

Modelagem de um indicador bibliométrico para análise da dispersão de conhecimentos

Guido Rummler

Professor titular. Universidade Estadual de Feira de Santana (UEFS).
Disciplinas: Metodologia da Pesquisa Científica e Estatística.
Pós-doutorado pela USP.
E-mail: rummler@ufba.br

Resumo

Quando os suportes de uma informação encontram-se distribuídos por diferentes partes, diz-se que a informação sofreu dispersão. Se ela se encontra citada na literatura, isso configura impacto bibliográfico. Assim também, a dispersão geográfica das fontes citantes representa a extensão geográfica do impacto daquela informação. Dessa forma, um indicador de dispersão pode constituir ferramenta útil para avaliação desses fenômenos. O presente artigo demonstra a modelagem de um indicador relativo de dispersão (IDS – Índice de Dispersão Segmentar), que considera a ocorrência desta, segundo diferentes níveis de distribuição (macro, meso e/ou microdispersão) referentes a unidades de análise bibliométricas, cienciométricas, informétricas, tecnométricas ou similares. Exemplos que ilustram possibilidades de aplicação do IDS no estudo comparativo de obras, periódicos, autores, métodos, técnicas, assuntos, vocábulos etc. são também apresentados.

Palavras-chave

Informação. Comunicação científica. Dispersão. Difusão. Indicadores. Bibliometria. Cienciométrica. Informetria. Índice de dispersão. Impacto bibliográfico.

A bibliometric indicator modelling for analysis of knowledge dispersion

Abstract

When information channels are distributed by different parts, it means that the information is dispersed. If this information is cited in the literature it means bibliographic impact. Therefore, the geographical dispersion of the sources that cites the information represents an extension of geographical impact. Thus, a dispersion indicator could represent a practical tool for the assessment of these events. The present article shows the modeling of a relative dispersion indicator (Index of Segmental Dispersion – ISD) that considered its occurrence related to different distribution levels (macro, meso and/or microdispersion) in terms of bibliometric, scientometric, informetric, technometric or similar units of analysis. Examples that illustrates possible applications of ISD in comparative studies of literary works, journals, authors, methods, technique, themes or words and so forth are also introduced.

Keywords

Information. Scientific communication. Dispersion. Diffusion. Indicators. Bibliometrics. Scientometrics. Informetrics. Dispersion index. Bibliographic impact.

1. INTRODUÇÃO

Sendo a difusão, definida por Rogers (1955), citado por Kortelainen (2001, p.133), “um processo através do qual uma inovação é comunicada, pelo tempo, através de certos canais entre os membros de um sistema social”, pode-se dizer que a dispersão do suporte que veicula a inovação constitui fator intrínseco desse processo. A eventual possibilidade de dimensionar o perfil de distribuição do suporte da informação (periódico, fascículo, artigo, por exemplo), ou o de sua citação, permitiria promover análises da dispersão físico-geográfica ou literária de obras, autores, questões, técnicas ou mesmo vocábulos, assim como comparar essa dispersão em ou entre áreas específicas de conhecimento. Em contrapartida, é interessante considerar que a dispersão de citações configura impacto bibliográfico do que é citado. Sua quantificação, portanto, configura uma medida desse impacto. Outrossim, a dispersão é um fenômeno explícito ou subjacente a diversas questões relacionadas com o “comportamento” de unidades de análise da bibliometria, cienciométrica (ou cientometria), informetria, webometria ou abordagens similares. A Lei de Bradford, ou Lei da Dispersão, assim mencionada por Quoniam e colaboradores (2001, p.24) e Vanti (2002, p.153), ou ainda referida por Matos (2002) como lei da concentração/dispersão da literatura científica, estabelece, conforme explicam Borges (2002), Glänzel (2005) e Ravichandra Rao (1998), a partir do ordenamento decrescente de uma população de periódicos quanto ao número de artigos referentes a determinado assunto, um núcleo de periódicos mais particularmente devotados a determinado assunto e alguns grupos ou zonas contendo o mesmo número de artigos que o núcleo, de acordo com o modelo matemático “ $1:n:n^2...$ ”. Essa expressão estabelece a relação existente entre o número de periódicos que constituem os sucessivos grupos daquela distribuição, representando uma proporcionalidade entre categorias de periódicos, e não uma medida de dispersão, a qual seria atributo de um indicador.

A CFHSS (2005) menciona a utilização por Godin de um “índice de dispersão” com a finalidade de medir o grau de difusão ou visibilidade de pesquisadores fora de seu próprio país. Esse indicador de dispersão compreende relações entre o número de comunicações feitas em congressos referentes a um país e o número de países em que os pesquisadores as apresentam. Assim, utilizando-se de dados referentes a diversos países, Godin (1998) estabelece uma listagem dos respectivos valores obtidos por esse índice de dispersão, a partir da qual faz ilações a respeito do que subtitula “dispersão do conhecimento”.

Entretanto, independentemente das denominações pelas quais sejam conhecidos os indicadores bibliométricos ou cientométricos (ou cientométricos), podemos encontrar, por exemplo, dentre os 46 classificados por Vinkler (1988) e outros agregados por Spinak (1998), vários indicadores em cujas definições permeia o fenômeno da dispersão. Entre os mais conhecidos, está o Fator de Impacto, de Garfield, largamente usado pelo Institut for Scientific Information (ISI), por ele fundado, conforme Vanti (2002), na Filadélfia. Trata-se de um indicador que, de acordo com o que consta em Campos (2003), Glänzel e Moed (2002) e Vinkler (2001), relaciona o número de referências a um periódico “x”, editado nos anos “A” e “A+1”, ocorridas em “n” artigos publicados em um elenco de “NP” periódicos, editados no ano “A+2”, com o número total de artigos publicados naquele periódico “x” dos anos “A” e “A+1”. A subjacência aí do fenômeno da dispersão pode ser mostrada, se considerarmos que os “n” artigos em que se encontram as citações a “x” poderiam estar publicados em NP/2 ou NP/3 ou NP/20 periódicos, portanto, menos ou mais “espalhados”, aspecto que, se considerado, implicaria uma suplementaridade de sentido àquele fator.

Tendo-se, em trabalhos anteriores, verificado a utilidade de expressar diferenças entre periódicos (ou outras unidades de análise) quanto à sua dispersão geográfica ou literária, na perspectiva de que esta, como já citado, constitui fator intrínseco da difusão, foi desenvolvido um algoritmo* para representação da dispersão em termos de relação a uma situação-padrão. Em continuidade àquele trabalho, desta feita considerando que o fenômeno da dispersão perpassa por níveis hierarquizados de distribuição, tais como citação em artigo – artigo em fascículo – fascículo em periódico, ou

obra em biblioteca – biblioteca em município – município em estado – estado em país, foi modelado, como tentativa de dar lastro a essa questão, o que aqui se denomina Índice de Dispersão Segmentar, ou IDS, assim como são ilustradas possíveis aplicações do mesmo para fins de avaliações bibliométricas, cientométricas ou similares.

2. PROCEDIMENTOS

O presente trabalho constitui-se de duas etapas. A primeira percorre fases caracterizadas por Trzesniak (1998): a) busca, nesse processo, de dimensões ou aspectos com características específicas que possam, direta ou indiretamente, conter a resposta desejada; b) padronização da metodologia de obtenção do indicador, representada pelo desenvolvimento de um algoritmo; c) interpretação e avaliação das características desse indicador.

A obtenção do indicador de dispersão (IDS – Índice de Dispersão Segmentar) baseia-se em representação bidimensional referente a relações entre pares de variáveis, correspondentes a níveis de distribuição (exemplo: citação em artigo; artigo em fascículo; fascículo em periódico). São consideradas, para efeito ilustrativo, sete variáveis (ou indicadores simples) genericamente representadas por “P”, “p”, “F”, “f”, “A”, “a”, “r”.

A cada nível de distribuição corresponde uma situação-padrão, de modo que se produza um índice que esteja relativizado à abrangência de cada avaliação. Da relação entre valores observados para cada variável e os correspondentes às situações-padrão, resultam coeficientes respectivos a cada nível. O somatório de tantos coeficientes quantos forem os níveis abrangidos pela análise corresponde ao IDS. E, a partir da atribuição de significados particulares às variáveis que compõem o índice, são ilustradas, na segunda etapa deste estudo, possíveis aplicações desse indicador.

3. MODELAGEM DO INDICADOR DE DISPERSÃO SEGMENTAR

3.1 Identificação de parâmetros de dispersão

Preliminarmente são indicadas, nas tabelas 1, 2 e 3, a seguir, algumas possíveis representações para as variáveis quantitativas (ou indicadores bibliométricos simples) que se relacionam com a dispersão de unidades de análise de diferentes naturezas, conforme diferentes contextos.

* Artigo “Proposition and testing of a dispersion indicator of printed periodicals” apresentado no 9th World Congress on Health Information and Libraries / 7th Latin American and Caribbean Congress on Health Sciences Information, em set./2005, Salvador-BA, Brasil.

TABELA 1

Identificação de variáveis relacionadas com a dispersão de uma obra, um autor ou um periódico, em uma área de conhecimento (ou campo, especialidade, disciplina etc.)

Variável	Representação
P	Número <i>total de periódicos</i> abrangidos por um estudo, os quais veiculam artigos referentes à determinada área.
p	Número de <i>periódicos positivos</i> , isto é, periódicos que veiculam artigos em cuja lista de referências consta a unidade de análise (autor, obra ou periódico)
F	Quantidade de <i>fascículos</i> que compõem o conjunto de periódicos “P”, publicados no período correspondente ao recorte temporal da avaliação.
f	Quantidade de <i>fascículos positivos</i> , isto é, fascículos em que constam artigos que referenciam a unidade de análise.
A	Número total de <i>artigos</i> correspondentes à área do estudo.
a	Número total de <i>artigos positivos</i> (artigos que referenciam a unidade de análise).
r	Número de <i>referências positivas</i> (número total de referências bibliográficas correspondentes à unidade de análise, existente no conjunto de <i>artigos positivos</i>)

TABELA 2

Identificação genérica de variáveis relacionadas com a dispersão de determinada questão ou assunto, tema, método, técnica, fenômeno, especialidade, vocábulo etc., no âmbito de periódicos referentes a uma determinada área de conhecimento

Variável	Representação
P, F e A	Mesma da tabela 1.
p	Número de <i>periódicos positivos</i> (periódicos que veiculam artigos que contemplam a unidade de análise).
f	Quantidade de <i>fascículos positivos</i> (fascículos em que constam artigos voltados à questão em estudo).
a	Número total de <i>artigos positivos</i> (artigos que contemplam a questão em estudo).
r	Número de <i>referências positivas</i> (número total de referências das obras que contemplam a questão em estudo)

TABELA 3

Identificação de variáveis relacionadas com a dispersão geográfica de uma unidade de análise bibliográfica

Variável	Representação
P	Número de regiões que constituem o território de referência da dispersão.
p	Número de <i>regiões positivas</i> , isto é, regiões em que se encontram os detentores (instituições ou pessoas etc.) da unidade de análise.
F	Número de unidades federativas que constituem o território de referência da dispersão.
f	Número de <i>unidades federativas positivas</i> , isto é, número daquelas em que se localizam os detentores da unidade de análise.
A	Número de localidades ou instituições abrangidas pelo estudo.
a	Número de <i>localidades ou instituições positivas</i> , isto é, número daquelas em que se situam detentores da unidade de análise.
r	Número de detentores (bibliotecas ou pessoas etc.) da unidade de análise.

Como exemplo hipotético de uma dentre as possíveis situações pertinentes às representações da tabela 1, tem-se o seguinte problema: Qual, dentre as duas revistas, *Jornal de Saúde Pública* (JSP) e *Revista de Saúde da Comunidade* (RSC), editadas nos anos 2000 e 2001, apresenta maior extensão de impacto no âmbito dos periódicos nacionais referentes à saúde pública, editados em 2002? Estabelecer ainda a evolução apresentada por estas duas revistas nas situações análogas dos anos seguintes.

Nesse caso, “ r_1 ” e “ r_2 ” representam o número de referências recebidas pelas respectivas revistas JSP e RSC dos anos 2000 e 2001; “ a_1 ” e “ a_2 ”, a respectiva quantidade de artigos em que se encontram essas referências. “A” corresponde ao número total de artigos publicados em 2002 nos “P” periódicos, e “ f_1 ” e “ f_2 ”, o respectivo número de fascículos em que se encontram os artigos que referenciam JSP e RSC. “F” corresponde ao número total de fascículos que constituem a edição de 2002 dos “P” periódicos amostrais. Para resolução do último item desse problema, serão comparadas ainda as duplas de índices de dispersão relativas a cada ano do período de análise da evolução.

Como exemplo hipotético de uma possível situação referente à tabela 2, tem-se: *Como o assunto nanotecnologia está se dispersando na medicina, na computação e na agricultura?*

Para essa análise comparativa, deverão ser selecionados um número “P₁”, “P₂” e “P₃” de periódicos representativos de cada área do estudo, editados em um período estabelecido. “F₁”, “F₂” e “F₃” corresponderão ao número de fascículos editados, e “A₁”, “A₂” e “A₃” ao respectivo número de artigos neles publicados. Assim, “p₁”, “p₂” e “p₃”; “f₁”, “f₂” e “f₃”, como “a₁”, “a₂” e “a₃”, corresponderão respectivamente ao número de periódicos que têm fascículos em que constam artigos que abordam o assunto da nanotecnologia. As variáveis “r₁”, “r₂” e “r₃” corresponderão ao número total de referências sobre obras de nanotecnologia citadas nos artigos das respectivas áreas. Para verificação de como está se dispersando a nanotecnologia nas áreas em questão, deverão ser levantados os escores das respectivas variáveis referentes a cada ano, tendo por finalidade calcular e comparar os índices de dispersão correspondentes.

No item 5.1, demonstra-se a aplicação do IDS em problema análogo ao anteriormente referido.

Como situação pertinente à identificação das variáveis dessa tabela 3, apresenta-se, no item 5.2, um exemplo explicativo com o respectivo cálculo do Índice de Dispersão Segmentar.

3.2 Representação gráfica de parâmetros de dispersão relacionados a diferentes níveis de distribuição

A figura 1 representa ternos de variáveis, considerando três possíveis níveis de distribuição de determinada unidade de análise: da macrodispersão, para a qual contribuem as variáveis P, p, f; da mesodispersão, relacionando as variáveis F, f, a; da microdispersão, relacionando A, a, r. Essas variáveis assumem representações diversas, conforme exemplificam as tabelas 1, 2 e 3.

Na figura 1, S₁, S₂ e S₃, representam áreas de quadriláteros cujos lados correspondem às medidas das variáveis, respectivas a cada nível de distribuição, e S_{1p}, S_{2p} e S_{3p} correspondem às respectivas situações-padrão, assim estabelecidas: f_p = P; a_p = F; r_p = A.

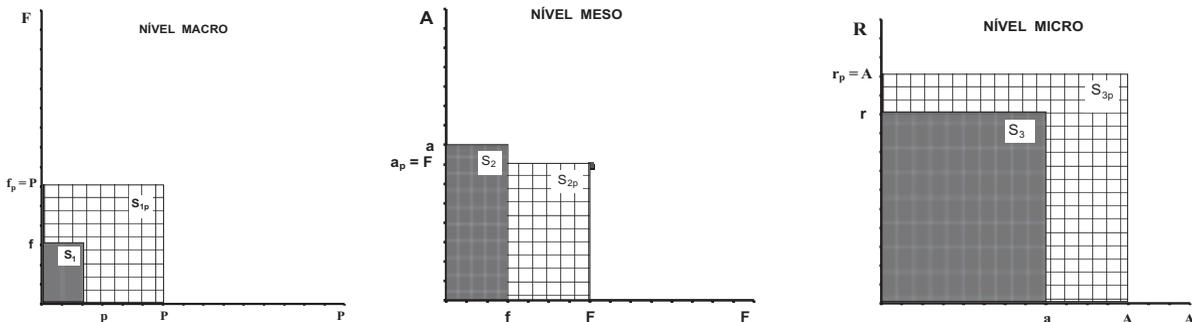
Interpretando essas igualdades de acordo com as variáveis, por exemplo, da tabela 1, tem-se que a situação padrão, f_p = P significa que existe um fascículo de cada periódico apresentando o objeto-alvo de estudo; a_p = F, que, em cada fascículo dos periódicos analisados, existe um artigo em que se encontra o objeto de estudo ou unidade de análise e r_p = A, significando que, em cada artigo analisado, ocorre uma referência relativa ao objeto de estudo.

De acordo com o representado na figura 1, as medidas das áreas de dispersão S₁, S₂ e S₃ são dadas pelos produtos p · f; f · a; a · r e das áreas de distribuição padrão S_{1p}, S_{2p}, S_{3p} por P · f_p; F · a_p; A · r_p. Considerando as igualdades da situação-padrão, tem-se:

$$S_{1p} = P \cdot P; S_{2p} = F \cdot F; S_{3p} = A \cdot A$$

FIGURA 1

Áreas de dispersão observadas (S₁, S₂, S₃) e respectivas áreas de dispersão-padrão (S_{1p}, S_{2p}, S_{3p}) estabelecidas pelas medidas de P, p, f (macrodispersão), F, f, a (mesodispersão) e A, a, r (microdispersão), correspondendo a variáveis bibliométricas identificadas nas tabelas 1, 2 e 3



3.3 Composição e definição do IDS

Para cada nível de distribuição representado na figura 1, estabelece-se um coeficiente de dispersão parcial (C_i). Este coeficiente corresponde a uma medida relativa que traduz a proporção entre os valores observados e os correspondentes à situação-padrão. Isto é, a relação entre as áreas dos quadriláteros representadas por “S” e “S_p”. Assim,

$$C_1 = \frac{S_1}{S_{1p}} = \frac{f_i \cdot p_i}{P_i \cdot P_i} \quad ; \quad C_2 = \frac{S_2}{S_{2p}} = \frac{f_i \cdot a_i}{F_i \cdot F_i} \quad e$$

$$C_3 = \frac{S_3}{S_{3p}} = \frac{a_i \cdot r_i}{A_i \cdot A_i}$$

correspondem, respectivamente, aos níveis macro, meso e micro de uma distribuição que abrange $k = 3$ níveis.

O somatório desses coeficientes compõem, por sua vez, o que aqui é definido como Índice de Dispersão Segmentar, isto é, $IDS_3 = C_1 + C_2 + C_3$

$$ou, \quad IDS_3 = \frac{f_i \cdot p_i}{P_i^2} + \frac{f_i \cdot a_i}{F_i^2} + \frac{a_i \cdot r_i}{A_i^2}$$

3.4 Relações entre as variáveis correspondentes aos níveis de distribuição

Considerando as sete variáveis exemplificadas, tem-se que:

- os valores-limite para as variáveis, “p”, “f” e “a”, correspondem, respectivamente, a “P”, “F” e “A”;
- em qualquer conjunto analisado, ter-se-á sempre $r \geq a \geq f \geq p$, assim como $A \geq F \geq P$;
- para cada um dos níveis de distribuição do objeto de estudo, quatro situações são possíveis de ocorrer quanto aos valores do coeficiente “C”:
 - se $f < P$ e $p < P$, então $0 < C_1 < 1$.
 - se $f > P$ e $p < P$, então $C_1 > 0$.
 - se $f > P$ e $p = P$, então $C_1 > 1$.
 - se $f = p = P$ (situação padrão), então $C_1 = 1$.

d) resultado análogo ocorre com referência à relação entre a-f-F e r-a-A, respectiva aos níveis de meso e microdistribuição.

3.5 Propriedades do IDS

- O IDS corresponde a um número fracionário ou inteiro, maior que zero, ou nulo quando $p = 0$.
- Para uma mesma variação de escore referente às diferentes variáveis da distribuição, a variação no IDS será tanto maior, quanto maior for o nível hierárquico de distribuição afetado (ver exemplos do item 5.2).
- O IDS de uma distribuição-padrão (escores em cada nível = escores da situação-padrão) é igual ao número de níveis considerados na análise. Isto é, $IDS_k = k$.

4. REPRESENTAÇÃO GENÉRICA DO IDS

Conforme definido, o IDS é representado pelo somatório de tantos coeficientes de dispersão quantos forem os níveis de distribuição abordados por um estudo sobre determinada unidade de análise. Assim, o Índice de Dispersão Segmentar abrangendo, por ordem crescente, k níveis de distribuição, fica assim representado:

$$IDS_k = C_1 + C_2 + \dots + C_k$$

$$ou, \quad IDS_k = \sum_{i=1}^k \frac{n_i \cdot n_{i+1}}{N_{i+1}^2}$$

onde,

n_i = número de elementos “positivos” da unidade amostral de menor ordem de grandeza do respectivo nível de distribuição, os quais contemplam a unidade de análise. (Exemplo: número de referências a determinado autor.)

N_{i+1} = número de elementos analisados da unidade amostral de ordem de grandeza imediatamente acima de n_i . (Exemplo: número de artigos analisados.)

n_{i+1} = número de elementos “positivos” da unidade amostral N_{i+1} , os quais contemplam a unidade de análise. (Exemplo: número de artigos analisados que referenciam o autor em estudo.)

5. EXEMPLOS HIPOTÉTICOS PARA CÁLCULO E CONSIDERAÇÕES SOBRE A APLICAÇÃO DO IDS

5.1 Dispersão e extensão de impacto bibliográfico

Objetivo: estabelecer, no âmbito dos periódicos nacionais indexados em determinada base e publicados em determinado biênio, por qual dos campos de conhecimento (CC₁, CC₂, ou CC₃) está mais dispersa a abordagem de uma questão “X”. Considerando-se que para as variáveis definidas na tabela 2 correspondam os dados da tabela 4, resultarão valores para o IDS, que, ordenados decrescentemente, hierarquizam os campos focalizados quanto à dispersão da questão “X”.

Observação: note-se que o presente exemplo abrange apenas dois níveis de distribuição da questão em estudo, uma vez que a unidade amostral de menor grandeza é representada pela variável “a” (número de artigos que abordam a questão “X”).

Portanto, a conclusão referente a esse exemplo mostra que a questão “X” apresenta-se mais dispersa no campo de conhecimento CC₁, seguindo-se os campos CC₂ e CC₃. De outro modo, esse exemplo também ilustra que, por ser o IDS um indicador relativo e *standartizado* às características de cada amostra (estabelecida pela “situação-padrão” na formulação do índice), permite comparar resultados independentemente de contextos particulares às unidades de análise, tais como heterogeneidade das áreas quanto ao número de periódicos que disponham para publicação, produtividade desses periódicos etc.

Em contrapartida e pelas razões antes definidas, se à variável “a” correspondessem artigos portadores, não da informação “X”, mas de citação dessa informação, colhida de qualquer fonte, o IDS representaria a “extensão de impacto” dessa informação. Nesse caso, a variável de menor grandeza, ao nível da microdistribuição, correspondente a “r”, representaria o número de fontes citadas dessa informação. A análise se estenderia para um terceiro nível de distribuição em que também seria considerado o número de artigos (“A”) analisados.

TABELA 4
Dados do exemplo 5.1

Campo de conhecimento	P	p	F	f	a	IDS ₂ *
CC ₁	10	5	60	20	30	1,167
CC ₂	15	8	90	12	15	0,449
CC ₃	20	5	120	16	20	0,222

$$*IDS_2 = \frac{p_i \cdot f_i}{P_i^2} + \frac{f_i \cdot a_i}{F_i^2} \quad (\text{Índice de dispersão segmentar referido a dois níveis de distribuição})$$

5.2 Dispersão e extensão de impacto geográfico

Objetivo: hierarquizar cinco periódicos (tabela 5) quanto à sua dispersão geográfica, considerando o número de bibliotecas (r) que os têm em seu acervo, número de municípios ou instituições (a) em que se encontram essas bibliotecas, número de unidades federativas (f) às quais pertencem esses municípios ou instituições e o número de regiões geográficas (p) às quais pertencem essas unidades federativas.

Observação: considere-se como território de referência da dispersão um país constituído por cinco regiões geopolíticas (P), nas quais se encontram distribuídas 27 unidades federativas (F), representando 60 municípios ou instituições (A) abrangidos por esse estudo, uma vez que neles se encontram as “r” bibliotecas que dispõem do periódico, as quais integram um sistema cooperativo

TABELA 5
Dados do Exemplo 5.2

	P	p	F	f	A	a	r	IDS*	Δ IDS**
Periódico 1	5	4	27	15	60	30	60	3,517	-
Periódico 2	5	5	27	15	60	30	60	4,117	0,600
Periódico 3	5	4	27	16	60	30	60	3,708	0,191
Periódico 4	5	4	27	15	60	31	60	3,555	0,038
Periódico 5	5	4	27	15	60	30	61	3,525	0,008
Situação padrão								3,000	

$$*IDS_3 = \frac{p \cdot f}{P^2} + \frac{f \cdot a}{F^2} + \frac{a \cdot r}{A^2}$$

** Δ IDS = diferença entre o IDS de cada periódico em relação ao IDS do Periódico 1.

de informação na constituição de um catálogo coletivo nacional de publicações seriadas.

Observação: conforme exemplificado, variações unitárias no escore de cada variável (valores sublinhados na tabela 5), sem alteração nas demais, produzem variações “ Δ IDS” tanto maiores quanto hierarquicamente maior for a categoria dimensional da variável em que ocorre o aumento do escore. Isso acontece, em geral, pela acentuada diferença entre os divisores, respectivos a cada nível, que tendem a ser progressivamente maiores a cada menor nível de distribuição. Isso decorre do fato empírico de que $A \geq F \geq P \dots$ estarem seus valores elevados à potência dois no cálculo do respectivo coeficiente.

Em contrapartida, se “r” constituísse o número de fontes que citam determinada informação e “a” o número de municípios em que estas se situam, o IDS estaria dimensionando “impacto geográfico” da referida informação.

6. POSSÍVEIS UTILIZAÇÕES DO IDS

Partindo do modelo descrito e da atribuição de representações às variáveis que compõem a respectiva equação, tais como as abordadas nas tabelas 1, 2 e 3, estabelece-se a possibilidade de executar comparações

referentes à dispersão literária ou físico-geográfica de distintas unidades de análise da bibliometria, cienciometria, informetria ou similares. É também considerada a possibilidade de o IDS dimensionar a “extensão de impacto” de uma unidade de análise, à medida que seja aplicado à análise de citações.

A tabela 6 sintetiza o que, nos limites do presente estudo, é dado concluir sobre possíveis aplicações, representações e inferências do índice proposto.

7. CONSIDERAÇÕES

A designação Índice de Dispersão Segmentar (IDS) foi cunhada visando a compatibilizar critérios sugeridos por Baena, Cases e Medina (2005) para denominação de indicadores quantitativos e qualitativos.

A esse IDS cabe ser atribuída a condição de um indicador, por se coadunar com a definição genérica de indicadores da ISO-1998 (STUBBS, 2004), e a definição de indicador cienciométrico de Vinkler (2001), quando aplicado em análises de cunho cienciométrico. No âmbito dos indicadores de ciência e tecnologia, é classificado como “complexo”, visto não ser obtido por uma medida de enumeração, mas porque seu valor é obtido por uma medida indireta, como é o caso dos índices (combinação de vários indicadores) e dos coeficientes (relação entre

TABELA 6

Representações e possíveis inferências relativas à aplicação do Índice de Dispersão Segmentar (IDS)

APLICAÇÃO	REPRESENTAÇÃO	VARIÁVEIS	POSSÍVEL INFERÊNCIA
Comparar áreas de estudo quanto à dispersão que nelas ocorra de determinado “X” *	Dispersão literária	Tabela 2	Penetrabilidade, ou pertinência, ou aplicabilidade de “X” em áreas de estudo.
Hierarquizar diversos “X” quanto às suas dispersões no âmbito das publicações referentes a áreas de conhecimento, disciplina etc.	Dispersão literária	Tabela 1	Associação ou relevância, ou contribuição, ou importância etc. de “X”, para áreas de conhecimento, disciplinas etc.
Comparar diversos “X” quanto às dispersões bibliográficas das fontes (artigos) que os citam.	Difusão literária ou “Extensão de impacto”	Tabela 2	Difusão ou impacto bibliográfico de “X”.
Hierarquizar diversos “X” quanto às suas dispersões por localidades abrangidas de um território geográfico.	Dispersão geográfica	Tabela 3	Potencial de difusão de “X” através de obras ou periódicos etc.
Comparar diversos “X” quanto às dispersões geográficas das fontes citantes.	Difusão geográfica ou “Extensão de impacto geográfico”	Tabela 3**	Difusão ou impacto de “X” em uma determinada área geográfica.

* Exemplos de “X”: obra(s), periódico(s), artigo(s), autor(es), tema(s), método(s) ou técnica(s), vocábulo(s) etc.

** As variáveis corresponderão às unidades físico-geográficas que forem eleitas no mapeamento das localidades em que se situam as fontes que citam “X”.

dois indicadores), conforme Sanchez (1997), citado por Liberal (2005). No âmbito da classificação de indicadores bibliométricos e cientométricos de Vinkler (1988, 2001), o IDS corresponde a um “indicador relativo”, quantitativo, modelado com base na comparação de dados de um conjunto aos dados de uma condição *standard*. Moed (1996), que classifica os indicadores como “absolutos” e “relativos”, apresenta como exemplo desse último a razão entre o “atual” e o “esperado” número de citações de alguma unidade de análise bibliométrica. Na modelagem do IDS, o “atual” corresponde aos valores observados respectivamente às diversas variáveis e o “esperado” corresponde aos valores respectivamente ao que aqui é denominado situação-padrão. É oportuno também citar que, embora se utilize a terminologia de “variáveis”, de acordo com Macias-Chapula (1998), estas constituem também indicadores quando ensejam significados avaliativos, fato que ele exemplifica dizendo “contar trabalhos não é difícil, dar sentido aos dados é mais complexo” (MACIAS-CHAPULA, 1998, p.137). Ainda quanto à identificação da natureza do índice aqui gerado, dentre as três categorias de indicadores estabelecidas quanto à sua possível finalidade, conforme já caracterizado nos exemplos 5.1 e 5.2, é também possível sua classificação, de acordo com Cunningham (1997, p.74), como “indicador usado para examinar impactos científicos”, quando aplicado a análises de citações.

Entretanto, o modelo aqui desenvolvido suporta ser utilizado para análises referentes às diferentes unidades de estudo em ciência da informação, genericamente expressas na proposição de classificação de Frota (1998) pelas categorias de “suportes”, “usuários”, “instituições” e “profissionais”, referente às quais se encontram algumas representadas nas tabelas 1, 2 e 3. Essa condição estabelece situações de análise que contemplam as três características desejáveis para um indicador mencionadas por Trzesniak (1998). A da “cobertura” (ou abrangência ou amplitude), visto que comporta aplicação a processos/sistemas de natureza diferente dentro de uma mesma área de conhecimento. A da “portabilidade” (ou transferibilidade), visto que seja extensível ao estudo de sistemas/processos de áreas contíguas. E, finalmente, da “invariância de escala”, à medida que mantém sua validade e interpretação, mesmo que as dimensões do processo ou sistema sejam acentuadamente diferentes.

Na sua aplicação, há de se considerar, primeiramente, que haja seleção adequada ao propósito da avaliação quanto ao conjunto a que se refere a variável “P”, que contém, por sua vez, os subconjuntos representados pelas demais variáveis. O conjunto dos elementos amostrais

representado por “P” constitui o conjunto de referência, que, como adverte Vinkler (2003, p.691), é “a parte crucial na formação de indicadores relativos”. Essa crucialidade está, por exemplo, entre os aspectos que são de reclamo de muitos estudiosos quanto à possível inadequação do Índice de Citação aplicado pelo ISI na avaliação da produção científica, tanto de países desenvolvidos, quanto em desenvolvimento. Por sinal, questões que vão de metodológicas a terminológicas, apontadas em diversos trabalhos (GLÄNZEL, 1996; LAZAREV, 1996; RAVICHANDRA RAO, 1996; McGRATH, 1996), são também exemplos de clamores pela *standartização* nas áreas da biblioteconomia e da cientometria (ou ciëntometria, conforme alguns), o que, por seu turno, parece próprio de áreas em desenvolvimento e consolidação, mormente quando constituem ferramentas de avaliação. Avaliar, porém, unidades de análise quanto à dispersão que estas apresentam pode oferecer conotações diversas. Assim, a dispersão físico-geográfica, ou virtual de uma unidade, refere sua distribuição no respectivo espaço de referência. Em contrapartida, a dispersão de citações correspondentes a um objeto de estudo, como, por exemplo, uma obra literária, incorpora o significado de “extensão de impacto”, conforme anteriormente caracterizado. Portanto, sendo o IDS utilizado para análise de dispersão dos suportes de informação, sua interpretação referirá “espalhamento” da informação. Uma vez utilizado em análises de citações, refere difusão de uma informação temporalmente associada à data em que foi usada como citação, o que corresponde, por sua vez, à representação de extensão de impacto da mesma.

O IDS, por hierarquizar a distribuição das citações atentando para o fato de que os níveis “macro”, “meso” e “micro” representam diferentes “estágios” de dispersão, não enseja uma referência de impacto dimensionalmente mais abrangente do que se este fosse calculado como se os diversos suportes referentes à comunicação constituíssem postos ponderalmente igualizados? Em trabalho recente, Frandsen (2005) considera a possibilidade de que a análise da concentração geográfica pode atuar como um suplemento ao Fator de Impacto – JIF (Journal Impact Factor). Em seu trabalho, ele se utiliza do índice de concentração HHI (Herfindahl-Hirschman Index) originário da área da economia, que serve para acrescentar, ao fator de impacto, informações nele originalmente não contidas.

Semelhante trabalho parece refletir que mesmo indicadores tão amplamente utilizados (ou impostos?) dêem margem a se considerar que os fenômenos que se

pretende com eles representar contêmham mais complexidade e requeiram que mais variáveis devam ser consideradas para sua avaliação.

CONCLUSÕES

O algoritmo aqui desenvolvido peculiariza a representação do fenômeno da dispersão por dois aspectos: individualiza hierarquicamente níveis referentes à distribuição da unidade de análise e permite que sejam considerados tantos destes níveis, quanto possíveis ou convenientes para essa representação. Isso decorre do caráter segmentar cumulativo do IDS, composto por tantos coeficientes, respectivos a tantos níveis de distribuição quantos forem elegidos para cada estudo. O IDS mostra-se uma ferramenta aplicável a estudos comparativos referentes à dispersão bibliográfica, dispersão físico-geográfica e “extensão de impacto” de unidades de análise bibliométricas, cienciométricas ou similares. Uma análise preliminar desse índice sugere que contém características gerais que são desejáveis para um indicador. Entretanto, diversos estudos que permitam testificar sua validação empírica ainda devem ser conduzidos.

Artigo submetido em 01/12/2005 e aceito em 10/07/2006.

REFERÊNCIAS

- BAENA, A. J. L.; CASES, M. V.; MEDINA, M. B. Indicadores cuantitativos y cualitativos para la evaluación de la actividad investigadora: ¿complementarios? ¿contradictorios? ¿excluyentes?. *Cuadernos IRC*. Disponível em: <http://www.uca.es/serv/consejo_social/doc/transferecia14.pdf>. Acesso em: 28 jun. 2005.
- BORGES, P. C. R. Métodos quantitativos de apoio à bibliometria: a pesquisa operacional pode ser uma alternativa?. *Ciência da Informação*, v. 31, n. 3, p. 5-17, 2002.
- CAMPOS, M. Conceitos atuais em bibliometria. *Arquivos Brasileiros de Oftalmologia*, n. 66, p. 18-21, 2003.
- CANADIAN FEDERATION FOR THE HUMANITIES AND SOCIAL SCIENCES - CFHSS. *Assessing the impact of research in the Humanities and Social Sciences in political, social and cultural areas*. Disponível em: <<http://fedcan.ca/english/fromold/perf-ind-literaturereviews-oise-part2.cfm>>. Acesso em: 14 jun. 2005.
- CUNNINGHAM, P. The evaluation of european programmes and the future of scientometrics. *Scientometrics*, v. 38, n. 1, p. 71-85, 1997.
- FRANSEN, T. F. Geographical concentration: the case of economics journals. *Scientometrics*, v. 63, n. 1, p. 69-85, 2005.
- FROTA, M. G. C. A delimitação das unidades de análise em ciência da informação. *Ciência da Informação*, v. 27, n. 3, p. 262-267, 1998.
- GLÄNZEL, W. The need for standards in bibliometric research and technology. *Scientometrics*, v. 35, n. 2, p. 167-176, 1996.

_____. *Bibliometrics as a research field: a course on theory and application of bibliometric indicators*. Disponível em: <http://www.norslis.net/2004/Bib_Module_KUL.pdf>. Acesso em: 25 maio 2005.

_____; MOED, H. F. Journal impact measures in bibliometric research. *Scientometrics*, v. 53, n. 2, p. 171-193, 2002.

GODIN, B. Measuring knowledge flows between countries: the use of scientific meeting data. *Scientometrics*, v. 42, n. 3, p. 313-323, 1998.

KORTELAJINEN, T. A. M. Studing the international diffusion of a national scientific journal. *Scientometrics*, v. 51, n. 1, p. 133-146, 2001.

LAZAREV, V. S. On chaos in bibliometric terminology. *Scientometrics*, v. 35, n. 2, p. 271-277, 1996.

LIBERAL, C. G. *Indicadores de ciência e tecnologia: conceitos e elementos históricos*. Disponível em: <http://www.fapesc.ret-sc.br/crciti/docs/gt_4B.doc>. Acesso em: 15 ago. 2005.

MACIAS-CHAPULA, C. A. O papel da informetria e da cienciométrica e sua perspectiva nacional e internacional. *Ciência da Informação*, v. 27, n. 2, p. 134-140, 1998.

MATOS, N. E. P. La bibliografía, bibliometría y las ciencias afines. *ACIMED*, v. 10, n. 3, 2002. Disponível em: <http://www.bvs.sld.cu/revistas/aci/vol10_3_02/Aci012002.htm>. Acesso em: 07 jun. 2005.

M_CGRATH, W. E. The unit of analysis (objects of study) in bibliometrics an scientometrics. *Scientometrics*, v. 35, n. 2, p. 257-264, 1996.

MOED, H. F. Differences in the construccion of SCI based bibliometric indicators among various producers: a first overview. *Scientometrics*, v. 35, n. 2, p. 177-191, 1996.

QUONIAM, L. *et al.* Inteligência obtida pela aplicação de data mining em base de teses francesas sobre o Brasil. *Ciência da Informação*, v. 30, n. 2, p. 20-28, 2001.

RAVICHANDRA RAO, I. K. Methodological and conceptual questions on bibliometric standards. *Scientometrics*, v. 35, n. 2, p. 265-270, 1996.

_____. An analysis of Bradford multipliers and a model to explain law of scattering. *Scientometrics*, v. 41, n. 1-2, p. 93-100, 1998.

SANCHEZ, O. M. *Dicionário de orçamento, planejamento e áreas afins*. 1. ed. Brasília: Prisma, 1997.

SPINAK, E. Indicadores cienciométricos. *Ciência da Informação*, v. 27, n. 2, p. 141-148, 1998.

STUBBS, E. A. Indicadores de desempenho: naturaleza, utilidad y construcción. *Ciência da Informação*, v. 33, n. 1, p. 149-154, 2004.

TRZESNIAK, P. Indicadores quantitativos: reflexões que antecedem seu estabelecimento. *Ciência da Informação*, v. 27, n. 2, p. 159-164, 1998.

VANTI, N. A. P. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão de conhecimento. *Ciência da Informação*, v. 31, n. 2, p. 152-162, 2002.

VINKLER, P. An attemnt of surveying and classifying bibliometric indicators for scientometric purposes. *Scientometrics*, v. 13, n. 5-6, p. 239-259, 1988.

_____. An attempt for defining some basic categories of scientometrics and classifying the indicators of evaluative scientometrics. *Scientometrics*, v. 50, n. 3, p. 539-544, 2001.

_____. Relations of relative scientometric indicators. *Scientometrics*, v. 58, n. 3, p. 687-694, 2003.