

ÍNDICE H PARA TORNEIOS ROUND ROBIN – UM ESTUDO DE CASO

Renata dos Santos Constant

Universidade Federal Fluminense
Rua Passo da Pátria 156, São Domingos, 24210-240, Niterói, RJ
renata.constant@hotmail.com

Luciana Gonçalves Machado

Universidade Federal Fluminense
Rua Passo da Pátria 156, São Domingos, 24210-240, Niterói, RJ
lucianagoncalves@id.uff.br

Deivison da Silveira Pereira

Universidade Federal Fluminense
Rua Passo da Pátria 156, São Domingos, 24210-240, Niterói, RJ
dspereira@id.uff.br

João Carlos Correia Baptista Soares De Mello

Universidade Federal Fluminense
Rua Passo da Pátria 156, São Domingos, 24210-240, Niterói, RJ
jcsmello@producao.uff.br

RESUMO

Este artigo tem como objetivo adaptar os conceitos bibliométricos do índice tipo h , propondo um novo índice para avaliação de torneios, tomando como base para o cálculo o número de competidores e as vitórias (ou derrotas). Os resultados obtidos com o índice proposto serão analisados de duas formas: a primeira através do núcleo h e a segunda através de um estudo de complementaridade dos resultados obtidos para vitórias e para as derrotas. O estudo de caso será realizado em um torneio do tipo *Round Robin* e para isso foi escolhida a primeira etapa da competição de Voleibol nas Olimpíadas de 2012.

PALAVRAS CHAVE. Índice h , Torneio, Jogos Olímpicos.

Tópicos: OA – Outras aplicações em PO

ABSTRACT

This article aims to adapt the bibliometric concepts of h -index, proposing a new index for evaluation of tournaments, taking as a basis for calculating the number of competitors and wins (or losses). The results obtained with the proposed index will be analyzed in two ways: the first through the h -core and the second through a study of the complementarity of results for wins and losses. The case study will be conducted in a tournament type Round Robin and it was chosen the first stage of Volleyball competition in the 2012 Olympics.

KEYWORDS. h -index, Tournament, Olympic Games.

Paper topics: OA - Other applications in OR

1. Introdução

A avaliação do trabalho de um pesquisador é um problema cada vez mais estudado na comunidade científica. Em 2005, Hirsch propôs o índice bibliométrico, denominado índice h , para avaliar e comparar a atividade de pesquisa dos cientistas de forma individual e de acordo com suas publicações. A popularidade desse indicador na área da ciência da informação é devida provavelmente por sua importância, por sua simplicidade e pela sua fácil compreensão [Franceschini e Maisano 2010].

Nesse artigo também serão estudadas outras aplicações do índice h , que extrapolaram sua utilização para além do desempenho individual do pesquisador. [Pereira et al 2015] propõem uma nova adaptação dos conceitos do índice h com a definição das localidades centrais da malha área brasileira utilizando essa técnica. Além desse estudo, existem outros envolvendo centralidade e princípios do índice h que podem ser encontrados em [Korn et al. 2009], [Zhao et al. 2011] e [Abbasi et al. 2013].

O presente artigo se propõe a sugerir uma nova aplicação do índice h para avaliações em torneios “*Round Robin*”, utilizando como insumo a relação entre os competidores e o número de vitórias e derrotas obtidas. Dessa forma, pretende-se analisar a importância de um competidor e a competitividade dentro um torneio.

Na próxima seção será apresentado o referencial teórico deste estudo e na seção 3 a proposta para adaptação do índice h para torneios. Na seção 4 será analisado o estudo de caso, além de seus resultados e, por fim, na seção 5 serão feitas as conclusões e propostas para estudos futuros.

2. Referencial Teórico

2.1. Índice h

O índice h foi desenvolvido por Hirsch, com o objetivo de quantificar a atividade científica e medir o impacto de pesquisadores, baseado nos seus *papers* mais citados. O índice h de um pesquisador é o número de documentos publicados com pelo menos h citações cada, isto é, o valor de h é o maior número de artigos que um autor possui com pelo menos o mesmo número de citações, conforme figura 1 [Hirsch 2015]. O conjunto formado pelas h publicações com pelo menos h citações formarão o chamado núcleo do índice h [Franceschini e Maisano 2010].

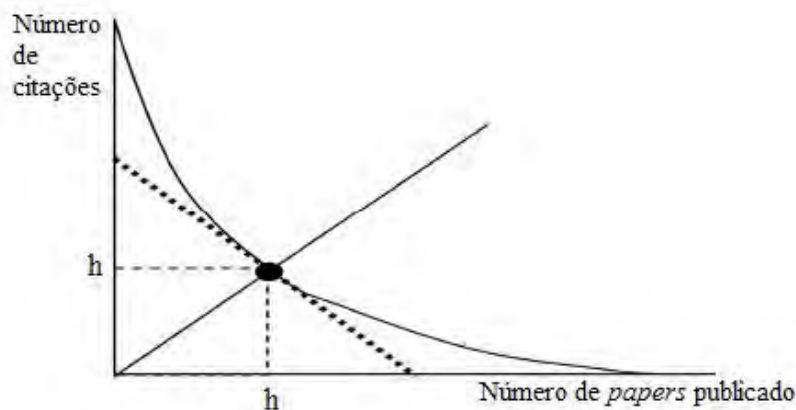


Figura 1 – Curva do índice h

O índice h é bastante utilizado em diversas áreas associadas à pesquisa científica e possui centenas de artigos relacionados a esse índice na literatura [Pereira et al. 2015]. [Rubem et al. 2014] realizaram uma análise do índice- h em grupos de pesquisadores e lembram ainda que o surgimento do índice- h em 2005 impactou significativamente a pesquisa bibliométrica, e após a

publicação do artigo seminal muitos pesquisadores se dedicaram a desenvolver variações, extensões e generalizações do índice de Hirsch. No entanto, esse índice está sujeito a algumas críticas sendo as principais destacadas abaixo [Franceschini e Maisano 2010]:

- Não leva em conta múltiplas coautorias e a autocitação, taxa que não é irrelevante entre os pesquisadores mais jovens [Burrell 2007];
- Não é útil para comparações entre pesquisadores de áreas diferentes [Antonakis e Lalive 2008], [Batista et al. 2006], [Braun et al. 2006];
- Não considera o tempo da publicação [Sidiropoulos et al. 2007];
- Não leva em conta as características das publicações, uma vez que não é possível uniformizar os canais de comunicação científicas [Castillo et al. 2007];
- O valor encontrado pelo índice pode ser afetado por citações feitas de forma incorretas [Bornmann e Daniel 2007], [Harzing e van der Wal 2008].

Além do índice h existem os chamados índices tipo h . A comparação de alguns destes índices com o índice h é feita por Brandão e Soares de Mello (2015).

No presente estudo serão analisados outros usos dados ao índice h , com o objetivo de ampliar o campo de aplicação dessa técnica.

2.2. Outros Usos do Índice h – Centralidades

A adaptação do índice h foi feita em outros estudos envolvendo, por exemplo, teoria de grafos, mais especificamente centralidade de redes. [Korn et al. 2009] definem, para analisar a centralidade em uma rede, o valor tipo h de cada vértice como a maior quantidade k de vizinhos que apresentam pelo menos grau k .

[Zhao et al. 2011] e [Abbasi et al. 2013] apresentam diferentes modelos de centralidade com base nos princípios do índice h , enquanto [Zhai et al. 2013] verificam a utilização do índice h em uma rede ponderada de comunicação envolvendo colaboração científica, no entanto é calculado o grau h de cada vértice, mas a centralidade depende da quantidade de vértices analisados no grafo. Esses trabalhos têm em comum o envolvimento do índice h com medidas de centralidade no âmbito da pesquisa científica.

[Pereira 2015] apresenta uma adaptação feita do índice h para determinação de centralidade na malha aérea brasileira, onde se utiliza uma adaptação do índice- h para determinar o impacto de cada vértice e é feita uma comparação com outros aeroportos a fim de definir uma ordenação de qual vértice seria o mais central.

Diferente de [Zhao et al. 2011] e [Abbasi et al. 2013] que depende da quantidade de vértices analisados, a proposta de [Pereira et al. 2015] simplifica a centralidade tipo h , visto que o valor tipo h de cada vértice não muda em relação no grafo analisado. Desta forma, em redes de transporte o impacto específico de um aeroporto não muda se comparado com outros aeroportos.

2.3. Outras Avaliações de Torneios

A avaliação de torneios envolvendo o índice h pode ser visto em [Kossi et al 2015] onde verificam uma competição em nível acadêmico e o índice é um dos itens a serem considerados para formulação do *ranking*.

Estudos envolvendo avaliação em esportes e torneios abordam diferentes perspectivas e analisam a eficiência de jogadores em torneios, como feito por [Lins et al 2003] e em [Soares de Mello et al 2009] onde analisaram o *ranking* em Jogos Olímpicos com base em DEA (*Data Envelopment Analysis*). Avaliações de desempenho em Jogos Olímpicos também foram feitas com análise de centralidades [Calzada-Infante and Lozano, 2016]. [Gutiérrez e Lozano 2014] avaliaram o desempenho das equipes de corrida da Fórmula 1 comparado com os investimentos feitos por elas.

[Brandao et al 2015] avaliaram o torneio da Taça Guanabara de 2014 utilizando técnicas de teoria dos grafos com ênfase em centralidades para buscar um *ranking* considerando a performance dos times.

Alguns autores verificam quais os melhores formatos para uma competição, neste sentido [Budden et al 2015] verificaram uma nova proposta de formato de jogo que reduz a variabilidade e melhora a performance do *ranking*. [Hirotzu et al 2015] analisam formatos de torneios considerando o número de partidas e a probabilidade de vitória na competição.

Outros autores verificam chances de um torneio ser manipulado. [Mattei et al 2015] propõem modelos probabilísticos para analisar torneios com rodadas diferentes e verificar as chances de manipulação ou suborno em torneios com poucas informações. [Pauly 2014] propõe modelo matemático para verificar estratégias usadas em times que perdem partidas específicas para conseguir ganhar a competição, olhando especialmente em torneios *Round Robin*.

2.4. Definição de um Torneio “Round Robin”

Existem dois tipos de torneio: o torneio “*Play-Off*” e o torneio “*Round Robin*”. O torneio “*Play-Off*” é aquele em que a cada jogo o vencedor elimina o perdedor, até que sobre apenas um vencedor. Já o torneio “*Round Robin*” é aquele onde todos os jogadores competem entre si uma única vez, o vencedor do torneio é aquele que tem o maior número de vitórias ou pontos, de acordo com as características específicas de cada torneio [Brandão et al. 2015].

Os torneios “*Round Robin*” são grafos orientados conforme demonstrado na figura abaixo:

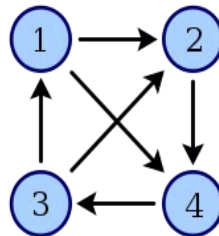


Figura 2 – Grafo de um Torneio *Round Robin*

Os torneios “*Round Robin*” são considerados mais justos, pois levam em consideração os resultados dos competidores em vários jogos, enquanto que nos torneios “*Play-Off*” uma performance pontual de um competidor influencia diretamente no resultado final do torneio [Brandão et al. 2015].

3. Adaptação do Índice h para Torneios

Assim como o índice h compara citações e *papers* para avaliar o nível de impacto do artigo e dos pesquisadores, o índice h proposto nesse artigo irá relacionar os competidores de um torneio com as vitórias e derrotas obtidas. Logo, o objetivo desse artigo é avaliar a importância de um competidor em um torneio e a competitividade deste.

No índice h proposto, a quantidade de competidores se equivale ao número de *papers* publicados no índice h original, e a quantidade de vitórias (ou derrotas) se equivale ao número de citações. Apesar de o novo índice utilizar a quantidade de competidores na sua composição, o índice pretende avaliar o torneio como todo e não seus competidores.

A figura 3 ilustra o índice h proposto, onde a relação *papers* x citações do índice h é alterada respectivamente para a relação competidores x vitórias. Nesse caso, o valor do índice h é encontrado quando o número de vitórias for igual ao número de competidores. E o núcleo do índice h será composto pelos times melhor posicionados no *ranking* que obtiverem mais de h vitórias.

O novo índice h considera, inicialmente, apenas torneios em que os resultados dos jogos se dão por uma relação binária de vitória ou derrota, não considerando casos de empate.

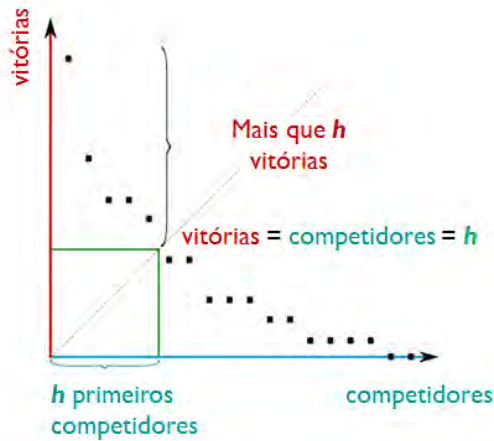


Figura 3 – Índice h proposto

4. Estudo de Caso

Para o estudo de caso desse artigo foi escolhida a primeira fase da competição de vôlei nos Jogos Olímpicos de 2012. Essa escolha foi feita devido às características de torneio “Round Robin” desse evento em que se tem 12 competidores divididos em dois grupos (A e B) com o mesmo número de integrantes, e todos os times dentro do grupo competem entre si uma única vez. Além disso, o vôlei não admite empate nos jogos. Para a fase final dessa competição, que não é objeto desse estudo, foram selecionados os quatro melhores de cada grupo.

Na tabela 1 é detalhado como foi calculado o índice h baseado no número de vitórias dos times em cada grupo. Com isso é possível analisar que em ambos os grupos o índice h obtido foi igual a 3. No caso do grupo A, como três países (Polônia, Argentina e Itália) tiveram o número de vitórias igual a 3, todas essas equipes foram incorporadas ao núcleo h . No caso do grupo B, como três equipes tiveram 4 vitórias (Estados Unidos, Brasil e Rússia) todas foram incluídas no núcleo. Cabe destacar que o objetivo neste estudo de caso não é fazer um *ranking* que discrimine os empates, por isso no caso de mais de um país ter o mesmo número de vitórias a tratativa dada é ou todos fazem parte do núcleo ou nenhum compõem o grupo do núcleo.

	Grupo A	Vitórias	Rank		Grupo B	Vitórias	Rank
Núcleo h	Bulgária	4	1	Núcleo h	Estados Unidos	4	1
	Polônia	3	2		Brasil	4	2
	Argentina	3	3		Rússia	4	3
	Itália	3	4		Alemanha	2	4
	Austrália	2	5		Sérvia	1	5
	Grã-Bretanha	0	6		Tunísia	0	6

Tabela 1 – Índice h das vitórias nos grupos A e B

Com o objetivo de verificar a existência de alguma relação entre as vitórias e as derrotas obtidas dentro de cada conjunto, foi calculado também o índice h das derrotas para cada grupo, conforme detalhado na tabela 2.

	Grupo A	Derrotas	Rank		Grupo B	Derrotas	Rank
Núcleo h	Grã-Bretanha	5	1	Núcleo h	Tunísia	5	1

Grupo A	Derrotas	Rank
Austrália	3	2
Polônia	2	3
Argentina	2	4
Itália	2	5
Bulgária	1	6

Grupo B	Derrotas	Rank
Sérvia	4	2
Alemanha	3	3
Estados Unidos	1	4
Brasil	1	5
Rússia	1	6

Tabela 2 – Índice h das derrotas nos grupos A e B

Comparando as tabelas 1 e 2, é possível notar que o núcleo h das vitórias é complementar ao núcleo h das derrotas para o exemplo desse estudo de caso. Ou seja, caso um país pertença ao núcleo h das vitórias de seu grupo, conseqüentemente ele não pertencerá ao núcleo h das derrotas. Além disso, é possível dizer que o grupo A é um grupo mais competitivo do que o grupo B, pois o núcleo h de vitórias do grupo A é composto por mais times do que quando comparado ao núcleo do grupo B.

Para testar a questão da complementaridade dos núcleos, foi realizada uma simulação adicionando três países fictícios em cada grupo. O objetivo dessa simulação é verificar a possibilidade de um mesmo time pertencer aos dois núcleos, vitória e derrota, ao mesmo tempo. Para adicionar os países fictícios, foi preciso fazer uma adaptação no número de vitórias e derrotas com o intuito de manter a coerência dos dados com relação aos resultados da competição (quando um time ganha, outro necessariamente tem que perder). Dessa maneira, a tabela 3 mostra os resultados obtidos na simulação para o grupo A.

Núcleo h de vitórias	Grupo A	Vitórias	Rank	Núcleo h de derrotas	Grupo A	Derrotas	Rank
	País 1	6	1		Austrália	6	1
País 2	6	2	Grã-Bretanha	6	2		
País 3	5	3	Itália	5	3		
Bulgária	4	4	Bulgária	4	4		
Polônia	4	5	Polônia	4	5		
Argentina	4	6	Argentina	4	6		
Itália	3	7	País 1	3	7		
Austrália	2	8	País 2	2	8		
Grã-Bretanha	2	9	País 3	2	9		

Tabela 3 – Simulação do índice h para o grupo A

Na tabela 3, os países destacados (Bulgária, Polônia e Argentina) fazem parte tanto do núcleo h de vitórias quanto do núcleo h de derrotas. No entanto, através da simulação foi possível notar que mesmo sendo possível uma equipe fazer parte dos dois núcleos ao mesmo tempo, esse é um resultado pouco provável. E ocorre principalmente devido aos empates que acontecem no número de vitórias ou de derrotas.

5. Conclusões

Através do estudo de caso apresentado na seção anterior é possível verificar que, se o núcleo h relacionado às vitórias tem muitos competidores, significa dizer que o torneio é fortemente competitivo, possuindo muitos competidores com capacidade de alcançar bons resultados no torneio. Nesse sentido, um alto núcleo h indica jogos mais acirrados, mais torcida, e conseqüentemente mais arrecadação para os times. Em contrapartida, um núcleo h alto relacionado às derrotas indica um torneio competitivo, porém de times fracos, com desempenho insatisfatório.

Uma das questões a serem validadas pelo artigo era a premissa da complementaridade do núcleo h das vitórias e das derrotas. No entanto, através da simulação foi possível verificar que não há necessariamente complementaridade dos dois grupos. Além disso, através da Tabela 3 é possível verificar que esse fato se deu principalmente devido ao empate no número de vitórias e derrotas encontradas no núcleo h .

Para trabalhos futuros se faz necessário desenvolver um método de desempate do índice h , promovendo assim um *ranking* mais heterogêneo e real dos competidores, permitindo uma melhor análise dos reais valores do índice h em torneios. Além disso, é importante aprimorar os estudos do índice h relacionado a jogos em geral, no intuito de estabelecer um modelo que lide também com a possibilidade de empates nos resultados das partidas, além das vitórias e derrotas.

Referências

Abbasi, A. (2013). h -Type hybrid centrality measures for weighted networks. *Scientometrics*. v. 96, p. 633–640.

Antonakis, J. e Lalive, R. (2008). Quantifying scholarly impact: IQ_p versus the Hirsch h . *Journal of the American Society for Information Science and Technology*. v. 59, n. 6, p. 956–969.

Batista, P.D., Campiteli, M.G., Kinouchi, O. e Martinez, A.S. (2006). Is it possible to compare researchers with different scientific interests?. *Scientometrics*. v. 68, n. 1, p. 179–189.

Bornmann, L. e Daniel, H.D. (2007). What do we know about the h index? *Journal of the American Society for Information Science and Technology*. v. 58, n. 9, p. 1381–1385.

Brandão, L. C., Del-Vecchio, R.R. e Soares de Mello, J.C.C.B. (2015). Graph Centrality Analysis for the Evaluation of the 2014 Guanabara Cup. In: *5th International Conference on Mathematics in Sport*, p. 11-17, Loughborough University.

Brandão, L.C.; Soares de Mello, J.C.C.B. (2015) Análises Comparativas entre Indicadores Bibliométricos. In: XLVII SBPO, p. 2210-2221. Porto de Galinhas.

Braun, T. e Glänzel, W., Schubert, A. (2006). A Hirsch-type index for journals. *The Scientist*. v. 69, n. 1, p. 169–173.

Budden, D., Wang, P., Obst, O., Prokopenko, M. (2014). Simulation Leagues: Analysis of Competition Formats. *RoboCup 2014: Robot World Cup XVIII, Lecture Notes in Computer Science*. v. 8992, p. 183-194.

Burrell, Q.L. (2007). Should the h -index be discounted? In: *Glänzel, W., Schubert, A., Schlemmer, B. (Eds.), In the Multidimensional World of Tibor Braun: A Multidisciplinary Encomium for his 75th Birthday*, p. 65–68, ISSI, Leuven.

Calzada-Infante, L., Lozano, S. (2016). Analysing Olympic Games through dominance networks *Physica A: Statistical Mechanics and its Applications*, doi:10.1016/j.physa.2016.07.001.

Castillo, C., Donato, D. e Gionis, A. (2007). Estimating number of citations using author reputation. In: *String Processing and Information Retrieval*, p. 107–117, Springer, Berlin.

Franceschini, F. e Maisano, D.A. (2010). Analysis of the Hirsch index's operational properties. *European Journal of Operational Research*, p. 494–504.

- Gutiérrez, E., Lozano, S. (2014). A DEA Approach to Performance-Based Budgeting of Formula One Constructors. *Journal of Sports Economics*. v. 15, n. 2, p 180-200.
- Harzing, A.W. e van der Wal, R. (2008). Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics*. v. 8, n. 11, p. 61–73.
- Hirotsu, N., Osawa, K., Miyaji, C. (2015). Calculation of probability of winning and number of games played for various tournament formats of the world baseball classic. *International Journal of Computer Science in Sport*. v. 14, n. 1, p. 87-101.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, v. 102, n. 46, p. 16569–16572.
- Korn, A., Schubert, A., Telcs, A. (2009). Lobby index in networks. *Physica A: Statistical Mechanics and its Applications*, v. 388, n. 11, 2221-2226.
- Kossi, Y., Lesueur, J.-Y., Sabatier, M. Academic competition and network externalities in scientific production: An econometric study on french economists. *Revue d'Economie Politique*. v. 123, n. 4, July 2013, p. 469-494.
- Lins, M.P.E., Gomes, E.G., Soares de Mello, J.C.C.B, Soares de Mello, A.J.R. (2003) Olympic ranking based on a zero sum gains DEA model. *European Journal of Operational Research*. v 148, p. 312–322.
- Mattei, N., Golsmith, J., Klapper, A., Mundhenk, M. (2015). On the complexity of bribery and manipulation in tournaments with uncertain information. *Journal of Applied Logic*. v. 13, n. 4, p.557-581.
- Pauly, M. (2014). Can strategizing in round-robin subtournaments be avoided? *Social Choice and Welfare*. v. 43, n. 1, p. 29-46.
- Pereira, D. S. (2015) *Análise da rede aérea brasileira com base na centralidade tipo H*. Dissertação de Mestrado em Engenharia de Produção, Universidade Federal Fluminense - Niterói, RJ. 84 f.
- Pereira, D. S., Brandão, L.C., Del Vecchio, R.R., Soares de Mello, J.C.C.B. (2015). Centralidade aérea baseado no índice tipo h. In: *XXIX Congresso de Pesquisa e Ensino em Transportes - Anpet, 2015*, Ouro Preto, MG. Anais do Congresso de Pesquisa e Ensino em Transportes, 2015, p. 2058-2068.
- Pereira, D.S., Costa, E.F., Wolff, M.G.C., Tavares, R.S. e Soares de Mello, J.C.C.B.(2015). Pós-Análise do índice H e Comparação com Métodos Ordinais Multicritério, *Pesquisa Operacional para o Desenvolvimento*, v. 7, n. 2, p. 173-189.
- Rubem, A.P.S., Moura, A.L., Soares de Mello, J.C.C.B. (2015). *Comparative analysis of some individual bibliometrics indices when applied to groups of researchers*. *Scientometrics*, v. 102, 1019-1035.
- Sidiropoulos, A., Katsaros, D. e Manolopoulos, Y. (2007). Hirsch h-index for disclosing latent facts in citation networks. *Scientometrics*. v. 72, n. 2, p. 253–280.
- Soares de Mello, J.C.C.B, Angulo-Meza, L., Silva, B.P.B. (2009) A ranking for the Olympic Games with unitary input DEA models. *IMA Journal of Management Math*. v. 20, n. 2, p.201-211.

Zhai, L., Yana, X., Zhang, G. (2013). A centrality measure for communication ability in weighted network. *Physica A*, v. 392, 6107–6117.

Zhao, S. X., Rousseau, R., Ye, F. Y., (2011). h-Degree as a basic measure in weighted networks. *Journal of Informetrics*, v. 5, n. 4, 668-677.