo, de modo a maximizar também, em segundo plano, os demais índices. Essa redistribuição é apresentada na Tabela 3.

Ordem da Publicação	Egghe	Rousseau	Leydesdorff	White	Martin	Van Raan
1	21	16	22	36	42	39
2	20	15	20	26	28	27
3	20	15	20	26	28	27
4	20	15	20	26	28	27
5	20	15	20	26	28	27
6	20	15	20	26	28	27
7	20	15	20	26	28	27
8	20	15	20	26	28	27
9	20	15	20	26	28	27
10	20	15	20	26	28	27
11	20	15	20	26	28	27
12	20	15	20	26	28	27
13	20	15	20	26	28	27
14	20	15	20	26	28	27
15	20	15	20	26	28	27
16	20	12	20	26	28	27
17	20		20	26	28	27
18	20		20	26	28	27
19	20		20	26	28	27
20	11		9	26	28	27
21				26	28	27
22				26	28	27
23				26	28	27
24				26	28	27
25				26	28	27
26				4	28	27
27					28	27
28					4	14
$h$	19	15	19	25	27	27
$g$	19	15	19	25	27	27
$A$	20.1	15.1	20.1	26.4	28.5	27.4
$R$	19.5	15.0	19.5	25.7	27.7	27.2

Tabela 3: Dispersão de citações

Observa-se que, na Tabela 3,  $h = g \cong A \cong R$  para todos os pesquisadores. Na verdade, os índices não são iguais por causa das citações extras, que não influenciam o índice  $h$ , mas afetam os índices reais  $A$  e  $R$ .

Além disso, observa-se que o valor de  $g$  das Tabelas 2 e 3 é o mesmo para todos os pesquisadores, o que ilustra a grande estabilidade do índice. Além disso, o índice  $R$  aumentou em todos os casos porque o núcleo  $h$  passou a incluir todas as citações de cada autor, exceto as citações da última publicação, que não sofreram modificação. Assim, o somatório de citações dentro do núcleo cresceu e, conseqüentemente, o índice  $R$  também. Como era esperado, o índice  $A$  reduziu para todos os pesquisadores.

Em seguida, as citações da Tabela 2 foram concentradas, i.e., redistribuídas de maneira que um menor número de publicações apresente uma maior parcela do total de citações, mas mantendo o mesmo número total de publicações citadas. Arbitrariamente, definiu-se o seguinte padrão de concentração: 38% do total de citações de cada autor para a primeira publicação, 26% para a segunda publicação, 16% para a terceira, 9% para a quarta, 5% para a quinta, 2% para a sexta e 1% para a sétima. Além disso, o número de publicações citadas não foi alterado, sendo, portanto, atribuída 1 citação para a oitava publicação em diante. Essa redistribuição é apresentada na Tabela 4.

Ordem da Publicação	Egghe	Rousseau	Leydesdorff	White	Martin	Van Raan
1	150	90	150	255	295	290
2	100	62	100	170	201	200
3	64	39	63	108	126	120
4	36	21	36	60	72	65
5	19	11	20	35	40	40
6	7	4	7	13	15	15
7	2	2	2	4	4	4
8	1	1	1	1	1	1
9	1	1	1	1	1	1
10	1	1	1	1	1	1
11	1	1	1	1	1	1
12	1	1	1	1	1	1
13	1	1	1	1	1	1
14	1	1	1	1	1	1
15	1	1	1	1	1	1
16	1	1	1	1	1	1
17	1		1	1	1	1
18	1		1	1	1	1
19	1		1	1	1	1
20	1		1	1	1	1
21	1			1	1	1
22				1	1	1
23				1	1	1
24				1	1	1
25				1	1	1
26				1	1	1
27					1	1
28					1	1
$h$	6	5	6	6	7	7
$g$	19	15	19	25	27	27
$A$	62.7	44.6	62.7	106.8	107.6	104.9
$R$	19.4	14.9	19.4	25.3	27.4	27.1

Tabela 4: Concentração das citações

Pode-se observar que o índice  $h$  ficou bastante reduzido, quando comparado às tabelas 2 e 3. O índice  $A$  cresceu significativamente em relação às tabelas anteriores, e o índice  $R$  também cresceu uma vez que a soma das citações dentro do núcleo é maior, ainda que o núcleo tenha reduzido. Para  $R$ , trata-se de situação semelhante àquela da quarta coluna da Tabela 1 de “Concentração Geral”.

Observa-se ainda que o índice  $g$  permaneceu o mesmo das Tabelas 2 e 3, o que ilustra a grande estabilidade do índice, sendo pouco dependente da atribuição das citações. Em outras palavras, as citações recebidas por um pesquisador beneficiam, em geral, o seu índice  $g$ , não sendo muito relevante para qual publicação aquela citação está associada.

## 6. Conclusão

O presente trabalho estudou o índice  $h$ , proposto por Hirsch (2005), para medir o desempenho de pesquisadores, bem como índices tipo  $h$ , desenvolvidos na literatura para corrigir desvantagens do índice original. Foram estudados apenas índices clássicos que aumentam sensibilidade a publicações muito citadas, quais sejam, os índices  $g$ ,  $A$  e  $R$ .

Foram analisadas teoricamente as relações de valor entre os índices estudados. Além das relações já apresentadas pela literatura, i.e.,  $A \geq g \geq h$ , demonstrou-se que  $A \geq R$ , mas que não existe relação fixa entre  $R$  e  $g$ . Além disso, foi desenvolvida a fórmula para calcular o número de citações associadas ao núcleo  $h$ , para que os índices  $g$  e  $h$  apresentem o mesmo valor.

Em seguida, analisou-se o comportamento dos índices frente a aumentos apenas no número de citações, situação recorrente na prática, principalmente após o final da carreira dos pesquisadores. Foi utilizado um exemplo numérico para ilustrar tal análise. Com esse exemplo, também foi possível verificar a relação variável entre  $R$  e  $g$ .

Além disso, estudou-se a estabilidade dos índices frente a variações na distribuição das citações, sem modificar o total de citações ou de publicações citadas. Verificou-se que o índice  $g$  é significativamente mais estável que todos os outros, mantendo-se constante mesmo com variações relevantes na distribuição de citações. Utilizando-se um exemplo ilustrativo, qual seja a base de dados de seis autores, apresentada em Egghe (2006b), também se verificou que, conforme esperado, os índices se aproximam quando as citações são distribuídas de maneira mais uniforme (distribuição dissipada), e se distanciam quando boa parte das citações é atribuída a um pequeno número de publicações (distribuição concentrada).

Estudos futuros podem estudar as implicações teóricas das propriedades verificadas por meio de exemplos numéricos no presente trabalho, principalmente quanto à estabilidade dos índices frente a mudanças na distribuição de citações.

## Agradecimentos

Ao CNPq, pelo apoio financeiro.

## Referências Bibliográficas

- Alonso, S., Cabrerizo, F.J., Herrera-Viedma, E., Herrera, F. (2010). hg-Index: A new index to characterize the scientific output of researchers based on the  $h$ - and  $g$ -indices. *Scientometrics*, 82(2), 391–400.
- Anderson, T., Hankin, R., Killworth, P. (2008). Beyond the Durfee square: Enhancing the  $h$ -index to score total publication output. *Scientometrics*, 76(3), 577–588.
- Antonakis, J., Lalive, R. (2008). Quantifying scholarly impact: IQp versus the Hirsch  $h$ . *Journal of the American Society for Information Science and Technology* 59 (6), 956–969.
- Batista, P.D., Campiteli, M.G., Kinouchi, O. (2006). Is it possible to compare researchers with different scientific interests? *Scientometrics*, 68(1), 179–189.
- Bornmann, L., Daniel, H.D. (2007). What do we know about the  $h$  index? *Journal of the American Society for Information Science and Technology* 58 (9), 1381–1385.
- Braun, T., Glänzel, W., Schubert, A. (2006). A Hirsch-type index for journals. *The Scientist* 69 (1), 169–173.
- BiHui, J. (2007). The AR-index: Complementing the  $h$ -index. *ISSI Newsletter* 3 (1), 6.
- BiHui, J., LiMing, L., Rousseau, R., Egghe, L. (2007). The  $R$ - and AR-indices: Complementing the  $h$ -index. *Chinese Science Bulletin* 52 (6), 855–963.
- Cabrerizo, F.J., Alonso, S., Herrera-Viedma, E., Herrera, F. (2010).  $q_2$ -Index: Quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core. *Journal of Informetrics*, 4(1), 23–28.

- Castillo, C., Donato, D., Gionis, A.** (2007). *Estimating number of citations using author reputation*. In: *String Processing and Information Retrieval*. Springer, Berlin, Heilberg, pp.107–117.
- Coccia, M.** (2008). Measuring scientific performance of public research units for strategic change. *Journal of Informetrics*, 183–94.
- Egghe, L.** (2006a). An improvement of the h-index: The g-index. *ISSI Newsletter*, 2(1), 8–9.
- Egghe, L.** (2006b). Theory and practise of the g-index. *Scientometrics*, 69(1), 131–152.
- Egghe, L.** (2010). The Hirsch index and related impact measures. *Annual Review of Information Science and Technology*, 44, 65–114.
- Egghe, L., Rousseau, R.** (2008). An h-index weighted by citation impact. *Information Processing and Management*, 44(2), 770–780.
- Franceschini, F., Maisano, D.A.** (2010). Analysis of the Hirsch index's operational properties. *European Journal of Operational Research*, 203(2), 494–504.
- Harzing, A.W., Van der Wal, R.** (2008). Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics* 8 (11), 61–73.
- Hirsch, J. E.** (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46), 16569–16572.
- Jin, B.** (2006). H-index: An evaluation indicator proposed by scientist. *Science Focus*, 1(1), 8–9.
- Jin, B., Liang, L., Rousseau, R., Egghe, L.** (2007). The R- and AR-indices: Complementing the h-index. *Chinese Science Bulletin*, 52(6), 855–863.
- Norris, M., Oppenheim, C.** (2010). The h-index: a broad review of a new bibliometric indicator. *Journal of Documentation*, 66(5), 681–705.
- Panaretos, J., Malesios, C.** (2009). Assessing scientific research performance and impact with single indices. *Scientometrics*, 81(3), 635–670.
- Rubem, A.P.S., Moura, A.L., Soares de Mello, J.C.C.B.** (2014). Comparative analysis of some individual bibliometric indices when applied to groups of researchers. *Scientometrics*, published online: 09 September 2014. DOI: 10.1007/s11192-014-1428-y
- Schreiber, M.** (2007). Self-citation corrections for the Hirsch index. *EuroPhysics Letters* 78. doi:10.1209/0295-5075/78/30002.
- Schreiber, M.** (2008). A modification of the h-index: The hm-index accounts for multiauthored manuscripts. *Journal of Informetrics*, 2(3), 211–216.
- Schreiber, M.** (2010a). Revisiting the g-index: the average number of citations in the g-core. *Journal of American Society for Information Science and Technology*, 61(1), 169–174.
- Schreiber, M.** (2010b). Twenty Hirsch index variants and other indicators giving more or less preference to highly cited papers. *Annalen der Physik*, 522(8), 536–554.
- Schreiber, M., Malesios, C. C., Psarakis, S.** (2011). Categorizing h-index variants. *Research Evaluation*, 20(5), 397–408.
- Sidiropoulos, A., Katsaros, D., Manolopoulos, Y.** (2007). Hirsch h-index for disclosing latent facts in citation networks. *Scientometrics* 72 (2), 253–280.
- Tol, R.S.** (2009). The h-index and its alternatives: An application to the 100 most prolific economists. *Scientometrics*, 80(2), 317–324.
- Waltman, L., Van Eck, N.J.** (2012). The inconsistency of the h-index. *Journal of the American Society for Information Science and Technology*, 63(2), 406–415.