

UMA MEDIDA DE DIFUSÃO DO CONHECIMENTO ATRAVÉS DAS CITAÇÕES

Eixo temático: Análise de Citação

Modalidade: Apresentação oral

1 INTRODUÇÃO

A distribuição das informações estatísticas oficiais produzidas pelo Instituto Brasileiro de Geografia e Estatística (IBGE) através de artigos que citam os documentos produzidos pelo Instituto e são difundidos em dezenas de periódicos científicos, em diversos idiomas e publicados em vários continentes favorece a disseminação das informações sobre a realidade brasileira nelas representadas. Este aspecto da difusão do conhecimento, entre outros, foi abordado por Zanotto (2011) e a distribuição geográfica das informações estatísticas foi apresentada por Zanotto, Vanz e Stumpf (2011). No presente texto pretende-se aprofundar a discussão sobre a difusão das informações publicadas pelo IBGE através das citações recebidas. Considerando-se que as citações representam uma medida de difusão das ideias científicas, buscou-se fundamentação na teoria dos fatores de difusão para apresentar uma aplicação prática.

2 METODOLOGIA

As citações à produção científica do IBGE foram identificadas na Web of Science em duas coletas que aconteceram em 2010 e 2011. Os dados coletados abrangem o período de 2001 a 2010, e totalizam 3.158 documentos citantes. Destes documentos são analisadas as variáveis de autores, instituições e países citantes ano a ano. Os autores citantes totalizam 10.716 nomes diferentes, perfazendo uma média de 1,24 citações por autor e frequência que variou de 1 a 19 citações às publicações do IBGE por autor. Estes autores, no total, pertencem a 1.757 instituições de filiações diferentes e das 7.587 ocorrências de países presentes no vínculo institucional dos autores, 6.168 (81%) destas se referem ao Brasil e as demais estão distribuídas entre 49 países diferentes. Neste contexto, mesmo sabendo que as publicações do

IBGE receberam um grande número de citações no período, advindas de diferentes autores, instituições e países, ainda permanece a dúvida se estes dados refletem uma difusão do conhecimento satisfatória. Com o objetivo de equacionar esta questão buscou-se na literatura os fatores de difusão protagonizados por Rousseau, Liu e Ye (2012). Nas seções que seguem é apresentado o fator de difusão e em seguida, a aplicação do mesmo às citações recebidas pelo IBGE.

3 FATORES DE DIFUSÃO

Os Fatores de Difusão foram estudados por Rowlands em 2002, quando o autor definiu a transdisciplinariedade de periódicos através de um número médio padronizado de citações num determinado tempo de citação. Para explicar o fenômeno, o autor usou a comparação metafórica do ato de jogar pedras em uma lagoa com as citações recebidas ao longo do tempo, desta forma chegou a uma medida da extensão das ondulações resultantes entre uma citação e outra, de uma determinada publicação num meio como a literatura científica (ROWLANDS, 2002¹ apud FRANDSEN; ROUSSEAU; ROWLANDS, 2006). A partir daí os estudos de fatores de difusão buscam capturar diferentes aspectos do processo de citação, avançando além das tradicionais medidas bibliométricas para os periódicos ou o conjunto de artigos, como o Fator de Impacto (IF) por exemplo (FRANDSEN; ROUSSEAU; ROWLANDS, 2006).

Em recente artigo, Rousseau, Liu e Ye (2012) consideram que apesar de as ideias científicas serem transmitidas também através de conversas formais e informais, e-mail, blogs e outros meios eletrônicos de comunicação, estas ideias normalmente convergem para um artigo científico e que o conhecimento contido em um artigo é difundido através das citações. Os autores também levam em conta que se um artigo recebe citações oriundas de vários países diferentes, o conteúdo deste artigo é difundido mais do que quando o artigo recebe a mesma quantidade total de citações proveniente de um mesmo país. Os estudos de difusão são decorrentes dessa visão, e introduzem um sistema de contagem fracionada em camadas, onde

¹ ROWLANDS, I. Journal diffusion factors: a new approach to measuring research influence. **Aslib Proceedings**, v. 54, n. 2, p. 77-84, 2002.

o conjunto de citações recebidas é analisado em relação aos indicadores autor, instituição e país citante, e todos eles são determinantes para se calcular a difusão.

A ferramenta estatística utilizada para equacionar o problema aplicado é o coeficiente de Gini, como apresentado na **Figura 1**:

Figura 1 – Coeficiente de Gini conforme Rousseau, Liu e Ye (2012)

$$G_e(X) = \frac{2 \sum_{j=1}^N jX_j}{\mu N^2} - \frac{1}{N}$$

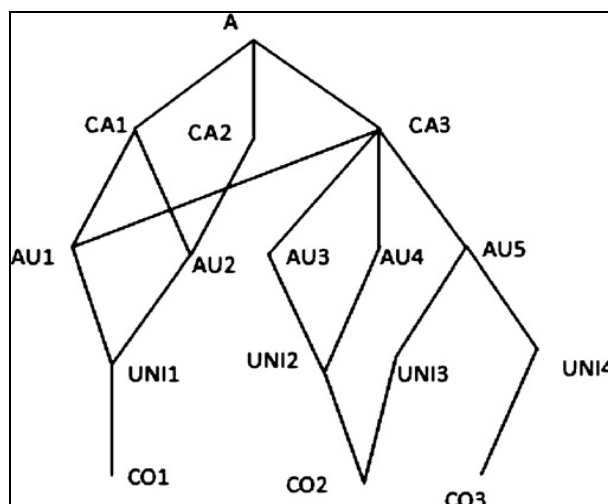
Fonte: Rousseau, Liu e Ye (2012)

Nota: $G_e(X)$ = coeficiente de Gini para o período;
 $X = (x_j)$ e $j = 1, \dots, n$ (matriz de números não negativos)
 jX = Variação das posições do período;
 N = Número de camadas analisadas;
 μ = Média do conjunto $\{x_j\}$.

O Coeficiente de Gini é uma medida de desigualdade desenvolvida pelo estatístico italiano Corrado Gini, em 1912. É comumente utilizado para calcular a desigualdade de distribuição de renda, entretanto, pode ser usado para qualquer distribuição. Ele consiste em um número entre 0 e 1, onde 0 corresponde à completa igualdade de renda (onde todos têm a mesma renda) e 1 corresponde à completa desigualdade (onde uma pessoa tem toda a renda, e as demais nada têm). O índice de Gini é o coeficiente expresso em pontos percentuais (é igual ao coeficiente multiplicado por 100) (WIKIPÉDIA, 2012).

O esquema a seguir apresenta o método de organização dos indicadores para análise do fator de difusão: citações, autores citantes, instituição de filiação e país de filiação do autor citante.

Figura 2 – Esquema do sistema em camadas



Fonte: Rousseau, Liu e Ye (2012)

As camadas analisadas correspondem, conforme **Figura 2** a: **Camada 1 (CA)**: ao conjunto de todas as citações (recebidas pelo artigo A) numa janela de citação dada; **Camada 2 (AU)**: ao conjunto de todos os autores citantes da amostra; **Camada 3 (UNI)**: ao conjunto de todas as universidades ou institutos a que os autores citantes pertencem; e **Camada 4 (CO)**: ao conjunto de todos os países aos quais pertencem essas instituições.

Tabela 01 - Exemplo da contagem fracionada das camadas de Autor, de Instituição e de País

Autores (AU)	Artigo 1 (CA1)	Artigo 2 (CA2)	Artigo 3 (CA3)	Soma
Autor 1	0,5		0,25	0,75
Autor 2	0,5	1		1
Autor 3			0,25	0,25
Autor 4			0,25	0,25
Autor 5			0,25	0,25
Total Autor	1	1	1	2,50
Universidades (UNI)	Artigo 1 (CA1)	Artigo 2 (CA2)	Artigo 3 (CA3)	Soma
Universidade 1	1	1	0,25	1
Universidade 2			0,5	0,5
Universidade 3			0,125	0,125
Universidade 4			0,125	0,125
Total Universidade	1	1	1	1,75
Países (CO)	Artigo 1 (CA1)	Artigo 2 (CA2)	Artigo 3 (CA3)	Soma
País 1	1	1	0,25	1
País 2			0,625	0,625
País 3			0,125	0,125
Total País	1	1	1	1,75

Fonte: Rousseau, Liu e Ye (2012)

Nota: Cada unidade (Artigo, Universidade, País) tem peso máximo = 1.

No exemplo dado pelos autores onde um Artigo A é citados por 3 outros artigos (CA), a soma da fração de autoria (AU) é 2,5; levando-se em conta que a contagem de cada artigo

citante contribui no total de 1 unidade de autor e se houver mais de um autor, as contribuições de cada um são fracionadas em partes iguais ($1/n$). A soma da fração das instituições (UNI) e países (CO) é respectivamente 1,75, ambas são resultados derivados das contribuições de autores. Então, cada artigo citante contribui com um total de exatamente 1 unidade (universidade/instituto ou país) e se existirem diferentes autores em digamos n universidades, cada autor recebe uma contribuição de ($1/n$) e se no entanto, um autor tem k endereços institucionais, cada instituto contribui através deste autor, uma pontuação de ($1/nk$). Finalizando a análise, a média da soma destas frações é $\mu = 2,25$, temos $N = 4$; $X = (3; 2,5; 1,75; 1,75)$ e, aplicada a fórmula chega-se ao índice de Gini ponderado como $G_e = 0,875$ (ROUSSEAU, LIU, YE, 2012).

3 APLICAÇÃO DO COEFICIENTE DE GINI (G_e) NOS DADOS DO IBGE

A partir do conjunto das citações recebidas pelas publicações do IBGE no período de 2001-2010, conforme descrito anteriormente, e a aplicação da análise dos fatores de difusão propostos por Rousseau, Liu e Ye (2012), obteve-se os dados da **Tabela 2**:

Tabela 02 – Produção científica do IBGE - Artigos Citantes, Autores, Instituições e Países – 2001 a 2010

Ano	Artigos Citantes (CA)	Número de Autores Diferentes	Soma Frações Autores Citantes (AU)	Número Instituições Diferentes	Soma Frações Instituições Citantes (UNI)	Número Países Diferentes	Soma Frações Países Citantes (CO)	G_e
2001	106	346	105,49	214	70,77	11	8,50	0,89
2002	110	381	109,24	250	70,93	13	8,42	0,89
2003	123	477	121,76	296	76,11	20	13,57	0,90
2004	177	573	176,00	390	97,33	17	10,83	0,89
2005	184	624	182,25	392	91,27	14	10,33	0,89
2006	222	845	221,22	529	122,51	26	18,15	0,89
2007	447	1738	444,96	1076	185,47	25	19,08	0,89
2008	654	2622	647,92	1615	243,75	26	21,42	0,89
2009	673	2775	671,42	1658	264,99	24	19,90	0,88
2010	462	1993	460,72	1161	254,26	30	22,33	0,89

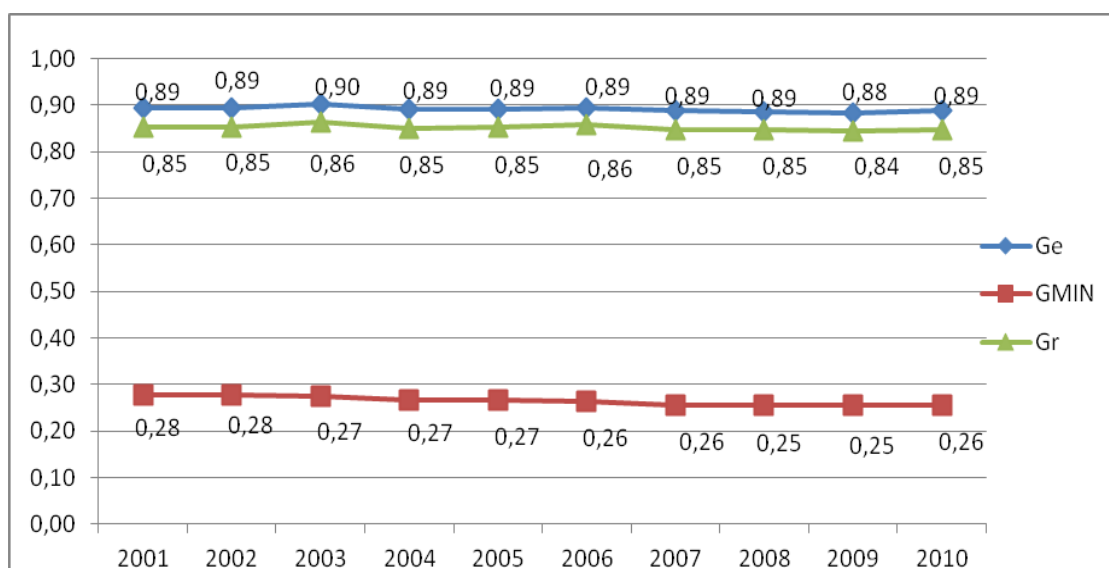
Fonte: Dados da pesquisa

Após aplicação da fórmula G_e obteve-se resultados bastante satisfatórios uma vez que o índice de Gini manteve-se entre 0,88 em 2009 e 0,90 em 2003, indicando um fator de

dispersão alto, ou seja, a existência de um grande número de autores, instituições e países diferentes que citam as publicações do IBGE. Outra consideração importante é a de que embora o número absoluto de citações seja variável e crescente no período, o índice G_e mantém-se no mesmo patamar durante a década estudada.

Quando é aplicado o índice de Gini relativo, denotado como $G_r = (G_e - MIN) / (1 - MIN)$, onde o valor mínimo é calculado por $MIN = (C + N^2 - 1) / ((C + N - 1) N)$ e temos N um valor fixo (igual ao número de camadas, ou seja $N=4$) e C que corresponde ao número de variações das unidades estudadas (citações recebidas) os resultados são apresentados no **Gráfico 1**.

Gráfico 1 : G_e , G_r e G_{MIN} no período de 2001-2010



Fonte: Dados da pesquisa

Nota: $G_{MIN} = (C + N^2 - 1) / ((C + N - 1) N)$

Os valores acima representados mostram alta equitabilidade, ou seja uma difusão bastante elevada em relação a todos os níveis considerados, observa-se que não existem grandes variações entre o indicador que mede o coeficiente de Gini (G_e) e o índice Gini relativo (G_r), contudo o índice mínimo ficou bem abaixo em todos os períodos. Para os padrões de difusão estudados, segundo Rousseau, Liu e Ye (2012) este resultado é bastante satisfatório em termos de uma boa difusão de conhecimento.

4 CONSIDERAÇÕES FINAIS

Partindo da ideia de que as citações representam a difusão do conhecimento e que o coeficiente de Gini padronizado (Ge) reflete a medida desta difusão demonstramos que é possível calcular e acompanhar a propagação de um conhecimento ao longo de um período.

A relação de que quanto maior a quantidade de autores, instituições e países diferentes melhor será o grau de difusão é bem clara quando analisamos os dados dos números absolutos de citações do IBGE. A aplicação do cálculo do fator de difusão comprovou a ideia aparente nos números absolutos: índice de Gini entre 0,88 em 2009 e 0,90 em 2003 indica um fator de dispersão alto, ou seja, a existência de um grande número de autores, instituições e países diferentes que citam as publicações do IBGE.

Da mesma forma que Rousseau, Liu e Ye (2012) consideram o estudo que definiu o coeficiente de Gini padronizado (Ge) preliminar, o presente estudo tem características de experimento compartilhado, sujeito a críticas e observações.

REFERÊNCIAS

- COEFICIENTE DE GINI. In: **WIKIPÉDIA**, a enciclopédia livre. Florida: Wikimedia Foundation, 2012. Disponível em:
<http://pt.wikipedia.org/w/index.php?title=Coeficiente_de_Gini&oldid=29857729>. Acesso em: 20 maio 2012.
- FRANDSEN, T.F.; ROUSSEAU, R.; ROWLANDS, I. Diffusion factors. **Journal of Documentation**, v. 62, n. 1, p. 58-72, 2006.
- ROUSSEAU, R.; LIU, Y.; YE, F.Y. A preliminary investigation on diffusion through a layered system. **Journal of Informetrics**, Amsterdam, v. 6, n. 2, p. 177-191, 2012.
- ZANOTTO, S.R.; VANZ, S.A.S.; STUMPF, I.R.C. A informação estatística oficial produzida pelo IBGE e a sua difusão geográfica. In: XII Encontro Nacional de Ciência da Informação e Biblioteconomia, 2011, Brasília. XII ENANCIB. Brasília, 2011.
- ZANOTTO, S.R. **Informação estatística oficial produzida pelo IBGE: apropriação pela comunidade científica brasileira no período de 2001 a 2009**. 2011. Dissertação (Mestrado em Comunicação e Informação) - Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Comunicação e Informação, Porto Alegre, 2011. 165 f.