

AINDA É POSSÍVEL PENSAR O ÍNDICE h COMO UM INDICADOR QUANTITATIVO, QUALITATIVO, FRACIONÁRIO?

RESUMO - Em 2005 o índice h é apresentado para comunidade científica como um indicador para estimar o desempenho científico, a partir da quantificação da produtividade e da citação de um cientista ao longo de sua carreira. Em uma perspectiva, a elegância desse indicador está na simplicidade em combinar duas grandezas, número de publicações e número de citações, em apenas uma variável. Mas, ao longo dos anos que se seguiram após a proposta deste índice, vários aspectos positivos e negativos foram apontados. Nesse sentido, este trabalho tem como objetivo principal apresentar alguma das principais características desse indicador.

Palavras-chave: Cientometria, índice h , indicadores.

IT IS STILL POSSIBLE TO THINK THE H-INDEX AS AN INDICATOR QUANTITATIVE, QUALITATIVE, FRACTIONAL?

ABSTRACT –In 2005 the h -index is presented to the scientific community as an indicator to assess the scientific performance, from the measurement of productivity and citation of a scientist over his career. In one perspective, the elegance of this indicator is in the simplicity of combining two quantities, number of publications and number of citations in just one variable. But over the years that followed the proposal of this index, several positive and negative aspects were pointed. In this sense, this work has as main objective to present some of the main features of this indicator.

Key-words: Scientometrics, h -index indicators.

Pablo Diniz Batista
Centro Brasileiro de

Pesquisas Físicas

Rua Xavier Sigaud 150 Urca
CEP 20000 Rio de Janeiro RJ

batista@cbpf.br

INTRODUÇÃO

No dia 3 de agosto de 2005, Jorge Hirsch entra para a história da cientometria quando disponibiliza para a comunidade científica um pequeno artigo apresentando um indicador para quantificar o desempenho individual de um cientista a partir de uma sutil combinação entre duas variáveis – número de citações e número de artigos publicados (HIRSCH, 2005). É curioso como esse novo indicador, índice h , atraiu rapidamente a atenção de muitos pesquisadores de diferentes áreas do conhecimento. Provavelmente, a rapidez com que essa proposta se propagou na comunidade científica esteja relacionada com o fato de que este primeiro artigo tenha sido comentado pouco tempo depois em revistas com grande visibilidade no meio científico, como a Nature (BALL, 2005). O impacto desse novo indicador na comunidade científica pode ser analisado a partir do número de citações que recebeu: 850 citações em pouco mais de cinco anos. A figura 1 mostra que, nos últimos anos, mais de 150 artigos por ano citaram o trabalho pioneiro de Hirsch, demonstrando, assim, que ainda existe um grande interesse por parte da comunidade científica.

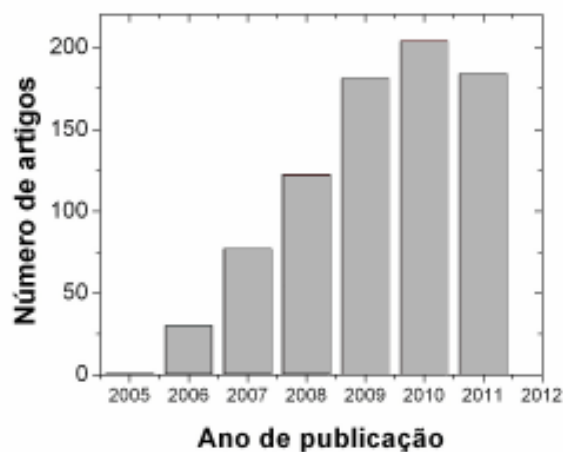


Figura 1 – Número de artigos publicados, em função do ano, citando o artigo em que Hirsch descreve o índice h . Esses dados foram obtidos do *Web Of Science*. (consulta em novembro 2011)

É possível observar que alguns desses trabalhos têm o objetivo de investigar a eficiência desse índice como um possível indicador para estimar o desempenho científico de cientistas, e, nesse sentido, é possível encontrar em artigos de revisão uma lista com suas vantagens e desvantagens (ALONSO S ET AL, 2009). De qualquer forma, não é novidade que outros trabalhos apresentem sugestões para possíveis modificações deste índice, mas também, não é arriscado dizer que grande parte dessas sugestões ainda preserva a idéia central apresentada por Jorge Hirsch. É diante deste cenário de grande interesse por um indicador inovador e aparentemente pouco complexo que o presente trabalho tem como objetivo principal apresentar algumas das principais características do índice h para que seja possível discutir alguns aspectos quantitativos e qualitativos relacionados à produção científica de um pesquisador. A título de curiosidade matemática será apresentado no final uma proposta para calcular a parte fracionária desse esse indicador, desde que o mesmo está definido apenas como um número inteiro.

A proposta do índice h como um indicador

O índice h se insere como um indicador de desempenho científico considerando um cenário em que os pesquisadores, ao longo de suas carreiras, disponibilizam em forma de artigos científicos o resultado de suas pesquisas a fim de garantir não somente a prioridade da descoberta, mas também de somar créditos que podem se converter em maior reconhecimento e prestígio entre seus pares (LATOIR & WOOLGAR, 1997). Dessa maneira, é fácil visualizar o processo através do qual um conjunto de novos artigos é publicado ao mesmo tempo em que artigos anteriores são citados. Tal processo deve-se ao fato de que as referências em um trabalho científico se constituem como uma característica marcante na construção e validação do novo conhecimento científico.

A visualização da dinâmica envolvendo as citações entre artigos, conhecida na comunidade científica como “rede de citações”, foi favorecida pela fundação da base de dados ISI, na década de 60, com sede na Filadélfia, EUA. Esta base foi uma das primeiras bases de dados bibliográficas, mantidas com recurso privado e cobertura multidisciplinar, concebida para organizar as informações contidas nas publicações científicas de diferentes

áreas do conhecimento, incluindo-se aí suas referências. Desta forma, o processo de indexação das publicações nesta base permite, entre outras análises, a contagem das citações em diferentes níveis: dos autores, de países, das publicações, das revistas, etc. Em suma, considerando a atual disponibilidade de bases de dados bibliográficas que permitem o acesso às citações de qualquer cientista e de seus trabalhos é possível então acompanhar a dinâmica envolvendo a publicação e citação de artigos. Entretanto, é importante chamar a atenção para o fato de que a representação na qual índice h esta sendo fundamentada é o mais simples possível, pois outras variáveis poderiam ser facilmente acrescentadas. Por exemplo, ao pensar no componente citação que constitui este índice devemos considerar que nem todos os cientistas e nem todos os artigos possuem a mesmas chances de serem citados.

E neste caso, não estamos considerando apenas a hipótese de que alguns artigos teriam uma qualidade superior a outros e isso explicaria as diferenças no número maior de citações que os artigos recebem. No entanto, como Bornmann & Daniel (2008) destacaram, há diversas razões para citar um artigo, a qualidade seria apenas uma delas. Apesar disso, ainda acredita-se que a maneira mais intuitiva para tentar caracterizar a produtividade de um pesquisador consiste simplesmente em contar o número de artigos publicados, enquanto que, para quantificar a qualidade de sua produção, é possível destacar o número de vezes em que cada autor tenha sido referenciado por outros trabalhos.

Dentro desse contexto podemos agora apresentar alguns pontos principais do índice h como um candidato a indicador. Primeiramente, uma característica intrigante é o fato de que o índice h não ter sido ainda formulado por uma equação matemática, como seria o caso do número total de citações ou a média de citações recebidas ao longo do tempo. Ao contrário, nas palavras de Hirsh, ***um cientista terá um índice h se h de seus N artigos tiverem ao menos h citações cada um***. Para compreender em detalhe como esse índice pode ser encontrado, considere, por exemplo, um pesquisador que tenha publicado 10 artigos científicos. A Tabela 1-A apresenta a lista desses artigos, listados por ordem decrescente em função do ano de publicação. O primeiro passo para o cálculo do índice h

consiste simplesmente em reordenar os artigos, desta vez, por ordem decrescente em relação ao número total de citações recebidas (Tabela 1-B).

Na Tabela 1-B, a segunda coluna representa o número total de citações recebidas por cada artigo, e a primeira coluna representa o número acumulado de artigos publicados. Finalmente, o cálculo do índice h , utilizando a Tabela 1-B, consiste em contar a quantidade de células em que o número total de citações é maior ou igual ao número de artigos publicados. Para o exemplo da Tabela 1, o índice h é igual a 8 para o pesquisador, ou em outras palavras, esse pesquisador possui 8 artigos científicos que receberam, cada um, ao menos 8 citações.

Tabela 1: Exemplo para o cálculo do índice h de um pesquisador com 10 artigos publicados.

Lista de artigos científicos					
Tabela 1-A			Tabela 1-B		
Ano	Artigo	Nº de Citações	Artigo	Nº de Citações	
2009	1	1	1	15	
2008	2	8	2	15	
2008	3	12	3	13	
2007	4	4	4	12	
2005	5	15	5	12	
2005	6	12	6	10	
2004	7	10	7	8	
2003	8	8	8	8	
2003	9	13	9	4	
2000	10	15	10	1	

Hirsh ainda destaca algumas vantagens na utilização do índice h em relação a outros índices normalmente adotados. Segundo observa, o número total de artigos publicados representa a produtividade do pesquisador, mas não é capaz de indicar a importância ou o impacto dos artigos publicados, enquanto que o número total de citações representa o impacto dos artigos, mas não a produtividade. Ele também chama atenção para um viés de interpretação nos casos em que o pesquisador tem baixa produtividade e alta citação ou o contrário. Para entender melhor esta questão, segue o

exemplo: imagine um pesquisador com somente cinco artigos publicados (1, 2, 3, 4, 5) e que os mesmos tenham recebido as seguintes citações: 100, 15, 10, 3 e 2. A partir desses dados, é possível calcular o número total de citações recebidas e a média de citações por artigo.

Nesse caso, a soma e a médias serão 130 e 26, respectivamente. Note que, o índice h para o pesquisador deste exemplo corresponderia a 3. Por outro lado, é fácil perceber que os dois valores, soma e média, foram inflacionados por apenas um artigo que recebeu um número de citações bem acima da média (100 citações) enquanto o índice h equaliza esta relação tão desbalanceada. Isso nos leva a concluir, que os primeiros índices são, de fato, sensíveis a artigos com um alto número de citações, já para o calculo do índice h este tipo de viés é ignorado, se a produtividade não acompanhar. Pesam para mais ou não pesam os chamados “artigos de revisão”, que são exemplos clássicos de artigos com freqüência recebem um alto número de citações devido ao papel que desempenham na comunidade científica, mas, por outro lado, esse número não está necessariamente correlacionado à qualidade do artigo.

A elegância do índice h pode ser compreendida pela simplicidade em combinar duas variáveis, a produtividade científica (número total de publicações) e o impacto das publicações (número total de citações), de um pesquisador que são expressas em apenas uma variável, o próprio índice. Ao contrário de estimar o desempenho científico a partir do número total de citações ou do número total de publicações, o índice h apresenta a vantagem de não aumentar drasticamente seu valor pela presença de artigos muito citados nem de um número grande de publicações. Assim, um pesquisador que tenha publicado um artigo com 100 citações tem o potencial para alcançar um índice h igual a 100, mas para isso precisa ter 100 artigos com no mínimo 100 citações cada, caso contrário a contribuição desse artigo será ponderada de acordo com o impacto dos outros artigos. O mesmo entendimento se aplica a um pesquisador com 100 publicações.

Uma das primeiras críticas ao índice h foram apresentadas por BATISTA et al (2006) e está relacionada aos diferentes padrões de colaborações realizadas pelos pesquisadores para o desenvolvimento das pesquisas científicas e que, em muitos casos, se refletem em

múltiplas autorias em um único artigo científico. Embora o índice h tenha como objetivo ser um indicador capaz de avaliar a produção científica de um pesquisador, ele não considera que os créditos de um artigo científico devam, a princípio, ser compartilhados, ou divididos, entre os autores.

Como calcular o índice h ?

O Institute for Scientific Information (ISI)¹ possui uma das mais importantes bases de dados contendo informações sobre os artigos publicados pela comunidade científica. Dentre todas as ferramentas disponibilizadas pelo ISI, o portal conhecido por Web of Science (WoS) tem se tornado indispensável para a comunidade científica, pois permite, de forma rápida e segura, a busca e acesso às diferentes formas bibliográficas, especialmente os artigos científicos, publicados em periódicos especializados. A utilização desta base de dados não se restringe apenas à busca de artigos em periódicos, sendo possível acompanhar toda a produção bibliográfica e toda a citação recebida ao longo dos anos de um pesquisador, grupo de pesquisa, instituição científica, país, tema, área de conhecimento, periódico, etc. Recentemente, esse portal passou a disponibilizar a busca das citações de cada pesquisador desconsiderando as auto-citações, pois estas podem influenciar no processo de citação (SCHREIBER, M., 2007). A figura 2 mostra o resultado de publicações e citações, catalogadas na base WoS, para o autor Batista, PB em que o cálculo para o índice h é gerado automaticamente.

¹ Desde XXX a base foi comprada por um grupo de empresas e passou a se chamar Thomson Reuters / ISI

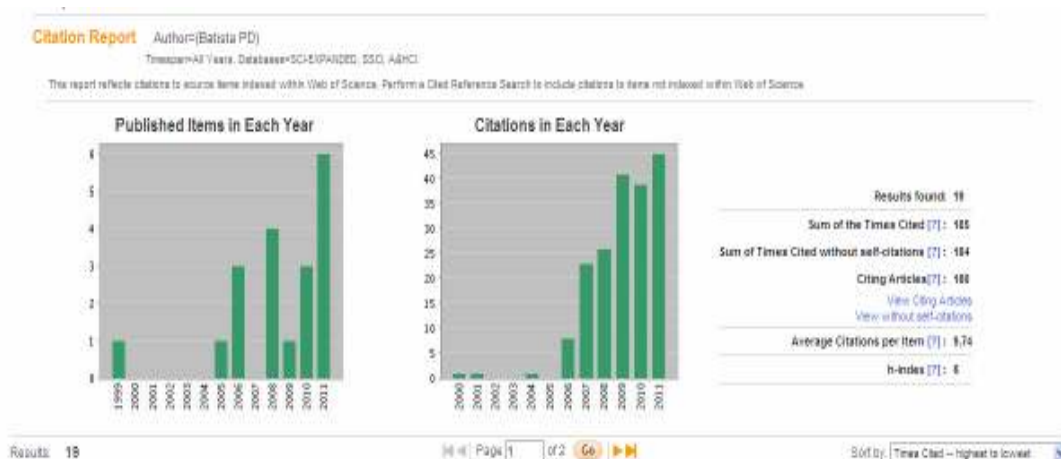


Figura 2: Relatório de produção e de citação gerado para o autor Batista, PD, na base Web Of Science. (consulta em novembro 2011)

Até pouco tempo, o portal WoS se constituía na única base de dados bibliográfica que disponibilizava as informações sobre citações e que permitia extrair um conjunto de dados sobre a produção científica em diferentes níveis de agregação. Desde 2004, uma nova base bibliográfica, a Scopus, passou também a oferecer estes serviços. Um análise das vantagens e desvantagens dessa base de dados comparado com a WoS podem se encontradas em LOKMAN I *et al*, 2006). Uma análise semelhante àquela da Figura 2 foi realizada na base Scopus e está apresentada na Figura 3. Notem que o índice *h* nesta base é menor do que na WoS, mostrando assim, que o valor para os indicadores também depende da base de dados utilizada.

linear com o valor de h . É fácil perceber, por meio da Tabela 2, que essas duas grandezas, citação mínima e h , estão relacionadas pela seguinte equação $T_{Cmin} = h^2$.

Iremos primeiramente mostrar que duas condições, relacionadas com o número de artigos e suas respectivas citações, devem necessariamente ser satisfeitas para que o índice h seja incrementado em uma unidade. A primeira delas, e a mais simples de perceber, é que o número total de artigos deve aumentar em ao menos uma unidade. A segunda, como mostram as duas últimas colunas da Tabela 2, diz que o número mínimo para o total de citações deve incrementar de $(2*h+1)$ unidades. Esse resultado pode ser obtido considerando que $\Delta T_{Cmin} (h_{i+1}-h_i) = (h+1)^2-h^2 = 2*h+1$.

Tabela 2: Apresenta a condição mínima, número de artigos e suas respectivas citações, para que um pesquisador tenha um índice h igual a 1 até 10.

Tabela 2												
h	Lista de artigos ordenados por N ^o de citação								T _{Cmin}	ΔT_{Cmin}	ΔT_{Cmin}	
h = 0	0									0	$\Delta(h_1-h_0) = 1$	$\Delta(h_2-h_1) = 3$
h = 1	1									1		
h = 2	2	2								4	$\Delta(h_3-h_2) = 5$	$\Delta(h_4-h_3) = 7$
h = 3	3	3	3							9		
h = 4	4	4	4	4						16	$\Delta(h_5-h_4) = 9$	$\Delta(h_6-h_5) = 11$
h = 5	5	5	5	5	5					25		
h = 6	6	6	6	6	6	6				36	$\Delta(h_7-h_6) = 13$	$\Delta(h_8-h_7) = 15$
h = 7	7	7	7	7	7	7	7			49		
h = 8	8	8	8	8	8	8	8	8		64	$\Delta(h_9-h_8) = 17$	$\Delta(h_{10}-h_9) = 19$
h = 9	9	9	9	9	9	9	9	9		81		
h = 10	10	10	10	10	10	10	10	10	10	100		

Como um exemplo fictício, o quadro 1 apresenta, em detalhe, os passos para que o índice h igual a 5 seja incrementado em uma unidade, passando assim, para 6. Para uma melhor discussão, vamos agrupar os artigos em dois grupos, sendo que o primeiro corresponderá sempre aos h artigos com no mínimo h citações (Grupo A), e o segundo será o restante dos artigos publicados. A primeira condição nos diz que ao menos mais um artigo precisa ser publicado e, como condição inicial, esse artigo entra no Grupo B sem nenhuma citação. Em seguida, o pesquisador deve receber um total de no mínimo onze novas citações ($\Delta=2*h+1$), de tal maneira que cada artigo do grupo A receba uma única

citação, enquanto que o artigo do grupo B receba seis citações. Nesse caso, esse artigo passa agora a participar do antigo grupo A, renomeado como grupo C.

Quadro 1: Passos para que o índice h igual a 5 seja incrementado em uma unidade.

$h = 5$	5	5	5	5	5			
	Grupo A					Grupo B		
	5	5	5	5	5	0		Um novo artigo é publicado
	Grupo A					Grupo B		
	+1	+1	+1	+1	+1	+6		Onze citações são recebidas
$h = 6$	6	6	6	6	6	6		
	Grupo C					Grupo D		

De uma maneira geral, cada um dos h artigos dentro do grupo A deve receber ao menos mais uma citação e ao menos um artigo deve ser publicado, caso ainda não exista artigos no grupo B. O primeiro artigo do grupo B deve receber ao menos $(h+1)$ citações. É possível agora propor um índice h fracionário (h_f) composto por três partes, que serão somadas como mostra a equação abaixo:

$$h_f = h + [(h_a * w_a) + (h_b * w_b)] / 2 \quad (1)$$

No exemplo proposto, h_a é a porcentagem do número de artigos do grupo A que possui um número de citação maior que h . E, h_b representa a porcentagem do número de citações que o primeiro artigo do grupo B deve ter para que participe do grupo C. Por último, w_a e w_b são os pesos atribuídos a essas duas partes. Lembrando que, nesse caso, o grupo C será composto por artigos com ao menos $(h+1)$ citações.

Finalmente, os pesos podem ser obtidos considerando o número total de citações que o pesquisador precisa receber para que seu índice h seja incrementado em uma unidade. Sabemos que o pesquisador deve receber ao menos $(\Delta=2*h+ 1)$ citações de tal maneira que os artigos do grupo A serão responsáveis por h citações, enquanto o primeiro artigo do grupo B por $(h+1)$ citações. Então,

$$w_a = h / (2 * h + 1) \quad (2)$$

enquanto

$$w_b = (h + 1) / (2 * h + 1) \quad (3)$$

Considerando as equações (2) e (3) é fácil mostrar que $w_a + w_b = 1$. Para que essas equações fiquem claras, o quadro 2 apresenta (em detalhe) o cálculo da parte fracionária do índice h de um pesquisador com um índice h igual a 5 e um total de 111. Nesse caso particular, é necessário que ao menos os dois últimos artigos do grupo A recebam no mínimo uma citação cada um, desde que três artigos do grupo A já contam com ao menos $(h+1)$ citações. Em outras palavras, 60 % dos artigos do grupo A possuem as condições mínimas necessárias para comporem o grupo C.

Quadro 2: Exemplo para o cálculo da parte fracionária do índice h para um pesquisador com 8 publicações e h igual a 5.

$h = 5$	50	35	10	5	5	3	2	1
	Grupo A					Grupo B		
				+ 1	+ 1	+ 3		
$h = 6$	50	35	10	6	6	6	2	1
	Grupo C						Grupo D	

Neste exemplo, um novo artigo não precisa ser publicado, pois, nesse momento, o grupo B é composto por três artigos. Mas, ao menos o primeiro artigo do grupo B deve receber no mínimo três citações para participar do grupo C. Isto mostra que o primeiro artigo do grupo B recebeu 50 % do número de citações necessárias para ser inserido no grupo C. Os dados referentes a estes exemplos estão sumarizados na Tabela 5. O índice h fracionário é facilmente encontrado substituindo esses valores na eq. 1. O pesquisador com índice h igual a 5 terá um $h_f = 5 + (0.270 + 0.275) / 2 = 5.2725$. Outro pesquisador ($h = 5$), tendo seis artigos publicados mas com cinco citações cada um, terá um índice

fracionário ($h_f = 5 + (0.0 + 0.45) / 2 = 5.225$) menor que o obtido no exemplo anterior. Dessa forma a parte fracionária do índice h pode ser utilizada como critério de desempate.

Tabela 5: Etapas para o cálculo da parte fracionária do índice h .

$h_a = 3/5 = 0.6$	$w_a = (5) / (2*5 + 1) = 0.45$	$h_a * w_a = 0.6*0.45 = 0.270$
$h_b = 3/6 = 0.5$	$w_b = (5 + 1) / (2*5 + 1) = 0.55$	$h_b * w_b = 0.5*0.55 = 0.275$

Conclusão

O número total de citações recebidas por um artigo é frequentemente utilizado como um indicador de sua qualidade, pois se acredita que quanto maior o número de citações, melhor a qualidade dos resultados nele apresentados. Entretanto, há uma vasta literatura que indica que existem diversos fatores envolvidos no processo de citação, haja vista que, citar ou não um artigo sempre será uma escolha pessoal dos autores (BORNMANN & DANIEL, 2008). Pesquisadores podem escolher, entre os vários artigos disponíveis para compor a sua lista de referência, aqueles publicados por colaboradores, ou por autores de mesma nacionalidade, ou ainda artigos que tenham um grande número de citações. Portanto, o número total de citações que um trabalho ou um autor recebe representa seu impacto na comunidade científica, mas não está necessariamente relacionado à qualidade do mesmo.

Nesse ponto, podemos propor uma reflexão para compreender por que outras proposições, que foram apresentadas ao longo do tempo e que de certa maneira se mostraram mais ousadas, não foram abraçadas pela comunidade científica. Dentre essas, é possível destacar, por exemplo, trabalhos que apresentam indicadores para quantificar a relevância de artigos científicos considerando diretamente toda a rede de citações. Um dos resultados interessantes desses trabalhos é a conclusão de que por meio desses indicadores seria possível, por exemplo, detectar a presença de artigos que foram relevantes para a construção do conhecimento científico, mas que, por outras razões, foram esquecidos durante o processo de citação (CHEN ET AL, 2007). Ainda é relevante destacar que a proposta de agregar duas grandezas em apenas uma medida vem sendo também aos poucos utilizada em outras áreas de pesquisa (KORN, A, 2009; NARUKAWA Y,

2009). Por exemplo, a proposta apresentada por CAMPITELI MG et al (2010) em estender o conceito desse indicador para uma análise quantitativa da centralidade de um nó em uma rede complexa. Assim, ao contrário de apresentar um artigo científico com 850 citações, é possível destacar que este artigo está sendo citado por 33 artigos com ao menos 33 citações cada um. Em outras palavras, um índice h igual a 33 pode ser atribuído ao artigo em contraponto ao número total de citações. Nesta proposta, o indicador de qualidade ou impacto está levando em consideração não apenas o número total de citações, como também a qualidade dessas citações.

Agradecimentos: O Autor desse trabalho agradece a Dr. Marcia Reis e Dr. Jacqueline Leta pela revisão e comentários durante a elaboração desse manuscrito.

Referência

ALONSO S, CABRERIZO, FJ, HERRERA-VIEDMA, HERRERA, F, **h-Index: A review focused in its variants, computation and standardization for different scientific fields**, *Journal of Informetrics* Vol 3 Issue 4 (2009) 273-289.

BALL P, **Index aims for ranking of scientists**, *Nature*, Vol 436 Issue 7053, (2005) 900-900.

BATISTA PD., CAMPITELI, M.G., KINOUCI O., MARTINEZ A.S **It is possible to compare researchers with different scientific interests?** *Sientometrics* (2006) 68:179-189.

BORNMANN, L, HANS-DIETER D., **What do citation counts measure? A review of studies on citing behavior**, *Journal of Documentation*, Vol 64. No. 1, (2008) 45-80.

BORNMANN, L, HANS-DIETER D., **What do we know about the h index**, *Journal of the American society for information science and technology* 58 (9): (2007)1381-1385.

CAMPITELI MG, HOLANDA J. ADRIANO, SCOLES R.C.P, SOARES H.D, KINOUCI O, **Hirsch index as a network centrality measure**, *arXiv.org* E-Print (Junho, 27, 2010). Disponível em <http://arxiv.org/abs/1005.4803>

CHEN P, XIE H, MASLIV S, REDNER S **Finding scientific gems with Google**. *Journal of Informetrics* 1:8–15 (2007).

HIRSCH JE (2005) **An index to quantify an individual's scientific research output.** *Proc Natl Acad Sci USA* **102**: 16569–16572.

KORN, A. SHUBERT, A.T, **Lobby index in networks,** *Pysica A* 388 (2009) 2221-2226

LATOUR, B.; WOOLGAR, S. A Credibilidade Científica. In: **A vida de Laboratório – A Produção dos Fatos Científicos.** Ed. Relume Dumará, Rio de Janeiro, 1997.

LOKMAN I.M, KIDUK Y., **Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science Versus Scopus and Google Scholar,** *Journal of the American Society for Information Science and Technology*, 58 (13):2105-2125, 2007.

NARUKAWA Y. TORRA V. **Multidimensional generalized fuzzy integral.** *Fuzzy Sets and Systems*, Vol 160. Issue 6, 16 March 2009, Pages 802-815.

SCHREIBER, M. **Self-citation corrections for the Hirsch index,** *Europhysics Letters* (2007) 78:3:30002