

# Análise qualitativa da produção científica individual na UNESP

Carlos Alberto Oliveira de Matos <sup>1</sup>

## 1 Introdução

J. E. Hirsch propôs índice  $h$ , “definido como o número de artigos com número de citação maior ou igual a  $h$ , como um índice útil para caracterizar a produção científica de um pesquisador” (HIRSCH, 2005).

O índice  $h$  é amplamente utilizado na comunidade científica em virtude do fato dele ser um índice representativo da produção intelectual ao longo da vida de um cientista expresso em um único número e de se basear em um cálculo simples usando os bancos de dados comuns na literatura. Bornmann e Daniel (2007) alertam que esses dois fatores acarretam o perigo de uso indevido do índice, descrevem suas vantagens e desvantagens e resumem os estudos sobre a validade do índice.

O fato deste artigo (HIRSCH, 2005) ter sido citado por 8386 publicações até o presente atesta a importância e a adequação do referido índice para a avaliação qualitativa da produção científica de um pesquisador.

O objetivo do presente trabalho é o de identificar aspectos relacionados à análise qualitativa da produção científica individual na UNESP com o intuito de fornecer subsídios para aprimorar o processo de avaliação docente.

## 2 Materiais e métodos

Web scraping é a denominação que se dá às técnicas de extração e coleta de dados de páginas da internet. “Consiste, basicamente, em especificar um site, determinar o trecho do código HTML que contém a informação desejada e, por fim, dispor essa informação num formato adequado para a análise dos dados (SAID, 2018).”

O presente trabalho levantou os seguintes dados de 30275 autores afiliados institucionalmente à UNESP: quantidade de documentos, número de áreas do conhecimento publicadas, total de citações, índice  $h$ , número de coautores e período de publicação (ano inicial e final). Este levantamento reflete a situação encontrada no dia 3 de abril de 2018.

Para o processamento e a análise dos dados foram utilizados os seguintes pacotes do R software versão 3.4.4 : base, purrr, rvest, xml2, httr, curl, stringr e plyr (R Core Team, 2018).

## 3 Resultados e discussão

A Figura 1 apresenta uma adaptação do índice  $h$  objetivando caracterizar a produção científica das instituições de ensino superior públicas do estado de São Paulo conjuntamente com as despesas efetuadas com a remuneração de docentes ativos obtidas no Censo da Educação Superior de 2017. Nas 67 unidades da FATEC mantidas pelo Centro Estadual de Educação Tecnológica Paula Souza o montante referente a esse valor aparece

---

<sup>1</sup>Universidade Estadual Paulista (UNESP), Campus Experimental de Itapeva. e-mail: *carlos.matos@unesp.br*

repetidamente citado como R\$ 872.823.664,70 que foi assumido como sendo o valor global da instituição. (MICRODADOS, 2018)

Há uma relação aproximadamente proporcional entre as despesas efetuadas com a remuneração de docentes ativos e o índice h. Essa tendência não é observada totalmente em virtude do comportamento dessas duas variáveis nos casos da UNESP e da FATEC.

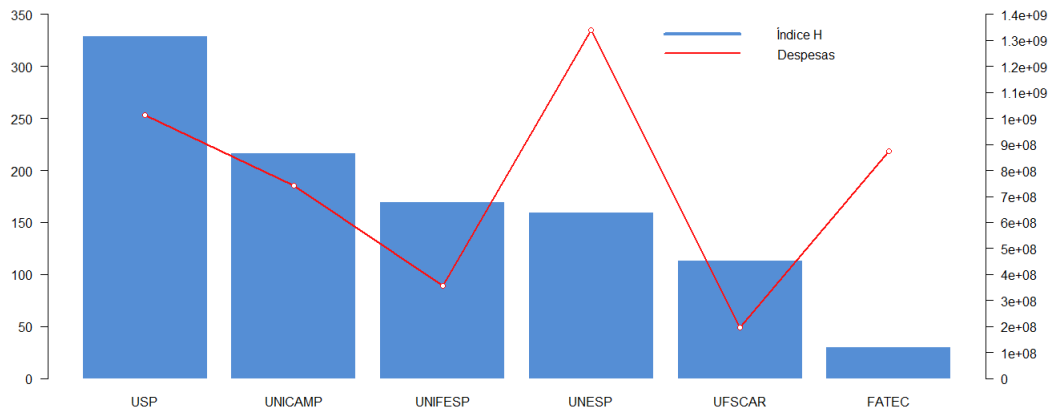


Figura 1: Índice h e despesas com remuneração de docentes ativos nas IES públicas do estado de São Paulo (Fonte: Autor)

O período durante o qual um autor publica estando afiliado institucionalmente à Unesp é muito baixo (Tabela 1). O terceiro quartil para esta variável indica que três quartos dos autores publicaram durante um período que varia entre 0 e 4 anos. O valor máximo (61 anos) se refere à produção intelectual do Prof. Dr. Abraham H. Zímerman.

Tabela 1: Medidas resumo referentes ao período de publicação.

Min.	Q1	Mediana	Média	Q3	Max.	NA
0.00	0.00	0.00	3.34	4.00	61.00	52

Fonte: Autor.

Do total de 30275 autores afiliados institucionalmente à Unesp, 6470 ( $\approx 21\%$ ) não tem sua produção intelectual caracterizada pelo índice h (Tabela 2). O terceiro quartil para esta variável é igual a 2 e pode ser categorizado como sendo muito baixo. O maior índice h (81) correspondia à produção intelectual do Prof. Dr. Sérgio F. Novaes.

Tabela 2: Medidas resumo referentes ao índice h dos autores afiliados à Unesp.

Min.	Q1	Mediana	Média	Q3	Max.	NA
1.00	1.00	1.00	2.49	2.00	81.00	6470

Fonte: Autor.

Existe uma relação linear significativa entre índice h e o período de tempo de afiliação à Unesp ( $F_{1,23758} \text{ g.l.} = 2,024 * 10^4$ ;  $\text{valor}P < 0,05$ ). O modelo  $\text{ind}H = 1.061642 + 0.3448558 * \text{período}$  apresenta um baixo coeficiente de determinação igual a 46% e indica que em média o índice h dos autores afiliados a Unesp sobe uma unidade a cada 3 anos (Figura 2). A análise dos resíduos da regressão constata a presença de heterocedasticidade e a necessidade de um modelo mais adequado.

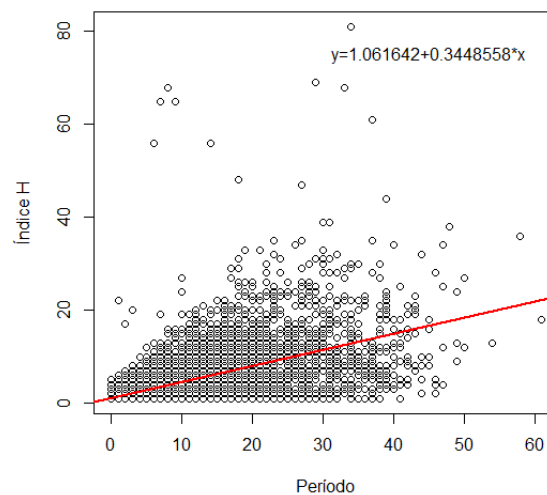


Figura 2: Relação entre o índice h e o período de publicação. (Fonte: Autor)

O detalhamento da produção intelectual por período é apresentado na Tabela 3. Desde a década de 70, que é quando o número de trabalhos científicos começa a ter importância quantitativa, a mediana do índice h atingiu o valor máximo igual a 3.

Tabela 3: Evolução temporal do índice h.

Década	n	Min.	Q1	Mediana	Média	Q3	Max.	NA
[ ,50)	1	NA	NA	NA	NA	NA	NA	1
[50,60)	10	1	4	5	9,78	8	36	1
[60,70)	57	1	2	2	4,78	5	27	3
[70,80)	368	1	1	3	6,29	9	44	95
[80,90)	769	1	1	3	6,76	9	81	154
[90,00)	2454	1	1	2	4,98	7	48	234
[00,10)	9011	1	1	1	2,74	3	68	581
[10,20)	17533	1	1	1	1,54	2	65	5394

Fonte: Autor.

O grau de multi e interdisciplinaridade da produção científica da Unesp é apresentado na Tabela 4. Três quartos dos autores publicam em até três das vinte e sete áreas do conhecimento categorizadas na base Scopus. Esse baixo índice reflete o privilégio para abordagens especializadas na produção científica nacional que transparece na Chamada CNPq

Nº 09/2018 Bolsas de Produtividade em Pesquisa cujo objetivo é “valorizar pesquisadores que possuam produção científica, tecnológica e de inovação de destaque em suas respectivas áreas do conhecimento e incentivar o aumento da produção científica, tecnológica e de inovação de qualidade”. Neste documento na maior parte das 34 vezes em que se menciona termos relacionados com a multi e interdisciplinaridade eles são acompanhados, ainda que sutilmente, por julgamentos de valor de natureza restritiva.(CHAMADA, 2018)

Tabela 4: Medidas resumo da quantidade de áreas do conhecimento nas quais os autores publicam.

Min.	Q1	Mediana	Média	Q3	Max.	NA
0.00	1.00	2.00	2.46	3.00	27.00	

Fonte: Autor.

A Figura 3 apresenta a relação crescente entre o índice h e a quantidade de áreas do conhecimento nas quais os autores publicam. Observa-se que a estratégia de publicação em poucas áreas adotada pela imensa maioria dos autores afiliados à Unesp é visivelmente ineficaz no que tange à qualidade da produção intelectual. Há 130 autores sobre os quais não existem referências quanto à(s) sua(s) área(s) de publicação e 48 autores com total ausência de informação.

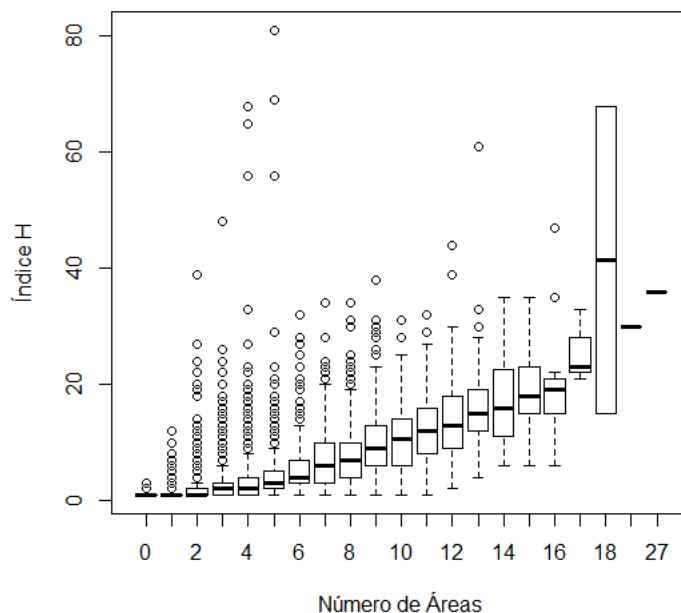


Figura 3: Comportamento do índice h em função da quantidade de áreas do conhecimento nas quais os autores publicam. (Fonte: Autor)

A Figura 4 apresenta as interações duplas entre áreas do conhecimento. Elas são maiores em cinco áreas: a Medicina; a Bioquímica, Genética e Biologia Molecular; as

Ciências Agrárias e Biológicas; a Engenharia e a Química. Elas são preocupantemente ínfimas em áreas que incluem a Economia, Econometria e Finanças. Considerando-se que a aplicação prática e inovadora de conhecimento novo gera produção e renda, a utilização de ferramentas desta área seria importante no sentido de justificar a importância econômica e social do investimento em pesquisa científica. As cinco melhores e as cinco piores interações duplas entre áreas do conhecimento são detalhadas na Tabela 5.

	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
01																										
02	94																									
03	4071	97																								
04	108	10																								
05	737	19	1047	38																						
06	1119	20	1513	42	1224																					
07	300	33	378	109	314	379																				
08	129	8	89	160	32	39	126																			
09	241	13	611	8	159	191	59	5																		
10	530	17	256	45	135	218	214	41	7																	
11	26	8	10	35	5	10	26	25	2	12																
12	207	10	214	96	313	312	269	34	17	143	15															
13	775	29	937	206	963	1077	981	121	337	358	32	576														
14	2157	42	1386	133	603	748	277	114	152	659	32	365	729													
15	284	21	553	13	97	116	67	15	176	14	4	17	154	137												
16	1965	41	1785	23	489	540	141	42	207	86	6	101	407	616	196											
17	827	27	909	55	891	1249	355	34	320	205	11	277	1587	584	121	303										
18	194	21	216	51	159	226	573	127	24	186	21	166	586	168	33	94	242									
19	3381	108	4417	60	660	952	346	98	1386	203	14	148	852	1189	1028	2292	707	207								
20	554	19	492	21	183	242	102	28	44	134	6	82	220	347	52	248	179	118	419							
21	459	46	733	11	101	140	64	29	69	29	3	13	126	170	183	342	93	44	905	94						
22	398	29	474	6	91	135	42	12	51	21	3	12	83	114	190	225	51	22	709	63	152					
23	1392	33	1723	15	516	907	160	32	183	75	5	72	381	767	222	1012	414	60	1877	226	458	244				
24	598	23	947	50	755	1308	500	52	294	313	14	302	1564	476	335	322	1587	535	861	219	123	59	394			
25	112	31	179	15	13	27	20	7	24	16	3	7	21	49	84	65	9	7	285	27	216	73	97	9		
26	356	255	352	102	122	163	298	91	65	172	53	90	339	303	216	163	129	197	632	80	124	150	133	231	151	
28	3273	46	1251	39	179	218	84	38	74	113	6	36	200	534	117	1162	242	49	1390	166	170	108	467	164	28	103

Figura 4: Interações duplas entre áreas do conhecimento. (Fonte: Autor)<sup>1</sup>

<sup>1</sup>1-Ciências Agrárias e Biológicas; 2-Artes e Humanidades; 3-Bioquímica, Genética e Biologia Molecular; 4-Negócios, Gestão e Contabilidade; 5-Engenharia Química; 6-Química; 7-Ciência da Computação; 8-Ciências da Decisão; 9-Odontologia; 10-Ciências da Terra e dos Planetas; 11-Economia, Econometria e Finanças; 12-Energia; 13-Engenharia; 14-Ciência Ambiental; 15-Profissões da Saúde; 16-Imunologia e Microbiologia; 17-Ciência dos Materiais; 18-Matemática; 19-Medicina; 20-Multidisciplinar; 21-Neurociência; 22-Enfermagem; 23-Farmacologia, Toxicologia e Farmacêutica; 24-Física e Astronomia; 25-Psicologia; 26-Ciências Sociais 27-Indefinido; 28-Veterinária

Tabela 5: Cinco interações mais fortes e fracas entre áreas do conhecimento.

03 X 19	01 X 03	01 X 19	01 X 28	16 X 19
4417	4071	3381	3273	2292
11 X 15	11 X 21	11 X 22	11 X 25	09 X 11
4	3	3	3	2

Fonte: Autor.

## 4 Conclusão

Este trabalho assume como pressuposto que o índice  $h$ , apesar das suas limitações, é válido e adequado como instrumento para a avaliação qualitativa da produção científica de um pesquisador. A natureza predominantemente diagnóstica deste trabalho demonstra uma relação desproporcional entre o índice  $h$  da instituição e suas despesas com a remuneração dos docentes ativos que também está presente no caso das Fatecs. O período durante o qual grande parte dos autores publica estando afiliado à Unesp varia entre 0 e 4 anos. Dos aproximadamente 80% dos autores com índice  $h$ , três quartos deles apresentam índice  $h$  entre 0 e 2. A relação linear entre o índice  $h$  e o período de publicação indica que o esse índice aumenta uma unidade a cada 3 anos. Não há diferença de comportamento significativa na análise das últimas cinco décadas. A maioria dos autores publicam trabalhos em até três áreas do conhecimento. A análise do da influência da multi e interdisciplinaridade na evolução do índice  $h$  mostra que esta é uma estratégia que carece de efetividade no que tange ao aumento da qualidade da produção intelectual caracterizada pelo índice  $h$ .

## Referências bibliográficas

BORNMAN, L.; DANIEL, H.-D. What do we know about the  $h$  index? *Journal of the American Society for Information Science and technology*, Wiley Online Library, v. 58, n. 9, p. 1381–1385, 2007.

CHAMADA CNPq N° 09/2018 Bolsas de Produtividade em Pesquisa. 2018. <<http://resultado.cnpq.br/3107104051551290>>. Acesso em: 12 mar 2019.

HIRSCH, J. E. An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences*, National Acad Sciences, v. 102, n. 46, p. 16569–16572, 2005.

MICRODADOS Censo da Educação Superior 2017. 2018. <[http://download.inep.gov.br/microdados/microdados\\_educacao\\_superior\\_2017.zip](http://download.inep.gov.br/microdados/microdados_educacao_superior_2017.zip)>. Acesso em: 2 nov 2018.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2018. Disponível em: <<https://www.R-project.org/>>.

SAID, S. *Web scraping: ou como raspar todas receitas de bolo de cenoura*. São Paulo: IBPAD Instituto Brasileiro de Pesquisa e Análise de Dados, 2018. Disponível em: <<https://www.ibpad.com.br/blog/web scraping-ou-como-raspar-todas-receitas-de-bolode-cenoura/>>. Acesso em: 9 abr 2018.

## Agradecimentos

À Fapesp - Fundação de Amparo à Pesquisa do Estado de São Paulo pelo Auxílio à Pesquisa - Participação em Reunião Científica e/ou Tecnológica.

À Prof<sup>a</sup>. Dr<sup>a</sup>. Luzia Aparecida Trinca responsável pela coordenação do Projeto Fapesp 2019/XXXXXX-X.