

## Orientação temática e coeficiente de correlação para análise comparativa entre dados altmétricos e citações: uma análise da revista DataGramZero

**Ronaldo Ferreira Araujo**

Doutor; Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG, Brasil;  
ronaldfa@gmail.com

**Gustavo Miranda Caran**

Doutorando; Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brasil;  
gmcaran@gmail.com

**Iara Vidal Pereira de Souza**

Doutoranda; Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brasil;  
iaravidalps@gmail.com

**Resumo:** Este trabalho propõe a verificação da relação entre a altmetria e os estudos tradicionais de citação por meio de uma orientação temática como traço identificador dos trabalhos que analisa, bem como no emprego do Coeficiente de Correlação de Pearson (CCP). Para tanto, categoriza-se 441 artigos da Revista DataGramZero segundo uma taxonomia que dispõe de dez temáticas de pesquisa para a ciência da informação, comparando as citações desses artigos com seus dados altmétricos com objetivo de qualificar a correlação entre esses dois conjuntos de dados. Houve destaque para a categoria: “aspectos teóricos e gerais da ciência da informação” tanto com mais artigos citados quanto em maior índice de atenção online. Verificou-se, dentre outros aspectos, que há uma correlação forte e diretamente proporcional ( $P1 = 0,8018$ ) entre as variáveis de tempo de publicação e número de citações, e uma correlação forte, porém inversamente proporcional ( $P2 = -0,7068$ ), entre as variáveis de tempo de publicação e número de menções altmétricas; bem como uma baixa correlação entre o número de menções e o número de citações por artigo. Ao serem calculados os coeficientes das médias ponderadas das duas variáveis por categoria temática, a correlação tornou-se moderada ( $P = 0,65$ ). A discussão dos resultados foi ao encontro de pesquisas anteriores, e reforçam as potencialidades no emprego do CCP para estudos métricos da Ciência da Informação.

**Palavras-chave:** Altmetria. Citação. Coeficiente de Correlação.

### 1 Introdução

A crescente utilização acadêmica de ferramentas de mídias sociais como o Facebook, o Twitter e o Mendeley têm despertado nas comunidades científicas

internacionais interessadas nos estudos métricos da informação a necessidade de rastrear o impacto acadêmico na web social por meio da criação de novos filtros e indicadores. Em resposta a esta necessidade, um grupo de pesquisadores (PRIEM et al., 2010) propôs a constituição de um novo campo dos estudos métricos, que tem sido chamado de *Altmetrics* ou *Altmétria*. Os proponentes da *altmétria* defendem que novas métricas da comunicação científica se fazem necessárias, tendo em vista a crise vivenciada pelos filtros tradicionais que determinam a qualidade da informação científica, como a revisão por pares (lenta, desestimula a inovação e não limita o volume de pesquisas publicadas); a contagem de citações (insuficiente, limitada a publicações formais, desconsidera contexto e razões de citação), e o fator de impacto (suscetível a manipulações e distorções) (PRIEM et al., 2010; ARAÚJO, 2015a).

As *altmetrics* se apresentam como complementos aos filtros baseados em citações, com o potencial de contribuir na superação de algumas de suas deficiências. Entre as vantagens da *altmétria* está a capacidade de verificar impactos em audiências dentro e fora da comunidade científica, uma vez que:

[...] o fluxo informacional que o artigo gera na web pode formar uma rede articulada por meio das interações em torno do artigo (cliquear, ler, compartilhar, comentar, favoritar). Tais dados podem dimensionar o impacto social desse artigo. (ARAÚJO, 2015b).

O manifesto pela *altmétria* traz em seus parágrafos finais uma chamada por trabalhos que demonstrem correlações entre as métricas alternativas e as tradicionais, como parte dos esforços de validação do novo campo por meio do diálogo com disciplinas mais consolidadas (PRIEM et al., 2010). Desde então, vários estudos têm surgido com intuito de indicar a complementaridade que os estudos altmétricos podem oferecer para os estudos métricos tradicionais.

O presente estudo é desenvolvido nessa perspectiva, e propõe a verificação da relação entre a *altmétria* e os estudos tradicionais de citação por meio de uma orientação temática como traço identificador dos trabalhos que analisa, bem como no emprego estatístico do Coeficiente de Correlação de Pearson – técnica capaz de indicar se há uma relação entre duas variáveis (métricas tradicionais e *altmetrias*, neste caso). Primeiramente, a presente

comunicação discute aspectos conceituais por meio de revisão da literatura, para em seguida descrever os procedimentos metodológicos da pesquisa empírica, os resultados encontrados e sua discussão.

Os resultados apresentados visam corroborar com estudos como o de Boon e Foon (2014), que também analisaram a correlação entre métricas tradicionais e altmetrias com o emprego do Coeficiente de Correlação de Pearson.

## **2 Altmetrics e dados de citação: à guisa de uma correlação**

Uma das métricas alternativas mais promissoras no que diz respeito à correlação com a contagem de citações são as medidas obtidas por meio de gerenciadores de referências online como o Mendeley e o Zotero. Tendo em vista que estas ferramentas são utilizadas principalmente por pesquisadores e estudantes para organizar as referências que utilizarão em suas publicações, é razoável considerar que seu uso é um indicador de futuras citações (Taraborelli, 2008). O Mendeley aparece com frequência como a principal fonte em estudos altmétricos (ver, por exemplo, PRIEM; PIWOWAR; HEMMINGER, 2012; ROBINSON-GARCÍA et al., 2014), o que pode ser atribuído ao tamanho e à variedade de sua base de usuários, bem como à relativa facilidade de obtenção de seus dados por meio de uma *Application Program Interface* (API) pública.

Um dos primeiros estudos a explorar possíveis relações entre métricas alternativas e indicadores tradicionais verificou correlações positivas, de fracas a moderadas, entre a contagem de citações e o número de leitores no Mendeley para artigos publicados em periódicos da *Public Library of Science* – PLOS (PRIEM; PIWOWAR; HEMMINGER, 2012). Outros indicadores altmétricos considerados na pesquisa, como menções no Twitter e Facebook, apresentaram correlações muito fracas ou inexistentes com as citações, indicando que estas métricas medem impactos distintos, ainda que relacionados.

Outro estudo comparando altmetrias e citações e analisando publicações cobertas pela Web of Science (COSTAS; ZAHEDI; WOUTERS, 2014) verificou uma relação positiva entre o número de indicadores altmétricos e a

contagem de citações de um determinado periódico: publicações com mais menções em fontes altmétricas tendiam a acumular mais citações. No entanto, a correlação no nível dos artigos é comparativamente muito fraca, o que pode ser explicado pela alta presença de publicações na amostra sem qualquer dado altmétrico – cerca de 85% do total.

Lutz Bornmann (2015) realizou uma meta-análise de estudos que exploram a correlação entre métricas alternativas e a contagem de citações, com foco em três métricas específicas: microblogs (Twitter), blogs e gerenciadores online de referências. A partir de uma revisão da literatura da área, o autor reuniu os coeficientes de correlação obtidos em diferentes estudos para realizar um cálculo agregado. Seus resultados mostram que a correlação entre métricas tradicionais e alternativas é em geral desprezível ( $r=0.003$ ) para dados do Twitter, baixa ( $r=0.12$ ) para dados de blogs, e baixa a moderada (CiteULike  $r=0.23$ ; Mendeley  $r=0.51$ ) para os gerenciadores eletrônicos de referência. Para Bornmann (2015), o valor de uma métrica alternativa é inversamente proporcional à sua correlação com a análise de citações, pois sua utilidade estaria justamente em revelar algo que escapa às medidas tradicionais. Sendo assim, o autor conclui que as menções no Twitter seriam as métricas alternativas mais valiosas – pois obteve o menor coeficiente de correlação em seus estudos.

Thelwall e outros (2013), em um amplo estudo, compararam dados altmétricos com citações de 208,739 artigos PubMed publicados entre 2010 e 2012. Os resultados mostram que seis fontes de dados altmétricos (Twitter, Facebook, *research highlights*, blogs, *mainstream media* e fóruns) apresentavam estreita associação com contagens de citação, pelo menos em áreas como ciências médicas e biológicas.

Em resumo, os estudos indicam que métricas alternativas e citações mensuram impactos distintos, mas relacionados em alguma medida. Estes indicadores se complementam, e nenhum deles é capaz de descrever isoladamente o quadro completo da comunicação científica (PRIEM; PIWOWAR; HEMMINGER, 2012). Além disso, é preciso considerar fatores disciplinares, linguísticos e geográficos, entre outros, que geram diferenças nas

práticas de comunicação científica entre comunidades diversas (ALPERIN, 2014).

Estas diferenças levam a variações nos padrões de acúmulo de citações, e também podem afetar o uso das ferramentas sociais e consequentemente a acumulação de indicadores altmétricos em diferentes fontes, disciplinas e regiões, o que aponta para a necessidade de realizar estudos altmétricos em contextos diversos.

### 3 Material e método

A presente pesquisa exploratória investiga as relações entre dados altmétricos e citações no âmbito do periódico DataGramZero. O estudo segue a orientação metodológica de Araújo (2014; 2015b) na busca de uma análise altmétrica por meio do uso de APIs para coleta de dados do Facebook <<https://developers.facebook.com/tools/explorer/>> e do Twitter <<https://dev.twitter.com/rest/public>>. Os dados de leitura no Mendeley foram obtidos por meio de busca manual <<http://mendeley.com/research-papers/search/>>. Estas três fontes altmétricas foram escolhidas por reunirem a maior concentração da atenção online que artigos científicos recebem, conforme o estudo de Robinson-García e outros (2014).

As consultas nas APIs foram parametrizadas pela URL de todos os artigos da revista DataGramZero, com sua representação quantitativa em termos de *shares* (compartilhamentos), *likes* (curtidas) e *comments* (comentários) para o Facebook e de *tweets* para o Twitter. Foram resolvidos e agrupados os que apresentaram duplicidade de URL (ARAÚJO, 2015b). No Mendeley, o termo utilizado no campo de busca foi o nome da revista com filtro para local de publicação/periódico [published\_in:"datagramazero"].

As citações, por sua vez, foram obtidas na base do Google Acadêmico por meio do Software *Publish or Perish* versão 4.17.0: pela consulta na aba *Journal impact* (impacto da revista); e no campo *Journal title* (título da revista) com o parâmetro “datagramazero”. Para ambos os conjuntos de dados, foram considerados para análise artigos publicados até a edição de abril de 2014.

Para a classificação da orientação temática dos artigos, recorreu-se à taxonomia elaborada por Oddone e Gomes (2004), aplicando as dez categorias de temas de pesquisa da Ciência da Informação. Após a coleta, tabulação e classificação temática dos artigos, foi calculado o Coeficiente de Correlação de Pearson ( $P$ ) entre as respectivas contagens de menções (variável alométrica) e citações (variável tradicional). Veja a fórmula a seguir (PESTANA; GAGEIRO, 2014):

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \cdot \text{var}(Y)}}$$

A análise baseada no Coeficiente de Correlação de Pearson permite identificar, através de métodos matemáticos, qual o grau de influência ou de similaridade no comportamento entre duas variáveis. Em síntese, o cálculo de Pearson indica se, à medida que uma variável for alterada, a outra variável terá um comportamento linear e previsível.

O coeficiente varia entre  $-1 \leq r \leq +1$ . Nesse intervalo ( $-1 \leq r \leq +1$ ), a correlação varia de sentido, de negativo para positivo, e de força, sendo fraca quando se aproxima de zero (à esquerda ou à direita), e moderada ou forte aproximando-se de  $-1$  ou  $+1$ , sendo  $r$  o coeficiente de Pearson (BARBETTA, 2003) – utilizando, dentre outras técnicas de representação, o Gráfico de Dispersão (VIEIRA, 2011).

#### **4 Resultados e discussão**

Foram analisados os dados de 441 artigos publicados entre os anos de 1999 e 2014, os quais somam 1.932 citações recebidas e 1.571 menções alométricas. Desse total, 241 artigos receberam algum tipo de menção (55%), e 162 artigos foram citados pelo menos uma vez (37%). Esses resultados evidenciam um volume de indicadores alométricos mais elevado em relação ao constatado por Costas, Zahedi e Wouters (2014). O número de menções se evidencia, em relação ao número de citações, de maneira mais abrangente (presente em um número

maior de artigos) e intensa (número maior de menções por artigo). A Tabela 1 apresenta a distribuição dos dados indicando o número de artigos, o total de citações e o total de menções por ano – com seus valores correspondentes em porcentagem. A Tabela 1 também apresenta o coeficiente de Pearson ( $P$ ) das variáveis citações anuais *versus* menções anuais.

**Tabela 1 - Artigos, citações e menções por ano.**

Ano	Artigos	Citações	(%)	Menções	(%)
1999	6	251	12,99	53	3,37
2000	23	119	6,16	47	2,99
2001	26	276	14,29	54	3,44
2002	29	137	7,09	40	2,55
2003	27	222	11,49	45	2,86
2004	29	321	16,61	170	10,82
2005	24	72	3,73	68	4,33
2006	27	127	6,57	77	4,9
2007	26	125	6,47	118	7,51
2008	31	128	6,63	94	5,98
2009	34	82	4,24	107	6,81
2010	34	46	2,38	149	9,48
2011	39	18	0,93	135	8,59
2012	43	8	0,41	131	8,34
2013	32	0	0	83	5,28
2014	11	0	0	200	12,73
<b>Total</b>	<b>441</b>	<b>1.932</b>	<b>100</b>	<b>1.571</b>	<b>100</b>
<b>Correlação Tempo Publicação x Citação:</b>				$P_1 = 0,8018$	
<b>Correlação Tempo Publicação x Menção:</b>				$P_2 = -0,7068$	
<b>Correlação Citação x Menção:</b>				$P_3 = -0,3427$	

Fonte: Dados da pesquisa.

No total, os artigos publicados no ano de 2004 reuniram o maior número de citações (321), o que parece condizente com o comportamento comumente visto nos estudos tradicionais de citação, nos quais artigos mais antigos tendem a acumular mais citações (BRAGA, 1974; SOLLA PRICE, 1974). Já os artigos publicados em 2014 acumularam o maior número de menções altmétricas (200). No entanto, mais da metade (55%) do total de menções se refere a artigos publicados antes de 2010. Esse dado parece ir de encontro a estudos anteriores,

em que a maioria dos dados altmétricos tendem a fazer referência a publicações recentes (COSTAS; ZAHEDI; WOUTERS, 2014; PRIEM; HEMMINGER, 2010; MARQUES, 2014; THELWALL, 2013).

Os resultados parecem apontar para duas tendências. Por um lado, o quantitativo de citações é considerado uma maneira de avaliar o interesse de pesquisadores individuais ou grupo de pesquisadores por determinados artigos ou periódicos (MEADOWS, 1999). Por outro lado, o valor altmétrico parece expressar a atenção online por meio da disseminação e outros tipos de interação que tais artigos geram na web. Por englobar mídias menos formais, atingem não apenas os interessados dentro da comunidade científica, mas também o grande público – cuja apropriação indica impacto social.

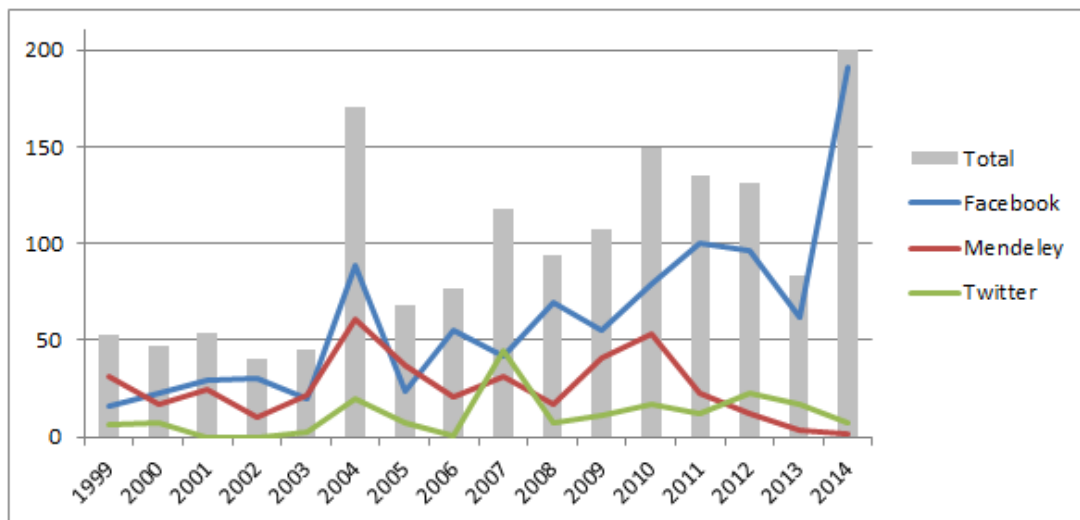
No intuito de estabelecer um parâmetro de correlação entre o período de tempo em que um artigo é publicado e o número de citações recebidas, foi aplicado o Coeficiente de Correlação de Pearson. O resultado desse cálculo indicou uma correlação forte e diretamente proporcional ( $P_1 = 0,8018$ ). Aplicado o mesmo cálculo entre as variáveis tempo de publicação *versus* número de menções, foi encontrada também uma correlação forte, porém inversamente proporcional ( $P_2 = -0,7068$ ). Desse modo, há uma evidência estatística que ratifica a hipótese de que quanto mais antigo um artigo, maior será o número de citações e menor o número de menções atuais.

Entretanto, ao ser aplicado o Coeficiente de Pearson nas variáveis citações e menções, percebeu-se uma correlação moderada/fraca ( $P_3 = -0,3427$ ). Tal evidência indica que há uma baixa preditividade na mensuração do número de citações a partir do número de menções consideradas em conjunto, e do número de menções a partir do número de citações.

A atenção online foi medida e pode ser visualizada no Gráfico 1, que apresenta a distribuição das menções recebidas por ano de publicação com indicativo do valor total (em barra) bem como das incidências (em linha) por mídia social.



**Gráfico 1** - Menções em mídias sociais por ano de publicação.



Fonte: Dados da pesquisa.

Quando se analisa o total de menções, percebe-se um movimento crescente com uma média de 98,19 menções por ano de publicação, os artigos publicados nos anos de 2000, 2002 e 2003 tendo os menores valores altmétricos, com 47, 40 e 45 menções, respectivamente. Os artigos publicados em 2004, 2010 e 2014, por outro lado, apresentam a maior audiência, alcançando respectivamente 170, 149 e 200 menções.

O Facebook, maior mídia social em número de usuários no mundo (1,32 bilhão de membros) e no Brasil (MÜLLER, 2014) lidera a audiência online alcançada pelos artigos com 981 menções (62,44%), seguido do Mendeley com 407 (25,91%), e do Twitter com 183 (11,65%). O resultado difere em ordem inversa ao estudo de Nascimento e Oddone (2015) que aplicaram a ferramenta Altmetric Explorer (da empresa Altmetric) para avaliar periódicos brasileiros na área de Ciência da Informação. Naquele estudo, as autoras registraram mais menções provenientes do Twitter, seguido do Mendeley e em menor expressão, do Facebook. Em sua análise de periódicos brasileiros disponíveis na base SciELO, que também utilizou a ferramenta Altmetric Explorer, Alperin (2014) encontrou apenas 596 artigos (2,81% de sua amostra) mencionados em um *post* público do Facebook, e 1.286 artigos (6,03% da amostra) com pelo menos uma menção no Twitter.

O Mendeley supera os dados do Facebook e do Twitter para os artigos publicados em 1999, 2003 e 2005. O único ano em que as menções do Twitter superaram os dados altmétricos do Facebook e do Mendeley foi em 2007. Os artigos publicados neste ano obtiveram 45 menções no Twitter, contra 40 do Facebook. O Facebook lidera a preferência para disseminação de artigos da revista publicados nos anos restantes. Como destaca Alperin (2014), é possível que os valores reais para o Facebook sejam maiores, uma vez que esse tipo de busca trabalha com postagens públicas, não contando atualizações com acesso restrito ou em grupos fechados.

Para iniciar a análise comparativa entre o conjunto de dados apresentado, os 441 artigos foram categorizados segundo a taxonomia de Oddone e Gomes (2004). As 10 categorias temáticas tiveram as médias de número de citações e número de menções calculadas, acompanhadas da variação de distribuição desses indicadores, utilizando o desvio padrão entre parênteses. A Tabela 2 apresenta tais resultados.

**Tabela 2** - Distribuição do conjunto de dados por categorização temática.

<b>Categorias temáticas</b>	<b>Nº Artigos</b>	<b>Média Menções (Dev. Padrão)</b>	<b>Média Citações (Dev. Padrão)</b>	<b>Correlação (P)</b>
T1 - Aspectos teóricos e gerais da ciência da informação	115	4,3 (7,68)	5,62 (17,44)	0,5710
T2 - Formação profissional e mercado de trabalho	23	1,26 (1,82)	3,04 (4,46)	-0,1887
T3 - Gerência de serviços e unidades de informação	55	3,24 (6,14)	4,31 (20,04)	0,0393
T4 - Estudos de usuário, demanda e uso da informação	36	6,44 (23,24)	4,42 (7,34)	-0,0197
T5 - Comunicação, divulgação e produção editorial	33	3,09 (4,6)	4,88 (14,28)	0,0952
T6 - Informação, cultura e sociedade	29	2,31 (3,17)	2,17 (5,62)	-0,0919
T7 - Legislação, políticas públicas de informação e de cultura	37	1,89 (4,54)	3,89 (7,68)	-0,0414
T8 - Tecnologias da informação	43	3,19 (6,16)	4,14 (8,98)	0,4779
T9 - Processamento, recuperação e disseminação da informação	59	4,22 (6,49)	4,64 (8,98)	0,4370
T10 - Assuntos correlatos e outros	11	1,18 (1,8)	0 (0)	-
<b>TOTAL</b>	<b>441</b>	<b>3,56 (8,91)</b>	<b>4,38 (13,31)</b>	<b>0,2306</b>

Fonte: Dados da pesquisa.

A aferição temática como traço identificador dos dados analisados tem como propósito verificar como eles se comportam e pode sinalizar similaridades e diferenças entre suas relações. No primeiro momento percebe-se que a categoria temática que dispõe sobre os aspectos teóricos e gerais da ciência da informação (T1) reúne o maior número de artigos publicados (115), e se destaca por obter a maior média de citações (5,62) e a segunda maior média de menções (4,3).

A categoria voltada para trabalhos que versam sobre o processamento, recuperação e disseminação da informação (T9), segunda maior em número de artigos (59) ocupa o terceiro lugar na média de menções (4,22) e de citações (4,64).

A categoria que discute os estudos de usuário, demanda e uso da informação (T4), embora esteja entre as que apresentam o menor número de artigos (36), registrou elevada audiência online, com a maior média de menções (6,44). A categoria temática que aborda a comunicação, divulgação e produção editorial (T5), também entre as que apresentam o menor número de artigos (33), é a segunda maior em média de citações (4,88).

Um aspecto importante a ser destacado diz respeito aos desvios padrão apresentados na Tabela 2. Todas as categorias temáticas investigadas apresentaram elevados índices de heterogenia, com resultados acima das médias (citações e menções). Percebe-se, portanto, que há uma significativa concentração de menções (22% dos artigos representam 80% de menções) e citações (12% dos artigos representam 80% das citações) em poucos artigos. O elevado desvio padrão fica também evidente ao considerarmos que 45% dos artigos não receberam menções e 63% não receberam citações – fazendo com que a mediana de citações e menções flutue entre 0 e 1.

Com a constatação do elevado desvio padrão somado à baixa correlação entre as duas variáveis, conforme a Tabela 1 ( $P_3 = 0,3427$ ) e a Tabela 2 ( $P_{total} = 0,2306$ ), é possível aferir que:

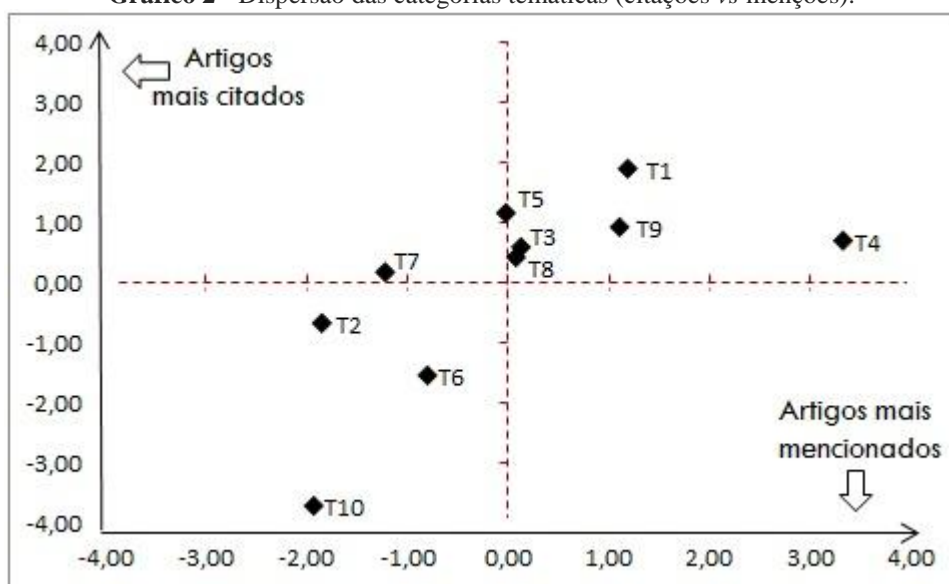
- a) artigos com números elevados de citações não terão, necessariamente, números elevados de menções, e vice-versa;
- b) artigos com números elevados de citações não terão,

necessariamente, baixos índices de menções, e vice-versa;

- c) categorias temáticas não são capazes de explicar, necessariamente, o comportamento do número de citações e de menções.

O Gráfico 2 ilustra a dispersão de cada categoria temática de acordo com a média de citações (eixo Y, vertical) e de menções (eixo X, horizontal). A média ponderada de todas as categorias (4,38 citações e 3,56 menções por artigo) representam o cruzamento dos eixos X e Y, e cada quadrante indica o comportamento padrão das categorias temáticas. Por exemplo: as categorias T10, T6 e T2 possuem baixo índice de citação e de menção, em relação à média; e T1, T3, T4, T8 e T9 obtiveram citações e menções acima da média. Curiosamente, não há categorias com citações abaixo da média e menções acima da média.

**Gráfico 2 - Dispersão das categorias temáticas (citações vs menções).**



Fonte: Dados da pesquisa.

A dispersão de pontos de dados no gráfico indica que a relação entre a alta citação e alta pontuação alométrica para os artigos de diferentes categorias temáticas é bem diversa. Os resultados por eixo temático apontaram uma baixa repercussão decorrente de citações e menções para artigos relacionados à atuação profissional e mercado de trabalho; à cultura, informação e sociedade; e a temas correlatos e outros.

Por sua vez, a temática relacionada aos *Estudos de usuário, demanda e uso da informação* se destaca por obter o melhor resultado em número de menções. A temática *Aspectos teóricos e gerais da ciência da informação* obteve a melhor média em número de citações. Ao calcular o coeficiente de Pearson para esses resultados, percebe-se uma correlação moderada ( $P = 0,65$ ), evidenciada pela disposição das categorias temáticas – predominantemente – orientadas de modo que temas com indicadores mais elevados de citações também tiveram indicadores mais elevados de menções.

Vale salientar que a correlação moderada entre citação e menção para as categorias temáticas ( $P = 0,65$ ) apresenta comportamentos diferentes quando analisados na Tabela 2 ( $P_{total} = 0,2306$ ). Isso pode ser explicado pela heterogeneia dos resultados citação *versus* menção em cada categoria – indicando propriedades/variáveis que também devem ser levadas em conta. Possivelmente, tais evidências reforçam a hipótese de que os indicadores tradicionais e altmétricos explicam impactos distintos (PRIEM; PIWOWAR; HEMMINGER, 2012).

No geral os resultados diferem dos encontrados na pesquisa de Boon e Foon (2014), na qual na aplicação do mesmo coeficiente de Pearson, ao indicar certa flutuação, apresenta uma fraca correlação. No entanto, há similaridade no fato de algumas temáticas ou áreas específicas apresentarem correlação mais moderada que outras.

## **5 Considerações finais**

O presente estudo possibilitou explorar um campo de estudos que ainda carece de futuros trabalhos para que seja possível compreender de maneira mais clara qual a contribuição dos indicadores altmétricos para as métricas tradicionais de impacto científico. As análises de correlação aqui realizadas serviram para reforçar hipóteses já lançadas na literatura (como correlações baixas e moderadas entre os dois indicadores), abrindo caminho para novas possibilidades de interpretação dos indicadores altmétricos.

O uso de técnicas estatísticas como o Coeficiente de Correlação de Pearson ainda parece ser pouco explorado na Ciência da Informação, sobretudo no Brasil, tendo em vista seu potencial para os estudos métricos da informação. A aplicação desse coeficiente pode ser uma estratégia importante para, por exemplo, identificar em um conjunto de indicadores quais são redundantes em relação ao fenômeno analisado (dois ou mais indicadores com correlação forte podem estar associados à influência mútua de propriedades em determinado fenômeno). Abre-se a possibilidade, também, de estabelecer modelos preditivos de fenômenos informacionais a partir de um indicador preexistente, além de auxiliar na explicação de relações de causalidade – causa e efeito entre variáveis.

A busca pela correlação entre as contagens de citação e os indicadores altmétricos é uma estratégia para validar estas novas medidas. No entanto, é preciso considerar a diversidade das fontes e métricas alternativas, dos usos que possibilitam, e das próprias comunidades científicas. É sabido que diferentes disciplinas apresentam padrões diferenciados de comunicação e de acúmulo de citações, e o mesmo pode ser verdadeiro também para a altmetria. A baixa correlação entre indicadores altmétricos e citações não significa, necessariamente, que a altmetria não tem validade, mas pode ser um indicativo de que estas novas métricas nos permitem observar impactos que não são capturados pelas métricas tradicionais. Nesse sentido, é possível que os estudos altmétricos sirvam de métricas para impactos alternativos da ciência, como a divulgação científica e o impacto social.

O presente trabalho tornou possível perceber que ainda é necessário se discutir muito a respeito dos estudos altmétricos. Algumas divergências entre os resultados aqui encontrados e a literatura referenciada trazem para o debate a possibilidade de vieses decorrentes dos instrumentos de coleta e das ferramentas de análise. O uso, por exemplo, do Coeficiente de Correlação de Postos de Spearman pode oferecer uma perspectiva complementar ao Coeficiente de Correlação de Pearson. Do mesmo modo, verificar as implicações do regime de informação em cada ambiente colaborativo na web (Facebook, Twitter etc.) e as diferenças entre ferramentas de coletas de dados (qualitativas e quantitativas) pode ser um campo fértil para futuros estudos altmétricos. Afinal, para que esse se

torne um campo reconhecido na comunidade científica, é preciso validar materiais e métodos capazes de representar a repercussão de pesquisas científicas na web, bem como seus desdobramentos sociais, científicos e tecnológicos.

## Referências

- ALPERIN, J. P. Exploring altmetrics in an emerging country context. In: **ALTMETRICS14: EXPANDING IMPACTS AND METRICS**, 2014, Bloomington. **Proceedings...** Bloomington, 2014. Disponível em: <<http://dx.doi.org/10.6084/m9.figshare.1041797>>. Acesso em: 28 jul. 2015.
- ARAÚJO, R. F. Mídias sociais e comunicação científica: análise altmétrica em artigos de periódicos da ciência da informação. **Em Questão**, Porto Alegre, v. 21, p. 96-109, 2015a.
- ARAÚJO, R. F. Da altmetria à análise de citações: uma análise da revista Datagramazero. **Datagramazero**, Rio de Janeiro, v. 16, n.1, fev., p. 1-20, 2015b.
- BARBETTA, P. A. **Estatística aplicada às Ciências Sociais**. 5. ed. Florianópolis: Editora da UFSC, 2003.
- BORNMANN, L. Alternative metrics in scientometrics: a meta-analysis of research into three altmetrics. **Scientometrics**, Dordrecht, v. 103, n. 3, p. 1123-1144, 2015.
- BRAGA, G. M. Informação, Ciência, Política Científica: o pensamento de Derek de Solla Price. **Ciência da Informação**, Brasília, v. 3, n. 2, p. 155-177, 1974. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/1634>>. Acesso em 29 set., 2012.
- COSTAS, R.; ZAHEDI, Z.; WOUTERS, P. Do “altmetrics” correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. **Journal of the Association for Information Science and Technology**, Hoboken, 28 jul. 2014. Disponível em: <<http://dx.doi.org/10.1002/asi.23309>>. Acesso em: 28 jul. 2015.
- BOON, C. Y.; FOON, J. W. J. Altmetrics is an indication of quality research or just hot topics. In: IATUL Conferences, 35th. Purdue University. Espoo, Filand, 2014. **Proceedings...** Espoo, Filand, 2014.
- MARQUES, F. Retuíte ou pereça. **Revista Pesquisa FAPESP**, São Paulo, n. 221, p. 46-47, jul. 2014. Disponível em: <<http://revistapesquisa.fapesp.br/wp->

<content/uploads/2014/07/pg046-047.pdf> >. Acesso em: 14 jul. 2014.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos, 1999. 268 p.

MÜLLER, L. **Facebook atinge 1,32 bilhão de usuários e 80% dos brasileiros na web**. TECMUNDO, 2014. Disponível em:

<<http://www.tecmundo.com.br/facebook/60937-facebook-chega-1-32-bilhao-usuarios-atinge-80-brasileiros-web.htm>>. Acesso em: 13 jul. 2015.

NASCIMENTO, A. G.; ODDONE, N. E. Uso de altmetrics para avaliação de periódicos científicos brasileiros em Ciência da Informação. **Ciência da Informação em Revista**, Maceió, v. 2, n. 1, 30 abr. 2015. Disponível em: <<http://www.seer.ufal.br/index.php/cir/article/view/1745>>. Acesso em 28 jul. 2015.

ODDONE, N.; GOMES, M. Y. F. S. F. Os temas de pesquisa em Ciência da Informação e suas implicações político-epistemológicas. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 5., 2004, Salvador. **Anais eletrônicos...** Salvador: Edufba, 2004.

PESTANA, M. H.; GAGEIRO, J. N. **Análise de dados para ciências sociais: a complementaridade do SPSS**. Lisboa: Silabo, 2014.

PRIEM, J.; HEMMINGER, B. H. Scientometrics 2.0: New metrics of scholarly impact on the social Web. **First Monday**, Illinois, v.15, n.5/7, jul., 2010. Disponível em: <<http://dx.doi.org/10.5210/fm.v15i7.2874>>. Acesso em 23 ago. 2014.

PRIEM, J.; TARABORELLI, D.; GROTH, P.; NEYLON, C.. **Altmetrics: a manifesto**. 2010. Disponível em: <<http://altmetrics.org/manifesto>>. Acesso em: 2 abr. 2014.

PRIEM, J.; PIWOWAR, H. A.; HEMMINGER, B. M. Altmetrics in the wild: using social media to explore scholarly impact. **CoRR**, New York, p. 1-17, 20 mar. 2012.

ROBINSON-GARCÍA, N.; TORRES-SALINAS, D.; ZAHEDI, Z.; COSTAS, R. New data, new possibilities: exploring the insides of Altmetric.com. **El Profesional de La Informacion**, Barcelona, v. 23, n. 4, p. 359–366, 2014. Disponível em: <<http://dx.doi.org/10.3145/epi.2014.jul.03>>. Acesso em: 2 mar. 2015.

SOLLA PRICE, D. J. Society's need in scientific and technical information. **Ciência da Informação**, Brasília, v. 3, n. 2, p. 97-103, 1974.

TARABORELLI, D. Soft peer review: social software and distributed scientific evaluation. In: INTERNATIONAL CONFERENCE ON THE DESIGN OF



COOPERATIVE SYSTEMS, 8th, Carry-le-Rouet, 2008. **Proceedings...** Paris, 2008. Disponível em: <<http://eprints.ucl.ac.uk/8279/>>. Acesso em: 28 jul. 2015.

THELWALL, M.; HAUSTEIN, S.; LARIVIÈRE, V.; SUGIMOTO, C. R. Do Altmetrics Work? Twitter and Ten Other Social Web Services. **PLoS ONE**, San Francisco, v. 8, n. 5, July 2013. Disponível em: <[doi: 10.1371/journal.pone.0064841](https://doi.org/10.1371/journal.pone.0064841)>. Acesso em 22 abr. 2014.

VIEIRA, S. **Introdução à bioestatística**. Rio de Janeiro: Elsevier, 2011. 345p.

## **Thematic orientation and correlation coefficient for comparison between altmetric data and citations: an analysis of the journal Datagramazero**

**Abstract:** The paper intends to verify the relationship between altmetrics and citation analysis using thematic orientation as a trace identifier for the works analysed, as well as Pearson's correlation coefficient. Therefore 441 articles from the journal DataGramaZero were categorized by a taxonomy featuring ten Information Science research themes. The category "Theoretical and general aspects of information science" was highlighted with more articles cited and a greater attention online. It was perceived that the two variables tend to behave similarly in relation to the time of publication, however, a low correlation between the number of mention and the number of citations for each article was noted. When the coefficients of the two variables weighted averages were calculated for thematic orientations, the correlation became moderate ( $P = 0,65$ ). The results were similar to previous findings in the literature, reinforcing the potential of the Pearson's correlation coefficient for metric studies in information science.

**Keywords:** Altmetrics. Citation. Correlation Coefficient.

Recebido: 28/01/2016

Aceito: 21/06/2016