



**UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”**

DEISE DEOLINDO SILVA

**MEDIDA DE DISPERSÃO PARA O ÍNDICE h :
proposta de um indicador do tipo h de Hirsch**

**Marília – SP
2018**

MEDIDA DE DISPERSÃO PARA O ÍNDICE h :
proposta de um indicador do tipo h de Hirsch

DEISE DEOLINDO SILVA

Tese apresentada junto ao Programa de Pós-Graduação em
Ciência da Informação da Universidade Estadual Paulista
“Júlio de Mesquita Filho”, Câmpus de Marília, como
requisito parcial para obtenção do título de Doutor em
Ciência da Informação.

Área: Informação, Tecnologia e Conhecimento.

Linha de pesquisa: Produção e Organização da Informação.

Orientadora: Profa. Dra. Maria Cláudia Cabrini Grácio.

Marília – SP

2018

Silva, Deise Deolindo.

S586m Medida de dispersão para o índice h : proposta de um indicador
do tipo h de Hirsch / Deise Deolindo Silva. – Marília, 2018.
93 f. ; 30 cm.

Orientadora: Maria Cláudia Cabrini Grácio.

Tese (Doutorado em Ciência da Informação) - Universidade
Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências,
2018.

Bibliografia: f. 86-90.

1. Bibliometria. 2. Indicadores de ciência. 3. Ciência da
informação – Métodos estatísticos. I. Título.

CDD 020.182

MEDIDA DE DISPERSÃO PARA O ÍNDICE h : proposta de um indicador do tipo h de Hirsch

Tese apresentada junto ao Programa de Pós-Graduação em Ciência da Informação, da Faculdade de Filosofia e Ciências da Universidade Estadual Paulista Júlio de Mesquita Filho, Câmpus de Marília, como requisito parcial para obtenção do título de Doutor em Ciência da Informação, sob a orientação da Profa. Dra. Maria Cláudia Cabrini Grácio.

Área de concentração: Informação, Tecnologia e Conhecimento.

Linha de Pesquisa: Produção e Organização da Informação.

Membros da Banca Examinadora:

Titular 1: Orientadora: Profa. Dra. Maria Cláudia Cabrini Grácio (UNESP/Marília)

Titular 2: Profa. Dra. Ely Francina Tannuri de Oliveira (UNESP/Marília)

Titular 3: Dr. Leandro Innocentini Lopes de Faria (UFSCar)

Titular 4: Dr. Rogério Mugnaini (USP)

Titular 5: Dr. Daniel Martinez Ávila (UNESP/Marília)

Suplente 1: Dr. Hércules de Araújo Feitosa (UNESP/Bauru)

Suplente 2: Dr. Fábio Mascarenhas (UFPE)

Suplente 3: Dr. Rene Faustino Gabriel Junior (UFRGS)

Marília, 2018

Dedico este trabalho

aquele que me concedeu a vida,

que me educou no caminho do bem,

que sempre estimou pelo meu aprimoramento intelectual

e que hoje é meu intercessor junto a Deus

Meu PAI:

Luiz Deolindo Apolinário.

AGRADECIMENTOS

Agradeço

primeiramente a Deus, por ter me concedido o dom da vida, do amor, da sabedoria e do discernimento, e ao Espírito Santo, por me reavivar sempre nas horas de desânimo e nesses momentos me conceder o dom da fortaleza.

à Maria, mãe de Deus, por ser minha mãe espiritual, ser meu espelho de mulher, cuja história de vida me ensinou a ouvir e a silenciar nos momentos mais difíceis.

ao meu pai, Luiz Deolindo Apolinário, que, durante toda sua trajetória de vida, soube amar sua família e também ao próximo, possuía uma sabedoria invejável e, mesmo não estando em nosso meio, sei que sempre sonhou em me ver alcançando esse tão almejado título.

à minha mãe, Josefa da Silva Deolindo, pelo amor incondicional, pelas palavras de incentivo, pelo carinho, por não me deixar desanimar diante dos obstáculos e por toda a educação concedida a mim.

ao meu esposo, Wanderley Alves Ferreira, pelo amor que nos une, por todo o incentivo e apoio, pela compreensão, pois iniciei o doutorado no início do nosso casamento e, mesmo assim, sempre me animou a prosseguir meus estudos e vibrou com cada conquista minha.

ao meu filho, Luiz Felipe Deolindo Ferreira, que me foi concedido no decorrer do doutorado e que me fez uma pessoa mais feliz e humana. Seu amor me deu forças para finalizar esse processo de doutoramento e me fez valorizar o que realmente é essencial em nossas vidas.

à minha irmã, Débora Deolindo Silva, por ser minha melhor amiga, por ser minha defensora e por me apoiar em todos os momentos da minha vida.

ao meu irmão, Jorge Luiz Deolindo Silva, por ser uma pessoa excepcional, sempre disposto a me ajudar, por ser um irmão muito especial, por alegrar todos a sua volta e pelo amor que nos une.

em especial à minha família – meus tios, primos, afilhados, cunhados e sobrinhos – pelo carinho que prezam por mim e aos meus amigos e colegas de trabalho pelo incentivo e apoio.

aos amigos do curso de pós-graduação em Ciência da Informação e do grupo de pesquisa em Estudos Métricos da Informação, pelo incentivo e carinho.

aos meus professores Maria José Jorente, Ely Francina, José Augusto, Leilah Bufrem, Lena Vânia, Maria das Graças, por toda transmissão de conhecimento.

de forma muito especial a Profa. Dra. Maria Cláudia Cabrini Grácio, por sua amizade, por sua orientação, por sua paciência, por entender meus problemas de horário, também, os problemas pessoais e, principalmente, pelas ideias e contribuições durante todo este trabalho. Espero que essa amizade se estenda por toda a vida!

à Faculdade de Tecnologia de Garça – FATEC/Garça, à Faculdade de Agronomia e Engenharia Florestal – FAEF/Garça e à Faculdade de Tecnologia de Pompeia – FATEC/Pompeia, pelo incentivo ao meu crescimento profissional.

ao convênio entre o Programa INOVA do Centro Paula Souza e à UNESP, por meio do qual fui selecionada para cursar o doutorado.

Conquiste a sabedoria, e ela o exaltar.

Abrace-a, e ela o honrará.

Provérbios 4, 8.

RESUMO

Conhecer o desempenho científico de um investigador, temática, disciplina, área do conhecimento ou país é uma atividade essencial para subsidiar e orientar políticas científicas institucionais ou governamentais. Para tanto, se faz necessário ter critérios de avaliação consistentes e fidedignos ao desempenho do avaliado e considerar na análise aspectos qualitativos e quantitativos. Entre os indicadores de impacto presentes na literatura científica, destaca-se o índice h , proposto por Jorge Hirsch em 2005, considerado um parâmetro avaliativo robusto por avaliar de forma simultânea os aspectos relativos à produtividade e ao impacto científico do avaliado. Apesar de ser uma medida simples de ser obtida e de seu desempenho ter se mostrado o mais significativo na representação do desempenho científico de um pesquisador, diversos trabalhos subsequentes apontam suas deficiências e limitações e propõem variações no intuito de minimizar ou dirimir os problemas e apontados. Neste contexto, esta Tese teve por objetivo analisar o índice h , suas propriedades, variações e ponderações e propor dois novos indicadores complementares ao índice h , denominados índices dc_i e dc_o , dedicados a medir a dispersão do índice h . Tem-se por hipótese que esses novos índices podem contribuir para avaliar a representatividade do índice h como indicador do impacto científico acumulado de um pesquisador e distinguir, de forma mais precisa, pesquisadores seletivos e produtivos. Além disso, contribui para estimar a possibilidade de o autor avaliado incrementar seu índice h , por meio da análise da dispersão do número de citações contidas no h -core e também fora dele. Para verificar a validade dos indicadores como medidas de dispersão, tomou-se como universo de análise o conjunto de 20 autores mais produtivos na revista *Scientometrics*, no período de 35 anos (1980-2014), equivalente à publicação de ao menos 20 artigos científicos. Considerou-se um segundo universo de análise, em âmbito brasileiro, relativo aos 116 pesquisadores com bolsa produtividade em Pesquisa Qualis1 (PQ1) em Matemática, e compararam-se o índice h , alguns indicadores do tipo h presentes na literatura e os índices dc_i e dc_o . Para ambos universos, levantaram-se na base *Scopus*, para cada pesquisador, o total de artigos publicados, o número de citações por artigo e o ano de publicação. A partir dos dados levantados, calcularam-se os índices de cada investigador. Conclui-se que, em ambas situações, os indicadores cientométricos (h e índices do tipo h), complementam-se e auxiliam na identificação de autores com alta produtividade em uma temática. As medidas de dispersão dc_i e dc_o mostraram-se capazes de caracterizar fidedignamente a dispersão do desempenho do pesquisador em torno do seu índice h e, conseqüentemente, a pertinência da representatividade desse índice como indicador do desempenho científico do pesquisador avaliado. Tais medidas de dispersão permitiram, ainda, distinguir de forma mais precisa o desempenho dos pesquisadores de um mesmo campo científico.

Palavras-chave: Indicadores bibliométricos. Índice h . Bibliometria.

ABSTRACT

Knowing the scientific performance of a researcher, thematic, discipline, area of knowledge or country is an essential activity to subsidize and guide institutional or governmental scientific policies. Therefore, it is necessary to have consistent and reliable evaluation criteria for the performance of the evaluated and to consider qualitative and quantitative aspects in the analysis. Among the impact indicators present in the scientific literature, the h-index, proposed by Jorge Hirsch in 2005, stands out as a robust evaluation parameter for simultaneously evaluating aspects related to productivity and to scientific impact of the evaluated researcher. Although h-index is a simple measure to obtain and its performance has been the most significant in the representation of the scientific performance of a researcher, several subsequent works pointed out their deficiencies and limitations and proposed variations in order to minimize or solve the problems. In this context, this Thesis aimed to analyze the h-index, its properties, variations and weights and propose two new complementary indicators to the h-index, called indexes dc_i and dc_o , dedicated to measure the dispersion of the h-index. It is hypothesized that these new indexes can contribute to evaluate the representativeness of the h-index as an indicator of the cumulative scientific impact of a researcher and to distinguish, more accurately, selective and productive researchers. In addition, it contributes to estimate the possibility of the evaluated author increasing his h-index by analyzing the dispersion of the number of citations contained in the h-core and also outside it. To verify the validity of the indicators as measures of dispersion, it was taken the group of 20 most productive authors of the *Scientometrics* journal, in the 35-year period (1980-2014), equivalent to the publication of at least 20 scientific articles. It was considered a second universe of analysis, in the Brazilian context, related to the 116 researchers with productivity scholarship in Qualis1 Research (PQ1) in Mathematics, and it were compared the h-index, some indicators of type h present in the literature and the indices dc_i and dc_o . For both universes, the total number of articles published, the number of citations per article and the year of publication were raised in the Scopus database for each researcher. From the data collected, the indices of each researcher were calculated. In conclusion, in both situations, the scientificometric indicators (h and h-type indices) complement and assist in the identification of authors with high productivity in a thematic. The dc_i and dc_o dispersion measures were able to reliably characterize the dispersion of the researcher's performance around his h-index and, consequently, the pertinence of the representativeness of this index as an indicator of the scientific performance of the evaluated researcher. These dispersion measures also allowed to distinguish in a more accurate way the performance of the researchers of the same scientific field.

Keywords: Bibliometric indicators. h-index. Bibliometry.

LISTA DE FIGURAS

Figura 1 – Representação gráfica do índice h	36
Figura 2 – Distribuição de frequências das citações recebidas pelos autores A, B, C e D.....	39
Figura 3– Interpretação geométrica do índice e	52

LISTAS DE GRÁFICOS

Gráfico 1 – Boxplot para o índice h dos pesquisadores PQ1 de acordo com os níveis.....	75
Gráfico 2: Pesquisadores PQ1 em Matemática subdivididos de acordo com o valor mediano do $CVdc_i$ e do $CVdc_o$	79

LISTA DE QUADROS

Quadro 1 – Representação e descrição dos indicadores	38
Quadro 2 – Escala classificatória para a dispersão do índice dc_i e dc_o	65
Quadro 3 – Pesquisadores PQ1 em Matemática.....	72

LISTA DE TABELAS

Tabela 1 - Ilustração de como encontrar o índice h de um pesquisador.....	31
Tabela 2 – Índice h dos autores E. Garfield e F. Narin	33
Tabela 3 – Indicadores do contexto de desempenho para os autores A, B, C e D	41
Tabela 4 – Cálculo do índice h e g para o pesquisador Egghe	45
Tabela 5 – Cálculo do índice ar para o pesquisador Egghe	48
Tabela 6 – Cálculo do índice h_w , utilizando os dados do autor Egghe	50
Tabela 7– Elementos para o cálculo do índice e	53
Tabela 8 – Elementos para o cálculo do dc_i e dc_o	64
Tabela 9 – Cálculo do dc_i e dc_o para os autores A, B, C e D.....	66
Tabela 10 – Estatísticas descritivas para os indicadores bibliométricos	73
Tabela 11 – Estatísticas descritivas para os indicadores bibliométricos, de acordo com o nível do pesquisador PQ1.	74
Tabela 12 – Estatísticas descritivas para os indicadores bibliométricos, conforme o nível de dispersão do núcleo Hirsch ($CVdc_i$)	77
Tabela 13 - Classificação dos pesquisadores de acordo com o $CVdc_i$ e $CVdc_o$	78
Tabela 14 – Estatísticas descritivas para os indicadores bibliométricos, conforme o índice h ..	81
Tabela 15 – Estatísticas descritivas para os indicadores bibliométricos, conforme o número de artigos.	82

LISTA DE ABREVIATURAS

C&T:	Ciência e Tecnologia
CA:	Comitês de Assessoramento
CAPES:	Coordenação de Aperfeiçoamento Pessoal de Nível Superior
CNPq:	Conselho Nacional de Pesquisa
$CVdc_i$:	Coeficiente de variação da dispersão das citações no interior do <i>h-core</i>
$CVdc_o$:	Coeficiente de variação da dispersão das citações fora do <i>h-core</i>
dc_i :	Dispersão das citações no interior do <i>h-core</i>
dc_o :	Dispersão das citações fora do <i>h-core</i>
FAPs:	Fundações Estaduais de Amparo à Pesquisa
g :	Índice g
h :	Índice h
PIB:	Produto Interno Bruto
PQ:	Produtividade em Pesquisa
PQ1:	Pesquisadores Produtividade em Pesquisa 1
PQ1A:	Pesquisadores Produtividade em Pesquisa 1A
PQ1B:	Pesquisadores Produtividade em Pesquisa 1B
PQ1C:	Pesquisadores Produtividade em Pesquisa 1C
PQ1D:	Pesquisadores Produtividade em Pesquisa 1D
PQ2:	Pesquisadores Produtividade em Pesquisa 2
TIC:	Tecnologia de Informação e Comunicação

SUMÁRIO

INTRODUÇÃO	17
OBJETIVO GERAL	20
OBJETIVOS ESPECÍFICOS	20
JUSTIFICATIVA	20
ORGANIZAÇÃO DO TRABALHO	21
1 INDICADORES BIBLIOMÉTRICOS	22
1.1 ORIGEM E EVOLUÇÃO DA BIBLIOMETRIA	22
1.2 INDICADORES BIBLIOMÉTRICOS E CIENTOMÉTRICOS: TRAJETÓRIA HISTÓRICO-CONCEITUAL	24
1.2.1 INDICADORES DE CITAÇÃO	27
1.2.2 CLASSIFICAÇÃO DOS INDICADORES CIENTOMÉTRICOS	29
2 ÍNDICE DO TIPO H DE HIRSCH	30
2.1 RELAÇÃO DO ÍNDICE H COM O NÚMERO TOTAL DE CITAÇÕES	34
2.1.1 <i>Análise de autores com mesmo índice h</i>	38
2.2 ÍNDICE H NAS DIFERENTES BASES DE DADOS	42
2.3 VARIAÇÕES DO ÍNDICE H	43
2.3.1 <i>Quociente m</i>	43
2.3.2 <i>Índice g</i>	44
2.3.3 <i>Índice $h^{(2)}$</i>	46
2.3.4 <i>Índice a</i>	47
2.3.5 <i>Índice r</i>	47
2.3.6 <i>Índice ar</i>	48
2.3.7 <i>Índice h normalizado (h^n)</i>	49
2.3.8 <i>Índice h_w</i>	49
2.3.9 <i>Índice m</i>	50
2.3.10 <i>Índice v</i>	51
2.3.11 <i>Índice π</i>	51
2.3.12 <i>Índice e</i>	52
2.3.13 <i>Índice w de Wu</i>	53

2.3.14 Índice q^2	53
2.3.15 Índice hg	54
2.3.16 Índice h^c	56
2.3.17 Índice h^t	57
2.3.18 Índice h dinâmico	58
2.3.19 Índice hpd	59
2.3.20 Indicador de vitalidade do impacto	60
2.3.21 Índice k	60
2.3.22 Índice s	61
2.3.23 Índice f	61
2.4 ESTUDOS ENVOLVENDO AS VARIAÇÕES DO ÍNDICE H	62
2.5 PROPOSTA DE UM INDICADOR DO TIPO HIRSCH	62
3. ANÁLISE BIBLIOMÉTRICA PARA OS PESQUISADORES PQ1 EM MATEMÁTICA	70
3.1 CARACTERÍSTICAS ASSOCIADAS AOS BOLSISTAS PRODUTIVIDADE EM PESQUISA	70
3.2 DESCRIÇÃO DO CORPUS DE PESQUISA	71
3.2.1 ANÁLISE BIBLIOMÉTRICA PARA OS PESQUISADORES PQ1 EM MATEMÁTICA	72
4. CONSIDERAÇÕES FINAIS	83
REFERÊNCIAS BIBLIOGRÁFICAS	86
APÊNDICE	91

INTRODUÇÃO

A socialização do conhecimento produzido é um fenômeno que cresce ao longo das últimas décadas, principalmente com o advento das Tecnologias da Informação e Comunicação (TIC). Nota-se um crescimento significativo do número de resultados de pesquisas científicas disseminados nos diferentes canais de comunicação científica, informais e formais, especialmente em bases de dados.

A produção científica tem sido sistematicamente analisada ao longo desse período, uma vez que reflete o grau de desenvolvimento de uma nação, por impactar de forma significativa em suas questões sociais, tecnológicas e econômicas. Neste contexto, governo e a comunidade científica entram em uma relação retroalimentadora, em que o primeiro exige o aumento da produtividade dos pesquisadores e estes aumentam a produção a fim de garantir os subsídios necessários aos seus programas educacionais e aos de pesquisa.

Em consequência, a avaliação da produção científica tornou-se um fator imprescindível, assim como a elaboração de critérios que avaliem os diferentes níveis de agregação: micro (pesquisadores), meso (periódicos científicos) e macro (país). Velho (1986) discute esse cenário e considera que, quanto mais produtivo o ambiente científico, mais frequentes e rigorosas serão as rotinas avaliativas.

Assim, por mostrar aspectos relativos ao grau de desenvolvimento de uma nação e proporcionar avanços em âmbitos social, político e econômico, a elaboração de critérios e indicadores para a análise e caracterização do desempenho científico para os diversos domínios do conhecimento e suas respectivas comunidades tornaram-se atividades imprescindíveis.

Em um mundo em que os recursos são limitados, a mensuração do desempenho científico, mesmo que potencialmente desconfortável, muitas vezes é necessária para fins avaliativos e comparativos, na seleção e análise do desempenho de docentes universitários, para a concessão de bolsas de pesquisa a programas de pós-graduação e a seus discentes, dentre outras situações (HIRSCH, 2005).

Os Estados Unidos e nações da Europa, que apresentam intensa atividade científica, têm adotado a avaliação da produção científica como uma prática comum utilizada pelas agências de fomento, ministérios e organismos ligados às políticas de Ciência e Tecnologia (C&T). As práticas em questão têm despertado o interesse de países, em especial, dos que têm crescimento da produção, como o Brasil (VANZ; STUMPF, 2010).

Nesse contexto, destaca-se a necessidade de esforços para a elaboração de mecanismos adequados de avaliação dos domínios científicos, assim como dispor-se de instrumentos e de

indicadores apropriados para a definição de suas diretrizes, alocação de investimentos e recursos, formulação de programas e avaliação de atividades relacionadas ao desenvolvimento científico e tecnológico, como forma de identificar as elites científicas, garantindo um investimento profícuo das agências de fomento à pesquisa (MUGNAINI *et al.*, 2004; THOMAZ *et al.*, 2011; VANZ, STUMPF, 2010).

Mugnaini *et al.* (2004) apontam a necessidade de se dispor de indicadores fidedignos¹ ao real comportamento científico dos pesquisadores, dos periódicos ou das instituições, entre outros, para a definição de diretrizes, alocação de investimentos e recursos, formulação de programas e avaliação de atividades relacionadas ao desenvolvimento científico e tecnológico no país.

Thomaz *et al.* (2011) afirmam que, diante da crescente demanda por insumos para financiamento de pesquisas científicas, tornou-se necessária a elaboração de mecanismos de avaliação da qualidade acadêmico-científica, como forma de prestigiar instituições e indivíduos capazes de produzir pesquisas que contribuam para a ampliação das fronteiras do conhecimento, garantindo um investimento profícuo das agências de fomento à pesquisa.

Com a mesma visão, Vanz e Stumpf (2010) consideram que o Brasil precisa desenvolver técnicas quantitativas de medição da produção científica e elaborar diferentes bases de dados e indicadores. Afirmam que as bases de dados darão o suporte para a visibilidade da produção científica nacional, e o desenvolvimento de indicadores permitirá uma análise mais fidedigna do desempenho científico das suas instituições, dos periódicos, das áreas de conhecimento, das relações de colaboração científica já existentes e dos investimentos em pesquisas.

Entre as abordagens para a caracterização e para a avaliação de um domínio científico, Hjørland (2002) destaca os estudos bibliométricos, em especial as análises baseadas nas citações recebidas pela produção científica de uma comunidade científica, por constituírem estudos que permitem uma ampla visualização da visibilidade, da influência e do impacto do conhecimento gerado, também, dos autores de maior reconhecimento nesta comunidade.

Os estudos bibliométricos compreendem o conjunto de estudos relacionados à avaliação da informação científica registrada, em diferentes suportes, baseados em recursos quantitativos, como procedimento de análise. Fundamentados na Sociologia da Ciência, na Ciência da Informação, na Matemática, na Estatística e na Computação, são estudos de natureza teórico-conceitual quando contribuem para o avanço do conhecimento da própria temática, propondo novos conceitos e indicadores, bem como reflexões e análises relativas à área. São de natureza

¹ Indicadores fidedignos têm a propriedade que diz respeito à consistência e à estabilidade de uma medida (FONSECA, 2010).

metodológica quando se propõem a dar sustentação aos trabalhos metateóricos nas áreas em que estão aplicados (OLIVEIRA; GRÁCIO, 2011).

As análises bibliométricas têm se mostrado procedimentos tangíveis e confiáveis. Seus indicadores, em geral, avaliam aspectos relativos à produção, à ligação e à citação e têm por finalidade explicitar, além da produtividade, o impacto e o reconhecimento de autores, de periódicos, de instituições, de grupos ou de países, nas diferentes áreas do conhecimento (OLIVEIRA; GRÁCIO, 2011).

No âmbito das questões relativas ao reconhecimento de um pesquisador na comunidade científica, a fim de mensurar o impacto acumulado da produção científica de um pesquisador, Hirsch (2005) desenvolveu um índice bibliométrico em cuja formulação consideram-se aspectos relativos à qualidade (impacto) e à quantidade (número de artigos produzidos), denominado índice h . Esse indicador foi definido da seguinte forma: Um pesquisador tem índice h , se h de seus N_p artigos têm pelo menos h citações cada, e os outros $(N_p - h)$ artigos têm não mais que h citações cada.

O autor supracitado disse que o índice h adquiriu rapidamente visibilidade e passou a ser critério de avaliação em diversas agências de fomento. Lima *et al.* (2012) relataram que o índice h tem sido adotado por diversos órgãos financiadores de pesquisa, com destaque para agências de algumas nações da Europa e da Oceania. No Brasil, a utilização do índice está acontecendo de maneira lenta. Destaca-se algumas fundações estaduais de amparo à pesquisa (FAPs), o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) estão utilizando-o como critério para alocação de recursos e benefícios, tais como as bolsas de produtividade científica concedidas pelo CNPq.

No entanto, existem limitações relativas a esse índice, apontam-se os fatos de ele não ser indicado para comparar pesquisadores de áreas diferentes do conhecimento e de poder atribuir vantagens para pesquisadores que têm maior tempo de dedicação à pesquisa.

Egghe (2010) aponta que novos pesquisadores têm clara desvantagem quando comparados aos que possuem carreira mais longa, pois estes tiveram um tempo maior para se dedicar à pesquisa e acumular citações. Além disso, autores estabelecidos podem aumentar seu índice h , por meio de um incremento no número de citações recebidas, mesmo se não houver novos artigos publicados, já que o índice h não é influenciado pelo número de publicações e sim pelas citações recebidas pelos artigos.

Nesse âmbito, como contextualizar o índice h a fim de contemplar diferenças de comportamento no desempenho científico, em função das citações recebidas pelos artigos

publicados, e de contribuir para a melhoria da avaliação do impacto científico dos pesquisadores?

Considera-se que a elaboração de uma medida de dispersão para o índice h possa oferecer uma resposta a essa questão e contribuir para o aprimoramento do processo avaliativo da produção científica de um pesquisador. Tem-se por hipótese que uma medida que caracterize a dispersão do número de citações contidas no h -core, e também fora dele, possa contribuir para distinguir de forma mais precisa o impacto científico acumulado dos pesquisadores e para identificar pesquisadores seletivos ou produtivos, bem como pode verificar a possibilidade de o autor incrementar seu índice h .

Objetivo Geral

Este trabalho teve por objetivo geral propor dois indicadores bibliométricos derivados do índice h de Hirsch, que mensurem e avaliem as caudas superior e inferior da distribuição das citações de um pesquisador, denominados dc_i e dc_o , respectivamente. A finalidade foi proporcionar indicadores mais fidedignos do impacto cumulativo do avaliado. Buscou-se contribuir para se discernir, de forma mais precisa, as diferenças entre o impacto científico acumulado de pesquisadores de um mesmo domínio científico.

Objetivos Específicos

Em âmbito específico, busca-se:

- a) analisar o índice h e suas variações considerando pesquisadores com tempos de atuação distintos e diferentes distribuições de citações;
- b) propor índices derivados do índice h de Hirsch, a partir da mensuração do número de citações dentro e fora do núcleo Hirsch e
- c) realizar uma análise comparativa entre os resultados do índice h , índice derivado de h proposto, índice g e outros observados na literatura, por meio da avaliação de pesquisadores com bolsa produtividade em Pesquisa Qualis1 (PQ1) em Matemática.

Justificativa

Essa pesquisa justifica-se pela contribuição para o aprofundamento da análise dos índices do tipo h de Hirsch e da relação entre eles, assim como para a compreensão desses indicadores como descritores da trajetória acadêmica de um pesquisador.

Considera-se, também, a preocupação crescente com a avaliação da produção científica e, por consequência, a necessidade de proporcionarem-se indicadores mais fidedignos ao real impacto e ao reconhecimento dos pesquisadores que constroem o conhecimento científico, propiciando parâmetros que tornem possível uma avaliação mais precisa, a fim de subsidiar as tomadas de decisões de forma mais consistente, viabilizando novas estratégias de crescimento.

Além disso, aponta-se a carência de estudos relacionados a indicadores bibliométricos que trabalhem de forma conjunta e aprofundada com metodologias estatísticas (VANZ, STUMPF, 2010; OLIVEIRA, GRÁCIO, 2011; MUGNAINI, *et al.*, 2004).

Desse modo, considera-se que esta pesquisa contribui para os estudos da área da Ciência da Informação, especialmente da Bibliometria e Cientometria, ao colaborar para compreensão, discussão e visualização do papel dos indicadores bibliométricos no processo avaliativo da produção científica.

Organização do trabalho

Inicialmente, realizou-se pesquisa bibliográfica acerca dos temas pertinentes ao trabalho, em especial, o estudo de índice h e suas variações.

Os levantamentos sobre as teorias associadas ao índice h de Hirsch teve por objetivo descrever, analisar sua formulação matemática e estudar as diferentes variações, generalizações e ponderações existentes para o indicador. Assim, no Capítulo 1 apresentaram-se os indicadores bibliométricos e, no Capítulo 2, os indicadores do tipo Hirsch.

Como aplicação prática, analisaram-se os índices para pesquisadores de uma área específica, com tempos de atuação eram diferentes. Com esta análise, objetivou-se investigar as vantagens e as desvantagens em se utilizar o índice h como indicador bibliométrico.

Após essas análises, encontrou-se um índice h do tipo Hirsch, a partir da contextualização do número de citações contidas na cauda superior e inferior da distribuição de citações do pesquisador, com base no instrumental estatístico.

Encontrados os indicadores, realizou-se uma análise comparativa entre os resultados do índice h , dos novos índices derivados de h , do índice g e de outros observados.

Por fim, adotou-se como universo de aplicação e de análise do comportamento dos novos indicadores, dc_i e dc_o , associados ao índice h , os pesquisadores com bolsa produtividade em Pesquisa Qualis1 (PQ1) em Matemática, a fim de **se** verificar as principais diferenças entre os indicadores do tipo h de Hirsch e a eficácia do desempenho dos novos índices propostos, em diferentes cenários.

1 INDICADORES BIBLIOMÉTRICOS

1.1 Origem e evolução da Bibliometria

Em 1934, Paul Otlet definiu a bibliometria como parte da bibliografia na qual se analisa a medida ou a quantidade aplicada ao livro (OTLET, 1986). O termo bibliometria deriva-se de *Biblion* (livro) e *metron* (medida), e foi utilizado inicialmente como um instrumento de avaliação quantitativa aplicado ao comportamento da literatura científica – quantidade de edições, de exemplares, de palavras contidas nos livros, de espaços ocupados nas bibliotecas (DANUELLO, 2014; ARAÚJO, 2006).

Originalmente, a finalidade da Bibliometria restringia-se a medir e a avaliar os custos, a extensão e a utilização das coleções, dessa maneira, esta atividade ficou conhecida como bibliografia estatística. Em 1948, Ranganathan introduziu o termo bibliotecometria para referir-se à atividade quantitativa no contexto das bibliotecas (DANUELLO, 2014).

Embora o termo Bibliometria tenha aparecido nos estudos de Otlet, consagrou-se e popularizou-se a partir de 1969, quando Alan Pritchard sugeriu que essa terminologia fosse mais adequada para substituir o termo bibliografia estatística. Pritchard a definiu como sendo os métodos estatísticos e matemáticos aplicados a livros ou outros meios de comunicação e sugeriu sua utilização em todos os estudos voltados à quantificação dos processos da comunicação escrita (PRITCHARD, 1969).

Com o passar dos anos, a Bibliometria deixou de analisar somente informações relacionadas a livros e voltou-se à análise de outros formatos da produção bibliográfica: artigos, periódicos, patentes e outros tipos de documentos. Além disso, passou a ocupar-se também da avaliação da produtividade dos autores, das instituições e dos estudos relacionados às citações (ARAÚJO, 2006).

Narin (1976) afirma que a Bibliometria poderia ser desdobrada em duas vertentes: a descritiva e a avaliativa. A primeira estuda determinadas características da literatura científica relacionadas, por exemplo, à distribuição geográfica e à evolução temporal da produção bibliográfica. A segunda analisa aspectos qualitativos da atividade científica, especificamente por meio das análises de citações e das publicações.

Thelwall (2008) relata, além destas, a existência da bibliometria relacional a qual busca apresentar as relações no âmbito da investigação. Pode-se destacar: a estrutura cognitiva dos campos de pesquisa, o surgimento de novas frentes de pesquisa nacionais ou internacionais e os padrões de co-autoria.

Nesse sentido, pode-se dizer que a Bibliometria analisa os aspectos quantitativos da produção científica, da disseminação e do uso da informação registrada, utiliza os resultados para elaborar previsões e auxilia tomadas de decisões em diferentes níveis – tanto ao pesquisador, quanto à instituição e ao governo, entre outros. Contemporaneamente, na Ciência da Informação, considera-se uma abordagem da Análise de Domínio e adota-se, assim, um paradigma social como fundamento teórico-metodológico de sustentação para seus estudos e indicadores destinados à análise objetiva da produção científica. Como recursos metodológicos, apoia-se em procedimentos matemáticos e estatísticos, os quais são capazes de quantificar os processos de comunicação escrita e podem contribuir para a visualização do comportamento de um campo do conhecimento científico (TAGUE-SUTCLIFFE, 1992; PRICE, 1963; GRÁCIO, OLIVEIRA, 2011; HJØRLAND, 2002; ARAUJO, 2006).

Logo, os estudos bibliométricos contribuem para se evidenciar o referencial teórico epistemológico de um campo científico e as relações existentes, seja dentro desse campo ou com os demais. Ao utilizar métodos quantitativos, a Bibliometria aproxima-se e intercepta-se à Cientometria, à Informetria e à Patentometria, mas diferencia-se quanto ao objeto e quanto ao objetivo do estudo (PRICE, 1963; GRÁCIO, OLIVEIRA, 2011; ROSAS, 2013).

Glänzel (2003) considera que as análises bibliométricas têm três fins e, a partir deles, definiu três grupos-alvo da Bibliometria:

- 1) **Bibliometria para especialistas em bibliometria:** pesquisas destinadas a contribuir à teoria e ao desenvolvimento metodológico da própria área,
- 2) **Bibliometria para disciplinas científicas:** pesquisas metateóricas, voltadas a contribuir ao conhecimento das diversas disciplinas e
- 3) **Bibliometria para a política e gestão científica:** estudos destinados a avaliar a pesquisa científica, com a finalidade de subsidiar as tomadas de decisão em política científica.

Considerando os objetivos propostos, esta pesquisa insere-se no grupo-alvo 1- Bibliometria para especialistas em bibliometria, ao se propor a contribuir para a discussão e refinamento do índice h e seus derivados, por meio do estabelecimento de novos indicadores que mensurem a dispersão da distribuição das citações de um pesquisador. Embora não seja objetivo principal desta pesquisa, ao analisar o conjunto de pesquisadores bolsistas produtividade em pesquisa, nível 1, esta pesquisa insere-se, ainda que tangencialmente, no grupo-alvo 2, contribuindo para o conhecimento do desempenho científico de pesquisadores da área da Matemática. Os indicadores propostos medem o desvio, ou a dispersão, das citações dentro e fora do núcleo Hirsch, podendo ser utilizados para a avaliação da pesquisa científica e

auxiliar em possíveis tomadas de decisões para órgãos destinados à política e gestão científica. Logo, considera-se que este estudo também está inserido no grupo-alvo 3.

1.2 Indicadores bibliométricos e cientométricos: trajetória histórico-conceitual

A produção científica nacional tem grande influência no desenvolvimento econômico e social de um país. Por isso, há a necessidade de se avaliar essa atividade e seus resultados, sendo um dos critérios mais relevantes, as publicações científicas (DORTA-GONZÁLEZ, P; DORTA-GONZÁLEZ, M., 2010). Essa avaliação tem por objetivo analisar de maneira quantitativa as publicações científicas, por meio de indicadores que caracterizem as variáveis relacionadas à produtividade e à visibilidade.

Santos e Kobashi (2005) consideram que os indicadores quantitativos não representam uma verdade absoluta sobre o estado da ciência e da tecnologia, mas configuram-se como aproximações da realidade ou como uma expressão incompleta dela. Sugerem que a utilização desses indicadores seja comparativa para ter significado, a fim de contextualizar e de relativizar os resultados, e que o excesso de confiança seja evitado. Consideram a seleção e a construção de indicadores adequados aos diferentes contextos serem uma tarefa extremamente complexa.

Diante dessa complexidade, parte dos esforços em Ciência, Tecnologia e Inovação concentram-se na elaboração de metodologias apropriadas para a formulação de indicadores adequados e representativos do universo (temática, disciplina, campo, área, país, entre outros) analisado.

Narin (1994) considera que o uso de indicadores quantitativos possibilita o entendimento do desenvolvimento científico e tecnológico contemporâneo. Sugere que existem três pressuposições básicas para a análise bibliométrica e estas podem ser utilizadas para avaliar a atividade científica, em três diferentes perspectivas: 1. indicadores de atividade; 2. indicadores de impacto e 3. indicadores de ligação.

No Brasil, a preocupação com o desenvolvimento de indicadores representativos vem se fortalecendo nas últimas décadas, dado o reconhecimento, por parte dos governos e da comunidade científica, da necessidade de dispor-se de instrumentos adequados para a definição de diretrizes, a alocação de investimentos e de recursos, a formulação de programas e a avaliação de atividades relacionadas ao desenvolvimento científico e tecnológico no país. Todavia, apesar de a atenção e de os esforços recentes no sentido de construírem-se metodologias, instrumentos e indicadores contextualizados, essa questão não está ainda plenamente solucionada (SPINAK, 1998; MUGNAINI *et al.*, 2004).

Um indicador em Ciência e Tecnologia (C&T) é uma medida, em geral quantitativa, usada para expressar, para quantificar ou para representar conceitos relacionados ao desempenho, ao processo ou ao grau de desenvolvimento científico ou tecnológico de um campo do conhecimento. Logo, mostram-se necessários e relevantes para a análise e para a avaliação das diferentes áreas da ciência (JANNUZZI, 2002; GRÁCIO, OLIVEIRA, 2012).

De acordo com Mugnaini *et al.* (2004), originalmente, os indicadores de C&T integravam os sistemas de indicadores econômicos, uma vez que existe uma dependência entre Produto Interno Bruto (PIB) de um país e seu dispêndio em C&T. Todavia, nas últimas décadas, observam-se o desenvolvimento e a ampliação conceitual e metodológica destes indicadores, de forma autônoma dos indicadores econômicos. Essa configuração resulta da evolução histórica na forma de entender o processo de produção do conhecimento científico e tecnológico.

Mugnaini *et al.* (2004) elucidam, ainda, que inicialmente os esforços em C&T eram mensurados por meio de indicadores de *input* (baseados em insumo), como o volume de investimento em pesquisa científica e tecnológica. A partir da década de 1960, os indicadores de *output* (focados em resultados) começaram a ser utilizados, devido à necessidade de dispor-se de medidas que permitissem aos tomadores de decisão avaliarem o retorno dos investimentos aplicados.

Nesse contexto, surge a cientometria, considerada um instrumento de Sociologia da Ciência, dedicada ao estudo e à análise da evolução, do comportamento e do impacto social das ciências. Abrangendo o sistema de pesquisa como um todo, utiliza indicadores que buscam realizar associações de causas e de efeitos dentro do sistema. A cientometria é, também, intitulada “ciência das ciências” (PRICE, 1963; MUGNAINI, *et al.*, 2004; SPINAK, 1998).

Na cientometria, os indicadores baseiam-se em procedimentos e em técnicas matemáticas e/ou em análises estatísticas, que têm um importante papel junto aos órgãos nacionais de avaliação em C&T por fornecerem medidas objetivas dos resultados relativos aos recursos investidos e aos avanços científicos obtidos, assim como subsídios para a visualização e análise de domínios científicos potenciais para o investimento de recursos financeiros subsequentes. Assim, esses indicadores emergiram com o objetivo de planejar, de monitorar e de avaliar as atividades em C&T (MUGNAINI, *et al.*, 2004; VELHO, 1997; SPINAK, 1998).

Os indicadores em C&T podem ser definidos como números sínteses que mensuram e expressam pontualmente características, comportamentos e tendências científicas de países, de domínios do conhecimentos, de periódicos, de instituições e de pesquisadores, entre outros, e destinam-se a: a. identificar e visualizar o desempenho científico e tecnológico dos países; b.

identificar trabalhos e projetos promissores de uma área; c. auxiliar as tomadas de decisões de gestores da política de C&T e d. analisar o domínio em que a comunidade científica está inserida e entender o ciclo de gestação, de reprodução e de disseminação da ciência, entre outros (SANTOS, KOBASHI, 2005; MUGNAINI, *et al.*, 2004).

Como ressaltado, Narin (1994) disse que, os indicadores podem ser agrupados em três grupos: **1. indicadores de atividade** que mensuram a atividade e desenvolvimento da ciência; **2. indicadores de impacto** que mensuram o impacto dessas atividades de pesquisa e **3. indicadores de ligação** que mensuram as ligações entre as organizações produtoras das atividades científicas e as ligações de conhecimento entre suas áreas ou temáticas.

Em concordância, Gregolin *et al.* (2005), disseram que os indicadores bibliométricos compõem um ferramental capaz de analisar quantitativamente a ciência e podem ser agrupados em indicadores de produção, de citação e de ligação, os quais compreendem:

- a) **Indicadores de produção científica:** baseiam-se na contagem do número de publicações, por país, por temática ou área do conhecimento, por tipologia documental, por ano, por instituição, pesquisador, entre outros. Os indicadores de produção baseados na intensidade da produção científica, notadamente, identificam a produtividade, mas não a qualidade das publicações. Como exemplos desses indicadores, citam-se, ainda, as taxas de crescimento da produção científica e tecnológica, porcentagens, meia-vida das publicações e as distribuições de produtividade (Lei de Lotka), de uso de vocabulários (distribuição de Zipf) e de classificações de periódicos (distribuição de Bradford).
- b) **Os indicadores de citação:** baseiam-se no número de citações e buscam mensurar a noção de impacto, a influência e a visibilidade de uma publicação, em nível micro (autores), meso (periódico e instituição, entre outros) ou macro (área do conhecimento e país, entre outros), ao atribuírem crédito aos autores (nível micro), que se estendem, consequentemente, para os níveis meso e macro. No entanto, devem ser compreendidos como parâmetros complexos, pois podem não ser equivalentes e nem estar correlacionados à qualidade científica.
- c) **Os indicadores de ligação (ou relacionais):** destinam-se à análise das intensidades de relações entre dois indivíduos (pesquisadores, palavras-chave, temáticas, instituições, periódicos, países, entre outros), sejam elas simétricas, como as relações de coautoria, de cocitação, de acoplamento bibliográfico ou de coocorrência de palavras-chave, ou assimétricas, como as relações citante-citado, relação autor-temática, autor-evento, entre outras. Esses indicadores permitem a visualização da proximidade ou da

similaridade entre os indivíduos envolvidos na relação analisada, contribuindo para se identificar e evidenciar, por exemplo, comunidades sociais, cognitivas e epistêmicas. Normalmente, representados por meio de redes, elaboradas pela abordagem de Análise de Redes Sociais e utilizam a análise de *clusters* para gerar os agrupamentos dos indivíduos envolvidos na relação analisada, a partir da proximidade ou da similaridade observada nesta relação.

Segundo Narin (1994), esses indicadores podem ser agrupados quanto a quatro aspectos distintos, a saber:

1. ***Questões relacionadas à política:*** caracteriza a produção científica e tecnológica das nações e regiões, por meio da análise das atividades científicas;
2. ***Análise estratégica:*** caracteriza a produção de publicações ou patentes dos centros de pesquisa e universidades, ou do rendimento tecnológico das empresas por meio da análise conjunta dessas atividades científicas;
3. ***Análise tática:*** trata do seguimento das atividades de pesquisa e de desenvolvimento realizadas em determinadas áreas ou que estejam relacionadas com problemas científicos e tecnológicos; e
4. ***Recuperação da informação convencional:*** identifica as atividades e as pessoas envolvidas na pesquisa e no seu desenvolvimento.

Considerando que o objeto de estudo desta pesquisa está no escopo dos indicadores de citação, por serem índices relacionados ao índice *h* de Hirsch, apresentam-se, de forma mais detalhada, na Seção seguinte, conceitos, definições e propriedades desta tipologia de indicadores.

1.2.1 Indicadores de citação

Baseada no conjunto de referências arroladas no final de cada trabalho científico, a análise de citação constitui importante parte da Bibliometria. Segundo Araújo (2006), a análise de citação permite identificar características específicas da comunicação científica:

- i. autores mais citados;
- ii. frente de pesquisa;
- iii. procedência geográfica e/ou institucional;
- iv. tipologia documental mais citada;
- v. meia-vida da literatura usada;

- vi. obsolescência da literatura;
- vii. idioma predominante na literatura citada;
- viii. *core* de periódicos que compõem um campo.

Logo, a análise de citação tem por objetivo medir o impacto e a visibilidade de determinados autores, instituições, periódicos ou país dentro de uma comunidade científica e possibilita a análise das fontes de informações utilizadas. Apesar de haver críticas relacionadas à sua utilização, os estudos de citação têm demonstrado confiabilidade como instrumento avaliativo do impacto das publicações (VANZ, CAREGNATO, 2003; GLÄNZEL, 2003).

Algumas questões associadas ao uso de indicadores de citação referem-se: às diferenças epistemológicas e de práticas de comunicação científica entre as diferentes áreas do conhecimento; às diferentes motivações para a citação e para a não citação; à própria natureza da publicação (se é de revisão ou original); à barreira linguística; à origem da publicação (de países periféricos ou *mainstream*) (VANZ, CAREGNATO, 2003; OLIVEIRA, GRÁCIO, 2011).

As autocitações também têm sido objeto de análise, de reflexão, de questionamentos e de críticas na comunidade científica. Glänzel (2003) aponta que a autocitação é condenada por alguns estudiosos por ser uma forma de reforçar sua posição na comunidade científica, ao amplificar os seus indicadores de citação, avultando e adulterando o real impacto das suas pesquisas na comunidade científica. Em contrapartida, há pesquisadores que as consideram como essenciais para a comunicação científica, ao revelarem o lastro científico do pesquisador e sua consistência na temática. Segundo ainda o autor, a falta quase absoluta de autocitação, durante um longo período de tempo, é tão patológica quanto uma presença intensa de autocitação, uma vez que a primeira pode indicar o caráter neófito do pesquisador na temática da publicação, e a segunda pode evidenciar um isolamento e a falta de comunicação do pesquisador.

P. Dorta-González e M. Dorta-González (2010) ressaltam que existem diversos indicadores clássicos, dentre eles, pode-se citar: número de artigos publicados, número total de citações e citações por artigo que, de forma individual, não refletem satisfatoriamente o êxito da carreira profissional de um pesquisador. Além disso, esses indicadores não são considerados robustos, pois podem ser afetados, tanto por artigos poucos citados, como por aqueles muito citados.

1.2.2 Classificação dos indicadores cientométricos

Conforme Vinkler (2010) os indicadores cientométricos podem ser classificados de acordo com o número de conjuntos de dados utilizados e, com a aplicação de referência, podem ser indicadores 1. simples, 2. compostos ou 3. complexos.

1. **Indicadores Simples:** esses indicadores utilizam um único aspecto dos sistemas cientométricos; são representados por uma única variável cientométrica e com um único nível hierárquico. Pode-se citar, como exemplo, o número total de citações de um conjunto de publicações.
2. **Indicadores Complexos:** esses indicadores utilizam dois ou mais conjuntos cientométricos ou consideram uma única variável cientométrica com mais de um nível hierárquico. Por exemplo, para o cálculo do índice h , deve-se considerar os artigos publicados [conjunto A] e as citações correspondentes [conjunto B].
3. **Indicadores Compostos:** esses indicadores são formados por vários indicadores simples ou complexos, de preferência com fatores de ponderação, e cada fator representa um aspecto especial de um dado sistema cientométrico. Eles consistem em índices parciais bem selecionados, por exemplo, podem caracterizar a atividade global de organizações (Instituição ou país) dentro de um sistema correspondente (universidade, mundo) contendo várias organizações.

Diante da necessidade da elaboração de indicadores bibliométricos e cientométricos que compreendam um significado maior e mais abrangente da atuação e do impacto científico de um pesquisador, Instituição ou país, verifica-se ser necessário o desenvolvimento de índices compostos e/ou complexos que avaliem adequadamente a atividade científica do avaliado. Logo, o Capítulo 2 apresenta diversos índices do tipo h de Hirsch, os quais são considerados indicadores de citação complexos.

2 ÍNDICE DO TIPO h DE HIRSCH

O físico argentino Jorge Hirsch, professor da Universidade da Califórnia, em San Diego, propôs, em 2005, o denominado índice h , um indicador bibliométrico destinado a medir, simultaneamente, o volume e o impacto da produção científica de um pesquisador. O desenvolvimento desse índice surgiu da indagação sobre como quantificar o impacto acumulado dos resultados das investigações de um pesquisador como reflexo da relevância científica da sua atuação (HIRSCH, 2005).

Uma das justificativas apresentadas por Hirsch para a proposição desse indicador foi o fato de os recursos serem limitados e a quantificação, mesmo que muitas vezes seja desconfortável, é necessária para fins avaliativos e comparativos. Destacou que o registro do conjunto de publicações de um indivíduo e de respectivas citações configura um corpus de informação relevante sobre o desempenho e o reconhecimento junto à comunidade científica. O autor propôs, então, o índice h , como um número particularmente simples de ser obtido e útil para caracterizar o impacto científico de um pesquisador.

Hirsch (2005) definiu o índice h da seguinte forma:

Definição 1 (Hirsch, 2005, p.1, tradução nossa): “Um cientista tem índice h , se h de seus N_p artigos têm ao menos h citações cada, e os outros $(N_p - h)$ artigos têm não mais que h citações cada”².

Em que:

h = valor do índice h ;

N_p = número de publicações.

Egghe (2010) destacou a seguinte definição equivalente para o índice h apresentada por Hirsch:

Definição 2 (Egghe, 2010, p.3, tradução nossa): *Ranquear os artigos de um autor e , em ordem decrescente, colocar o número de citações recebidas. O índice h do autor é o maior valor $r = h$ que o trabalho teve no rank $1, 2, \dots, h$, tendo h ou mais citações*³.

²A scientist has index h if h of his/her N_p papers have at least h citations each, and the other $(N_p - h)$ papers have no more than h citations each.

³If we rank the papers of an author in decreasing order of the number of citations they received then this author's h -index is the highest rank $r=h$ such that the papers on ranks $1, 2, \dots, h$ each have h or more citations.

Desse modo, para determinar o índice h , é necessário construir uma sequência numérica⁴ decrescente das citações dos artigos de um pesquisador. Formalmente, a sequência de citações recebidas pelo conjunto de artigos publicados por esse pesquisador pode ser descrita por:

$$(N_1, \dots, N_i, \dots, N_C),$$

em que:

C = total de artigos publicados pelo pesquisador, com $C \geq 1$;

N_i = número de citações recebidas pelo i -ésimo artigo publicado pelo pesquisador, com $N_i \geq N_{i+1}$, para $1 \leq i \leq C$;

C e $N \in \mathbb{N}$ (conjunto dos números naturais).

Para determinar o valor do índice h , deve-se encontrar o maior valor de i , aqui rotulado por h , tal que $h \leq N_h$. Assim, todos os artigos de número de ordem menor que h têm pelo menos h citações e aqueles com número de ordem i maior que h , isto é, os posteriores, não têm mais que h citações cada. A fim de ilustrar a forma de obter o índice h , a Tabela 1 simula a obtenção do índice h de um pesquisador hipotético.

Tabela 1 - Ilustração de como encontrar o índice h de um pesquisador

Posição (i) do artigo na sequência de artigos	Nº de Citações (N_i)
1	60
2	57
3	42
4	38
5	38
6	35
7	26
8	25
9	25
10	20
11	15
12	12
13	11
14	11
15	10
16	9
17	9
18	8
19	7
20	7

Fonte: elaborada pela autora.

Para a Tabela 1, observa-se que $C = 20$ e para $h = 12$, temos que $h = N_h$ e $h = 12$ é o maior valor de i tal que $h \leq N_h$. Portanto, o índice h do autor hipotético é igual a 12, significando

⁴ Uma sequência (x_n) de números reais diz-se crescente se tem $x_1 < x_2 < \dots < x_n < \dots$, isto é, $x_n < x_{n+1}$, para todo $n \in \mathbb{N}$ (LIMA, 1993, p.121).

que este autor tem 12 artigos com pelo menos 12 citações cada e os outros 8 artigos (N_{13} , N_{14} , ..., N_{20}) não tem mais que 12 citações cada.

Observa-se, assim, uma característica relevante do índice h : ser uma medida capaz de combinar quantidade e qualidade da produção acadêmica e, por esse motivo, é considerado um indicador eficaz (MARQUES, 2013; EGGHE, 2010).

Desse modo, o índice h identifica o núcleo mais produtivo das publicações científicas de um pesquisador em termos daqueles trabalhos que receberam maiores número de citações. Esse grupo de artigos é composto pelos primeiros documentos h e é chamado de núcleo Hirsch, termo inserido por Rousseau em 2006. Outra terminologia utilizada para descrever os artigos que compõem o núcleo Hirsch é h -core. Os documentos que compõem esse grupo têm alto impacto, no que diz respeito à carreira do cientista (JIN *et al.*, 2007, p.855; BURRELL, 2007b, p.170; ROUSSEAU, 2006).

A Tabela 2 apresenta uma situação presente em Egghe (2010), em que se comparam pesquisadores com quantidade de artigo e de citações diferentes, mas índices h iguais.

Observa-se, na Tabela 2, que tanto E. Garfield como F. Narin possuem índice $h = 27$. Todavia, para todos os 26 artigos entre a 1ª posição e 26ª posição, Garfield recebeu mais citações do que Narin, evidenciando que o índice h não diferencia a distribuição das citações acima do valor de h dos pesquisadores e, desse modo, não beneficia os artigos altamente citados. Neste contexto, Garfield tem 14 artigos com mais de 100 citações e Narin somente 1, todavia esse fato não influenciou o valor do índice h . Além disso, somando-se o total de citações dos artigos de ordem de 1 a 27 (N_1 a N_{27}) de cada autor, observa-se que, até o valor do índice h (27), Garfield acumulou 3048 citações, ao passo que Narin acumulou 1406 citações, correspondentes a menos da metade (46%) das citações recebidas por Garfield.

Essa característica do índice h associa-se à sua robustez⁵ em dois aspectos: não é influenciada por um conjunto de artigos com poucas citações, nem por artigos altamente citados (EGGHE, 2010).

⁵ Por analogia à definição de teste robusto presente em Vieira (2003), considerou-se uma medida robusta como aquela em que pequenas modificações nos dados não alteram o resultado do índice.

Tabela 2 – Índice h dos autores E. Garfield e F. Narin

E. GARFIELD		F. NARIN	
Posição do artigo (i)	Nº de Citações (Ni)	Posição do artigo(i)	Nº de Citações (Ni)
1	625	1	112
2	149	2	95
3	138	3	86
4	132	4	82
5	132	5	73
6	129	6	71
7	127	7	70
8	111	8	63
9	109	9	59
10	108	10	55
11	107	11	55
12	105	12	53
13	104	13	52
14	101	14	52
15	96	15	44
16	91	16	41
17	89	17	38
18	88	18	37
19	87	19	35
20	85	20	33
21	80	21	33
22	67	22	29
23	63	23	28
24	41	24	28
25	29	25	28
26	28	26	27
27	27	27	27
28	26	28	26
⋮	⋮	⋮	⋮

Fonte: Egghe (2010).

Entre os diversos artigos que discutem os pontos positivos em se utilizar este indicador, citam-se Egghe (2010), Marques (2013) e Hirsch (2005), os quais salientam as seguintes propriedades do índice h :

1. Capacidade de combinar quantidade e impacto da pesquisa em um único indicador;
2. Facilidade de ser obtido e de se compreender;
3. Possibilidade de caracterizar a produtividade científica de um pesquisador com objetividade,
4. Possibilidade de utilizá-lo na tomada de decisões sobre promoções, alocação de verbas e atribuição de prêmios.

5. Desempenho melhor do que o de outros indicadores bibliométricos utilizados para a avaliação da produtividade científica de um pesquisador (fator de impacto, número de artigos, número de citações, citações por artigo e número de artigos altamente citados), quando utilizado de forma isolada.
6. Identifica pesquisadores que produzem de forma consistente bons trabalhos durante um intervalo de tempo e aqueles que escrevem artigos altamente citados, durante um curto período de tempo e depois se estagnam em produção científica.

Por outro lado, Egghe (2010), Marques (2013) e P. Dorta-Gonzalez e M. Dorta-Gonzalez, P. (2010) apontaram algumas limitações do índice h , a saber:

1. Não é indicado para comparar pesquisadores de disciplinas diferentes;
2. Não é indicado para comparar pesquisadores com tempo de titulação diferente, pois tem alta correlação positiva, tanto com o número total de citações, como com o número de publicações dos investigadores, o que tende a favorecer autores com carreiras mais longas do que aqueles com titulação mais recente;
3. Pode ser influenciado pelas autocitações;
4. Dá a livros o mesmo peso que dá aos artigos, tornando complicado comparar pesquisadores de áreas em que há a cultura de publicar os resultados de pesquisa em livros, como as humanidades;
5. Não considera o contexto das citações: não faz distinção entre um artigo de autoria individual ou de um pequeno grupo de colaboradores e um artigo com centenas de autores, cuja participação individual é difícil avaliar;
6. Dificuldade em obter todas as publicações de um autor, o que dificulta o cálculo do índice h , além de poder apresentar problemas relacionados à homografia;
7. Não é adequado para a comparação de pesquisadores de áreas científicas distintas, uma vez que cada uma tem diferentes práticas de publicação e citação e, portanto, o número de citações depende de distintos parâmetros bibliométricos entre áreas, que não estão relacionados com a qualidade.

2.1 Relação do índice h com o número total de citações

Hirsch (2005) descreveu uma relação entre o valor do índice h e o número total de citações (N_c). O limite inferior do número total de citações de um pesquisador é dado por h^2 , uma vez que todos os primeiros h artigos (aqueles com maiores quantidades de citação)

apresentam a relação $N_i \geq N_h \geq h$, para $1 \leq i \leq h$. Logo, para cada $i \leq h$, $N_i \geq h$, então, o número total de citações (N_c) será maior que $h \times h = h^2$. Assim, a relação pode ser escrita da seguinte forma:

$$N_c = \sum N_i = ah^2, \text{ para } 1 \leq i \leq C. \quad (1)$$

Para o exemplo da Tabela 1, tem-se que: $N_c = 465$, $h = 12$ e $h^2 = 144$. Logo,

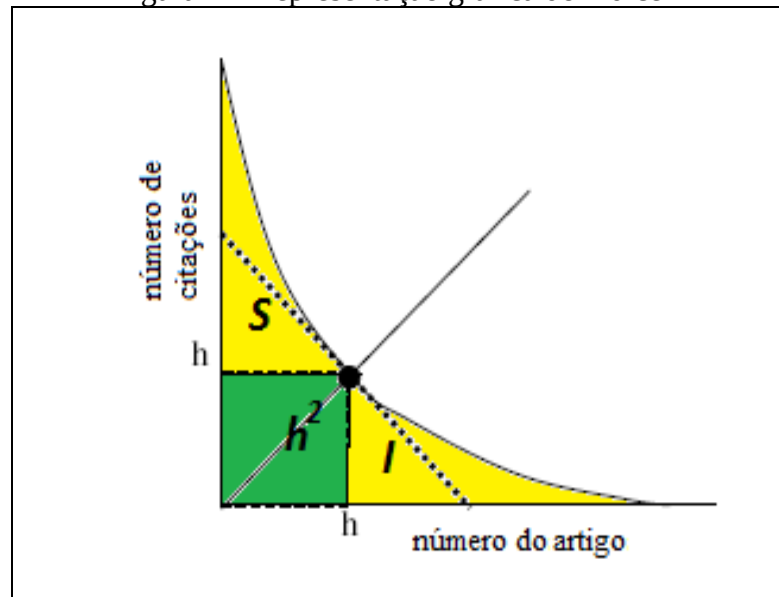
$$N_c = ah^2$$

$$465 = a.144$$

$$a = 465/144 = 3,23.$$

Hirsch (2005) encontrou empiricamente um valor para o coeficiente a entre 3 e 5. No modelo linear, o valor mínimo para o coeficiente a na equação (1) é $a = 2$. Esse valor é considerado quando o número de novas citações por ano (c) for igual ao número de artigos publicados por ano (p). Nesse caso, os artigos com pelo menos h citações e aqueles com menos de h citações contribuem igualmente para o total de citações (N_c). O valor de a será maior para ambos os casos: $c > p$ e $c < p$. Para $c > p$, há mais contribuições para o número total de citações a partir dos artigos altamente citados, ao passo que, para $c < p$, as contribuições dão-se pelos artigos esparsamente citados (HIRSCH, 2005).

A Figura 1 apresenta a interpretação gráfica do índice h , em que o eixo horizontal do plano cartesiano refere-se ao número de ordem, na sequência de artigos publicados por um pesquisador, e o eixo vertical corresponde às citações recebidas pelos respectivos artigos. Para cada par ordenado (número de ordem dos artigos em ordem crescente e respectivo número de citações - em ordem decrescente), assinala-se no plano cartesiano um ponto correspondente. A intersecção a 45 graus da linha com a curva determina o índice h .

Figura 1 – Representação gráfica do índice h 

Fonte: adaptado de Hirsch (2005).

O número total de citações corresponde à área total abaixo sob a curva, correspondente à soma das áreas $S+h^2+I$. Assumindo a segunda derivada não negativa em todos os pontos, a área mínima é dada pela distribuição indicada pela linha pontilhada, quando $\alpha = 2$ na Equação 1.

Em trabalho publicado em 2006, em função da incipiência dos estudos relativos ao recém proposto índice h , Glänzel (2006) apontou que havia incertezas relacionadas à interpretação do índice h , em decorrência da falta de experiência com esse indicador e, também, devido aos aspectos matemático-estatísticos ainda não terem sido completamente estudados, naquele momento. Desde então, diversos estudos têm sido propostos, dedicados à compreensão das propriedades matemáticas desse índice.

Em 2010, P. Dorta-Gonzalez e M. Dorta-Gonzalez (2010) apresentaram uma formulação matemática que leva em consideração as citações recebidas pelas publicações e a influência que essas têm em incrementar os valores futuros do índice h .

Os autores propuseram o raciocínio seguinte: considere as publicações de um autor, coloque em ordem decrescente o número de citações, com N_p representando o número total de publicações deste autor e c_i o número de citações recebidas pela publicação i ($c_1 \geq c_2 \geq \dots \geq c_{N_p}$), com $1 \leq i \leq N_p$. Para $C_j = \sum_{i=1}^j c_i$, a soma das citações para as j publicações mais citadas, com $j \leq N_p$, $C_{N_p} = \sum_{i=1}^{N_p} c_i$, define a soma total de citações do autor e $\bar{C} = C_{N_p}/N_p$ a média de citações por artigo.

A representação do par “posição do artigo na sequência ordenada e respectivo número de citações”, rotulado por (i, c_i) , resulta na distribuição das citações e a linha que une esses pontos consiste na curva das citações, observada na Figura 1.

O índice h é o maior inteiro que satisfaz $c_h \geq h$, ou seja,

$$h = \text{máximo valor de } i \in Np \text{ tal que } c_i \geq i. \quad (2)$$

Geometricamente (Figura 1), é o ponto de intersecção da curva das citações com a bisetritz do primeiro quadrante (linha a 45 graus do eixo horizontal) que sinaliza o autor ter h trabalhos com pelo menos h citações cada.

O índice h fornece um limite inferior $H = h^2$ do número total de citações, não incluindo o volume de citações nas caudas da distribuição de citações: cauda superior “S”, composta por artigos mais citados e cauda inferior “I” por publicações menos citadas (Figura 1). Observam-se as seguintes relações:

$$C_{Np} = H + S + I, \quad (3)$$

$$S = C_h - H, \quad (4)$$

$$I = C_{Np} - C_h. \quad (5)$$

O peso relativo das caudas da distribuição de citações é dada por $C_{Np} \div H$. De acordo com estimativas Hirsch, pode-se dizer que, se $C_{Np} \div H < 3$, a distribuição tem caudas pouco pesadas, enquanto se $C_{Np} \div H > 5$ a distribuição tem caudas pesadas.

Segundo P. Dorta-Gonzalez e M. Dorta-Gonzalez (2010), o índice h beneficia aqueles pesquisadores que têm caudas menos pesadas. Isso porque, à medida que aumenta o peso relativo da cauda, a proporção de eventos considerados no cálculo deste índice é menor e consideravelmente prejudica pesquisadores com uma proporção das caudas S/I elevada. Quando $S/I > 1$, o peso da cauda superior é maior do que a inferior, o que pode indicar que é um investigador seletivo (possui artigos altamente citados), ou seja, ele pode ter menos publicações, mas são de grande impacto. Quando ocorre o inverso, isto é, $S/I < 1$, a relação entre caudas poderia indicar que se tem um pesquisador altamente produtivo (as citações recebidas pelos artigos que compõem o núcleo Hirsch são bem distribuídas).

Além dessa análise, os autores supracitados fazem uma comparação do índice h com a mediana (m). Quando a distribuição das citações é simétrica em relação à bisetritz, tem-se que $m < h$ e $S = I$. O caso $m \ll h$ (onde $\ll$ indica que muito inferior) corresponde a uma distribuição em que $S \gg I$, enquanto $m \gg h$ corresponde a $S \ll I$.

Sabe-se que o índice h não consegue discriminar os diferentes perfis na distribuição de citação, uma vez que não leva em consideração o peso das caudas, mesmo para valores de m distante de h . Pode acontecer de duas distribuições terem o mesmo valor para h , uma com $S \gg I$ e outra com $S \ll I$. Nesses casos, a medição dos parâmetros S e I complementam a análise. Assim, aumenta a capacidade de discriminação nos casos em que exista uma dúvida sobre o desempenho dos investigadores (DORTA-GONZALEZ, P; DORTA-GONZALEZ, M., 2010).

Quadro 1 – Representação e descrição dos indicadores

Indicador	Descrição	Fórmulas	% de citações
N_p	Número de publicações	-	-
$i=1, 2, \dots, N_p$	Índice de publicações em ordem decrescente do número de citações	-	-
c_i	Citações da publicação i	-	-
C_j	Número de citações acumuladas até a j -ésima publicação	$C_j = \sum_{i=1}^j c_i$	-
H	Quadrado de Hirsch	h^2	$\frac{h^2}{C_{Np}} \times 100$
S	Cauda superior	$C_h - H$	$\frac{C_h - H}{C_{Np}} \times 100$
I	Cauda inferior	$C_{Np} - C_h$	$\left(1 - \frac{C_h}{C_{Np}}\right) \times 100$
C_h	Número de citações dos h artigos	$C_h = \sum_{i=1}^h c_i = S + H$	$\frac{H + S}{C_{Np}} \times 100$
C_{Np}	Total de citações	$\sum_{i=1}^{Np} c_i = S + H + I$	100
$\bar{C}$	Média de citações	$\bar{C} = \frac{C_{Np}}{N_p}$	-
S/I	Relação entre as caudas	-	-
C_{Np}/H	Peso relativo das caudas	-	-
M	Mediana do número de citações	-	50

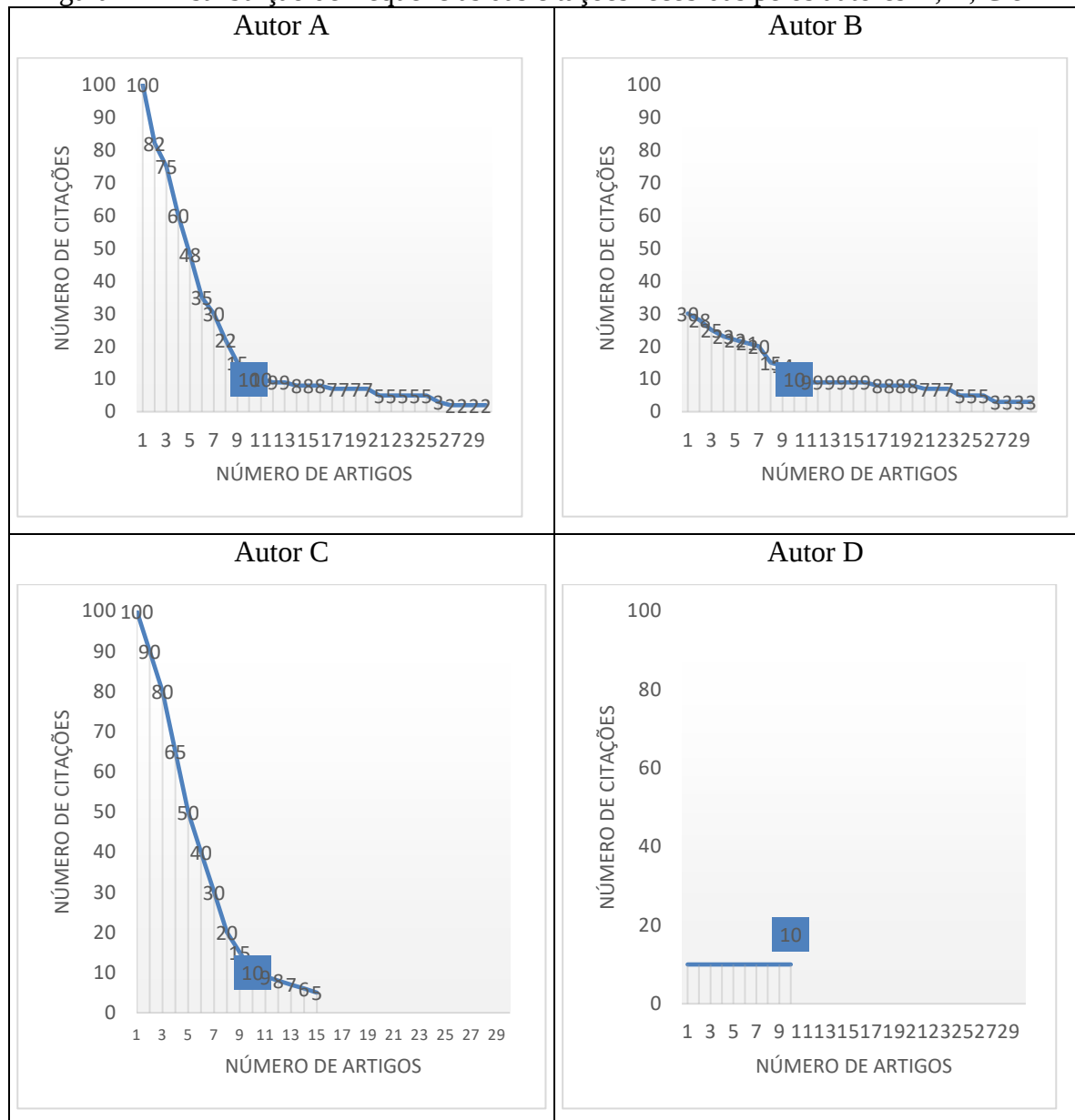
Fonte: adaptado de Dorta-Gonzalez, P. e Dorta-Gonzalez, M. (2010).

2.1.1 Análise de autores com mesmo índice h

A fim de exemplificar a relação entre número de citações e o valor do índice h , na Figura 2 apresentam-se situações hipotéticas relativas à distribuição das citações recebidas pelas publicações e o valor do índice h de quatro autores fictícios: A, B, C e D. O índice h desses 4 pesquisadores é $h = 10$, mas o número de publicações (N_p) e as citações recebidas pelas publicações ($c_i \geq c_j$ com $1 \leq i \leq j \leq N_p$) diferem significativamente entre os 4 autores fictícios. O autor A tem 30 artigos publicados ($N_p = 30$) e possui artigos altamente citados, com $c_1 = 100$ citações, como também artigos com poucas citações. O autor B também tem 30 artigos

publicados, mas seu artigo expoente (c_1) recebeu 30 citações. O cientista C tem 15 artigos publicados ($N_p = 15$), mas tem trabalhos altamente citados, com $c_1 = 100$ citações, $c_2 = 90$ citações, $c_3 = 80$ citações, e muito poucos artigos pouco citados. O autor D tem 10 artigos ($N_p = 10$) e cada um deles recebeu 10 citações ($c_i = 10$, para $1 \leq i \leq 10$).

Figura 2 – Distribuição de frequências das citações recebidas pelos autores A, B, C e D



Fonte: elaborada pela autora.

Observa-se na Figura 2 que, embora os 4 pesquisadores fictícios apresentem diferentes distribuições das citações recebidas, todos têm índice $h = 10$, desconsiderando tanto os artigos altamente citados como as distintas quantidades de artigos publicados por cada pesquisador, assim como não conseguindo diferenciar o desempenho desses pesquisadores.

Diante desse cenário, os debates concentram-se sobre quão o índice h é uma medida significativa do desempenho de um cientista (BORNMANN, *et al* 2008; EGGHE, 2010), visto que, observando somente o resultado do índice h , os autores fictícios da Figura 2 têm desempenhos iguais.

As distribuições de frequências das citações dos autores A, B e C, apresentadas na Figura 2, mostram que o comportamento de tais pesquisadores pode ser modelado pela distribuição probabilística Exponencial Estendida ou pela Lei de Potências.

Redner (1998) ressaltou que, para um número pequeno de citações, a Exponencial Estendida é adequada (número de citações menor que 500), enquanto a um número grande de citações os dados ajustam-se melhor por meio da Lei de Potências. Além disso, observou que o comportamento das citações pode ser modelado pela mistura das duas funções.

A distribuição probabilística da Exponencial Estendida, com parâmetro α , pode ser descrita pela equação (6):

$$f(x) = \alpha e^{-\alpha x}, x > 0, \quad (6)$$

em que,

α é a taxa da distribuição e deve satisfazer $\alpha > 0$.

A função de probabilidade da Lei de Potências é descrita por:

$$f(x) = Cx^{-\alpha}, \quad (7)$$

em que,

C: constante de normalização, tal que, as probabilidades para todos os valores de x devam somar 1.

α : parâmetro escala, geralmente estimado após a aplicação do logaritmo natural ($\ln$).

Em relação aos autores fictícios presentes na Figura 2, considerando que os autores A e C somam 593 citações e 535 citações, respectivamente, a distribuição de probabilidade que melhor modela o comportamento do impacto científico desses pesquisadores é a Lei das Potências, expressa pela função (7). O autor B soma 342 citações e, portanto, sua distribuição de probabilística pode ser representada pela Exponencial Estendida.

A distribuição do autor D assemelha-se a uma distribuição Uniforme Discreta ($f(x) = 1/N_p$). Nesse caso, conforme Bornmann *et al.* (2008), o desempenho científico do pesquisador é considerado constante (evento raro).

De acordo com a abordagem de P. Dorta-Gonzalez e M. Dorta-Gonzalez (2010), esses autores têm o mesmo índice h , mas diferentes indicadores de desempenho, os quais complementam o contexto da sua atuação e do impacto científico. A Tabela 3 apresenta essa relação.

Tabela 3 – Indicadores do contexto de desempenho para os autores A, B, C e D

Indicador	Autores			
	A	B	C	D
N_p	30	30	15	10
C_{Np}	593	342	535	100
h	10	10	10	10
H	100	100	100	100
C_h	477	208	500	100
S	377	108	400	0
I	116	134	35	0
S/I	3,25 (autor seletivo)	0,81 (autor produtivo)	11,43 (autor seletivo)	-
C_{Np}/H	5,93 (cauda pesada)	3,42 (cauda normal)	5,35 (cauda pesada)	1 (cauda leve)
M	4	8	4	5

Fonte: elaborada pela autora.

A partir da Tabela 3, observa-se que os totais de citações recebidas são diferentes entre os autores. Além disso, os autores A e C possuem razão $S/I > 1$, portanto, são considerados seletivos. Complemente-se que eles têm distribuição de citação com cauda pesada, pois a relação C_{Np}/H foi maior que 5. Tais valores indicam que esses pesquisadores têm publicações de grande impacto, uma vez que o referido comportamento é refletido nesses indicadores.

O autor B foi considerado produtivo, pois tem razão $S/I < 1$ e a sua distribuição de citação tem cauda considerada normal, pois varia entre 3 e 5, significando que o pesquisador não tem um artigo altamente citado. O autor D é considerado de desempenho constante, pois todos os artigos publicados têm o mesmo número de citações. Destaca-se, todavia, que probabilisticamente a situação hipotética desse autor fictício é rara.

Ao analisar a mediana da distribuição das citações dos autores, verifica-se que os investigadores A e C têm $m = 4 < h = 10$ e $S > I$, ou seja, metade das citações recebidas pelos autores concentram-se nos quatro primeiros artigos e se distanciam do valor do índice h (o índice h de A se encontra no percentil 80 e de C no percentil 93). O pesquisador B tem $m = 8 < h = 10$ e $S < I$, nota-se que há uma proximidade maior do valor da mediana e do índice h (o qual se localiza no percentil 60). A proximidade da mediana caracteriza o perfil do pesquisador como sendo mais produtivo do que seletivo.

Existe, na literatura, a discussão em torno da interpretação do índice h . Egghe (2010) questionou sobre o que ele mensura, por não ser uma média, não ser um percentil, não ser uma

fração e ser, em seu ponto de vista, uma medida de performance, de impacto e de visibilidade totalmente nova. Acredita-se que o índice h seja uma medida de posição⁶.

2.2 Índice h nas diferentes bases de dados

Existem na literatura alguns estudos comparativos no que se refere ao valor do índice h nas diferentes bases de dados.

Bar-Ilan (2008) compara os índices h de um conjunto de pesquisadores israelenses com base nas citações obtidas da *Web of Science*, *Scopus* e *Google Scholar*. Os resultados obtidos por meio do *Google Scholar* são consideravelmente diferentes dos resultados coletados via *Web of Science* e *Scopus*.

Jasco (2008a, 2008b, 2008c) analisam os valores dos indicadores obtidos em diferentes fontes de informações. Ressaltam que os aspectos práticos para a determinação do índice h precisam ser examinados, porque algumas características de conteúdo, banco de dados e *software* podem influenciar fortemente os valores do indicador. Além disso, destacam que o processo de rastreamento e de contagem de citações é muito demorado para se analisar e corrigir manualmente. Concluíram que existe um manejo não cuidadoso dos dados obtidos por meio do *Google Scholar* e que a base de dados *Scopus* é a mais realista e justa para a representação do desempenho científico de um pesquisador.

Franceschet (2010) apresentou um estudo de caso para acadêmicos da Ciência da Computação, utilizando, de forma comparativa, como fonte de dados *Web of Science* e *Google Scholar*. O estudo conclui que o *Google Scholar* calcula indicadores significativamente mais altos do que a *Web of Science* e as classificações baseadas no índice h mostram um grau moderado de variação entre as bases.

Silva e Grácio (2017) realizaram uma pesquisa analisando os índices h para 20 pesquisadores representativos na temática “Estudos Métricos da Informação” nas fontes de dados — *Scopus*, *Web of Science* e *Google Acadêmico* —, com o objetivo de identificar e analisar possíveis diferenças entre os valores desses indicadores e examinar a congruência dos índices nessas bases. Identificaram que o valor do índice h não diferiu estatisticamente entre as bases de dados *Scopus* e *Web of Science*, considerando um nível de significância de 5%. Por outro lado, esse índice diferiu significativamente dos valores obtidos a partir do *Google Acadêmico*.

⁶ Medidas de posição: é um valor que descreve ou representa os dados de acordo com sua localização (SEBASTIÃO, 2017).

Dessa maneira, destacam-se a importância e a influência da base de dados escolhida pelo avaliador na determinação do valor do índice h . Assim, o avaliador deve escolher criteriosamente o banco de dados que utilizará para a obtenção do índice h , a fim de que possa retratar o mais fidedignamente possível o pesquisador avaliado.

2.3 Variações do índice h

Atualmente, existem estudos direcionados à obtenção de variações, de ponderações e de generalizações do índice h , em função das limitações e das questões apontadas nas seções anteriores, uma vez que esse indicador tem se mostrado relevante para a análise bibliométrica de pesquisadores, de grupos, de instituições ou de países e sido utilizado como critério de mérito e de excelência por agências de fomento e por instituições para distribuição de verbas e reconhecimento acadêmico-científico.

Bornmann *et al.* (2008) e Egghe (2010) apresentam diversas variações do índice h , todas têm o intuito de minimizar ou de dirimir os problemas e questões levantadas. A seguir apresentam-se algumas variações do índice h .

2.3.1 Quociente m

Hirsch (2005) desenvolveu o quociente m , pois identificou que o índice h poderia ser influenciado pelo tempo e que o pesquisador dedica-se à pesquisa.

O quociente m é obtido pela equação:

$$\text{Quociente } m = \frac{h}{y} \quad (8)$$

em que, h = índice h e y = número de anos desde a publicação do primeiro artigo.

No entanto, determinar o número de anos que o pesquisador dedica-se à pesquisa, muitas vezes, não é uma tarefa fácil, por isso, Silva e Grácio (2015a) propuseram relativizar o índice h pelo tempo (t) de publicação que o autor tem na revista científica ou na base de dados analisada. Dessa forma, elimina-se a influência do tempo no valor do índice h e propicia uma visualização mais fidedigna do impacto e da inserção do pesquisador na comunidade científica.

O indicador relativo h/t é definido da seguinte forma: *dado o índice h do autor e o tempo t de publicação, obtém-se a razão h/t , segundo a qual o autor publicou, em média, h/t artigos com pelo menos h/t citações cada, por ano* (SILVA; GRÁCIO, 2015a).

As autoras analisaram um universo de 20 autores, considerados mais produtivos na revista *Scientometrics*, ou seja, eles deveriam ter pelo menos 20 artigos científicos publicados, no período de 35 anos (1980-2014). Concluíram que a razão h/t pode complementar a avaliação de um autor quando são analisados pesquisadores com diferentes tempos de publicação em uma determinada revista, visto que esse quociente elimina a influência do tempo, pois traz o índice h por ano, ou seja, o índice h médio.

2.3.2 Índice g

Egghe (2006, p. 8) desenvolveu uma variação de h denominada *índice g* , definida por: “um conjunto de artigos tem índice g se g é o posto mais alto, de modo que os artigos principais têm juntos, pelo menos g^2 citações”.

A métrica g tem a propriedade *índice $g \geq$ índice h* , pois g considera em sua estrutura a quantidade de citações recebidas pelo grupo de artigos mais citados, ao passo que o índice h não as considera (EGGHE, 2010). Por essa característica, o autor considera o índice g superar uma das desvantagens de h .

Silva e Grácio (2015b) exemplificaram o cálculo do índice h e g para o pesquisador Egghe. Para esse pesquisador, recuperaram-se 47 artigos na base *Scopus* e as respectivas quantidades de citações recebidas. Para o cálculo do índice g , é necessário que se tenham a frequência acumulada das citações (C_i) e o número de cada artigo elevado ao quadrado (i^2). O índice g é o posto (número do artigo) mais alto, de modo que os artigos principais têm, juntos, pelo menos g^2 citações. A Tabela 4 apresenta estas informações.

Neste exemplo (Tabela 4), Egghe apresenta índice $h = 11$ e índice $g = 32$, pois escreveu 32 artigos que têm, conjuntamente, pelo menos 1024 citações e índice $h = 11$, confirmando a propriedade apresentada por Egghe (2010): *índice $g \geq$ índice h* .

Uma das características desse indicador é atribuir maior peso ao número de citações recebidas pelo artigo. Todavia, também o índice g pode atribuir maiores valores para pesquisadores com maior tempo de dedicação à pesquisa.

Tabela 4 – Cálculo do índice h e g para o pesquisador Egghe

Ano de publicação	Índice h		Índice g	
	I	c_i	C_i	i^2
2006	1	524	524	1
2006	2	167	691	4
1992	3	48	739	9
2002	4	39	778	16
1996	5	31	809	25
2000	6	27	836	36
1993	7	23	859	49
2000	8	17	876	64
1988	9	14	890	81
2008	10	13	903	100
1986	11	12	915	121
2004	12	11	926	144
1995	13	11	937	169
2012	14	10	947	196
1991	15	10	957	225
2001	16	9	966	256
2010	17	8	974	289
2008	18	8	982	324
2002	19	8	990	361
1999	20	8	998	400
2002	21	7	1005	441
2010	22	6	1011	484
2013	23	5	1016	529
1999	24	5	1021	576
2009	25	4	1025	625
1998	26	4	1029	676
1997	27	4	1033	729
1987	28	4	1037	784
2007	29	3	1040	841
2005	30	3	1043	900
2013	31	2	1045	961
2011	32	2	1047	1024
2011	33	2	1049	1089
2011	34	2	1051	1156
2010	35	2	1053	1225
2009	36	2	1055	1296
2005	37	2	1057	1369
1999	38	2	1059	1444
2013	39	1	1060	1521
2013	40	1	1061	1600
2006	41	1	1062	1681
2014	42	0	1062	1764
2011	43	0	1062	1849
2010	44	0	1062	1936
2005	45	0	1062	2025
1994	46	0	1062	2116
1992	47	0	1062	2209

Fonte: adaptado de Silva e Grácio (2015b).

Silva e Grácio (2015b) também propuseram relativizar o índice g pelo tempo de publicação no veículo de comunicação científica que disseminou sua publicação. Para o índice g , o indicador relativo g/t é definido por: *dado o índice g do autor e o tempo t de publicação no veículo de comunicação, obtém-se a razão g/t , segundo a qual o pesquisador publicou, em média, g/t artigos por ano, os quais têm, juntos, pelo menos g^2/t citações.*

Desta forma, g/t determina as citações recebidas conjuntamente pelos artigos publicados, por ano. Egghe (2010) diz que o índice h não consegue distinguir, dentre os autores com mesmo índice, aqueles que têm artigos mais citados. Burrell (2007a) afirma que, para autores com elevado tempo de dedicação à pesquisa, os indicadores em questão são proporcionais.

Shing e Kumar (2014) apontam que, na literatura, a métrica g vem ganhando popularidade e visibilidade na avaliação da produtividade científica. Esses autores desenvolveram uma análise comparativa do desempenho dos índices h e g , com investigadores da área de Física, na qual observaram que o valor de g é 1,5 vezes maior que o de h .

Silva e Grácio (2015b) realizaram uma análise comparativa entre o índice h e o índice g , a fim de verificar seus desempenhos, propriedades e relações, como indicadores de impacto na comunidade científica, tomando como universo de aplicação os investigadores mais produtivos em “Estudos Métricos”, temática consignada à área de Ciência da Informação, e obtiveram que g é 1,76 vezes maior que o índice h . Além disso, concluíram que a razão g/h pode auxiliar e/ou complementar a avaliação de um autor, quando são analisados pesquisadores com diferentes tempos de publicação em um determinado periódico, visto que esse quociente elimina a influência do tempo.

2.3.3 Índice $h^{(2)}$

O indicador foi proposto por Kosmulski (2006) e é descrito como *o maior número natural tal que seus $h^{(2)}$ trabalhos mais citados receberam cada um, pelo menos, $[h^{(2)}]^2$ citações*. Segundo o autor, um índice $h^{(2)}$ de 20, por exemplo, significa que o pesquisador analisado publicou pelo menos 20 artigos, dos quais cada um foi citado pelo menos 400 vezes.

Para os dados da Tabela 4, identifica-se que $h^{(2)} = 5$ e $[h^{(2)}]^2 = 25$, o artigo da posição $i = 5$ é o maior número natural, que tem pelo menos 25 citações (no caso, este artigo tem 31 citações). Em outras palavras, o autor analisado tem 5 artigos, com pelo menos 25 citações cada.

É notável que esse indicador beneficia aqueles artigos altamente citados, ou seja, aqueles artigos expoentes do autor. Além disso, esse índice tem a seguinte propriedade matemática: valor $h^{(2)} \leq h$.

Jin *et al.* (2007) ressaltam que a principal vantagem do índice $h^{(2)}$ em relação ao índice h é que ele reduz o problema de precisão, uma vez que necessita de um conjunto de publicações

menor e, por isso, exige menos trabalho para verificar a precisão dos dados das publicações, especialmente no caso em que existam homógrafos.

2.3.4 Índice a

Jin (2006) desenvolveu o índice a , o qual considera as citações acumuladas até a posição do índice h , ponderado pelo valor de h .

$$\text{índice } a = \frac{1}{h} \sum_{i=1}^h c_i, \quad (9)$$

em que c_i = número de citações recebidas pelo artigo i , para $1 \leq i \leq h$.

Para os dados presentes na Tabela 4, a soma das citações até h é 915 e $h = 11$. Logo, o índice a é dado por $915/11 = 83,18$ citações/artigo.

O índice a é interpretado como o impacto médio do núcleo Hirsch (h -core). Burrell (2007b) mostra que o índice a depende do tempo de dedicação à pesquisa do pesquisador.

2.3.5 Índice r

Jin *et al.* (2007) observaram que o índice a penaliza os pesquisadores com maiores índices h , pois o cálculo de a envolve uma divisão por h . Dessa forma, em vez de dividirem por h , propuseram a utilização da raiz quadrada da soma de citações do núcleo Hirsch. Por remeter à raiz quadrada, designaram o novo indicador por índice r . Em termos formais, o índice r é obtido por:

$$\text{índice } r = \sqrt{\sum_{i=1}^h c_i}, \quad (10)$$

em que h = índice h , e c_i = número de citações recebidas pelo artigo i , para $1 \leq i \leq h$.

Assim como o índice a , o índice r mede a intensidade das citações do núcleo Hirsch, por isso, pode ser sensível a artigos altamente citados. Bornmann *et al.* (2008) ressaltam que o índice r é sensível à influência de alguns artigos expoentes do autor.

Para os dados da Tabela 4, tem-se que a soma das citações é 915. Logo, o índice $r = \sqrt{915} = 30,25$ citações^{1/2} (nota-se que a unidade de medida não tem uma interpretação real).

2.3.6 Índice *ar*

Jin (2007) propôs o índice *ar*, que constitui uma adaptação do índice *r*, ao relativizar esse índice pela idade das publicações no núcleo Hirsch. Segundo o autor, o valor desse índice pode tanto aumentar como diminuir ao longo do tempo. Considerou esse fato como uma característica benéfica e necessária para um índice de avaliação de pesquisa, ao ser sensível às mudanças de desempenho do cientista.

O índice *ar* é definido como a raiz quadrada da soma do número de citações por ano dos artigos incluídos no *h-core*.

$$\text{índice } ar = \sqrt{\sum_{i=1}^h \frac{c_i}{a_i}}, \quad (11)$$

em que a_i = número de anos desde a publicação do artigo i , h = índice h e c_i número de citações do artigo i , para $1 \leq i \leq h$.

A Tabela 5 apresenta os elementos necessários para calcular o índice *ar*, conforme Egghe. Para este exemplo, o índice $ar = \sqrt{74,68} = 8,64$ (*citações/ano*)^{1/2}. Assim como o índice *a*, a unidade de medida desse indicador não é facilmente interpretada.

Tabela 5 – Cálculo do índice *ar* para o pesquisador Egghe

Ano de publicação	Número do artigo (i)	Número de anos até 2017(a _i)	C _i	C _i /a _i
2006	1	11	524	47,64
2006	2	11	167	15,18
1992	3	25	48	1,92
2002	4	15	39	2,60
1996	5	21	31	1,48
2000	6	17	27	1,59
1993	7	24	23	0,96
2000	8	17	17	1,00
1988	9	29	14	0,48
2008	10	9	13	1,44
1986	11	31	12	0,39
Total			915	74,68

Fonte: elaborada pelo autor.

2.3.7 Índice h normalizado (h^n)

Em função do fato de o índice h ser uma métrica que depende do número de publicações do pesquisador, Sidiropoulos *et al.* (2007) propuseram uma versão para este indicador, denominado índice h normalizado, relativizando-o pelo número de publicações. Enunciam o índice h normalizado (h^n) da seguinte maneira: *um pesquisador tem índice normalizado $h^n = h/N_p$, se h de seus artigos N_p receberam pelo menos h citações cada, e o restante ($N_p - h$) artigos não receberam mais de h citações.*

Aplicando este indicador para os dados da Tabela 4, temos índice $h = 11$ e $N_p = 47$. Logo, o índice $h^n = 11/47 = 0,23$ citações/artigo.

2.3.8 Índice h_w

Egghe e Rousseau (2008) desenvolveram o índice h_w (índice h ponderado pelo impacto da citação) que, assim como o índice ar , é sensível às mudanças de desempenho do pesquisador em sua vida científica. O índice h_w é definido como:

$$\text{índice } h_w = \sqrt{\sum_{i=1}^{r_0} c_i} \quad (12)$$

em que r_0 = a maior posição (menor valor) de i tal que $r_w(i) = (C_i \div h) \leq c_i$, isto é, o total de citações até o artigo i (rotulado por C_i) dividido pelo índice h é menor que o número de citações do artigo i (rotulado por c_i), para $1 \leq i \leq h$.

A fim de exemplificar a obtenção do índice h_w , a Tabela 6 apresenta o cálculo para os dados da Tabela 4, relativos ao pesquisador Egghe, em que $h = 11$, c_i = nº de citações para o artigo i e C_i = total de citações acumuladas até o artigo de posição, para $1 \leq i \leq 11$.

Para este exemplo, $r_0 = 2$, pois o artigo que ocupa a segunda posição (menor valor de $i = 2$) satisfaz a relação $r_w(i) \leq c_i$, pois $r_w(2) = 62,81 < 167$.

Logo, o índice h_w para Egghe é:

$$\text{índice } h_w = \sqrt{\sum_{j=1}^2 c_j} \sqrt{524 + 167} = \sqrt{691} = 26,29 \text{ citações}^{1/2}.$$

Tabela 6 – Cálculo do índice h_w , utilizando os dados do autor Egghe

Posição	citações (c_i)	C_i	$r_w = C_i/h$
1	524	524	524/11 = 47,64
2	167	691	691/11 = 62,81
3	48	739	739/11 = 67,18
4	39	778	778/11 = 70,73
5	31	809	809/11 = 73,55
6	27	836	836/11 = 76,00
7	23	859	859/11 = 78,09
8	17	876	876/11 = 79,64
9	14	890	890/11 = 80,91
10	13	903	903/11 = 82,09
11	12	915	915/11 = 83,18

Fonte: elaborada pelo autor.

Os autores ressaltaram que o índice h_w é maior ou igual ao índice h . Para o exemplo da Tabela 6, Egghe tem $h_w = 26,29 > h = 11$. Além disso, verifica-se que a unidade de medida desse indicador, assim como os índices a e ar , não é interpretável.

2.3.9 Índice m

Bornmann *et al.* (2008) definem o índice m como o número mediano de citações recebidas por artigos do núcleo Hirsch. A opção pela mediana é baseada no fato de que, geralmente, a distribuição das citações é assimétrica, portanto, a medida de tendência central mais indicada é a mediana.

Considerando os dados da Tabela 4, a mediana dá-se para o artigo da posição $i = 6$, o qual tem 27 citações. Logo, $m = 27$ citações.

Segundo Bornmann *et al.* (2008), pode-se dividir os índices do tipo h de Hirsch em dois grupos. O primeiro é composto por: índice h , quociente m , índice g e índice $h^{(2)}$, que descrevem o núcleo da produtividade de um cientista. Os índices índice a , índice m , índice r , índice ar e o índice h_w retratam o impacto dos artigos do núcleo.

2.3.10 Índice v

Riikonen e Vihinen (2008) definiram o índice v como a porcentagem de artigos que formam o índice h . Destacam que um único indicador não consegue avaliar todos os aspectos de um pesquisador. Por isso, sugerem que os analistas utilizem o índice h e v para verificar o reconhecimento dos cientistas e a amplitude de sua produtividade.

Os autores apontam que o índice h cresce muito lentamente com um aumento no número de publicações, todavia o índice v consegue identificar variações na proporção de artigos altamente citados para pesquisadores com valores semelhantes de índice h .

Para os dados da Figura 2, relativa aos quatro autores fictícios com mesmo valor para o índice h , mas com comportamento distinto tanto na distribuição de citação e como na quantidade de publicações científicas, os autores A e B com 30 artigos publicados, cada um, e índice $h = 10$, logo o índice v destes dois autores é $0,33$, obtido a partir da razão $h/Np = 10/30$, o que significa que o índice h é composto por 33% das publicações destes pesquisadores. O autor C tem 15 artigos e $h = 10$, então $v = 10/15 = 0,67$, ou seja, o índice h do autor C é composto por 67% de sua publicação e, por fim, D utiliza 100% de sua produção para compor seu índice h .

2.3.11 Índice π

Vinkler (2009) propôs o índice π a fim de poder avaliar, de forma comparativa, pesquisadores de temáticas semelhantes, dando ênfase aos artigos mais influentes (maiores quantidades de citação). Para P o número de artigos publicados por um pesquisador e denotando-se por P_π o número de artigos no conjunto daqueles altamente citados, em que $P_\pi = \sqrt{P}$, o índice π é igual a um centésimo do número total de citações do conjunto P_π dos artigos mais citados, denominado por $C(P_\pi)$. O cálculo do índice π pode ser apresentado por:

$$\pi = 0,01 \cdot C(P_\pi) = 0,01 \cdot C(\sqrt{P}) . \quad (13)$$

Galton (1971 *apud* VINKLER, 2009), sugeriu a aplicação da raiz quadrada do número total de publicações como limiar de eminência, pois a raiz quadrada do número de pessoas de um país, de profissionais ou de outras categorias pode representar o número de indivíduos que compõem a elite científica de uma área do conhecimento, conforme enunciado por Price, na Lei da Raiz Quadrada ou do Elitismo, em 1963.

2.3.12 Índice e

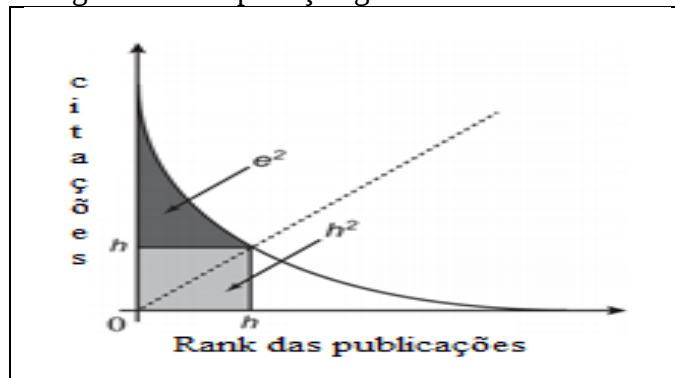
Zhang (2009) definiu o índice e como forma de complementar as informações relativas ao desempenho de um pesquisador, avaliado pelo índice h . O autor aponta que para as citações ignoradas, uma vez que as citações adicionais dos artigos que formam o núcleo Hirsch são completamente ignoradas no índice h . Em função disso, propõe o índice e , levando-se em conta as citações em excesso recebidas por todos os artigos do h -core. Denota por e^2 a seguinte formulação matemática:

$$e^2 = \sum_{j=1}^h (c_j - h) = \sum_{j=1}^h c_j - h^2, \quad (14)$$

em que, c_j representa as citações recebidas pelo j -ésimo artigo e e^2 denota o excesso de citações recebidas pelo núcleo Hirsch.

A Figura 3 apresenta a interpretação geométrica do índice e .

Figura 3— Interpretação geométrica do índice e .



Fonte: Zhang (2009).

O índice e depende do valor de h , visto que e^2 representa o excesso de citações recebidas por todos os artigos no h -core, além de citações h^2 .

Observe que quanto maior o valor de e , maior o excesso de citações e , portanto, têm-se mais perdas de informações sobre o comportamento das citações, quando se utiliza somente o índice h . Similarmente, quando o índice h é usado para avaliar cientistas individuais, quanto menor o valor de e , mais confiável é o índice h . Considerando um caso extremo (como o autor fictício D da Figura 2), quando e é igual a zero (bastante improvável em situações reais), h descreve completamente a distribuição das citações para as publicações no h -core. Caso contrário, quando $e \neq 0$, o índice h sempre perderá informações sobre as citações, o que é complementado por e (ZHANG, 2009).

Com a finalidade de exemplificar a obtenção do valor do índice e , consideraram-se os dados da Tabela 1, construiu-se a Tabela 7.

Tabela 7– Elementos para o cálculo do índice e .

Nº do Artigo	Nº de citações (c_i)	C_j	$(c_j - h)$
1	60	60	48
2	57	117	45
3	42	159	30
4	38	197	26
5	38	235	26
6	35	270	23
7	26	296	14
8	25	321	13
9	25	346	13
10	20	366	8
11	15	381	3
12	12	393	0
Soma			249

Fonte: elaborada pelo autor.

Tem-se, assim, que o autor tem o índice e :

$$e^2 = \sum_{j=1}^h (c_j - h) = 249$$

$$e = \sqrt{249} = 15,78.$$

2.3.13 Índice w de Wu

Wu (2010) definiu o índice w de forma semelhante ao h , mas focando apenas nos artigos altamente citados. O autor enuncia o índice w como: *Se w artigos de um pesquisador têm pelo menos $10w$ citações cada e os outros artigos têm menos de $10(w+1)$ citações, o pesquisador tem índice w .*

Para os dados do autor Egghe apresentado na Tabela 4, o índice w tem valor igual a 3, pois os três primeiros artigos têm mais que 30 citações cada e os outros não têm mais que 40 citações cada.

Alguns índices do tipo Hirsch aglomeram informações de dois índices e são denominados *índices de agregação*. Entre eles, citam-se os índices q^2 e hg .

2.3.14 Índice q^2

Considerando que cada índice do tipo Hirsch propõe-se a medir diferentes aspectos da produção científica dos pesquisadores, Cabrerizo *et al* (2009) propuseram a junção de alguns

deles, a fim de haver uma medida simples e mais completa do desempenho científico de um pesquisador, e desenvolveram o índice q^2 , baseado na média geométrica do índice h e da mediana do número de citações do h -core (o índice m).

$$q^2 = \sqrt{h \times m}. \quad (15)$$

Os autores propositores do índice q^2 justificam a construção desse índice em função do índice h e do índice m , em função de h ser um índice robusto e descrever o número de artigos (dimensão quantitativa) no núcleo produtivo de um pesquisador. Já o índice m é a medida de tendência mais indicada para retratar o impacto dos artigos do núcleo Hirsch (dimensão qualitativa), em distribuições não simétricas, como geralmente ocorre nas distribuições de citação para o conjunto de publicações de um autor. Segundo ainda os autores, por considerar diferentes dimensões da produção científica de um pesquisador, o índice q^2 permite uma visão mais global dessas publicações.

Considerando as informações da Tabela 4, os índices de Egghe são: $h = 11$ e $m = 27$. Logo, o índice $q^2 = \sqrt{11 \times 27} = 17,23$ citações.

2.3.15 Índice hg

Alonso *et al* (2010) propuseram um novo índice para caracterizar a produção científica dos pesquisadores com base nos índices h e g e o denominaram *índice hg* .

O indicador caracteriza a produção científica dos pesquisadores, baseando tanto em h como em g , a fim de manterem-se as vantagens de ambas as medidas, bem como minimizarem-se suas desvantagens. Os autores concordam que h e g incorporam várias propriedades interessantes sobre as publicações de um pesquisador e por isso devem ser considerados para medir sua produção científica.

O índice hg de um pesquisador é calculado como a média geométrica de seus índices h e g , assim:

$$hg = \sqrt{h \times g}. \quad (16)$$

Este índice tem as seguintes propriedades:

- (a) $h \leq hg \leq g$,
- (b) $hg - h \leq g - hg$.

Para verificar a propriedade (a) considerou-se a seguinte relação: $0 \leq h \leq g$, então:

$$h = \sqrt{h \cdot h} \leq \sqrt{h \cdot g}, \text{ pois } g \geq h. \quad (16.1)$$

Por outro lado,

$$g = \sqrt{g \cdot g} \geq \sqrt{h \cdot g}, \text{ também porque } g \geq h. \quad (16.2)$$

Logo, de (13.1) e (13.2), tem-se:

$$h \leq \sqrt{h \cdot g} \leq g. \quad (16.2)$$

Para a verificação da propriedade (b) considere:

Como $0 \leq h \leq g$, então $0 \leq \sqrt{h} \leq \sqrt{g}$. Logo, $0 \leq \sqrt{g} - \sqrt{h}$.

Novamente, como $\sqrt{g} - \sqrt{h} \geq 0$, então:

$$\begin{aligned} (\sqrt{g} - \sqrt{h})^2 &\geq 0 \\ (\sqrt{g})^2 - 2 \cdot \sqrt{h} \cdot \sqrt{g} + (\sqrt{h})^2 &\geq 0 \\ g - 2 \cdot \sqrt{h \cdot g} + h &\geq 0 \\ g - \sqrt{h \cdot g} - \sqrt{h \cdot g} + h &\geq 0 \\ g - \sqrt{h \cdot g} &\geq \sqrt{h \cdot g} - h \leftrightarrow \sqrt{h \cdot g} - h \leq g - \sqrt{h \cdot g} \end{aligned} \quad (16.3)$$

Com base na propriedade (b), o índice hg corresponde a um valor mais próximo de h do que de g . Alonso *et al.* (2010), ponderaram que esta propriedade pode evitar o problema da influência que um artigo de alto impacto pode acarretar no índice g .

Os autores defendem algumas vantagens ao se utilizar o índice hg :

- a. É muito simples de calcular, uma vez que os índices h e g tenham sido obtidos.
- b. Fornece maior granularidade na escala de medida em relação aos índices h e g . Isso é especialmente interessante quando comparado com o índice h , pois aumentar h é difícil e é comum encontrar pesquisadores que têm o mesmo índice h , embora tenham número muito diferente de publicações e de citações. O índice hg fornece uma maneira mais refinada de comparar cientistas.
- c. O índice hg é avaliado na mesma escala que os índices h e g (ambos representam o número de artigos que cumprem uma condição sobre suas

citações). Assim, o índice hg é fácil de entender e de comparar aos índices existentes.

- d. Considera as citações dos artigos altamente citados (o índice h é insensível a esses artigos), mas reduz significativamente o impacto desses artigos (os autores consideram como um inconveniente do índice g), nesse sentido, garante um melhor equilíbrio sobre o impacto dos melhores artigos do autor.

Franceschini e Maisano (2011) relataram que, embora a construção de indicadores bibliométricos compostos pareça adequada, em muitos casos pode levar a resultados equivocados. Os autores supracitados analisaram as propriedades das escalas de medida do índice hg , sugeridas por Alonso *et al.* (2010), e apresentaram três críticas principais, a saber: 1. a escala hg é o resultado de uma composição dos índices h e g , que são definidos em escalas ordinais distintas, 2. as classes de equivalência de hg são questionáveis e a taxa de substituição entre h e g pode mudar arbitrariamente, dependendo dos valores específicos de h e g e 3. o aumento aparente na granularidade de hg , em relação a h e g , é ilusório e questionável. Esses argumentos foram verificados por meio de diversos exemplos.

Para os dados da Tabela 4, tem-se que os índices h e g de Egghe são: 11 e 32, respectivamente. Logo, o índice $hg = \sqrt{11 \times 32} = 18,76$. Observa-se que o valor obtido é mais próximo de $h = 11$. Como o autor analisado tem artigos altamente citados, o índice g torna-se relativamente maior do que h e hg , pois g considera em sua estrutura os efeitos dessas altas citações.

Roberts (1979 *apud* FRANCESCHINI e MAISANO, 2011) ressalta que os indicadores compostos são sínteses de dois ou mais indicadores em que se utiliza alguma operação matemática (adição, multiplicação, média, potência, dentre outras). Discorre que, ao se compor indicadores, suas escalas de medição devem ser levadas em consideração e a interpretação deve ser muito cuidadosa.

A seguir, apresentam-se algumas variações do índice h que consideram o tempo de dedicação à pesquisa dos autores.

2.3.16 Índice h^c

Em função de o índice h não levar em consideração a “idade” do artigo em seu cálculo, Sidiropoulos *et al.* (2007) refletem sobre a seguinte situação: suponha que um pesquisador tenha vários artigos com alto impacto e tenha um índice h relativamente grande, mas após um período

de tempo torne-se inativo; esse fato não é distinguido pelo índice h . À vista disso, ressaltaram a necessidade de definir um indicador que agregue informação relativa ao tempo de publicação dos artigos.

Para um artigo i , a fim de relativizar seu impacto em função de idade (tempo de sua publicação), os autores propuseram o seguinte escore $S^c(i)$:

$$S^c(i) = \gamma \cdot \frac{c_i}{[Y(atual) - Y(i) + 1]^\delta} \quad (17)$$

em que: $Y(atual)$ é o ano atual, isto é, o ano em que se está calculando o escore para o artigo i ; $Y(i)$ é o ano de publicação do artigo i ; c_i = número de citações recebidas pelo artigo i .

Observa-se que, quanto maior o valor estabelecido para δ , menor relevância atribui-se às citações recebidas pelos artigos mais antigos, isto é, artigos com maior tempo de publicação. De acordo com os autores, para $\delta = 1$, o escore $S^c(i)$ é o número de citações recebidas pelo artigo i , dividido pela “idade” do artigo. O coeficiente γ utilizado na formulação do escore dá maior peso às citações recebidas pelos artigos mais recentes. Desse modo, para $\gamma = 4$ e $\delta = 1$, observam-se as seguintes ponderações às citações dos artigos: para um artigo publicado no corrente ano, suas citações contam quatro vezes; para um artigo publicado há 4 anos, suas citações contam apenas uma vez; para um artigo publicado há 6 anos, suas citações contam 4/6 vezes, e assim por diante.

Dessa maneira, um artigo antigo, perde gradualmente o seu “valor”, mesmo que ainda receba citações. Logo, o escore dá maior peso aos artigos mais recentes, por isso, definiram o índice h^c da seguinte forma:

Um pesquisador tem índice h^c atual se h^c de seus Np artigos obtiver um escore de $S^c(i) \geq h^c$ cada, e os outros artigos ($Np - h^c$) obtêm um escore de $S^c(i) \leq h^c$ (SIDIROPOULOS *et al.*, 2007).

A dificuldade em se utilizar esse indicador está em atribuir adequadamente o valor de γ e δ , uma vez que os autores não propuseram uma forma de obtê-los.

2.3.17 Índice h^t

Sidiropoulos *et al.* (2007) propuseram o índice h^t a fim de responderem à seguinte situação problema: suponha que um pesquisador contribuiu com artigos de alto impacto em

uma determinada década e tem um índice h relativamente alto. Se as citações a esses artigos persistirem na atualidade, há o caso de um autor influente, cujas contribuições continuam a influenciar os autores mais recentes. Por outro lado, se, atualmente, esses artigos não são citados, é um indicativo que o tema neles presente está ultrapassado ou desatualizado.

Outro aspecto relevante no envelhecimento das citações é que essas podem revelar tendências, isto é, o cientista cujo trabalho é considerado pioneiro e estabelece uma nova linha de pesquisa.

Para retratar essa situação, Sidiropoulos *et al.* (2007) propuseram uma versão que se opõe ao índice h contemporâneo: em vez de atribuir a cada artigo do pesquisador um peso decadente dependendo de sua idade, a cada citação de um artigo é atribuído um peso que decresce exponencialmente, expresso em função da “idade” da citação. Assim, estima-se o impacto do trabalho de um pesquisador em função de tempo particular. Nesse ensejo, o interesse concentra-se em verificar se o artigo ainda recebe citações. A seguinte equação é definida:

$$S^t(i) = \gamma \times \sum_{\forall x \in C(i)} (Y(atual) - Y(x) + 1)^{-\delta}. \quad (18)$$

em que, γ , δ , $Y(i)$ e $S(i)$ para um artigo i são definidos como no índice contemporâneo h^c .

Nesse caso, o índice de tendência h^t pode ser definido por: *um pesquisador tem índice h^t se h^t de seus N_p artigos obtiver um escore de $S^t(i) \geq h^t$ cada, e o restante $(N_p - h^t)$ artigos obter $S^t(i) \leq h^t$* (SIDIROPOULOS *et al.*, 2007).

Os autores ressaltaram que para valores de δ de próximo de zero, o impacto da penalidade de tempo é reduzido e, para $\delta = 0$ e $\gamma = 1$ essa variante coincide com o índice h .

Destaca-se que os autores não evidenciaram fórmulas para definir as constantes γ e δ presentes no índice, apresentando nas aplicações presentes no artigo apenas a intuição para o estabelecimento desses parâmetros.

2.3.18 Índice h dinâmico

Rousseau e Ye (2008) propuseram o índice h dinâmico, que depende de informações do núcleo Hirsch, mais especificamente do número de citações recebidas por artigos pertencentes ao h -core e do recente aumento de h . O índice h dinâmico necessita de três elementos relacionados ao tempo, a saber: o tamanho e o conteúdo do h -core, o número de citações recebidas e a velocidade h .

Os autores desse índice apresentaram a seguinte situação problema: suponha que dois pesquisadores tenham o mesmo índice h e o mesmo número de citações no h -core, mas o primeiro não tem alteração no valor de seu índice h há algum tempo, enquanto o índice h do segundo autor vem subindo durante o intervalo de tempo analisado. Para avaliações que considerem a intensidade da atividade científica, o segundo cientista é, provavelmente, o mais produtivo.

Rosseau e Ye (2008) propuseram:

$$R(T) \times v_h(T), \quad (19)$$

em que: $R(T)$ denota o índice r , definido por Jin *et al.*, (2007), calculado por meio da raiz quadrada da soma de todas as citações recebidas por artigos pertencentes ao h -core no tempo T e assumindo um modelo contínuo para h [$h(t)$], então, v_h é a derivada $dh(T)/dt$.

De acordo com os autores, é necessário determinar um ponto de partida ($T = 0$), o qual sugerem não ser o início da carreira do autor pesquisado. Por exemplo, pode-se tomar $T = 0$ como sendo 5 anos atrás (ou qualquer outro momento apropriado). Precisa-se determinar, também, uma forma de encontrar v_h , cuja obtenção envolve modelagem matemática. Além disso, os autores sugerem que, se o índice for utilizado para fins avaliativos, as autocitações devem ser removidas.

A dificuldade principal em se utilizar esse indicador concentra-se na obtenção de $v(t)$, visto que consideram a distribuição probabilística de h contínua, enquanto, por definição, o índice h deveria ser modelado por uma distribuição discreta.

2.3.19 Índice *hpd*

Kosmulski (2009, p.345) definiu o índice *hpd* da seguinte forma: *Um pesquisador tem índice hpd se hpd de seus artigos tiver pelo menos citações hpd por década cada, e seus outros artigos têm menos de (hpd + 1) citações por década cada.*

Segundo o autor, esse índice permite comparar o desempenho científico de pesquisadores com diferentes tempos de dedicação à pesquisa, pois analisa a produção e o impacto de um pesquisador durante uma década. Admite que esse intervalo de tempo (10 anos) pode beneficiar ou prejudicar o pesquisador analisado, todavia considera que o período de tempo de um ano também é arbitrário.

2.3.20 Indicador de vitalidade do impacto

A fim de propor uma definição operacional de “excelência em pesquisa”, Rons e Amez (2009) apresentaram um indicador bibliométrico que busca refletir a vitalidade do impacto da produção de um pesquisador, com base na mudança de volume ao longo do tempo das publicações citadas.

$$IV(y_1, n) = \frac{\left[n \frac{\sum_{i=1}^n P(y_i)/i}{\sum_{i=1}^n P(y_i)} - 1 \right]}{\left[\sum_{i=1}^n \frac{1}{i} - 1 \right]} \quad (20)$$

com $n > 1$, $y_{i+1} = y_i - 1$ e $\sum P(y_i) > 0$.

A razão será maior do que 1 para publicações quando aumentam as citações naquele determinado período de tempo, será menor que 1 quando diminuem naquele determinado intervalo de tempo. Será igual a 1 quando o trabalho do cientista receber uma quantidade igual de citações todos os anos.

2.3.21 Índice k

Ye e Rousseau (2010) propuseram o índice k , a fim de retratarem não apenas as citações no h -core, mas também, aquelas presentes nas caudas na distribuição de citações. Tinham como objetivo verificar como dá-se a relação cauda-núcleo em função do tempo ($C_T(t)/C_H(t) = (C(t) - C_H(t))/C_H(t)$) e comparar o impacto das citações nessa relação.

De acordo com os autores, é necessário determinar, para cada ano t , o número acumulado de publicações do um determinado ano até o ano t e o correspondente número acumulado das citações recebidas por essas publicações durante esse mesmo período tempo, para se definir o termo da série temporal em que se baseia o índice $k(t)$.

Considera-se que as variáveis $C(t)$, $P(t)$, $C_H(t)$, e $C_T(t)$ signifiquem, respectivamente, conjunto de citações, conjunto de publicações, conjunto de citações recebidas pelo h -core e o conjunto de citações recebidas pela cauda, no período de tempo t . Logo, define-se o índice k por meio da seguinte expressão:

$$k(t) = \frac{\frac{C(t)}{P(t)}}{\frac{C_T(t)}{C_H(t)}} = \frac{C(t)C_H(t)}{P(t)C_T(t)} = \frac{C(t)C_H(t)}{P(t)(C(t) - C_H(t))}. \quad (21)$$

Ao aplicar o índice k em situações práticas, Ye e Rousseau (2010) concluíram que, para uma determinada série temporal, esse indicador diminuiu na maioria das aplicações de acordo com um modelo de Lei de Potências, enquanto a relação cauda-núcleo aumentou.

2.3.22 Índice s

De Visscher (2010) definiu o índice s como a taxa de citação por artigo, por meio da modelagem probabilística, usando o modelo exponencial, a fim de oferecer uma estimativa do número de citações por artigo no tempo infinito, dividido por 10. Desse modo, segundo o autor, o índice s é uma medida que analisa o impacto projetado para um artigo de um pesquisador analisado.

O autor considera esse índice complementar a outros índices que medem o impacto global, pois pode distinguir entre maximizar a quantidade ou a qualidade da produção de um autor, e destaca que a sua principal vantagem é reduzir o viés inerente aos artigos com mais idade, visto que esses tiveram mais tempo para acumular citações do que os artigos recentes.

2.3.23 Índice f

Franceschini e Maisano (2010) propuseram o índice f como o intervalo de tempo entre os artigos com pelo menos uma citação mais 1, considerando, assim, o tempo gasto para preparar o primeiro artigo incluído no conjunto. O índice f é definido por:

$$f = Y_{max} - Y_{min} + 1, \quad (22)$$

em que Y_{max} é o ano do artigo mais recente com pelo menos uma citação e Y_{min} é o ano do artigo mais antigo que recebeu pelo menos uma citação.

Consideram que esse índice complementa h com as informações relativas à idade de publicação. Uma de suas principais características é que não compromete a simplicidade original e a imediação da compreensão do índice h .

Para os dados da Tabela 4, identifica-se que o artigo mais recente, que recebeu pelo menos uma citação, data de 2013 e o artigo mais antigo, com pelo menos uma citação, data de 1986. Portanto, o índice f é igual a:

$$f = (2013 - 1986) + 1 = 28 \text{ anos.}$$

Logo, Egghe tem índice $h = 11$ e obteve esse valor em 28 anos.

2.4 Estudos envolvendo as variações do índice h

Uma busca⁷ realizada, em 07/02/2017, na base de dados *Scopus*, sobre a incidência do índice h e de algumas variações deste indicador, a fim de se observar a continuidade e reconhecimento desses indicadores na literatura científica, recuperaram-se 324 trabalhos, dentre os quais 282 são artigos científicos, publicados em maior quantidade no periódico *Scientometrics*. Em 2009, ocorreu a maior incidência de publicações, com 46 artigos. Além disso, aproximadamente 60% dos trabalhos foram publicados entre os anos 2009 e 2013, com os dois anos subsequentes apresentando uma menor frequência de publicação sobre o assunto. No ano de 2016, a incidência de publicação voltou a crescer (é duas vezes maior em relação ao ano anterior).

Para o índice g , recuperaram-se 52 trabalhos, dos quais, aproximadamente, 58% foram publicados entre os anos de 2008 e 2010 e 46 (88%) foram disseminados na forma de artigo. O periódico *Journal of Informetrics* publicou 33% dos artigos que abordam este índice.

Sobre o índice r , foram recuperados 11 artigos científicos e a maior incidência de publicação (64%) ocorreu nos anos de 2008 e 2009. Para o índice a , recuperaram-se 8 trabalhos dos quais 62,5% foram publicados no período 2007 a 2010. Os demais índices não tiveram impacto significativo nos periódicos indexados na base de dados *Scopus*. Verifica-se que os índices mais comentados e estudados são os índices h e g .

2.5 Proposta de um indicador do tipo Hirsch

Sabe-se que o índice h não beneficia artigos altamente citados e nem considera aqueles com poucas citações. Ele pode ser considerado como uma medida de posição, logo para uma análise estatística ficar completa necessita-se confrontar uma medida de tendência central e uma de variabilidade. As medidas de tendência central são casos especiais de medidas de posição (localização) e representam a centralidade de um conjunto de dados. Citam-se, entre elas, a

⁷Utilizaram-se os seguintes termos de busca: “*h-index*” and Hirsch, “*g-index*” and Hirsch, “*h2-index*” and Hirsch, “*a-index*” and Hirsch, “*m-index*” and Hirsch, “*r-index*” and Hirsch, “*ar-index*” and Hirsch, “*weighted h-index*” and Hirsch.

média, a mediana e a moda. As medidas de dispersão determinam a variabilidade (ou a dispersão) dos dados em relação à medida de localização do centro da amostra em análise.

Nesse cenário, propõe-se, como medida complementar a h , o par de índices denominados dc_i (dispersão das citações no interior do h -core) e dc_o (dispersão das citações fora do h -core).

O índice dc_i é definido pela soma das distâncias (desvio) entre o número de citações recebidas pelos artigos do núcleo Hirsch e o índice h , dividida por h . Em termos matemáticos, formula-se dc_i por:

$$dc_i = \frac{\sum_{j=1}^h (cit_j - h)}{h}, \quad (23)$$

em que, cit_j : número de citações do j -ésimo artigo, para $1 \leq j \leq h$; h : índice h .

Em função da sua construção, considera-se que dc_i representa a distância média entre as citações dos artigos do h -core e h . Destaca-se que o numerador de dc_i é equivalente ao valor e^2 do índice e .

O índice dc_o é definido pela soma das distâncias (desvio) entre o número de citações recebidas pelos artigos que estão fora do núcleo Hirsch e o índice h , dividida por h . Em termos matemáticos, formula-se dc_o por:

$$dc_o = \frac{\sum_{j=h+1}^{N_p} (h - cit_j)}{N_p - h}, \quad (24)$$

em que, cit_j : número de citações do j -ésimo artigo, para $h + 1 \leq j \leq N_p$; N_p : número de publicações, h : índice h .

Em função da sua construção, dc_o determina a distância média de citações necessárias para chegar até h . Logo, pode-se considerá-lo como um índice que mede a expectativa de crescimento do índice h .

A Tabela 8 apresenta uma ilustração dos procedimentos necessários para a obtenção dos índices dc_i e dc_o , com base nos dados da Tabela 1.

Tabela 8 – Elementos para o cálculo do dc_i e dc_o

	Nº de Artigos	Nº de Citações	($cit - h$)
	1	60	48
	2	57	45
	3	42	30
	4	38	26
	5	38	26
	6	35	23
	7	26	14
	8	25	13
	9	25	13
	10	20	8
	11	15	3
Índice h	12	12	0
Soma			249
			($h - cit$)
	13	11	1
	14	11	1
	15	10	2
	16	9	3
	17	9	3
	18	8	4
	19	7	5
	20	7	5
Soma			24

Fonte: elaborada pelo autor.

Neste caso, o valor de dc_i será:

$$dc_i = \frac{\sum_{j=1}^h (cit_j - h)}{h} = \frac{249}{12} = 20,75 \text{ citações/artigo.}$$

Ou seja, os artigos que compõem o núcleo Hirsch têm $dc_i = 20,75$ citações/artigo, ou seja, em média, cada artigo distancia-se 20,75 citações de h . Nota-se que cada artigo perdeu em torno de 20,75 citações, pois essas não são contabilizadas no cálculo do índice h .

O valor de dc_o será:

$$dc_o = \frac{\sum_{j=h+1}^{N_p} (h - cit_j)}{N_p - h} = \frac{24}{20 - 12} = 3 \text{ citações/artigo.}$$

Ao considerar os artigos que não contemplam o índice h , verifica-se que, em média, cada artigo necessita receber 3 citações para acompanhar o valor de h .

Se dc_i for um valor baixo, significa que em média, existem poucas citações excedendo h , caso contrário o pesquisador tem artigos influentes que receberam muitas citações.

Na tentativa de fazer uma analogia com o Coeficiente de Variação propõe-se o cálculo de:

$$CV_{dci} = \frac{dc_i}{h} \times 100 \quad e \quad CV_{dco} = \frac{dc_o}{h} \times 100.$$

Como a distribuição das citações não é simétrica e o cálculo do índice dc_i não é centrado na média, e sim em h , identifica-se que todos os valores são maiores que ou iguais a h . Em contrapartida, para dc_o as citações serão menores ou iguais a h . O Quadro 2 traz a escala classificatória para dispersão de CV_{dc_i} e CV_{dc_o} .

Quadro 2 – Escala classificatória para a dispersão do índice dc_i e dc_o

Classificação	CV_{dc_i}	CV_{dc_o}
Muito baixa	$0\% \leq CV_{dc_i} \leq 115\%$	$0\% \leq CV_{dc_o} \leq 15\%$
Baixa	$115\% < CV_{dc_i} \leq 130\%$	$15\% < CV_{dc_o} \leq 30\%$
Média	$130\% < CV_{dc_i} \leq 170\%$	$30\% < CV_{dc_o} \leq 70\%$
Alta	$CV_{dc_i} > 170\%$	$CV_{dc_o} > 70\%$

Fonte: elaborado pelo autor.

Calcula-se a seguir CV_{dc_i} para os dados do exemplo:

$$CV_{dc_i} = \frac{dc_i}{h} \times 100 = \frac{20,75}{12} \times 100 = 173\%.$$

Para essas informações, o valor de CV_{dc_i} é alto (173%), ou seja, a medida de dispersão proposta tem seu valor 73% maior que o índice h , o que significa o pesquisador ter artigos altamente citados e esses influenciaram a variabilidade da cauda superior da distribuição de citações.

Similarmente, se dc_o for pequeno significa que, em média, os artigos que não compõem o h -core têm número de citações próximo ao valor de h e, se esses artigos receberem citações, o autor pode incrementar seu índice h .

$$CV_{dc_o} = \frac{dc_o}{h} \times 100 = \frac{3}{12} \times 100 = 25\%.$$

Para os dados desse autor, verifica-se que CV_{dc_o} é 25%, o que caracteriza uma dispersão baixa. Logo, pode-se concluir que os artigos a não compõem o núcleo Hirsch têm um número de citações próximas, caracterizando uma pequena variabilidade na cauda inferior da distribuição de citações. Isso pode caracterizar, também, que o valor do índice h não é tão representativo, visto que há artigos na cauda inferior da distribuição de citações com citações próximas a h .

A Tabela 9 apresenta os valores obtidos para dc_i e dc_o para os autores A, B, C e D, discriminados na Figura 2.

Tabela 9 – Cálculo do dc_i e dc_o para os autores A, B, C e D

Indicadores Autores	N_p	h	dc_i	CV_{dc_i}	dc_o	CV_{dc_o}
A	30	10	37,7	377%	4,2	42%
B	30	10	10,8	108%	3,3	33%
C	15	10	40	400%	3,0	30%
D	10	10	0	0%	-	-

Fonte: elaborada pelo autor.

Ao se analisar o autor A, conclui-se que, em média, os artigos a comporem o *h-core* distanciam-se de h , 37,7 citações cada. Nota-se que esse valor é 377% maior que o valor do índice h e caracteriza-se uma alta dispersão da distribuição de citações dos artigos pertencentes ao *h-core*, ou seja, o pesquisador A tem artigos expoentes inseridos em seu núcleo Hirsch.

Além disso, observa-se que os autores A e C têm valores parecidos para dc_i , isso significa ambos terem artigos altamente citados e esse fato indicar que o comportamento das citações do núcleo Hirsch desses autores são semelhantes. Tal resultado remete à classificação obtida por meio dos indicadores de contexto de P. Dorta-Gonzalez e M. Dorta-Gonzalez (2010), o qual classificou esses pesquisadores (A e C) como seletivos.

Faz-se importante ressaltar que os autores A e C têm quantidades de publicações diferentes e, apesar de tal diferença, a dispersão das citações na cauda superior desses autores é semelhante.

O autor B apresentou $dc_i = 10,8$ citações/artigo, ou seja, cada artigo distancia, em média, 10,8 citações de h . Observa-se que os artigos integrantes do núcleo Hirsch distanciam-se de h , em média, 8%. Isso significa que esse pesquisador não tem artigos altamente citados. Ao confrontar esses resultados com a abordagem de P. Dorta-Gonzalez e M. Dorta-Gonzalez (2010) conclui-se que esse autor é considerado produtivo.

Desse modo, em uma situação hipotética de concorrência por auxílio financeiro de uma agência de fomento, na avaliação dos candidatos A e B, além do índice h e o número de citações recebidas pelos artigos desses avaliados no núcleo Hirsch, poderiam ser considerados seus respectivos valores de dc_i . Neste caso, a concessão do auxílio financeiro seria para o autor A.

O autor D apresentou $dc_i = 0$, o que significa todos os valores das citações do *h-core* serem iguais ao índice h , nesse caso, não há variabilidade na distribuição de citações.

Ao analisar os resultados para dc_o do autor A, verifica-se que os artigos que não contemplam o índice h têm, em média, 4,2 citações cada. Ou seja, em média, os artigos necessitam de aproximadamente 4 citações cada, para igualar ao valor h . Destaca-se que o dc_o do autor D não existe, pois, o pesquisador não tem artigos fora do núcleo Hirsch.

Com o objetivo de verificar o comportamento dos índices dc_i e dc_o em situações em que os autores tenham índices h semelhantes, mas tempos de publicação distintos, utilizaram-se os dados do trabalho de Silva e Grácio (2015b), as quais tomaram como universo 20 pesquisadores considerados representativos na temática, Estudos Métricos, na revista *Scientometrics*, correspondendo a ter pelo menos 20 artigos científicos publicados no período de 35 anos (1980-2014). De 6 a 10 de fevereiro de 2015, para cada pesquisador, levantaram, na base *Scopus*, informações referentes aos artigos publicados, número de citações por artigo e o ano (t) de publicação. A partir desses dados, calcularam o índice h e verificaram-se o tempo (t) que os autores tinham artigos nessa revista.

O autor Egghe publicou $N_p = 47$ artigos e possui $h = 11$ e tempo $t = 28$ anos. O pesquisador Moya-Anegón produziu $N_p = 25$ artigos, tem $h = 11$ e tempo $t = 16$ anos. Garg tem $N_p = 21$ artigos, $h = 11$ e $t = 29$ anos. Verifica-se que, apesar de terem publicados artigos em quantidades e períodos de tempo diferentes, o índice h deles é igual. No entanto, ao se analisarem os índices dc_i e dc_o , observam-se comportamentos distintos na distribuição de citações.

Egghe apresentou $dc_i = 72,18$ citações/artigo, o que representa uma dispersão alta na distribuição das citações no núcleo Hirsch, ou seja, esse pesquisador tem artigos altamente citados no h -core, a saber, os trabalhos mais citados têm 524 e 167 citações, ao passo que, em referência aos outros, a frequência de citações diminui substancialmente. Ao se analisar a cauda inferior da distribuição de citações, encontra-se $dc_o = 6,92$ citações/artigo e $CV_{dc_o} = 63\%$ e infere-se que haja dispersão média.

Moya-Anegón obteve $dc_i = 17,36$ citações/artigo, o que representa uma dispersão moderada, ou seja, esse pesquisador não teve um artigo expoente tal que fizesse inflacionar a dispersão no h -core. Para a cauda inferior da distribuição de citações, encontrou-se $dc_o = 6,07$ citações/artigo e $CV_{dc_o} = 55\%$, assim, conclui-se a dispersão também ser moderada.

Garg possui $dc_i = 6,27$ citações/artigo, o que representa uma dispersão muito baixa, ou seja, esse pesquisador não tem um artigo altamente citado em seu núcleo Hirsch e tem uma dispersão moderada nos artigos que compõem a cauda inferior da distribuição de citações ($dc_o = 4,20$ citações/artigo).

Portanto, conclui-se que, apesar Moya-Anegón ter tempo de dedicação à pesquisa diferente dos outros analisados, esses autores apresentaram índices h iguais e um comportamento distinto na distribuição das citações no núcleo Hirsch. Identificou-se que os artigos mais citados de Egghe foram publicados em 2006 e estão relacionados à publicação do

índice g . Para Moya-Anegón, o trabalho mais citado data de 2007. Já Garg não apresentou um artigo que fora altamente citado.

Inferre-se que as medidas dc_i e dc_o podem auxiliar um avaliador em situações em que os analisados tenham índices h semelhantes e tempos de publicação diferentes, por se analisar o comportamento da dispersão das distribuições de citações.

Os avaliadores das agências de políticas científicas poderiam utilizar esses índices para orientá-los em uma possível tomada de decisão, ou seja, esses indicadores serviriam como critérios desempate. Neste caso, o autor Egghe apresenta o melhor desempenho por ter tido dispersão alta no núcleo Hirsch, caracterizando ter artigos de alto impacto.

Os autores Rousseau e Vinkler, no trabalho de Silva e de Grácio (2015b), apresentaram $h = 14$ e tempos de publicação semelhantes, 27 e 26 anos, respectivamente. Entretanto, o número de publicações nesse período foi diferente, Rousseau publicou $N_p = 49$ artigos e Vinkler $N_p = 27$ artigos. Observa-se que o comportamento da distribuição das citações da cauda superior dos analisados diferiu.

Para o primeiro, $dc_i = 24,14$ citações/artigo, o que determinou uma dispersão alta ($CV_{dc_i} = 172\%$), caracterizando a esse autor ter artigos de impacto no h -core, o qual foi publicado em 2006. Em contrapartida, Vinkler apresentou $dc_i = 17,36$ citações/artigo, o que determinou uma baixa variação de torno de h ($CV_{dc_i} = 129\%$), além disso, o trabalho mais citado foi publicado em 1986. Nota-se uma diferença substancial em relação ao período em que os artigos expoentes dos pesquisadores foram publicados. Esse tempo difere 20 anos e caracteriza o segundo autor não ter publicações de alto impacto após 1986.

Com o auxílio da medida de dispersão dc_i , conseguiu-se identificar que, apesar de terem índices h iguais, tempos de publicação semelhantes, a análise da distribuição de citações foi importante para determinar o impacto das publicações dos autores analisados e, assim, melhor avaliar os autores, ou seja, conseguiu-se distinguir que mesmo tendo variáveis que se assemelham, como o índice h e o tempo de publicação, o desvio das citações possibilitou se identificar um aspecto do impacto científico em que um dos pesquisadores (Rousseau) apresentou melhor desempenho, a saber: maior dispersão das citações acima do valor do índice h , equivalente a dizer que o desempenho do pesquisador está sub-representado pelo índice h .

Para verificar a validade dos indicadores como novas medidas de dispersão, utilizou-se a mesma amostra do trabalho de Silva e de Grácio (2015b) e calcularam-se a correlação entre dc_i e o desvio padrão dos artigos que compõem o núcleo Hirsch. Conclui-se que são altamente correlacionados ($r = 0,94$). Similarmente, determinou-se a correlação entre dc_o e o desvio

padrão para os artigos da cauda inferior da distribuição de citações e as conclusões foram as mesmas ($r = 0,95$).

Esse resultado garante que os indicadores propostos podem ser utilizados como medidas de variabilidade, assim, podem complementar o ferramental estatístico destinado à área de Bibliometria e auxiliar especialistas em uma possível avaliação.

3. ANÁLISE BIBLIOMÉTRICA PARA OS PESQUISADORES PQ1 EM MATEMÁTICA

3.1 Características associadas aos bolsistas Produtividade em Pesquisa

Segundo o Conselho Nacional de Pesquisa (2018), um bolsista de Produtividade em Pesquisa (PQ) é um pesquisador com titulação mínima de doutor, que se destaca entre seus pares, por sua produção científica atender critérios estabelecidos pelos Comitês de Assessoramento (CA) da sua área de pesquisa do CNPq. Esses pesquisadores bolsistas integram uma das duas categorias presentes na concessão das bolsas: PQ1 e PQ2, de acordo com o tempo de doutoramento, com os pesquisadores PQ1 devendo ter no mínimo 8 oito anos de doutorado por ocasião da implementação da bolsa e ter demonstrado capacidade de formação contínua de recursos humanos e os Pesquisadores PQ2, no mínimo 3 anos.

Para os pesquisadores PQ1, são estabelecidos 4 níveis (A, B, C ou D), com base comparativa entre os seus pares e nos dados dos últimos 10 anos. A diferenciação entre os níveis A, B, C e D é baseada em critérios estabelecidos pelos CA, os quais podem inserir requisitos que julgarem importantes para a área de pesquisa, além daqueles mínimos exigidos pelo CNPq. Devem, todavia, privilegiar a qualidade e o conjunto da obra do pesquisador (CNPq, 2018).

Espera-se ainda que os pesquisadores PQ1 tenham gradual inserção nacional e internacional, por meio de palestras e assessorias *ad hoc* a revistas nacionais e internacionais e de órgãos de financiamento à pesquisa, bem como envolvimento em atividades de gestão científica, incluindo a organização de eventos, a participação em comitês assessores estaduais ou nacionais, em sociedades científicas, em revistas científicas, em assessoria de órgãos de governo estaduais ou nacionais, e em conferências proferidas a convite e/ou em plenárias de congressos.

Entre os bolsistas PQ1, o nível A é reservado a pesquisadores que tenham excelência contínua na produção científica e na formação de recursos humanos e que liderem grupos de pesquisa consolidados. O perfil deste nível de pesquisador deve, na maior parte dos casos, extrapolar os aspectos unicamente de produtividade para incluir aspectos adicionais que mostrem uma significativa liderança dentro da sua área de pesquisa no Brasil e capacidade de explorar novas fronteiras científicas em projetos de risco.

Os pesquisadores PQ1B e PQ1C, além de uma crescente contribuição à formação de recursos humanos e à produção de Ciência e Tecnologia, devem contribuir na organização de grupos de pesquisa e programas de graduação e pós-graduação de sua instituição, e uma

significativa participação em atividades de pesquisa em suas instituições e, particularmente, no caso dos pesquisadores PQ1B, em órgãos de fomento à pesquisa.

Na categoria PQ2, os pesquisadores bolsistas apresentam produtividade científica destacada e exercer atividade de orientação em nível de pós-graduação.

3.2 Descrição do corpus de pesquisa

Optou-se por analisar o conjunto de pesquisadores PQ1 em Matemática pela aproximação com a formação acadêmica da autora da pesquisa e pela consolidação desta área em âmbito nacional, com publicações e pesquisadores de referência e impacto científico nacional e internacional.

Durante o mês de janeiro de 2018, buscaram-se no portal da CAPES os nomes dos pesquisadores PQ1 em Matemática. Identificaram-se 116 pesquisadores, sendo 21 classificados como PQ1A, 36 como PQ1B, 28 como PQ1C e 31 como PQ1D. Em seguida, foram coletadas as informações referentes à produção científica de cada autor na base de dados *Scopus*.

O Quadro 3 contém os nomes dos pesquisadores PQ1 de acordo com os níveis A, B, C e D.

Quadro 3 – Pesquisadores PQ1 em Matemática.

PQ1A	PQ1B	PQ1C	PQ1D
Alcides Lins Neto	Abramo Hefez	Aldo Procacci	Alexander Eduardo A. Mendoza
Alfredo N. Iusem	Adalberto P. Bergamasco	Boyan Slavchev Sirakov	Ali Tahzibi
Aron Simis	Alagacone Sri Ranga	Carlos Kubrusly	Amilcar Pacheco
Artur Avila	Alexandre N. de Carvalho	Claudianor Oliveira Alves	Andrés Koropecski
Benar F. Svaiter	Andre Nachbin	Claudio Gorodski	Carolina Bhering de Araujo
Carlos G. T. A. Moreira	Antonio Carlos Gardel Leitão	Daniel Marinho Pellegrino	Claudio Aguinaldo Buzzi
Carlos Tomei	Artur Oscar Lopes	Dimitar Kolev Dimitrov	Ederson Moreira dos Santos
Cesar L. C. Manco	Bruno César A. Scárdua	Flavio Ulhoa Coelho	Edson Vargas
Dan Marchesin	Claudia A. Sagastizabal	Frederic Gerard C. Valentin	Emanuel Augusto de S. Carneiro
Enrique R. Pujals	Daciberg Lima Gonçalves	Gregório Pacelli F. Bessa	Everaldo Souto de Medeiros
Hermano Frid Neto	Dessislava H. Kochloukova	Gustavo Alberto P. Menzala	Fábio Armando Tal
Israel Vainsencher	Eduardo de Sequeira Esteves	Hossein Movasati	Geraldo Márcio de A. Botelho
Jorge G. Hounie	Gerson Petronilho	Jaime Edilberto M. Rivera	Giovany de J. M. Figueiredo
Keti Tenenblat	Harold William Rosenberg	Jairo Zacarias Goncalves	Henrique Fernandes de Lima
Lorenzo J. Diaz Casado	Helena Judith N. Lopes	Jorge Herbert Soares de Lira	Jaime Angulo Pava
Luiz Carlos P. R. Velho	Henrique Bursztyn	Levi Lopes de Lima	Joao Xavier da Cruz Neto
Marcelo M. V. Silva	Hilário Alencar da Silva	Luis Adrian Florit	José Fábio Bezerra Montenegro
Marcos Dajczer	Jaime Bruck Ripoll	Luiz Henrique de Figueiredo	Ketty Abaroa de Rezende
Paolo Piccione	Joao Marcos Bezerra do O	Marcelo Moreira Cavalcanti	Lev Birbrair
Paulo D. Cordaro	Jorge Passamani Zubelli	Marcos Benevenuto Jardim	Lucas Catão de Freitas Ferreira
Ruy Exel Filho	Jorge V Bacellar S. Pereira	Murilo Francisco Tome	Marcelo Jose Saia
	José Alberto Cuminato	Pedro Luiz Queiroz Pergher	Marcos P. A. Cavalcante
	Luiz Antonio B. San Martin	Rafael Oswaldo R. Rodriguez	Nicolau Corcao Saldanha
	Manoel Jose M. S. Lemos	Ricardo Sá Earp	Olimpio Hiroshi Miyagaki
	Marcelo da Silva Montenegro	Robert David Morris	Pedro Antonio Santoro Salomão
	Marcio Gomes Soares	Roberto Andreani	Ricardo Martins da Silva Rosa
	Marco Antonio Teixeira	Sonia Maria Gomes	Ronaldo Alves Garcia
	Maria Aparecida Soares Ruas	Vilmar Trevisan	Salvador Addas Zanata
	Mikhail Solodov		Valentin Raphael Henri Ferenczi
	Milton da Costa Lopes Filho		Valeria Neves D. Cavalcanti
	Pavel Shumyatsky		Vitor Domingos M. de Araújo
	Renato de Azevedo Tribuzy		
	Ruy Tojeiro de Figueiredo Jr		
	Sandra Augusta Santos		
	Severino Collier Coutinho		
	Vyacheslav Futorny		

Fonte: elaborado pela autora.

3.2.1 Análise bibliométrica para os pesquisadores PQ1 em Matemática

Para cada autor, armazenaram-se as citações recebidas por seus artigos e os anos de suas publicações. Com essas informações, calcularam-se: o tempo em anos (t) que o pesquisador

publica na base, o índice h , o índice g , os índices relativos h/t e g/t , os índices propostos dc_i e dc_o e os coeficientes de variação $CVdc_i$ e $CVdc_o$.

As informações individuais de cada pesquisador e os respectivos índices podem ser verificados no Apêndice. A Tabela 10 apresenta estatísticas descritivas para os indicadores de impacto científico analisados, a partir da distribuição dos dados dos pesquisadores PQ1 em Matemática.

Tabela 10 – Estatísticas descritivas para os indicadores bibliométricos

<i>Indicadores</i> <i>Estatísticas</i>	<i>t</i>	<i>h</i>	<i>h/t</i>	<i>g</i>	<i>g/t</i>	<i>dc_i</i>	<i>CVdc_i</i>	<i>dco</i>	<i>CVdco</i>	<i>Nº artigos</i>
Média	24,7	11,1	0,6	18,1	0,8	17,0	145%	7,8	69%	47,8
Coeficiente de Variação	36%	51%	150%	59%	100%	140%	85%	55%	13%	62,0%
Máximo	53	32	9,3	64	7,24	240,1	1044%	24,8	91%	149
Mínimo	9	4	0,12	4	0,15	0,4	11%	2,1	36%	8

Fonte: elaborada pela autora.

Verifica-se que, em média, os pesquisadores PQ1 em Matemática produziram aproximadamente 48 artigos, têm índice h igual a 11,1 obtidos no período aproximado de 25 anos. Logo, produziram 0,6 artigo com citação por ano. O índice g médio dos autores analisados está em torno de 18, ou seja, os artigos do núcleo Hirsch receberam conjuntamente, em média, 324 citações.

Observa-se que os pesquisadores têm um comportamento distinto em relação ao tempo dos artigos publicados na base *Scopus*: o autor mais recente tem publicações indexadas há 9 anos e o mais antigo há 53 anos. Os valores dos índices h e g também variam, na escala de 4 a 32 e de 4 a 64, respectivamente.

A média dos desvios de citações ($dc_i = 17$) revela uma tendência de os artigos do núcleo Hirsch dos pesquisadores terem, em média, dispersão moderada ($CVdc_i = 145\%$). Verifica-se, ainda, que há pesquisadores com artigos altamente citados, entre os quais, destaca-se Robert David Morris, com o maior desvio de citações ($dc_i = 240,1$), o que determinou $CVdc_i$ de 1044%. Este autor é muito citado em artigos relacionados à Análise Combinatória e à Teoria dos Grafos e recebeu premiações em publicações relacionadas à essas temáticas.

Entretanto, existem pesquisadores poucos citados e com índice h relativamente baixo. Entre eles, cita-se Amilcar Pacheco, com 17 publicações indexadas há 27 anos e índice $h = 4$. O autor Andrés Koropechi, também com índice $h = 4$, todavia resultante de 14 artigos publicados em um período de 9 anos. Ambos os autores têm coeficiente de variação do desvio de citações do núcleo Hirsch muito baixo (19% e 11%, respectivamente), ou seja, os artigos do

h-core receberam poucas citações. No entanto, ao analisar a medida relativa *h/t* pode-se dizer que o segundo autor tem maior frequência de citação em relação ao primeiro.

Observa-se que a dispersão média das citações fora do núcleo Hirsch para os pesquisadores foi moderada. Ou seja, em média, a chance de incrementar o valor do índice *h* é moderada.

Com a finalidade de verificar se há diferença significativa no comportamento dos índices bibliométricos analisados para os pesquisadores PQ1 nos diversos níveis (A, B, C e D) elaboraram-se a Tabela 11 e o Gráfico 1.

Tabela 11 – Estatísticas descritivas para os indicadores bibliométricos, de acordo com o nível do pesquisador PQ1.

PQ1 (Nº pesquisadores)	Indicadores Estatísticas	<i>t</i>	<i>h</i>	<i>h/t</i>	<i>g</i>	<i>g/t</i>	<i>dc_i</i>	<i>CVdc_i</i>	<i>dc_o</i>	<i>CVdc_o</i>	Nº artigos
PQ1A (21)	Média	31,0	14,7	0,5	24,7	1,2	22,5	166%	10,4	68%	60,4
	CV%	32%	45%	60%	45%	125%	52%	57%	48%	15%	50,0%
	Máximo	47	32	1,18	53	7,24	46,1	424%	24,8	83%	114
	Mínimo	15	5	0,12	7	0,17	4,8	52%	3,0	36%	8
PQ1B (36)	Média	26,7	11,4	0,5	17,7	0,7	13,3	117%	8,3	71%	52,4
	CV%	28%	45%	40%	49%	57%	67%	46%	49%	13%	56%
	Máximo	53	25	1,04	47	1,96	48,1	238%	23	91%	138
	Mínimo	15	5	0,16	7	0,24	3,08	26%	3,0	51%	12
PQ1C (28)	Média	24,6	11,6	0,5	19,9	0,9	24,0	169%	7,9	68%	48,8
	CV%	30%	50%	60%	67%	67%	185%	120%	54%	9%	64%
	Máximo	44	26	1,14	64	2,1	240,1	1044%	18,8	79%	149
	Mínimo	14	5	0,19	7	0,22	2,1	31%	3,3	56%	20
PQ1D (31)	Média	18,1	7,8	0,8	12,7	0,8	11,4	141%	5,4	69%	33,1
	CV%	34%	46%	200%	54%	50%	91%	69%	52%	14%	71%
	Máximo	30	18	9,3	31	1,77	48,2	419%	14,4	86%	106
	Mínimo	9	4	0,15	4	0,15	0,4	11%	2,1	45%	12

Fonte: elaborada pela autora.

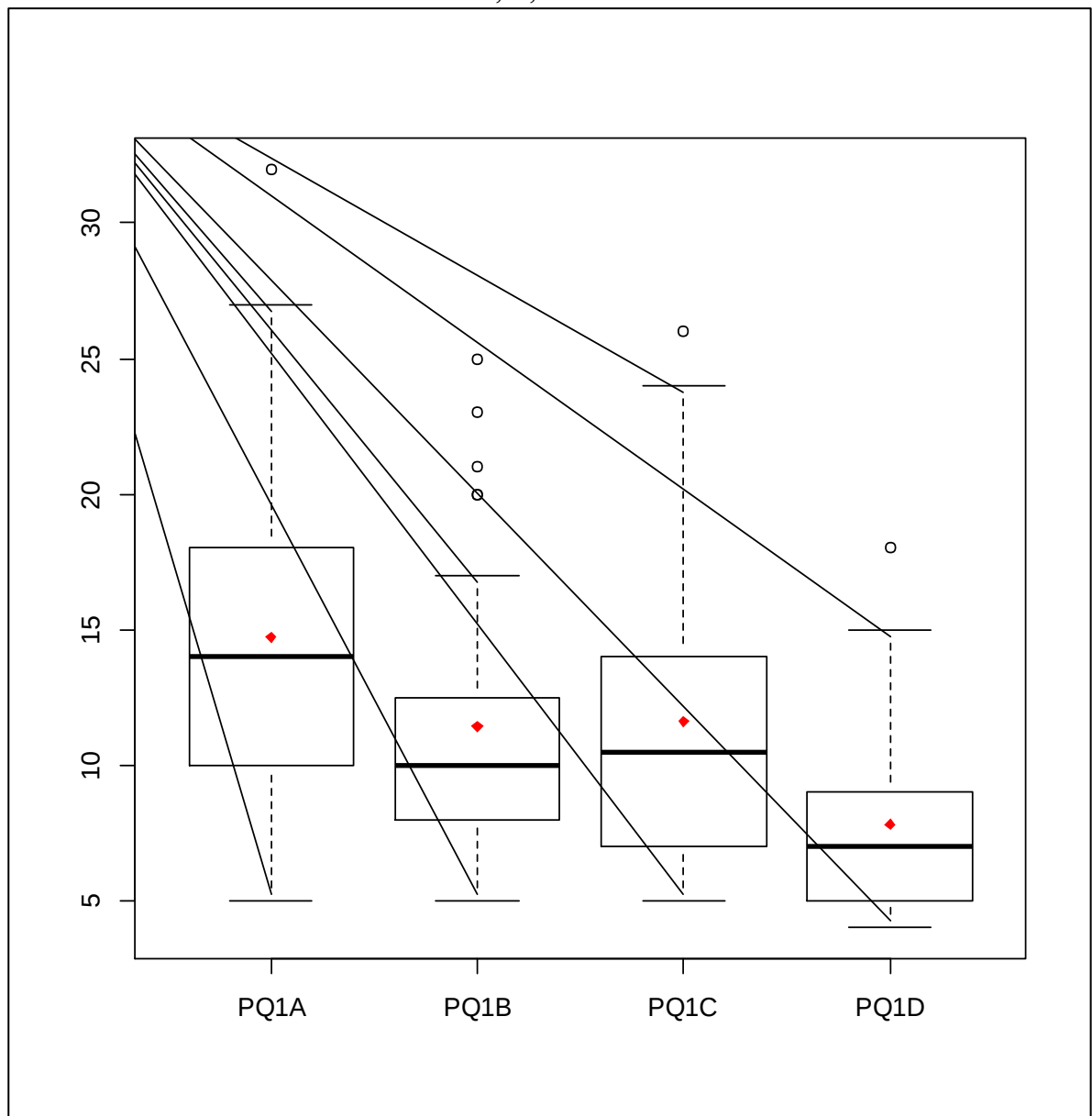
Verifica-se que, em média, os pesquisadores PQ1A em Matemática publicaram aproximadamente 60 artigos e têm índice *h* médio igual a 14,7 obtidos no período de 31 anos. Logo, produziram, em média, meio artigo com citação por ano. O índice *g* desses autores está em torno de 24,7, ou seja, os artigos do núcleo Hirsch receberam, em média, 610 citações.

Os pesquisadores PQ1B têm, em média, *h* = 11,4 e *g* = 17,7 obtidos ao longo de 29 anos, a publicação média foi de aproximadamente 52 artigos. Os autores classificados como PQ1C, em um período de 25 anos obtiveram índice *h* = 11,6, *g* = 19,9 e publicação média de

49 artigos. Os pesquisadores PQ1D têm os menores índices h e g e a menor média de publicação ($h = 7,8$ e $g = 12,7$ e 33 artigos), obtidos em um período de 18 anos.

Por meio da análise desses indicadores bibliométricos, verifica-se que existe uma tendência de os pesquisadores classificados no nível PQ1A terem maiores índices h , g e número de artigos publicados em relação aos demais níveis, além de terem maior tempo de dedicação à pesquisa.

Gráfico 1 – Boxplot para o índice h dos pesquisadores PQ1 de acordo com os níveis A, B, C e D.



Fonte: elaborada pela autora.

Por meio do Gráfico 1, identifica-se que o valor para o índice h mínimo é similar em todas as categorias de pesquisadores PQ1 em Matemática. Os pesquisadores PQ1A são aqueles que possuem os maiores valores para o índice h e também é o grupo que tem maior variabilidade em relação a distribuição dos valores desse indicador.

A distribuição dos valores do índice h para os pesquisadores PQ1C também apresentou alta variabilidade, mas ao comparar com o grupo A identificou-se que o valor mediano para o grupo C coincide com o primeiro quartil do grupo A e que o valor máximo do grupo A aproximou-se do valor discrepante do grupo C. Além dessas características, pode-se dizer que 75% dos pesquisadores PQ1C tem índice h abaixo da mediana do grupo A.

Os pesquisadores PQ1B determinaram uma distribuição do índice h com menor dispersão, no entanto, existem 4 autores com valores para o índice h considerados discrepantes. Similarmente a distribuição do índice h para os pesquisadores PQ1D também segue comportamento, no entanto, a amplitude de variação é menor quando comparado com o grupo B.

No entanto, esses resultados podem estar relacionados à excelência e também ao tempo de atuação do pesquisador, pois os níveis 1A, 1B, 1C, 1D são considerados categorias ordinais. Considerando que um bolsista PQ1A, já pertenceu às outras categorias: PQ1B, PQ1C e PQ1D e há um tempo mínimo de permanência em cada estrato. Por meio da Tabela 11 confirma-se que os pesquisadores PQ1A tendem a ter mais tempo dedicados à vida acadêmica e os maiores valores para os índices h e g .

Ao analisar o desvio das citações do núcleo Hirsch (dc_i), identifica-se que, em média, a dispersão é moderada em todos os níveis ($CVdc_i = 145\%$). Esse resultado evidencia que há autores altamente citados e outros que receberam poucas citações. Similarmente, a média dos desvios das citações fora do h -core (dc_o) também determinou dispersão moderada (68%).

Benar F. Svaiter é pesquisador PQ1A e tem o maior índice h ($h = 32$), alcançado durante o período de 27 anos. O desvio das citações do núcleo Hirsch é moderada ($dc_i = 144\%$) e fora do núcleo é alta ($dc_o = 78\%$), ou seja, ele tem baixa probabilidade de incrementar o valor de seu índice h . Esse autor trabalha com as temáticas: Otimização, Programação não-linear, Métodos Estocásticos.

Propõe-se discriminar as informações dos indicadores h , dc_i e dc_o por meio da terna (h ; dc_i ; dc_o). Assim, para o autor Marcelo Miranda Viana da Silva a terna seria (18; 42,1; 9,6), significando que ele tem índice h igual a 18, alto desvio das citações do núcleo Hirsch (42,1) e moderada dispersão fora do h -core (9,6), obtidos em 28 anos. Pode-se concluir que o autor teve 18 artigos com pelo menos 18 citações cada, além disso, a dispersão do núcleo Hirsch é alta

caracterizando que ele tem citações expoentes nos artigos que compõem o núcleo; o valor do dc_o reforça a representatividade do valor do índice h , pois há moderada probabilidade de incrementar tal índice.

Paolo Piccioni publicou 104 artigos e tem a seguinte terna (13; 6,7; 8,5), obtida em um período de 22 anos. Este autor tem dispersão muito baixa na cauda superior da distribuição de citações e moderada na cauda inferior. Ruy Exel Filho publicou 67 artigos em um período de 30 anos e tem a seguinte terna (19; 35,3; 11,2), ou seja, índice $h = 19$, dispersão alta no h -core e moderada fora do núcleo Hirsch. Observa-se que, embora o primeiro autor tenha maior número de publicações, seu índice h e a variabilidade das citações no h -core é menor em comparação com o segundo autor. Conclui-se que, nesse caso, o pesquisador Rui Exel Filho apresenta evidências de ser mais reconhecido na área, apesar de ter menos publicações.

A Tabela 12 apresenta as estatísticas descritivas para os indicadores bibliométricos de acordo com o nível de dispersão contidas nos núcleos Hirsch.

Tabela 12 – Estatísticas descritivas para os indicadores bibliométricos, conforme o nível de dispersão do núcleo Hirsch ($CVdc_i$)

$CVdc_i$ (Nº pesquisadores)	Indicadores Estatísticas	t	h	h/t	g	g/t	dc_i	$CVdc_i$	dc_o	$CVdc_o$	Nº artigos
Dispersão Muito Baixa (59 pesquisadores)	Média	25,9	10,4	0,4	14,7	0,7	8,2	75%	7,3	71%	50,9
	CV%	38%	50%	50%	56%	129%	71%	36%	51%	10%	59%
	Máximo	53	27	1,14	43	7,24	28,2	114%	18,2	86%	138
	Mínimo	9	4	0,12	4	0,15	0,4	11%	2,2	53%	12
Dispersão Baixa (4 pesquisadores)	Média	25,8	17,5	0,80	28,0	1,3	21,4	123%	15,5	73%	82,0
	CV%	47%	27%	50%	26%	46%	24%	3%	43%	11%	56%
	Máximo	43	24	1,13	38	1,87	28,3	127%	23	81%	149
	Mínimo	15	13	0,30	21	0,49	15,7	118%	8,7	66%	43
Dispersão Moderada (16 pesquisadores)	Média	23,3	11,9	1,06	19,8	0,9	17,0	142%	8,5	71%	51,4
	CV%	23%	65%	208%	65%	56%	64%	8%	68%	14%	72%
	Máximo	31	32	9,30	53	1,96	46,1	161%	24,8	91%	137
	Mínimo	13	6	0,19	10	0,32	8,3	131%	3,6	52%	12
Dispersão Alta (37 pesquisadores)	Média	23,2	11,2	0,5	22,0	1,0	31,4	266%	7,5	65%	37,4
	CV%	34%	47%	40%	54%	50%	120%	59%	47%	17%	50%
	Máximo	44	25	1,05	64	2,1	240,1	1044%	16,8	89%	87
	Mínimo	9	4	0,16	8	0,28	9	171%	2,1	36%	8

Fonte: elaborada pela autora.

Ao organizar os pesquisadores por graus de dispersão (muito baixa, baixa, moderada e alta) dos dc_i , observa-se que o tempo médio de publicação, nos quatro níveis, são similares. Existe uma certa semelhança em relação ao índice h médio para os pesquisadores com dc_i muito

baixo, moderado e alto. Somente quatro pesquisadores foram classificados com dispersão baixa e tiveram o valor do índice h relativamente maior em comparação com os demais níveis.

Identificaram-se 37 autores com alta dispersão para dc_i . Em média, esse grupo de autor apresentou o menor índice de produtividade, correspondente a 37,4 artigos, o que significa que os artigos que publicam são de alto impacto, uma vez que receberam muitas citações. Esse fato faz com que os valores de dc_i inflacione. É importante frisar que têm pesquisadores com índices h variando de 4 a 25, ou seja, apesar de terem alta dispersão de citações no h -core, este fato não determina ter índice h alto.

A Tabela 13 apresenta a classificação dos pesquisadores de acordo com o $CVdc_i$ e $CVdc_o$.

Tabela 13 - Classificação dos pesquisadores de acordo com o $CVdc_i$ e $CVdc_o$

$CVdc_o \backslash CVdc_i$	Moderado		Alto		Total	
	autores	%	autores	%	autores	%
Muito baixo	26	(22,4%)	34	(29,3%)	60	(51,7%)
Baixo	2	(1,7%)	2	(1,7%)	4	(3,4%)
Moderado	7	(6,0%)	9	(7,8%)	16	(13,8%)
Alto	22	(19,0%)	14	(12,1%)	36	(31,1%)
Total	57	(49,1%)	59	(50,9%)	116	(100,0%)

Fonte: elaborada pela autora.

Verifica-se que aproximadamente 55% dos autores têm $CVdc_i$ muito baixo ou baixo. Isso significa que eles não têm artigos com citações inflacionando a cauda superior da distribuição de citações.

Além disso, para todos os pesquisadores o $CVdc_o$ ficou entre moderado e alto, o que significa baixa probabilidade de incrementar o valor do índice h . Desse modo, pode-se dizer que o índice h desses pesquisadores é representativo do seu desempenho científico.

Observa-se que 36 autores (31,1%) têm $CVdc_i$ alto, caracterizando grande dispersão na distribuição das citações no núcleo Hirsch. Desses autores, 14 têm também alta dispersão na cauda inferior das citações. Este fato evidencia que eles possuem citações oscilando entre 0 e h nessa cauda.

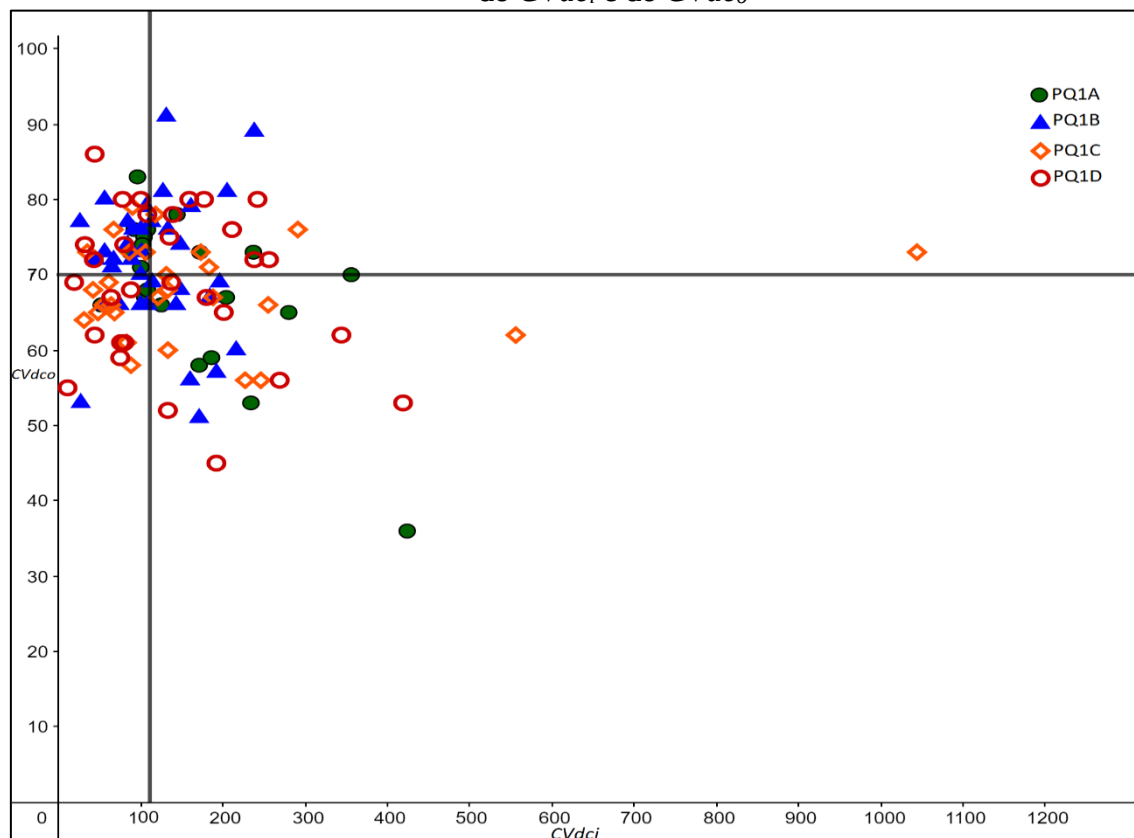
Com o objetivo de verificar a dispersão dos pesquisadores de acordo com o valor de $CVdc_i$ e $CVdc_o$ elaborou-se o Gráfico 2. Os quadrantes foram definidos pelos valores medianos para $CVdc_i$ e $CVdc_o$ que foram, respectivamente, 111% e 70%.

Ou seja, metade dos pesquisadores tem dispersão muito baixa na cauda superior da distribuição de citações e a outra metade é composta por pesquisadores com comportamentos

heterogêneos, pois contemplam pesquisadores com dispersão variando entre muito baixa e alta. O valor mediano para $CVdc_o$ determinou que 50% dos pesquisadores tem dispersão moderada e alta e os outros 50% variabilidade a alta.

Designou-se ao 1º Quadrante os pesquisadores com $CVdc_i > 111\%$ e $CVdc_o > 70\%$, ao 2º Quadrante aqueles com $CVdc_i \leq 111\%$ e $CVdc_o > 70\%$, ao 3º Quadrante os com $CVdc_i \leq 111\%$ e $CVdc_o \leq 70\%$ e ao 4º Quadrante os pesquisadores com $CVdc_i > 111\%$ e $CVdc_o \leq 70\%$.

Gráfico 2: Pesquisadores PQ1 em Matemática subdivididos de acordo com o valor mediano do $CVdc_i$ e do $CVdc_o$



Fonte: elaborada pela autora.

Foram classificados 24 pesquisadores pertencentes ao 1º Quadrante (4 PQ1A, 5 PQ1B, 7 PQ1C e 8 PQ1D), ou seja, eles têm $CVdc_i$ e $CVdc_o$ maiores que o valores medianos. Isto significa que eles têm alta dispersão na calda inferior da distribuição de citações, indicando que existe pouca probabilidade de incrementar os valores dos seus índices h . Em relação à classificação de $CVdc_i$ esses pesquisadores têm dispersão oscilando de muito baixa à alta, 55% deles tem alta variabilidade, 33% moderada, 8% baixa e 4% muito baixa. Desse modo, é um grupo de pesquisadores que varia consideravelmente quanto à dispersão das citações dos artigos que compõem o núcleo Hirsch.

Ao 2º Quadrante pertencem 32 pesquisadores (7 PQ1A, 13 PQ1B, 5 PQ1C, 7 PQ1D) os quais têm dispersão muito baixa no núcleo Hirsch e alta fora desse núcleo, indicando que não têm artigos altamente citados no núcleo Hirsch e o valor do índice h representativo. Dessa forma, considera-se que os pesquisadores integrantes desse quadrante têm valor do índice h estável, ou seja, o índice h é um indicador representativo de seus desempenhos e com baixa probabilidade de mudança em seu valor.

Ao analisar o 3º Quadrante observou-se que 26 pesquisadores possuem essas características: têm dispersão muito baixa no núcleo Hirsch e variabilidade classificada de moderada à alta fora do h -core. Ou seja, esses pesquisadores não têm artigos altamente citados no núcleo Hirsch e o valor do índice h representativo, são 3 pesquisadores PQ1A, 6 PQ1B, 9 PQ1C, 8 PQ1D.

No 4º Quadrante existem 34 pesquisadores (8 PQ1A, 9 PQ1B, 9 PQ1C, 8 PQ1D), eles têm $CVdc_o$ variando entre moderado e alto e $CVdc_i$ oscilando entre muito baixo e alto (3% com dispersão muito baixa, 6% baixa, 24% moderada e 67% alta). Esses pesquisadores têm índice h representativo e diferentes comportamentos na distribuição de citações da cauda superior.

Conclui-se que 52% dos pesquisadores PQ1 em Matemática não têm artigos altamente citados na cauda superior da distribuição de citações. Este resultado pode representar alguma característica particular da área, que pode estar relacionada à temática ou ao tempo de atuação do pesquisador. Além disso, identificou-se que nenhum pesquisador foi classificado com dispersão baixa ou muito baixa na cauda inferior da distribuição de citações. Este resultado reforça que o índice h de cada pesquisador é representativo e existe baixa probabilidade de incrementá-lo.

A Tabela 14 sintetiza as informações referentes aos indicadores bibliométricos ordenados pelo valor do índice h . Subdividiram-se os pesquisadores em três grupos: **1.** índices h variando de 1 a 10; **2.** índices h variando de 11 a 20; **3.** índices h maiores que 20.

Os autores com índice h menor ou igual a 10 representam 56% dos analisados. Em média, eles publicaram 34 artigos em 23 anos e possuem a seguinte terna: (7,2; 9,3; 5,1). Assim, apresentam uma tendência de índice h próximo a 7, dispersão moderada no h -core e também fora do núcleo Hirsch. Os autores com índice h entre 11 e 20, em média, publicaram aproximadamente 58 artigos em 26 anos e possuem a seguinte terna: (14,4; 21,1; 10,1). Pode-se dizer que o índice h está em torno de 14, dispersão moderada no núcleo Hirsch e fora dele. Os autores com índice h maiores que 20, em média, publicaram 104 artigos em 30 anos e possuem a seguinte terna: (25,1; 58,1; 17,5), ou seja: índice h médio = 25, dispersão alta no h -core e moderada fora do núcleo Hirsch.

Tabela 14 – Estatísticas descritivas para os indicadores bibliométricos, conforme o índice *h*

Índice <i>h</i> (Nº pesquisadores)	Indicadores Estatísticas	<i>t</i>	<i>h</i>	<i>h/t</i>	<i>g</i>	<i>g/t</i>	<i>dc_i</i>	<i>CVdc_i</i>	<i>dc_o</i>	<i>CVdc_o</i>	Nº artigos
1 a 10 (65 pesquisadores)	Média	23,1	7,2	0,5	11,1	0,6	9,3	132%	5,1	70%	34,4
	CV%	85%	25%	220%	32%	150%	78%	77%	29%	13%	64%
	Máximo	42	10	9,3	21	7,24	44,5	556%	8,04	91%	138
	Mínimo	9	4	0,12	4	0,15	0,4	11%	2,1	36%	8
11 a 20 (43 pesquisadores)	Média	25,9	14,4	0,6	24,0	1,0	21,1	147%	10,1	68%	57,6
	CV%	36%	21%	33%	25%	40%	54%	52%	32%	12%	41%
	Máximo	47	20	1,14	40	2,1	50,9	344%	23	81%	107
	Mínimo	13	11	0,25	14	0,36	3,08	26%	5,6	51%	22
>20 (8 pesquisadores)	Média	30,9	25,1	0,9	44,3	1,5	58,1	240%	17,5	69%	104,3
	CV%	32%	13%	22%	24%	33%	128%	137%	20%	10%	28%
	Máximo	53	32	1,18	64	1,96	240,1	1044%	24,8	78%	149
	Mínimo	22	21	0,43	33	0,62	17,1	0,74	13,8	0,57	64

Fonte: elaborada pela autora.

Logo, conclui-se que aqueles autores com maiores índices *h* também foram os que receberam mais citações nos artigos que compõem o núcleo Hirsch. No entanto, somente 8 pesquisadores estão classificados nesse grupo e, em decorrência, esse valor representa 7% do total de pesquisadores.

A Tabela 15 apresenta um resumo estatístico para os indicadores bibliométricos organizados de acordo com o número de publicações. Os autores que publicaram até 40 artigos foram classificados como grupo 1, aqueles com publicações entre 41 e 80 artigos, designados como grupo 2 e, total de publicações acima de 80 artigos, reunidos no grupo 3.

Do total de pesquisadores, 63 compõem o grupo 1 e, em média, publicaram 27 artigos em um período de 22 anos, aproximadamente. Esses autores têm a seguinte terna (8,0; 12,6; 5,3), o que significa que publicaram, em média, oito artigos com pelo menos 8 citações cada, a dispersão no núcleo Hirsch e fora dele é moderada, evidenciando que, em média, não existem trabalhos com altas citações. O grupo 2 é composto por 36 pesquisadores que, em média, publicaram 56 artigos e possuem a seguinte terna (12,9; 23,4; 9,3), ou seja: índice *h* próximo a 13, variabilidade dentro do *h-core* moderada e alta fora do núcleo Hirsch. Os pesquisadores do grupo 3, composto por 17 pesquisadores, publicaram em média 107 artigos em um período de 29 anos. A terna desses autores é: (18,8; 20,0; 13,5), indicando: índice *h* aproximadamente 19, desvio das citações no núcleo Hirsch muito baixa e baixa chance de incrementar o índice *h*, pois o *CVdc_o* é alto.

Tabela 15 – Estatísticas descritivas para os indicadores bibliométricos, conforme o número de artigos.

<i>Nº artigos (Nº pesquisadores)</i>	<i>Indicadores Estatísticas</i>	<i>t</i>	<i>h</i>	<i>h/t</i>	<i>g</i>	<i>g/t</i>	<i>dc_i</i>	<i>CVdc_i</i>	<i>dc_o</i>	<i>CVdc_o</i>	<i>Nº artigos</i>
1 a 40 (63)	Média	22,3	8,0	0,53	13,1	0,65	12,6	151%	5,3	67%	27,3
	CV%	33%	4%	215%	51%	57%	86%	71%	36%	15%	29%
	Máximo	42	18	9,30	35	1,81	48,2	556%	10,2	91%	40
	Mínimo	9	4	0,12	4	0,15	0,4	11%	2,1	36%	8
41 a 80 (36)	Média	26,8	12,9	0,53	21,8	1,06	23,4	156%	9,3	72%	56,0
	CV%	34%	33%	45%	48%	109%	165%	165%	106%	8%	20%
	Máximo	47	23	1,13	64	7,24	240,1	1044%	16,8	81%	78
	Mínimo	12	7	0,25	10	0,26	3,5	32%	5,3	58%	41
>80 (17)	Média	29,1	18,8	0,71	29,0	1,10	20,0	97%	13,5	72%	106,8
	CV%	38%	36%	44%	43%	50%	69%	43%	36%	8%	17%
	Máximo	53	32	1,18	53	1,96	48,1	192%	24,8	80%	149
	Mínimo	11	9	0,31	13	0,36	3,08	26%	7,2	57%	85

Fonte: elaborada pela autora.

Diante das análises realizadas com o grupo de pesquisadores PQ1 em Matemática, pode-se concluir que os índices propostos, dc_i e dc_o , conseguem distinguir os diferentes comportamentos nas distribuições de citações da cauda superior e inferior. Ainda, identificam pesquisadores com artigos altamente citados e oferecem subsídios para a análise relativa à possibilidade de um pesquisador incrementar o valor do índice h . Desse modo, constituem índices bibliométricos de impacto científico que podem contribuir nas tomadas de decisões em avaliações de agências de fomento e de políticas científicas.

4. CONSIDERAÇÕES FINAIS

Devido à relevância de se conhecer o desempenho científico de um investigador, de temáticas, de disciplinas, de áreas do conhecimento ou de países, faz-se necessário o desenvolvimento de critérios de avaliação consistentes e fidedignos ao desempenho do avaliado.

Dentre os indicadores de impacto, destaca-se o índice h , considerado robusto por avaliar, de forma simultânea, os aspectos relativos à produção e o impacto científico do avaliado. Trata-se de uma medida simples de ser obtida e seu desempenho tem-se mostrado significativo na avaliação do desempenho científico de um pesquisador. No entanto, diversos estudos apontaram deficiências e limitações desse indicador e, na literatura científica, existem diversas variações que buscam minimizar ou dirimir os problemas apontados.

Nesse cenário, esta tese teve por objetivo analisar o índice h , suas propriedades, suas variações e suas ponderações. Identificou-se que há diversidade de variações desse índice, que também apresentam problemas relacionados, principalmente, à escala interpretativa do indicador. Além disso, existem indicadores que envolvem uma modelagem matemática mais avançada, o que contradiz à simplicidade da obtenção do índice h .

Propuseram-se, nessa pesquisa, dois novos indicadores complementares ao índice h , denominados índices dc_i e dc_o , dedicados a medir a dispersão contida no h -core e na cauda inferior da distribuição de citações.

A partir do cálculo desses índices, para pesquisadores em âmbito internacional - analisou-se um conjunto de 20 autores com atuação destacada na área de Biblometria e Cientometria, em função de serem os mais produtivos na revista *Scientometrics*, no período de 1980-2014, considerou-se que os indicadores cientométricos (h e as medidas de dispersão de h), complementam-se e auxiliam na identificação de autores com alta produtividade em uma temática.

Desse modo, as medidas de dispersão propostas, dc_i e dc_o , mostraram-se capazes de caracterizar fidedignamente a dispersão do desempenho do pesquisador em torno do seu índice h e, conseqüentemente, a pertinência da representatividade desse índice, como indicador do desempenho científico do pesquisador avaliado. Além desse aspecto, permitiram distinguir de forma mais precisa o desempenho dos pesquisadores de um mesmo campo científico.

Em âmbito nacional, a partir do cálculo do índice h , índice g e os novos índices propostos dc_i e dc_o para os pesquisadores com bolsa produtividade em Pesquisa Qualis1 (PQ1) em Matemática, verificou-se que ter um maior número de trabalhos publicados não é um fator

determinante para o autor alcançar maiores valores para esses índices, que medem o impacto de um pesquisador.

Identificaram-se situações em que há autores com mesmos índices h , mas com diferentes números de publicações, de tempo dedicados à pesquisa e com distintos desvios de citações dentro e fora do núcleo Hirsch.

Verificou-se, ainda, uma tendência de os pesquisadores PQ1 pertencentes ao nível A terem os maiores valores para: número de publicações, tempo dedicado à pesquisa e índices h e g . Em contrapartida, os pesquisadores PQ1D têm a tendência de apresentar os menores valores para o tempo, número de publicações e índices h e g .

Ao estratificar os pesquisadores PQ1 pela dispersão dentro do núcleo Hirsch, observou-se, em todos os níveis de variabilidade, similaridade no valor do índice h e no tempo de publicação.

Além desses resultados, identificou-se que quanto maior o valor do índice h maiores são: o tempo de dedicação à pesquisa, o número de publicações, os índices h e g e o desvio de citações dc_i .

Observou-se, também, que os novos indicadores contribuíram para estimar a possibilidade do avaliado incrementar seu índice h , por meio da análise da dispersão do número de citações contidas no h -core e, também, fora dele.

Conclui-se que os novos índices dc_i e dc_o podem contribuir para avaliar a representatividade do índice h como indicador do impacto científico acumulado de um pesquisador e distinguir, de forma mais precisa, pesquisadores seletivos e produtivos.

Desse modo, pode-se concluir que, em cenários em que são necessárias avaliações do desempenho científico, como por exemplo, em agências de fomento, os indicadores propostos dc_i e dc_o podem contribuir para as análises que alicerçam as tomadas de decisão, em especial, como critério desempate em situações de similaridade quanto aos outros indicadores de impacto científico dos avaliados.

É importante ressaltar que, para desenvolvimento deste trabalho, algumas dificuldades foram encontradas, entre as quais, destacam-se: a coleta das informações na base de dados, a exportação das informações, o tratamento da planilha de dados, o cálculo dos índices h , g , h/t , g/t , dc_i , dc_o , $CVdc_i$ e $CVdc_o$ para cada autor, utilizando somente os recursos do *Software* Excel.

Quanto às limitações dessa pesquisa, ressalta-se que os resultados obtidos para esta tese foram baseados em informações da base de dados *Scopus* e, eventualmente, podem diferir caso seja utilizada outras bases informacionais.

Além dessa limitação, destaca-se que para verificar a efetividade dos indicadores propostos foram utilizadas duas amostras: a primeira composta por um pequeno grupo de pesquisadores atuantes na temática Estudos Métricos; e a segunda composta de pesquisadores da área de Exatas, o que pode levar a diferenças entre o comportamento dos indicadores estudados e dos propostos de acordo com a área, temática ou disciplina em estudo.

Como propostas para pesquisas subsequentes, sugerem-se: 1. aplicar os indicadores propostos em outros conjuntos de dados; 2. elaborar o desvio de citações geral. 3. propor o desvio de citações para o índice g . 4. elaborar um sistema computacional para calcular os índices bibliométricos em estudo.

Finalizando, considera-se que esta tese atingiu seu objetivo ao apresentar indicadores bibliométricos derivados e complementares ao índice h de Hirsch, que permitem mensurar e avaliar as caudas superior e inferior da distribuição das citações de um pesquisador, os quais contribuem para avaliar a adequação do índice h como representativo do impacto científico do pesquisador avaliado, ao possibilitar distinguir, de forma objetiva, diferenças entre o impacto científico acumulado de pesquisadores de um mesmo domínio científico.

REFERÊNCIAS BIBLIOGRÁFICAS

ALONSO S, CABRERIZO FJ, HERRERA-VIDEIRA E, HERRERA F hg-index: A new index to characterize the scientific output of researchers based on the h- and g-indices. **Scientometrics**. v. 82, n. 2, p. 391-400, 2010.

ARAÚJO, C. A. Bibliometria: evolução histórica e questões atuais. **Em Questão**, Porto Alegre, v. 12, n. 1, p. 11-32, jan/jun, 2006.

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

BURRELL, Q. L. Hirsch's h-index and Egghe's g-index. In D. Torres-Salinas & H. F. Moed (Eds.), **Proceedings of ISSI 2007.11ª Conferência Internacional da Sociedade Internacional de Cienciometria e Informetria**, p. 162-169, 2007a.

_____. On the h-index, the size of the Hirsch core and Jin's A-index. **Journal of Informetrics**. v. 1, n. 2, p. 170–177, 2007b.

BORNHANN, 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**, New York, v. 59, n. 5, p. 830-837, 2008.

CABRERIZO, F.J., ALONSO, S., HERRERA-VIDEIRA, E., HERRERA, F. q²-Index: Quantitative and Qualitative Evaluation Based on the Number and Impact of Papers in the Hirsch Core. **Journal of Informetrics**. v. 4, n. 1, p. 23-28, 2009.

CONSELHO NACIONAL DE PESQUISA, (2018). Produtividade em Pesquisa – PQ. Disponível em: <http://www.cnpq.br/documents/10157/5f43cefd-7a9a-4030-945e-4a0fa10a169a>. Acesso em: 15/02/2018.

DANUELLO, J. C. Estudo da produção científica dos docentes de pós graduação em Fonoaudiologia, no Brasil, para uma análise do domínio. **Tese (Doutorado)** – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2014.

DE VISSCHER A. An Index to Measure a Scientist's Specific Impact. **Journal of the American Society for Information Science and Technology**. v. 61, n. 2, p. 319-328, 2010.

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. An improvement of the h-index: the g-index. **ISSI Newsletter**, v. 2, n. 1, p. 8-9, 2006.

EGGHE, L. The Hirsch index and related impact measures. **Annual review of information science and technology**, White Plains, v. 44, p. 65-114, 2010.

EGGHE L, ROUSSEAU R An h-index weighted by citation impact. **Information Processing and Management**. v. 44, n. 2, p.770-780, 2008.

FONSECA, G. L. B. Qualidade dos indicadores educacionais para avaliação de escolas de redes públicas de ensino básico no Brasil. **Dissertação** (Mestrado). Universidade Federal de Juiz de Fora, 2010.

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, 243–258, 2010.

FRANCESCHINI, M. MAISANO, D. A. Analysis of the Hirsch index's operational properties. **European Journal of Operational Research**. v. 203, n. 2, p. 494-504, 2010.

FRANCESCHINI, M., MAISANO, D. A. Criticism on the *hg*-index. **Scientometrics**. v. 86, n. 2, p. 339–346, 2011.

GLÄNZEL, W. **Bibliometrics as a research field: a course on theory and application of bibliometric indicators**. [S.l.]: Coursehandouts, 2003.

GLÄNZEL, W. On the h-index – A mathematical approach to a new measure of publication activity and citation impact. **Scientometrics**. v. 67, n. 2, p. 315-321, 2006.

GRÁCIO, M. C. C. A; OLIVEIRA, E. F. T. Produção e comunicação da informação em CT&I – GT7 da ANCIB: análise bibliométrica no período 2003/2009. **Liinc em Revista**, Rio de Janeiro, v.7, n.1, p. 248–263, mar 2011.

GRÁCIO, M. C. C. A; OLIVEIRA, E. F. T. A inserção e o impacto internacional da pesquisa brasileira em “estudos métricos”: uma análise na base Scopus. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 5, n. 1, 2012.

GREGOLIN, J. A. R. FARIA, L. I. L.; HOFFMANN, W. A. M.; QUONIAM, L.; QUEYRAS, J. Análise da produção científica a partir de indicadores bibliométricos. In: Francisco Romeu Landi; Regina Gusmão. (Org.). **Indicadores de Ciência, Tecnologia e Inovação em São Paulo 2004**. 1ed.São Paulo: Fapesp, 2005, v. 1, p. 1-44.

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, v. 102, n. 46, p. 16569-16572, Nov. 2005.

HJÖRLAND, B. Domain analysis in information science: eleven approaches – traditional as well as innovative. **Journal of Documentation**, London, v. 58, n. 4, p. 422- 462, 2002.

JACSO P. (2008a) The plausibility of computing the h-index of scholarly productivity and impact using reference-enhanced databases. **Online Information Review**. v. 32, n. 2, p. 266-283, 2008.

JACSO P. (2008b) The pros and cons of computing the h-index using Google Scholar. **Online Information Review**. v. 32, n. 3, p. 437-452, 2008.

JACSO P. (2008c) Testing the calculation of a realistic h-index in Google Scholar, Scopus, and Web of Science for F. W. Lancaster. **Library Trends**. v. 56, n. 4, p. 784-815, 2008.

JANNUZZI, P. Considerações sobre o uso, mau uso e abuso dos indicadores sociais na formulação e avaliação de políticas públicas municipais. **Revista de Administração Pública**. p. 51-72, 2002.

JIN, B. *h*-index: an evaluation indicator proposed by scientist. **Science Focus**, v.1, n. 1, p. 8–9, 2006.

JIN, B. The AR-index: complementing the h-index. **ISSI Newsletter**. v.3, n. 1, p. 6, (2007).

JIN, B., LIANG, L., ROUSSEAU, R., EGGHE, L. The R- and AR indices: Complementing the h-index. **Chinese Science Bulletin**, v. 52, n. 6, p. 855–863, 2007.

KOSMULSKI, M. A new Hirsch-type index saves time and works equally well as the original h-index. **ISSI Newsletter**, v. 2, n. 3, p. 4–6, 2006.

KOSMULSKI M New seniority-independent Hirsch-type index. **Journal of Informetrics**. v. 3, n. 4, p. 341-347, 2009.

LIMA, E. L. Espaços métricos. Instituto de Matemática Pura e Aplicada, **CNPq**, Rio de Janeiro, ed. 3, 1993.

LIMA, R. A. de; VELHO, L. M. L. S.; FARIA, L. I. L. de. Bibliometria e "avaliação" da atividade científica: um estudo sobre o índice-h. **Perspectivas em Ciência da Informação**, v. 17, n. 3, p. 3-17, jul/set. 2012. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-99362012000300002. Acesso em: 11 jul. 2018.

MARQUES, F. Os limites do índice-h: supervalorização do indicador que combina qualidade e quantidade da produção científica gera controvérsia. **Pesquisa Fapesp**, São Paulo, n. 207, p. 35-39, maio 2013.

MUGNAINI, R.; JANNUZZI, P. M.; QUONIAM, L. Indicadores bibliométricos da produção científica brasileira: uma análise a partir da base Pascal. **Revista Ciência e Informação**. v. 33, p. 123-131, 2004.

NARIN, Francis. Evaluative bibliometrics: the use of publication and citation analysis in the evaluation of scientific activity. **Cherry Hill**, N. J.: Computer Horizons, 1976.

_____. Patent bibliometrics. **Scientometrics**, v. 30, p. 147-155, 1994.

OLIVEIRA, E. F. T.; GRÁCIO, M. C. C. Indicadores bibliométricos em ciência da informação: análise dos pesquisadores mais produtivos no tema estudo métricos na base Scopus. **Perspectivas em Ciência da Informação**, v. 16, n.4, p. 16 – 28, out/dez, 2011.

OTLET, Paul. **O livro e a medida: bibliometria**. In: _____. **Bibliometria: teoria e prática**. São Paulo: Cultrix, p. 19-34, 1986.

PRICE, D. S. **Little Science, Big Science**. New York: Columbia University. Press, 1963.

PRITCHARD, A. Statistical bibliography or bibliometrics? **Journal of Documetation**, v. 25, n.4, p. 348-349, Dec. 1969.

REDNER, S. How popular is yourpaper? An empirical study of the citation distribution. **The European Physical Journal B** – [S.I.], v. 4, n. 2, p. 131-134, ago, 1998.

RIIKONEN P, VIHINEN M. National research contributions: A case study on Finnish biomedical research. **Scientometrics**. v. 77, n. 2, p. 207-222, 2008.

RONS N., AMEZ L. Impact vitality: an indicator based on citing publications in search of excellent scientists. **Research Evaluation**. v. 18, n. 3, p. 233-241, 2009.

ROSAS, F. S. Indicadores de impacto, visibilidade e colaboração para a produção científica da Pós-graduação brasileira: um estudo nos programas de excelência na área de Zootecnia. **Dissertação** (Mestrado) – Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2013.

ROUSSEAU, R. (2006). New developments related to the Hirsch index. **e-prints in library & information science**.

Disponível em: http://eprints.rclis.org/7616/1/Hirsch_new_developments.pdf. Acesso em: 06 de fev 2017.

ROUSSEAU R, YE, F.Y. A proposal for a dynamic h-type index. **Journal of the American Society for Information Science and Technology**. v. 59, n. 11, p.1853-1855, 2008.

SANTOS, R. N. M.; KOBASHI, N. Y. Aspectos metodológicos da produção de indicadores em ciência e tecnologia. In: **ENCONTRO NACIONAL DE CIÊNCIA DA INFORMAÇÃO**, 4., 2005, Salvador. Anais.... Salvador: Ufba, 2005.

SEBASTIÃO, D. Teoria da atividade e lousa digital no ensino superior: perspectivas para aprendizagem dos conceitos matemáticos. **Dissertação** (Mestrado) – Universidade do Extremo Sul Catarinense, Programa de Pós-Graduação em Educação, Criciúma, SC, 2017.

SIDIROPOULOS A, KATSAROS D, MANOLOPOULOS Y (2007) **Generalized** Hirsch h-index for disclosing latent facts in citation networks. **Scientometrics**. v. 72, n. 2, p. 253-280, 2007.

SILVA, D. D., GRÁCIO, M. C. C. Relativizando os índices h e g: estudo aplicado ao domínio dos Estudos Métricos. In: ISKO BRASIL, 2015a, Marília: Universidade Estadual Paulista. **Organização do conhecimento e diversidade cultural**, 2015, p. 526 – 533.

_____. Indicadores bibliométricos de impacto: estudo comparativo entre os índices h e g aplicados aos pesquisadores mais produtivos na temática Estudos Métricos. In: ENCUESTRO IBÉRICO EDICIC, 7., 2015b, Madrid. **Desafios y oportunidades de las Ciencias de la Información y Documentación en la era digital**. Madrid: Universidad Complutense, 2015. 10 p. Disponível em: http://edicic2015.org.es/ucmdocs/actas/art/230-Deolindo_indicadores-G-H.pdf.> Acesso em: 26 ago. 2016.

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**. v. 23, Edição Especial 5 EBBC, p. 196-212. 2017.

SHING, M.; KUMAR, S. Application of h and g indices to Quantify Scientific Productivity of Physicists at JNU, India, **Library Philosophy and Practice (e-journal)**, p. 1028 (2014). Disponível em: <https://digitalcommons.unl.edu/libphilprac/1028>. Acesso em: 30 jan 2015.

SPINAK, E. Indicadores cientométricos. **Ciência da Informação**, Brasília, v. 27, n. 2, p. 141-148, maio/ago. 1998.

TAGUE-SUTCLIFFE, J. An introduction to infometrics. **Information Processing & Management**. Oxford, v. 28, n. 1, p. 1-3, 1992.

THELWALL, M. Bibliometrics to webometrics. **Journal of Information Science**, v. 34, n. 4, p. 605–621, 2008.

THOMAZ, P. G; ASSAD, R. S.; MOREIRA, L. F. P. Uso do fator de impacto e do índice h para avaliar pesquisadores e publicações. **Arquivo Brasileiro de Cardiologia**, p. 90-93, 2011.

VANZ, S. A. S.; STUMPF, I. R. C. Procedimentos e ferramentas aplicados aos estudos bibliométricos. **Informação e Sociedade**, v. 20, n.2, p. 67-75, maio/ago. 2010.

VANZ, S. A. S.; CAREGNATO, S. E. Estudos de Citação: uma ferramenta para entender a comunicação. **Em Questão**, Porto Alegre, v. 9, n. 2, p. 295-307, jul/dez, 2003.

VELHO, A. A avaliação do desempenho científico. **Cadernos USP**, n. 1, p. 22-40, 1986.

VELHO, L. M. L. S. **Indicadores de C&T no Brasil: antecedentes e estratégia**. Brasília: [s.n.], 1997.

VIEIRA, S. **Bioestatística: tópicos avançados**. Rio de Janeiro: Elsevier, 2003.

VINKLER P. The π -index: a new indicator for assessing scientific impact. **Journal of Information Science**. v. 35, n. 5, p. 602-612, 2009.

VINKLER P. The evaluation of research by scientometric indicators. **Chandos Publishing**. Oxford Cambrigde, New Delhi, 2010.

YE F.Y., ROUSSEAU R. Probing the h-core: an investigation of the tail-core ratio for rank distributions. **Scientometrics**. In press, 2010.

WU Q. The w-Index: A Measure to Assess Scientific Impact by Focusing on Widely Cited Papers. **Journal of the American Society for Information Science and Technology**. v. 61, n. 3, p. 609-614, 2010.

ZHANG C. T. The e-Index, Complementing the h-Index for Excess Citations. **PLoS ONE**. v.4, n. 5, e5429, 2009.

APÊNDICE

Indicadores bibliométricos para os autores PQ1 em Matemática.

PQ (Nº autores)	<i>Indicadores</i> <i>Autores</i>	<i>t</i>	<i>h</i>	<i>h/t</i>	<i>g</i>	<i>g/t</i>	<i>dci</i>	<i>CVdci</i>	<i>dco</i>	<i>CVdco</i>	Nº <i>artigos</i>
PQ1A (21)	Alcides Lins Neto	15	5	0,33	8	0,53	21,2	424%	3,0	36%	8
	Alfredo N. Iusem	32	27	0,84	43	1,34	28,2	105%	18,2	67%	112
	Aron Simis	47	18	0,38	27	0,57	16,5	92%	13,7	76%	75
	Artur Avila	15	17	1,13	28	1,87	21,3	125%	11,3	66%	64
	Benar F. Svaiter	27	32	1,18	53	1,96	46,1	144%	24,8	78%	114
	Carlos G. T. de A. Moreira	21	10	0,48	15	0,71	10,8	108%	7,6	76%	46
	Carlos Tomei	35	9	0,25	20	0,57	32,0	356%	6,3	70%	36
	Cesar L. C. Manco	42	9	0,21	17	0,40	21,3	237%	6,6	73%	30
	Dan Marchesin	45	19	0,42	28	0,62	16,5	87%	13,9	73%	85
	Enrique R. Pujals	20	14	0,70	29	1,45	39,1	280%	9,1	65%	52
	Hermano Frid Neto	28	13	0,46	23	0,82	22,4	172%	13,9	73%	57
	Israel Vainsencher	41	5	0,12	7	0,17	4,8	96%	4,1	83%	26
	Jorge G. Hounie	40	10	0,25	16	0,40	10,8	108%	6,8	68%	71
	Keti Tenenblat	47	16	0,34	24	0,51	16,6	104%	11,9	75%	78
	Lorenzo J. Diaz Casado	28	16	0,57	28	1,00	27,3	171%	9,3	58%	42
	Luiz Carlos Pacheco R. Velho	22	13	0,59	24	1,09	26,5	204%	8,7	67%	40
	Marcelo M. Viana da Silva	28	18	0,64	35	1,25	42,1	234%	9,6	53%	39
	Marcos Dajczer	37	18	0,49	27	0,73	18,4	102%	13,4	74%	95
	Paolo Piccione	22	13	0,59	19	0,86	6,7	52%	8,5	66%	104
	Paulo Domingos Cordaro	29	8	0,28	12	0,29	8,0	100%	5,7	71%	28
	Ruy Exel Filho	30	19	0,63	35	1,17	35,3	186%	11,2	59%	67
PQ1B (36)	Abramo Hefez	32	5	0,16	9	0,28	10,8	216%	3,0	60%	19
	Adalberto P. Bergamasco	30	9	0,30	13	0,43	8,0	89%	6,5	72%	33
	Alagacone Sri Ranga	25	7	0,28	10	0,40	9,1	131%	6,4	91%	12
	Alexandre N. de Carvalho	27	21	0,78	33	1,22	22,6	108%	13,8	66%	76
	Andre Nachbin	25	14	0,56	19	0,76	5,6	27%	7,4	53%	32
	Antonio Carlos Gardel Leitão	19	11	0,58	19	1,00	17,6	160%	6,2	56%	29
	Artur Oscar Lopes	31	11	0,35	19	0,61	15,7	143%	7,3	66%	64
	Bruno César Azevedo Scárdua	20	7	0,35	11	0,55	9,3	133%	5,3	76%	63
	Claudia Alejandra Sagastizabal	22	20	0,9	30	1,40	20,0	100%	13,3	66%	43
	Daciberg Lima Gonçalves	39	12	0,31	14	0,36	3,1	26%	9,3	77%	107
	Dessislava H. Kochloukova	21	8	0,38	10	0,26	3,6	45%	5,8	72%	66
	Eduardo de Sequeira Esteves	22	10	0,45	13	0,59	5,6	56%	8,0	80%	31
	Gerson Petronilho	28	9	0,32	11	0,39	5,0	55%	6,0	66%	37
	Harold William Rosenberg	53	23	0,43	33	0,62	17,1	74%	15,2	66%	95
	Helena Judith N. Lopes	21	11	0,52	17	0,81	11,1	101%	8,0	73%	46
	Henrique Bursztyn	16	11	0,69	19	1,19	18,8	171%	5,6	51%	22
	Hilário Alencar da Silva	24	9	0,38	16	0,67	21,4	238%	8,0	89%	20

	Jaime Bruck Ripoll	29	6	0,21	10	0,34	6,8	114%	4,1	69%	40
	Joao Marcos Bezerra do O	21	20	0,95	31	1,48	21,2	106%	15,8	79%	98
	Jorge Passamani Zubelli	27	12	0,44	21	0,78	17,8	148%	8,2	68%	48
	Jorge V Bacellar S. Pereira	15	9	0,60	16	1,07	16,6	184%	6,0	67%	33
	José Alberto Cuminato	30	13	0,43	24	0,80	25,5	196%	9,0	69%	43
	Luiz Antonio B. San Martin	33	10	0,30	16	0,48	11,3	113%	7,7	77%	45
	Manoel Jose M. S. Lemos	28	7	0,25	10	0,36	7,0	100%	5,3	76%	52
	Marcelo da Silva Montenegro	17	12	0,70	22	1,29	24,7	205%	9,8	81%	61
	Marcio Gomes Soares	31	6	0,19	10	0,32	9,7	161%	4,7	79%	20
	Marco Antonio Teixeira	41	17	0,41	25	0,61	15,5	91%	13,0	76%	92
	Maria Aparecida Soares Ruas	25	8	0,32	14	0,56	11,9	148%	5,9	74%	45
	Mikhail Solodov	24	25	1,04	47	1,96	48,1	192%	14,2	57%	87
	Milton da Costa Lopes Filho	19	11	0,58	16	0,84	11,1	100%	7,7	70%	37
	Pavel Shumyatsky	25	10	0,40	13	0,52	5,6	56%	7,3	73%	138
	Renato de Azevedo Tribuzy	36	8	0,22	11	0,30	6,8	84%	6,2	77%	28
	Ruy Tojeiro de Figueiredo Jr	25	8	0,32	11	0,44	5,2	65%	5,7	71%	44
	Sandra Augusta Santos	23	16	0,69	25	1,09	20,2	127%	13,0	81%	43
	Severino Collier Coutinho	29	5	0,17	7	0,24	4,2	84%	3,7	74%	38
	Vyacheslav Futorny	29	10	0,34	13	0,45	6,7	67%	7,2	72%	99
PQ1C (28)	Aldo Procacci	24	11	0,46	17	0,71	11,6	106%	8,1	73%	47
	Boyan Slavchev Sirakov	16	15	0,94	29	1,81	34,1	227%	8,4	56%	33
	Carlos Kubrusly	44	11	0,25	23	0,52	32,0	291%	8,3	76%	63
	Claudianor Oliveira Alves	22	24	1,09	38	1,73	28,3	118%	18,8	78%	149
	Claudio Gorodski	23	7	0,30	8	0,35	2,1	31%	4,5	64%	29
	Daniel Marinho Pellegrino	14	16	1,14	22	1,57	8,9	55%	10,5	66%	94
	Dimitar Kolev Dimitrov	26	11	0,42	14	0,54	5,3	48%	7,1	65%	51
	Flavio Ulhoa Coelho	26	12	0,46	17	0,65	10,8	90%	9,5	79%	45
	Frederic Gerard C. Valentin	19	12	0,63	24	1,26	29,6	246%	6,8	56%	35
	Gregório Pacelli Feitosa Bessa	25	7	0,28	10	0,40	6,1	88%	4,0	58%	28
	Gustavo Alberto Perla Menzala	43	13	0,30	21	0,49	15,7	121%	8,7	67%	72
	Hossein Movasati	17	6	0,35	8	0,47	3,7	61%	4,2	69%	29
	Jaime Edilberto Munoz Rivera	27	26	0,96	43	1,59	34,2	131%	18,3	70%	137
	Jairo Zacarias Goncalves	34	10	0,29	12	0,35	3,5	35%	7,3	73%	42
	Jorge Herbert Soares de Lira	15	6	0,40	11	0,73	11,0	183%	4,3	71%	34
	Levi Lopes de Lima	22	5	0,23	7	0,32	3,4	68%	3,3	65%	20
	Luis Adrian Florit	23	9	0,39	13	0,56	7,4	83%	5,5	61%	30
	Luiz Henrique de Figueiredo	27	8	0,30	21	0,78	44,5	556%	5,0	62%	24
	Marcelo Moreira Cavalcanti	19	20	1,05	40	2,10	50,9	255%	13,2	66%	60
	Marcos Benevenuto Jardim	16	7	0,44	10	0,62	6,0	86%	5,1	73%	37
	Murilo Francisco Tome	22	17	0,77	28	1,27	22,7	133%	10,2	60%	39
	Pedro Luiz Queiroz Pergher	31	6	0,19	7	0,22	2,5	42%	4,1	68%	34
	Rafael Oswaldo R. Rodriguez	27	6	0,22	8	0,30	3,8	64%	4,0	66%	34

PQ1D (31)	Ricardo Sá Earp	26	8	0,31	14	0,54	13,9	173%	5,8	73%	21
	Robert David Morris	35	23	0,66	64	1,83	240,1	1044%	16,8	73%	64
	Roberto Andreani	20	13	0,65	24	1,20	24,5	188%	8,8	67%	41
	Sonia Maria Gomes	23	7	0,30	11	0,48	9,3	133%	4,8	68%	25
	Vilmar Trevisan	24	9	0,37	12	0,50	6,0	67%	6,8	76%	49
	Alexander E. A. Mendoza	13	7	9,3	11	0,85	9,4	135%	5,2	75%	35
	Ali Tahzibi	14	7	0,50	11	0,79	9,3	133%	3,6	52%	20
	Amilcar Pacheco	27	4	0,15	4	0,15	0,8	19%	2,8	69%	17
	Andrés Koropecski	9	4	0,44	5	0,56	0,4	11%	2,2	55%	14
	Carolina Bhering de Araujo	15	6	0,40	9	0,60	6,0	100%	4,8	80%	16
	Claudio Aguinaldo Buzzi	16	8	0,50	11	0,69	7,0	88%	5,4	68%	26
	Ederson Moreira dos Santos	9	5	0,56	9	1,00	9,0	180%	3,4	67%	19
	Edson Vargas	25	4	0,16	10	0,40	16,8	419%	2,1	53%	14
	Emanuel Augusto S. Carneiro	9	9	1,00	13	1,44	6,9	76%	5,5	61%	28
	Everaldo Souto de Medeiros	14	9	0,64	17	1,21	21,8	242%	7,2	80%	30
	Fábio Armando Tal	18	6	0,33	11	0,61	12,7	211%	4,6	76%	30
	Geraldo M. de A. Botelho	19	15	0,79	18	0,95	4,8	32%	11,1	74%	62
	Giovany de J. M. Figueiredo	13	13	1,00	23	1,77	23,1	177%	10,4	80%	72
	Henrique Fernandes de Lima	11	9	0,82	13	1,18	7,0	78%	7,2	80%	103
	Jaime Angulo Pava	22	9	0,41	17	0,77	18,1	201%	5,8	65%	31
	Joao Xavier da Cruz Neto	18	6	0,33	11	0,61	14,3	238%	4,3	72%	14
	José Fábio Bezerra Montenegro	22	5	0,23	9	0,41	9,6	192%	2,2	45%	21
	Ketty Abaroa de Rezende	30	7	0,23	10	0,33	5,6	80%	4,3	61%	27
	Lev Birbrair	27	7	0,26	9	0,33	3,0	43%	5,0	72%	30
	Lucas Catão de Freitas Ferreira	12	8	0,67	11	0,92	6,4	80%	5,9	74%	56
	Marcelo Jose Saia	25	6	0,24	9	0,36	6,5	108%	4,7	78%	26
	Marcos P. de A. Cavalcante	10	4	0,40	5	0,50	3,0	75%	2,4	59%	12
	Nicolau Corcao Saldanha	24	4	0,17	8	0,33	10,2	256%	2,9	72%	20
	Olimpio Hiroshi Miyagaki	26	18	0,69	31	1,19	28,6	159%	14,4	80%	106
	Pedro Antonio S. Salomão	14	5	0,36	6	0,43	2,2	44%	4,3	86%	15
	Ricardo Martins da Silva Rosa	22	12	0,55	25	1,14	32,2	269%	6,7	56%	32
	Ronaldo Alves Garcia	25	6	0,24	10	0,40	8,3	139%	4,7	78%	29
	Salvador Addas Zanata	14	5	0,36	6	0,43	2,2	44%	3,1	62%	23
	Valentin Raphael H. Ferenczi	20	11	0,55	15	0,75	7,0	64%	7,4	67%	30
	Valeria N. D. Cavalcanti	20	14	0,70	31	1,55	48,2	344%	8,7	62%	33
	Vitor D. Martins de Araujo	17	9	0,53	15	0,88	12,3	137%	6,2	69%	36

Fonte: elaborada pela autora.