

# IMPACTO E VISIBILIDADE DE PUBLICAÇÕES SOBRE WEB SEMÂNTICA: dados de citação e de atenção *online*

## ACADEMIC IMPACT AND VISIBILITY OF RESEARCH ON SEMANTIC WEB: analyzing citations and online attention

Ronaldo Ferreira de Araújo<sup>1</sup>  
José Eduardo Santarem Segundo<sup>2</sup>  
Crispulo Travieso-Rodríguez<sup>3</sup>  
Gustavo Miranda Caran<sup>4</sup>

### RESUMO

O artigo discorre sobre o campo da Web Semântica e analisa a produção e o impacto de suas publicações a partir de seus indicadores bibliométricos de citação e altmétricos de atenção *online* e a correlação entre eles. Foram considerados dados de 5,000 pesquisas publicadas entre 2014 e 2019 recuperados na base de dados Dimensions. A principal área de estudo das publicações é *Information and Computing Sciences* concentrando 79% das pesquisas e o maior agrupamento em co-ocorrência de termos indica predominância de discussões relacionadas a estrutura fundamental de desenvolvimento da Web Semântica. Quanto ao desempenho, o conjunto das pesquisas obteve um impacto acadêmico de 14.945 citações e de atenção *online* de 3.929 menções em fontes da web social. O coeficiente de correlação aplicado entre os indicadores de altmetria e de citações demonstrou relações predominantemente fracas para a maioria das variáveis, com exceção do Mendeley ( $p = 0,73$ ). O estudo destaca a complementaridade entre técnicas bibliométricas e altmétricas e reforça a necessidade de abordagens qualitativas que evidenciem de forma mais adequadas às suas correlações.

**Palavras-chave:** Web semântica. Bibliometria. Altmetria. Análise de citação. Atenção *online*.

### ABSTRACT

This study is focused on Semantic Web and based on the analysis of production and impact of published research, according to bibliometric indicators of citation, altmetrics indicators of online attention and their correlation. Using data from 5.000 scientific papers by Dimensions database published from 2014 to 2019, it was found that the main subject area was *Information and Computing Science*, which gathered 79% of the total sample. Attending to co-occurrence of terms, the principal cluster identified proves the relevance of discussion topics on the structure for developing the semantic web. As a whole, it was counted 14.945 cites regarded to academic impact, and 3.929 mentions in social media regarding online attention. Correlation coefficient referred to citation and altmetrics showed weak linear relationships for most of the variables, except for Mendeley ( $p$ -value = 0,73). Finally, it is underlined the need to combine both bibliometrics and altmetrics methods and the convenience of carrying out qualitative approaches that explain their mutual relationships.

**Keywords:** Semantic Web. Bibliometrics. Altmetrics. Citation analysis. Online attention.

*Artigo submetido em 27/02/2021 e aceito para publicação em 02/03/2021.*

- 1 Pesquisador-doutor do Laboratório de Estudos Métricos da Informação na Web (Lab-iMetrics) da Universidade Federal de Alagoas, Brasil. ORCID <https://orcid.org/0000-0003-0778-9561>. E-mail: ronaldfa@gmail.com
- 2 Docente permanente no Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Julio de Mesquita Filho, Brasil. ORCID <https://orcid.org/0000-0003-3360-7872>. E-mail: santarem@usp.br
- 3 Docente no Departamento de Biblioteconomía y Documentación da Universidade de Salamanca, Espanha. ORCID Departamento de Biblioteconomía y Documentación da Universidade de Salamanca, Espanha. E-mail: ctravieso@usal.es
- 4 Doutor em Ciência da Informação pelo convênio Instituto Brasileiro de Informação em Ciência e Tecnologia e Universidade Federal do Rio de Janeiro, Brasil. ORCID <https://orcid.org/0000-0002-1199-5002>. E-mail: gmcaran@gmail.com

## 1 INTRODUÇÃO

O conceito de Web Semântica tem sido um dos pontos de inflexão nos estudos sobre Organização do Conhecimento, na medida em que motivou o repensar de muitos pressupostos básicos dessa disciplina documental. Desde o seu surgimento na prática e na bibliografia especializada, hoje em dia é impossível separar as formas de classificar o conteúdo dos documentos e sua posterior recuperação, independentemente do suporte, das características técnicas que o sustentam e representam, ou seja, sua conexão com os postulados teóricos das Ciências Documentárias é clara (ROBREDO, 2010; MACHADO et al, 2017).

Mas depois de alguns anos de presença ininterrupta em debates e pesquisas, fica evidente que o conceito da Web Semântica continua evoluindo, com a necessária gestão, interoperabilidade e reutilização dos dados da web (PASTOR-SÁNCHEZ, 2016).

Por relacionar fortemente questões tecnológicas nas estruturas de representação da informação e organização do conhecimento, a Web Semântica tem sido impactada por uma grande quantidade de estudos e pesquisas na Ciência da Informação (CI). E uma das mais utilizadas formas de demonstrar o impacto de um conceito dentro de uma disciplina científica, que também tem sido tradição nos estudos da CI, é a realização de estudos bibliométricos, capazes de relacionar as tendências da pesquisa e sua evolução com base em resultados publicados.

Além da aplicação da Bibliometria para desenvolver *rankings* quantitativos que enfatizem o papel de autores e instituições, um de seus principais usos é detectar mudanças no ritmo e no impacto dos temas estudados, que no final facilita a caracterização do panorama de pesquisa de uma disciplina e a detecção das abordagens mais utilizadas (REISS et al, 2013). É, portanto, uma ferramenta muito poderosa para identificar áreas emergentes e analisar tendências. Sua aplicação na melhoria e adaptação de políticas científicas, também no nível de sua gestão, está sendo amplamente adotada.

No entanto, os estudos bibliométricos que abordaram diretamente a evolução das pesquisas sobre Web Semântica ainda são relativamente escassos. Isto responde tanto ao grau de especificidade do tema em questão ou, por outro lado, deriva de aspectos da sua transversalidade, uma vez que abordagem para sua análise varia consideravelmente dependendo da disciplina que a estuda. Em um dos primeiros estudos conduzidos sob a ótica bibliométrica, com base da literatura produzida nas bases *Web of Science* e *Scopus*, Ding (2010) além de destacar as revistas, autores e artigos mais relevantes, enfatiza a questão de sua transversalidade temática e estimula a comunidade científica a

fortalecer a colaboração entre disciplinas afins para avançar no conhecimento sobre a Web Semântica, frente aos desafios futuros.

Entre a transversalidade e a influência disciplinar St-Germain e Mongeon (2018) analisam a Web Semântica a partir de cinco agrupamentos de áreas científicas (Engenharia e Tecnologia, Saúde e Ciências da Vida, Ciências da Informação, Ciências Naturais e Ciências Sociais e Humanas). Os autores destacam expressamente a presença da Web Semântica nas pesquisas em Ciências da Informação, lembrando que o interesse dessa disciplina por este objeto de estudo começa por volta de 2008, um pouco mais tarde do que em outras disciplinas tecnológicas.

Outros estudos analisam a produção científica sobre a Web Semântica em contextos específicos como no estudo de Zhang, Hua e Yuan (2018), que o inclui dentro do campo dos dados abertos ou de Figueiredo e Almeida (2017) que o faz com ênfase nos estudos de ontologias. Por fim, vale citar outros estudos que, a partir da avaliação científica, também contêm análises sobre Web Semântica ou priorizando algum aspecto específico, seja temático como *Linked data* e web de dados (GANDON, 2018) e de colaboração científica e gênero entre os autores que pesquisam o tema (CHAMBERS; MILOJEVIC; DING, 2014).

Atualmente, a evolução da comunicação científica e importância dada à sua divulgação nos mais diversos meios de comunicação, têm evidenciado a necessidade de se tratar de forma complementar os indicadores bibliométricos com outros tipos de medições realizadas a partir da influência da informação científica nas mídias sociais e os traços de sua mediação e interação entre pesquisadores (TRAVIESO; ARAÚJO, 2018). Nesse sentido, análises bibliométricas de citação têm sido complementadas com dados altmétricos de menções e outras atividades de mídias sociais, ampliando o quadro avaliativo dos estudos de comunicação científica que buscam avaliação de resultados de pesquisa (ARAÚJO; CARAN; SOUZA, 2016).

Promovida e mantida por possibilidades tecnológicas das redes digitais e uma demanda crescente para demonstrar impacto além da comunidade científica, as altmetrias têm recebido grande atenção como “[...] democratizantes potenciais do sistema de recompensa científica e indicadores de impacto social” (HAUSTEIN, 2016, p.413) com potencial de oferecer uma visão mais rica da circulação e do impacto das pesquisas (NEYLON, 2015).

O presente estudo, dedicado a publicações sobre Web Semântica, é desenvolvido nessa linha de análise e conduzido a partir de uma ótica integradora da aplicação dos estudos bibliométricos de citação, que costumam indicar o impacto acadêmico, aos altmétricos, com apelo ao impacto social e

de visibilidade (HAMMARFELT, 2014; TRAVIESO; ARAÚJO, 2018) considerando os propósitos distintos e as particularidades de cada um.

## 2 WEB SEMÂNTICA

A Web Semântica, proposta por Tim Berners-Lee, James Hendler e Ora Lassila, em artigo publicado na *Scientific American* em 2001 propõe, essencialmente, que se estruture os dados da Web de forma que eles possam ter significado e principalmente que se tornem passíveis de interpretação por máquinas, através de agentes computacionais.

O termo “semântica” encontra justificação terminológica no verbo grego *semaíno* (transliteração de *σημαίνω*), introduzido por Aristóteles para denotar a função específica do signo linguístico, remetendo à noção de “significar algo” a partir de uma linguagem estruturada (ABBAGNANO, 2007). Considerando que a semântica é o estudo do significado, denota-se que constituir ligações semânticas é relacionar termos a conceitos, dessa forma a Web Semântica propõe justamente isso, estabelecer o significado entre recursos que se interligam na Web, dando a eles uma estrutura de domínio sobre um tema.

Para disponibilizar dados numa estrutura semântica é necessário pensar em partes do modelo descrito por Berners-Lee em 2001, um conjunto de conceitos e especificações, estruturados em formas de camadas que apresenta a Web Semântica.

O conjunto de conceitos e tecnologias que estruturam a Web Semântica tem passado por evoluções significativas desde 2001, quando a ideia inicial foi apresentada. Há uma percepção clara que o modelo ainda é baseado conceitualmente no *Resource Description Framework* - RDF (também chamado de linguagem), que deve ser representado computacionalmente por uma linguagem computacional, como foi por algum tempo o *eXtensible Markup Language* (XML) mas posteriormente substituído por outras ferramentas capazes de executar o mesmo papel como json, n3, turtle, entre outros.

O RDF é uma estrutura de uso geral para representar informações na Web. O RDF tem como princípio fornecer interoperabilidade aos dados, contribuindo com a representação e posterior recuperação de informações. Um dos principais objetivos do RDF é criar uma rede de recursos semanticamente relacionados, a partir de dados distribuídos.

A partir do fenômeno da conexão com significado, proposta pelo RDF, caracterizam-se modelos mais estruturados e complexos que são as ontologias. São elas os principais elementos da Web

Semântica na construção de informações relacionadas que apresentem significado. Santarem Segundo (2015, p. 226) afirma que:

Utilizar ontologias é uma das maneiras de se construir uma relação organizada entre termos dentro de um domínio, favorecendo a possibilidade de contextualizar os dados, tornando mais eficiente e facilitando o processo de interpretação dos dados pelas ferramentas de recuperação da informação.

Ao abordar a Web Semântica, trata-se de uma mistura de interoperabilidade; padronização, organização e reuso da informação; inferências e de serendipidade. Como já fora observado, um dos elementos que compõe, ou está diretamente ligado às ontologias é a construção de axiomas que permitem a inferência. A inferência é citada em grande parte das pesquisas que tratam de Web Semântica como sendo o grande diferencial na construção de ambientes semânticos.

As regras que possibilitam a inferência dão mais qualidade à análise dos dados, visto que permitem conhecer novas relações dentro de um conteúdo e em alguns casos também permitem detectar possíveis inconsistências. Berners-Lee, Hendler e Lassila (2001, tradução nossa) dizem que “[...] para a Web Semântica funcionar, os computadores devem ter acesso a coleções estruturadas de informações e conjuntos de regras de inferência, que eles podem usar para conduzir o raciocínio automatizado”.

A serendipidade está relacionada a descobertas feitas ao acaso, sendo percebida em como as tecnologias da Web Semântica, e sua materialização, os dados ligados (*Linked Data*) são capazes de trazer à tona ou mesmo possibilitar a ligação semântica entre dados de fontes diversas espalhadas pelo mundo.

Desde 2001 até o presente momento, as tecnologias e conceitos da Web Semântica deixaram de ser utopia para virar realidade, entretanto muitos estudos ainda continuam sendo realizados. As principais *big techs* do mundo utilizam os conceitos da Web Semântica em parte de seus produtos tecnológicos. A todo esse contexto está cada vez mais latente a questão de disponibilizar e consumir dados enriquecidos semanticamente.

Em seu desenvolvimento, pesquisas buscaram contribuir com a reflexão da Web Semântica como objeto de estudo analisando a construção do conhecimento científico sobre temática e as disciplinas dedicadas em sua consolidação (PINHEIRO, 2008). E para outros, mais do que um tema ou

assunto de pesquisa, a Web Semântica se estabeleceu como um campo de pesquisa por direito próprio (HITZLER; JANOWICZ, 2011).

Essa perspectiva se fundamenta, por exemplo, na concepção de Hitzler e Janowicz (2011) por aspectos como: a conformação multidisciplinar e heterogênea da Web Semântica, área cujos estudiosos mantêm laços estreitos com disciplinas vizinhas que fornecem métodos ou áreas de aplicação para seu trabalho; o aumento de pesquisadores que se dedicam ao tema, suas especialidades e especificidades de atuação; o crescimento de eventos sobre a temática; bem como, o surgimento de revistas especializadas com seu enfoque.

### 3 MATERIAL E MÉTODO

A presente pesquisa se caracteriza como um estudo quantitativo descritivo de abordagem bibliométrica e altmétrica, que buscou compreender o desempenho da produção científica sobre Web Semântica. Os estudos bibliométricos de citação e os altmétricos mensuram impactos distintos, mas que se relacionam, e se complementam. Ao serem trabalhados em conjunto tais métricas podem descrever um quadro mais completo da comunicação científica (ARAÚJO, 2015) e do impacto da pesquisa científica, considerando tanto o impacto acadêmico, geralmente medido pelos tradicionais estudos de citação, quanto o possível impacto social, que vem sendo atribuído às métricas alternativas e de atenção *online*, capazes de medir a influência e visibilidade da pesquisa em fontes da web social.

A produção científica sobre Web Semântica foi coletada por consultas manuais pela expressão “*Semantic Web*” buscada junto ao título e resumo das publicações de 2014 a 2019 de artigos, capítulos de livros e anais de eventos presentes na base de dados *Dimensions* (<https://www.dimensions.ai>). Os dados da produção (número do *Digital Object Identifier* - DOI, autor, título, tipo de publicação, ano e citação) foram exportados em formato CSV. A lista de DOI de cada publicação foi importada para o *Altmetric Explorer* para análise dos seus dados altmétricos.

A *Dimensions* é uma plataforma criada pela *Digital Science* em 2018 com foco em dados de publicações científicas bem similar ao *Web of Science* (WoS, Clarivate Analytics) e *Scopus* (Elsevier), tendo a cobertura substancialmente maior que estas (BodeE et al, 2018; VISSER; JAN VAN ECK; WALTMAN, 2020). A base registra mais títulos de periódicos e representatividade de países e idiomas com dados de publicações oriundos da CrossRef, agência oficial de registro do

DOI, indicando também maior cobertura de dados de citações quando comparada com a *Web of Science* e *Scopus* (VISSER; JAN VAN ECK; WALTMAN, 2020). A base inclui dados adicionais, como subvenções e ensaios clínicos e conta com recurso de esquema de classificação de campo, que não é baseado em classificações de periódicos, como em WoS ou Scopus, mas em aprendizado de máquina (BORNMANN, 2018).

O *Altmetric Explorer*, por sua vez, é um serviço fornecido pela Altmetric.com (<http://www.altmetric.com>), que reúne dados de mídias sociais, como Twitter e Facebook, de gerenciadores de referência on-line como Mendeley, além de menções em postagens de *blogs* e em matérias de mídia *online* e portais de notícias. As fontes utilizadas pelo *Altmetric Explorer* contam como uma indicação de “interesse público” e seus dados são calculados posteriormente para aferição de seu “valor altmétrico” geral ou *Altmetric Score* (HAMMARFELT, 2014).

Ao todo foram identificadas 5,270 publicações e consideradas as 5 mil mais relevantes (*Sort by Relevance*). A classificação de relevância dos resultados da pesquisa na base considera maior precisão na busca com cuidados relacionados à grafia, contexto e maior completude dos dados da publicação. O conjunto dessas publicações então é descrito segundo sua distribuição por ano, tipo, autoria, revista, categorias e agrupamento temático. Para este último, os dados foram submetidos ao algoritmo do *software VosViewer*.

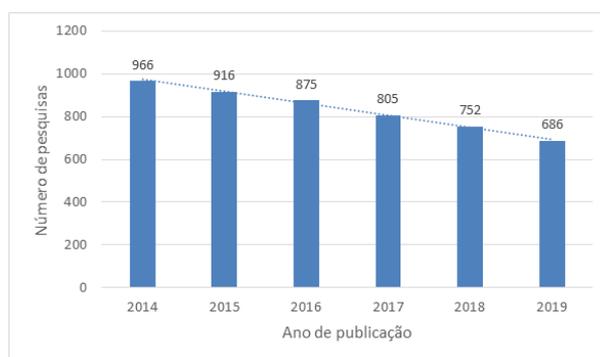
Por fim, os trabalhos são analisados segundo seu impacto acadêmico, a partir das citações recebidas e impacto social, considerando a visibilidade e atenção *online* atingida pela circulação em fontes da web social, bem como a relação desses impactos por meio da aplicação do Coeficiente de Correlação de Pearson (CCP), realizado pela fórmula CORREL do *Microsoft Excel 365*, representado pela expressão matemática (MICROSOFT, 2019):

$$\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$$

## 4 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Ao todo foram analisados os dados de 5,000 pesquisas publicadas entre 2014 e 2019, com média de 833 artigos por ano. A distribuição desses estudos por data de publicação pode ser visualizada no Gráfico 1 com indicação dos anos e o quantitativo de pesquisas.

**Gráfico 1.** Distribuição das publicações sobre Web Semântica por ano (2014-2019)

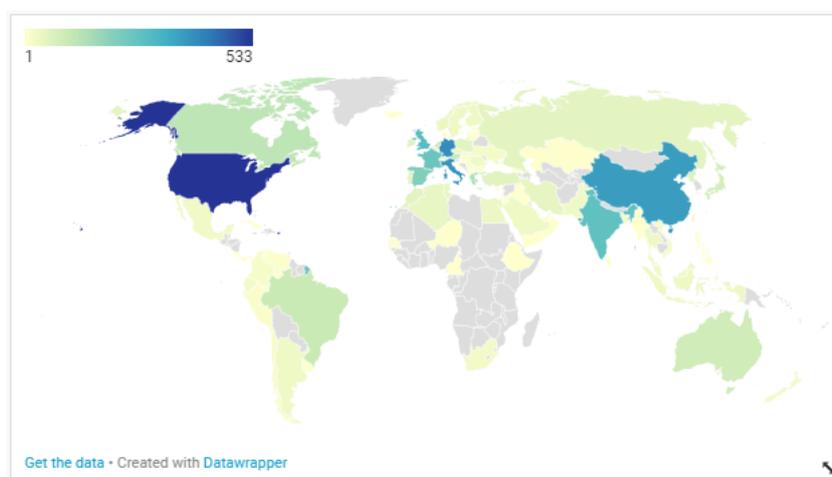


**Fonte:** Dados da pesquisa (2020)

O ano de maior número registrado foi 2014 com 966 pesquisas e o com menor foi 2019 com 686. Logo a partir do primeiro ano considerado para análise percebe-se pela linha de tendência que houve um decréscimo de publicações durante o período analisado com uma média de queda anual de 5,2%. Essa queda pode não representar necessariamente que se produz menos sobre Web Semântica, mas refletir questões sobre a transversalidade e as dimensões interdisciplinares do tema como apontado na literatura (ST-GERMAIN; MONGEON, 2018; ZHANG; HUA; YUAN, 2018; GANDON, 2018) o que por sua vez, pode resultar em variações terminológicas.

Apenas 9,7% das publicações sobre Web Semântica foram escritas em autoria única e 90,3% em co-autoria indicando que publicações sobre o tema no período analisado se dão majoritariamente por algum tipo de colaboração científica com mais de 60% dos trabalhos na faixa entre 2 a 4 autores. Foram registrados mais de 90 países que produziram sobre o tema, a Figura 1 indica a distribuição dessa produção, levando em consideração a organização de vínculo do primeiro autor.

**Figura 1.** Mapa de distribuição das publicações sobre Web Semântica por países (2014-2019)



**Fonte:** Dados da pesquisa (2020)

No total, dez países concentram 2.986 pesquisas, o que corresponde a 63,63%. Desses, o principal é os Estados Unidos com 533 trabalhos, seguido da Alemanha com 399, Itália com 387, China com 368 e Reino Unido com 294, sendo respectivamente 11,36%, 8,50%, 8,25%, 7,84% e 6,26%. O resultado é um pouco similar ao de outras pesquisas que analisaram temas ligados à tecnologia como de publicações sobre inteligência artificial (GONTIJO, 2021). Ainda sobre os países, destaca-se que o Brasil é o único da América Latina entre os registram mais de 100 pesquisas.

A tipologia das publicações, por sua vez, indicou que pesquisas sobre Web Semântica tem uma penetração na literatura para tipos distintos de canais de comunicação científica e com certo equilíbrio entre eles. A maioria são de artigos de periódicos, com 2.181, seguido de trabalhos publicados em anais de evento, com 1.507 e capítulos de livro, com 1.312, representando respectivamente 43,62%, 30,14% e 26,24%.

Os 2.181 artigos identificados foram publicados por um total de 882 periódicos, o que demonstra uma grande variedade de fontes que publicam sobre o tema. Apenas 13 (1,47%) destes publicam 15 ou mais artigos, deixando dois grandes grupos com extratos menores, sendo 41 (4,65%) com 7 a 14 artigos e 828 (93,8%) com 1 a 6 artigos. A Tabela 1 apresenta os principais periódicos e o quantitativo de pesquisas publicadas por eles.

**Tabela 1.** Artigos sobre Web Semântica por periódicos (2014-2019)

#	Revistas	Artigos	(%)
1	Procedia Computer Science	74	3,39
2	Semantic Web	64	2,93
3	Journal of Web Semantics	41	1,88
4	International Journal of Computer Applications	31	1,42
5	Journal of Biomedical Semantics	30	1,38
6	Assistive technology research series	29	1,33
7	International Journal on Semantic Web and Information Systems	20	0,92
8	Automation in Construction	18	0,83
9	Knowledge-Based Systems	18	0,83
10	Applied Mechanics and Materials	16	0,73
11	Journal of Information Science	16	0,73
12	Cataloging & Classification Quarterly	15	0,69
13	Future Generation Computer Systems	15	0,69
14	Não identificado	30	1,38
15	Revistas com 7 a 14 artigos (n=41)	391	17,93
16	Revistas com 1 a 6 artigos (n= 828)	1373	62,95
<b>Total</b>		<b>2181</b>	<b>100</b>

**Fontes:** Dados da pesquisa (2020)

Os resultados apontam para um núcleo pequeno de periódicos que tratam o tema de forma mais abrangente e extensa, enquanto núcleos periféricos com diversos periódicos em que a produtividade de publicação de artigos é reduzida, o que se assemelha a lei de dispersão quando se analisa a literatura periódica em uma área específica (SÁ; LEITÃO, 2018). Importante ressaltar que do maior núcleo, com 828 revistas, 542 (24,85%) registram apenas um único artigo no período analisado. É neste mesmo núcleo que se encontra títulos brasileiros como *Brazilian Journal of Information Science*, *Brazilian Archives of Biology and Technology*, *Perspectivas em Ciência da Informação* e *Transinformação*.

Do núcleo menor, com maior concentração de artigos, vale destacar que três dos principais periódicos levam “web semântica no nome” como é o caso das revistas *Semantic Web*, *Journal of Web Semantics* e *International Journal on Semantic Web and Information Systems*. O surgimento destas revistas tem sido apontado na literatura como parte do movimento de consolidação da Web Semântica como campo de pesquisa próprio (HITZLER; JANOWICZ, 2011).

A forte relação com a Ciência da Computação e Ciência da Informação pode ser percebida quando se analisa o foco e escopo dos principais periódicos listados. A base *Dimensions* aferiu 69 categorias únicas do campo de pesquisa para as 5.000 publicações sobre Web Semântica analisadas. Embora algumas publicações possuam mais de uma categoria (31,4% com duas; 5,2% com três; e 1% com quatro), a maioria delas estão classificadas em apenas uma categoria (68,5%). Para representação, consta na Tabela 2 a distribuição das publicações por categorias, sendo consideradas, para as que possuem mais de uma, apenas a principal.

**Tabela 2.** Estudos sobre Web Semântica por categorias de pesquisa (2014-2019)

<b>Campos de pesquisa</b>	<b>Publicações</b>	<b>(%)</b>
Information and Computing Sciences	3954	79,08
Engineering	148	2,96
Linguistics	112	2,24
Public Health and Health Services	40	0,80
Specialist Studies In Education	36	0,72
Medical and Health Sciences	30	0,60
Psychology and Cognitive Sciences	22	0,44
Mathematical Sciences	14	0,28
Applied Mathematics	14	0,28
Civil Engineering	14	0,28
Sociology	10	0,20
Cultural Studies	10	0,20
Campos de pesquisa com 1 a 9 estudos (n=57)	189	3,78
Campos não identificados	4	0,08
<b>Total Geral</b>	<b>5000</b>	<b>100</b>

**Fontes:** Dados da pesquisa (2020)

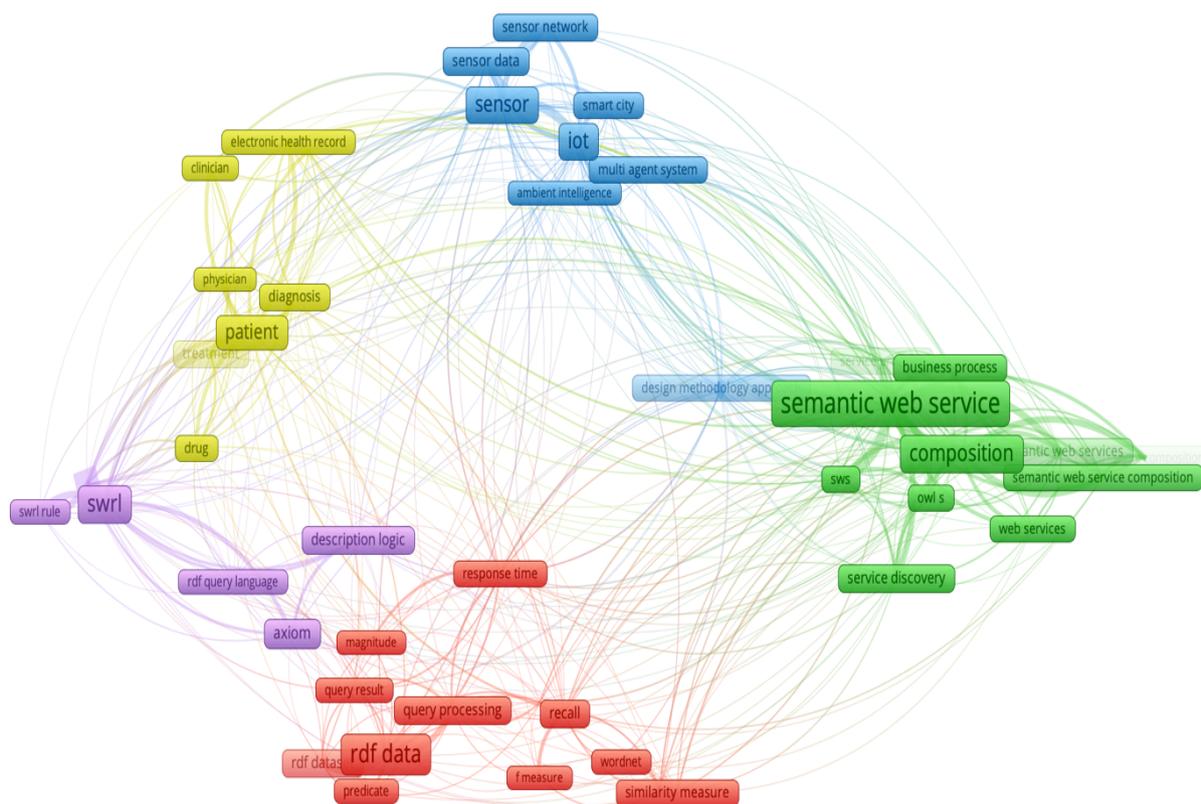
Embora Web Semântica seja um campo multidisciplinar e transversal, a Ciência da Computação e a Ciência da Informação costumam ser as áreas de conhecimento consideradas de maior proeminência em pesquisas e produção científica dessa temática (DING, 2010; ST-GERMAIN; MONGEON, 2018). Os resultados obtidos confirmam isso uma vez que a principal categoria de pesquisa das publicações sobre Web Semântica é *Information and Computing Sciences* com 3.954 trabalhos (79%). Vale ressaltar que todas as tecnologias desenvolvidas no âmbito da Web Semântica “trazem cada vez mais componentes que a situam como tema de fronteira entre a Ciência da Computação e a Ciência da Informação interligando instrumentos, ferramentas e conceitos presentes nas pesquisas de ambas as matrizes disciplinares” (PINHEIRO, 2008, p.42).

Em seguida, com uma boa distância temos campos como *Engineering e Linguistics* com 148 (2,96%) e 112 (2,24%) publicações respectivamente. Embora em menor proporção percebe-se outras áreas representadas como *Public Health and Health Services e Specialist Studies In Education* com 40 (0,8%) e 36 (0,72%) e nas últimas posições *Sociology e Cultural Studies*, cada uma delas com 10 pesquisas (0,2%). As categorias apresentadas também refletem a transversalidade e a influência disciplinar da Web Semântica inclusive se assemelha aos cinco agrupamentos de áreas científicas percebidas por St-Germain e Mongeon (2018) como Engenharias e Tecnologia, Saúde e Ciências da Vida, Ciências da Informação, Ciências Naturais e Ciências Sociais e Humanas.

Assim ao analisar a distribuição das publicações por esses campos, levando em considerações as conformações das contribuições disciplinares que a Web Semântica possui (DING, 2010; ST-GERMAIN; MONGEON, 2018) podemos considerar que as áreas das primeiras posições (Ciências da Informação e Computação, Engenharia e Linguística) desempenham importante papel nas discussões de suas bases teóricas de fundamentação e implementação, e que as demais encontram-se no domínio de sua aplicação ou experimento.

A análise de conteúdo contribui para compreensão da distribuição das pesquisas pelas categorias e reforça as temáticas dos trabalhos como demonstrado na Figura 2. O agrupamento temático foi alcançado por meio do uso do *software* VosViewer, usando os termos do título e do resumo das publicações, com método de contagem binário de palavras, sem apoio de um tesauro, estabelecendo a quantidade mínima de ocorrência dos termos em ( $n > 25$ ) e limitando a quantidade de termos em 50 para a apresentação gráfica. Foram identificados cinco agrupamentos (*clusters*) principais.

**Figura 2.** Agrupamento temático das pesquisas sobre Web Semântica (2014-2019)



**Fonte:** Dados de pesquisa (2020)

Por meio da Figura 2 é possível notar a formação de 5 grupos temáticos, representados pelas cores vermelha, verde, azul, roxa e amarelo. No contexto da Web Semântica, os grupos formados demonstram, claramente, a relação entre os trabalhos e principalmente o foco e evolução dos estudos na área, além da consolidação no efetivo uso.

Ao analisar o agrupamento posicionado na parte inferior da Figura 2, representado pelo *cluster* de cor vermelha, é possível identificar termos como “rdf data”, “rdf dataset”, “query processing”, “recall” entre outros. Esse agrupamento é o maior em quantidade de termos, são quatorze dos cinquenta, com significativa relevância para o contexto. Os termos mais citados “rdf data” e “rdf dataset” juntos foram citados 218 vezes.

Os principais termos nos remetem a estrutura fundamental de desenvolvimento da Web Semântica, onde se organiza estruturalmente qualquer proposição de desenvolvimento sobre o tema. Os termos “rdf data”, “rdf dataset”, “query processing”, “recall”, “rdf dataset”, entre outros que compõem esse agrupamento, caracterizam a necessidade sempre presente da discussão de como estruturar o dado semântico e também de como recuperá-lo.

Esse agrupamento (vermelho) demonstra que o *framework* RDF continua sendo o principal conceito para estruturação de dados, e que, mesmo após 20 anos da proposição da Web Semântica, continua ocupando papel fundamental no desenvolvimento, evolução e uso da mesma.

O segundo maior agrupamento, observado no *cluster* de cor verde, é formado por doze termos, e torna-se muito importante destacar aqui que os termos “semantic web service” e “semantic web services” não foram unidos entre eles (por serem singular e plural), entretanto, juntos, são responsáveis por 253 ocorrências, representando dentre todos os termos, o mais citado.

Esse segundo agrupamento é caracterizado fortemente por pesquisas que englobam estrutura computacional para que os dados, em formato semântico, possam ser consumidos. Termos como “semantic web services”, “service discovery” e “service provider” estão muito relacionados a definição de ambientes computacionais que permitam, principalmente, o acesso ao dado já estruturado e enriquecido semanticamente. O agrupamento, dessa forma, demonstra o uso efetivo da Web Semântica, a materialização da proposição de Berners-Lee em 2001, visto que quando tratam-se de disponibilizar dados para consumo, compreende-se que a Web Semântica já está acontecendo.

O terceiro agrupamento temático representado no *cluster* em azul, na parte superior da Figura 2, nos apresenta o contexto de uso da Web Semântica, mas principalmente a demonstração de que a Web Semântica tem sido fundamental para o desenvolvimento das cidades inteligentes e também tem se relacionado muito bem com estudos que tratam de análise de dados e principalmente de inteligência artificial. É possível dizer que esse agrupamento representa muito do que há por vir sobre Web Semântica.

O termo “iot” com 133 ocorrências e o termo “sensor” com 125 ocorrências, além da grande quantidade de ocorrências de termos ligados a “sensor”, e também o aparecimento do termo “smart city”, caracterizam fortemente a grande influência da internet das coisas, dos dados oriundos de sensores, da geração constante de dados e da pervasividade intrínseca à coleta e posterior enriquecimento semântico dos dados. Os termos “iot” e “smart city” representam os termos mais citados nos artigos mais recentes, favorecendo a compreensão de que os estudos sobre Web Semântica pode estar sendo forte influenciador no desenvolvimento das cidades inteligentes.

O quarto agrupamento, representado no *cluster* em amarelo demonstra o grande uso da Web Semântica na área da saúde. É notório que o domínio que mais tem se apropriado dos conceitos, técnicas e ferramentas da Web Semântica é a área de saúde. Os termos “paciente”, “diagnosis” e “treatment” são os termos de maior ocorrência nesse agrupamento, que apesar de contar com apenas

8 termos, indica claramente como a área tem se apropriado das relações semânticas para evoluir no enriquecimento de dados relacionados a diagnósticos e tratamento como forma de evolução da ciência.

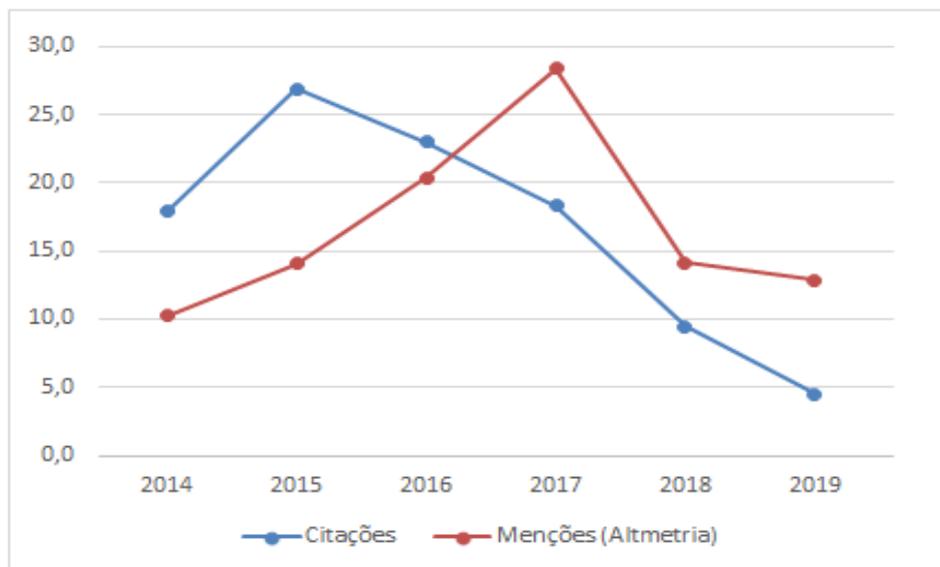
A área de saúde tem se caracterizado como a que mais produz metodologias e principalmente artefatos e repositórios de dados enriquecidos com características semânticas, dessa forma, esse agrupamento vem confirmar que mesmo em uma busca baseada em termos relacionados puramente à Web Semântica, um domínio foi suficientemente capaz de se destacar como um dos agrupamentos temáticos formalizados. Vale ressaltar ainda que a Saúde está entre os campos de pesquisa com mais concentração das publicações analisadas conforme Tabela 2, estão presentes nos valores registrados na *Public Health and Health Services* e *Medical and Health Sciences*.

Não mais importante, o quinto agrupamento temático, representado pelo *cluster* de cor roxa, e posicionado mais à esquerda na Figura 2, completa o conjunto de aproximações temáticas com termos como “wsrl” e “semantic web rule language” (um é sigla do outro) e ainda o termo “axiom”. Esse agrupamento representa o desenvolvimento e estudos relacionados às questões que permitem a inferência com dados da Web Semântica. Os axiomas e a linguagem de desenvolvimento de inferências podem ser considerados temas que foram conclamados por Berners-Lee em 2001 como os resultados a serem obtidos por meio da Web Semântica. Assim, esse agrupamento caracteriza fortemente os estudos de inferência e atrelado a isso a possibilidade da serendipidade, outro propósito da Web Semântica.

Conforme a literatura, a serendipidade se refere a descobertas não planejadas, condizendo com a capacidade que as tecnologias da Web Semântica e principalmente o *Linked Data* trazem à tona, bem como as conexões que possibilitam através da ligação semântica entre dados de fontes diversas espalhadas ao redor do mundo. Enquanto a inferência diz respeito à capacidade de se deduzir ou tomar decisões, baseadas na consolidação de uma verdade de uma proposição que não é conhecida, mas é tida a partir de sua relação direta com outras verdades existentes, podendo ser considerada uma das principais buscas na implementação da Web Semântica (SANTAREM SEGUNDO; CONEGLIAN, 2016).

O total de publicações analisadas apresenta impacto acadêmico de 14.945 citações e visibilidade e atenção *online* de 3.929 menções em fontes da web social, com média de 16,7 e 16,5, respectivamente. No geral, esses resultados evidenciam um volume médio de indicadores de citação bem próximo aos de altmetria e sua distribuição média ao longo do tempo (Gráfico 2) revela, mesmo com certa flutuação, que publicações mais antigas (2014 a 2016) recebem mais citações e que publicações mais recentes (2017 a 2019) recebem mais dados altmétricos.

**Gráfico 2.** Média de citações e menções *online* ao longo do tempo

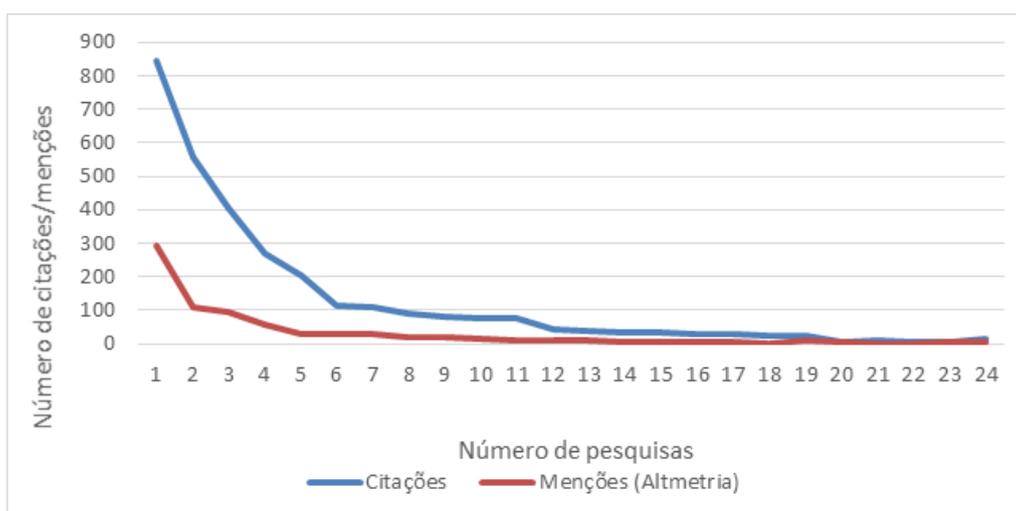


**Fonte:** Dados de pesquisa (2020)

Vistas individualmente, o desempenho dos indicadores parece um pouco incomum em alguns momentos, o que pode ser observado para as citações ter o menor desempenho da série no seu primeiro ano (2014) e para altmetria, a redução do seu desempenho nos dois últimos anos (2018 e 2019) embora ainda superior ao de citação. Por outro lado, o fato de publicações mais antigas terem mais citações e publicações mais recentes terem mais valores altmétricos é reconhecido como particularidades dessas métricas, sendo recorrente em estudos similares (ARAÚJO; CARAN; SOUZA, 2016; TRAVIESO; ARAÚJO, 2018).

O Gráfico 3 apresenta a distribuição dos documentos de acordo com as suas quantidades de citações e menções. Para fins de ilustração, o gráfico traz o intervalo de distribuição mais representativo, de 0 até 24 citações (n=4842; 96,8%) ou menções (n=4967; 99,3%).

**Gráfico 3.** Distribuição da quantidade de artigos por nº de citações e menções



**Fonte:** Dados de pesquisa (2020)

A distribuição dos artigos de acordo com as frequências de citação e menção se concentram nos intervalos inferiores, em uma curva senoide descendente, sem a apresentação de picos em seu desenho. Isso indica um caráter adequado para aplicação do coeficiente de correlação. Considerando o total de documentos coletados (N=5000), 34,4% (n=1722) não apresentaram índice de citações, enquanto 84% (n=4200) não apresentaram índices de menções. Por sua vez, 13 documentos (0,26%) apresentaram nº de citações ou de menções  $\geq 100$ , tendo sido excluídos por fugirem do padrão normal de distribuição dos dados coletados. Como resultante, foram considerados para análise de correlação 4987 documentos.

A análise de correlação entre as métricas alternativas de atenção *online* das menções em fontes da web social e as tradicionais baseadas em citação tem sido recorrente nos estudos do campo da altmetria (BORNMANN, 2015; ARAÚJO; CARAN; SOUZA, 2016; TRAVIESO; ARAÚJO, 2018). Com o intuito de investigar possíveis associações entre os diferentes indicadores altmétricos e o número de citações, o presente trabalho aplicou o Coeficiente de Correlação de Pearson (CCP), um cálculo que indica o grau de relação direta ou inversamente proporcional entre duas séries de variáveis. Por meio de um coeficiente de intervalo -1 até 1 ( $-1 \leq x \leq 1$ ), o CCP foi dividido em 3 intervalos: correlação forte ou muito forte ( $-1 \leq x \leq -0,7$  ou  $0,7 \leq x \leq 1$ ); correlação moderada ( $-0,7 < x \leq -0,5$  ou  $0,5 \leq x < 0,7$ ); e correlação fraca ou desprezível ( $-0,5 < x \leq 0$  ou  $0,5 < X \leq 0$ ). Coeficientes com valores positivos negativos ( $-1 \leq x < 0$ ) indicam correlação inversamente proporcional, enquanto valores positivos indicam correlação diretamente proporcional (BARBETTA, 2013; PESTANA; GAGEIRO, 2014).

A análise dos CCPs foi aplicada entre todos os 20 indicadores, sendo um deles referente ao número de citações e outros 19 altmétricos, destes foram excluídos cinco (Weibo, LinkedIn, Pinterest, Q&A e Syllabi) que não registraram nenhuma menção às pesquisas sobre Web Semântica analisadas. Os resultados das análises cruzadas entre os 15 indicadores analisados estão na Figura 3, apresentados os CPPs entre os indicadores das linhas e das colunas.

A figura apresenta o CCP entre cada um dos índices altmétricos por tipo de fonte e do índice bibliométrico de citação. Um gradiente na cor vermelha foi aplicado de acordo com a força da correlação – seja direta ou inversamente proporcional.

**Figura 3.** CCP entre variáveis\* altmétricas e de citação

	Al	No	Bl	Po	Pa	Tw	Pe	Fb	Wk	G+	Re	Fk	Vi	Me	Ci
Al	1,00	0,79	0,63	0,21	0,01	0,77	0,05	0,40	0,21	0,32	0,33	0,13	0,03	0,19	0,19
No	0,79	1,00	0,41	0,16	0,00	0,31	0,00	0,32	0,08	0,12	0,16	0,00	0,00	0,09	0,06
Bl	0,63	0,41	1,00	0,07	0,00	0,38	0,05	0,27	0,09	0,26	0,49	0,10	-0,01	0,12	0,17
Po	0,21	0,16	0,07	1,00	0,00	0,12	0,00	0,05	0,08	0,04	0,07	0,24	0,00	0,11	0,09
Pa	0,01	0,00	0,00	0,00	1,00	0,00	0,00	-0,01	0,01	0,01	0,00	0,00	0,00	0,02	0,03
Tw	0,77	0,31	0,38	0,12	0,00	1,00	0,04	0,34	0,19	0,34	0,32	0,19	0,07	0,18	0,17
Pe	0,05	0,00	0,05	0,00	0,00	0,04	1,00	0,05	0,00	0,39	0,00	0,00	0,00	0,01	0,01
Fb	0,40	0,32	0,27	0,05	-0,01	0,34	0,05	1,00	0,02	0,11	0,10	0,03	0,02	0,12	0,07
Wk	0,21	0,08	0,09	0,08	0,01	0,19	0,00	0,02	1,00	0,12	0,11	0,06	-0,01	0,13	0,18
G+	0,32	0,12	0,26	0,04	0,01	0,34	0,39	0,11	0,12	1,00	0,37	0,00	-0,01	0,07	0,06
Re	0,33	0,16	0,49	0,07	0,00	0,32	0,00	0,10	0,11	0,37	1,00	0,00	0,00	0,03	0,04
Fk	0,13	0,00	0,10	0,24	0,00	0,19	0,00	0,03	0,06	0,00	0,00	1,00	0,00	0,13	0,34
Vi	0,03	0,00	-0,01	0,00	0,00	0,07	0,00	0,02	-0,01	-0,01	0,00	0,00	1,00	0,02	0,03
Me	0,19	0,09	0,12	0,11	0,02	0,18	0,01	0,12	0,13	0,07	0,03	0,13	0,02	1,00	0,73
Ci	0,19	0,06	0,17	0,09	0,03	0,17	0,01	0,07	0,18	0,06	0,04	0,34	0,03	0,73	1,00

\*Al= Altmeria; No = Notícias; Bl = Blogs; Po = Documentos políticos; Pa = Patentes; Tw = Twitter; Pe = Revisões; Fb = Facebook; Wk = Wikipedia; G+ = Google+; Re = Reddit; Fk = F1000; Vi = Vídeos; Me = Mendeley; Ci = Citações.

**Fonte:** Dados da pesquisa (2020)

Ao se analisar a relação entre as fontes de altmetria e citações, os resultados demonstraram relações predominantemente fracas para a maioria das variáveis, com exceção do Mendeley (Me x Ci,  $p = 0,73$ ). O resultado tanto do comportamento do Mendeley e das demais fontes é similar ao de outras pesquisas que buscaram medir a correlação entre esses índices (BORNMANN, 2015) ou que revisaram a literatura a respeito (SUGIMOTO et al., 2017).

Por se tratar de uma plataforma de rede social voltada para a comunidade científica, a correlação direta e mais forte entre o Mendeley e dados de citações era esperado nos resultados. O número de leitores no Mendeley se baseia no quantitativo de inclusões de publicações nas bibliotecas de usuários que usam o serviço de referência *online*, finalidade essa fortemente associada à marcação (por parte de usuários da plataforma) de publicações relevantes para leituras e utilização das mesmas para citações futuras (MOHAMMADI; THELWAL; KOUSHA, 2015).

Por sua vez, ao se analisar a correlação interna entre os índices altmétricos de cada tipo de fonte, os resultados indicaram relação franca ou insignificante entre todos eles. Os que se saíram um pouco melhor foram entre Blog e Reddit (BI x Re,  $p = 0,49$ ) e Blog e Notícia (BI x No,  $p = 0,41$ ). Tais resultados podem estar relacionados a duas causas. Primeiramente, o mecanismo de coleta de dados altmétricos pode não ser capaz de cobrir suficientemente as menções de publicações científicas entre os diferentes tipos de fontes. Ou seja, a cobertura das mídias de um determinado artigo pode não ser capaz de encontrar a sua vasta difusão entre os diferentes tipos de mídias. Conseqüentemente, “zonas altmétricas não rastreáveis” podem ocorrer em alguns tipos de mídias, enviesando o índice.

Em segundo lugar, pela variedade das fontes de altmetria e a visão diversificada que cada uma pode dar para o fluxo do conhecimento compartilhado e principalmente o que cada uma delas significa (NEYLON, 2015). A questão é que cada tipo de menção altmétrica pode se referir a diferentes intenções ou finalidades do usuário. Por exemplo, em que medida a menção de uma publicação no Facebook tem motivações e finalidades similares às do Mendeley ou aos de *sites* de notícias? É possível que os índices altmétricos de cada meio (ou agrupamento dos mesmos) explique diferentes naturezas de repercussão social de publicações científicas, assumindo papéis complementares.

## 5 CONSIDERAÇÕES FINAIS

A Web Semântica tem se consolidado não apenas como tema de estudo, mas se firmado como campo de pesquisa, seja pelo acúmulo da produção científica que registra ou mesmo outros fatores como o surgimento de revistas especializadas que dedicam a sua investigação. A pesquisa possibilitou caracterizar esse campo no período analisado e inclusive confirmar a concentração de seus trabalhos periódicos especializados criados para este fim, verificando o desempenho dessa produção.

A diminuição observada quanto ao número de publicações no período analisado pode ser devida à transversalidade do objeto de estudo, o que condiciona também uma certa dispersão e,

sobretudo, uma certa disparidade terminológica que assume. Portanto, uma das possíveis linhas de pesquisas futuras pode ser justamente ampliar a busca bibliográfica por meio de outros termos, tarefa que pode ser explorada e enriquecida com as relações semânticas que foram detectadas neste estudo.

Quanto à diversidade dos tipos de publicação, vale ressaltar a concentração de pesquisas sobre a Web Semântica publicadas em anais de evento e capítulos de livro que juntos chegam a 56,38% superando os artigos de periódicos. Esse resultado confirma a relevância dos eventos científicos e dos espaços editoriais para veiculação de resultados de pesquisa sobre o tema, sobretudo dos eventos, o que já era apontado em pesquisas anteriores (DING, 2010).

O destaque da categoria *Information and Computing Sciences* como campo de pesquisa com maior concentração significativa de publicações (79%) e mesmo a ordem de distribuição entre as demais identificadas além de confirmar o caráter multidisciplinar da Web Semântica evidencia a participação das mesmas e suas contribuições que vão desde a reflexão de fundamentos e implementação (Ciência da Computação, Ciência da Informação, Engenharia e Linguística), ao domínio de sua aplicação ou experimento (Saúde e Ciências Médicas, Exatas, Humanas e Sociais Aplicadas). O que pode ser verificado na análise dos agrupamentos temáticos nas incidências e co-ocorrência dos termos mais frequentes usados nos títulos e resumos das publicações.

Para a análise de desempenho das publicações e sobretudo exploração de complementaridade entre dados de citação e de menções em fontes da web social, a análise de correlação é um importante instrumento de indicação quanto à existência de comportamentos diretamente ou inversamente proporcionais entre os fenômenos alométricos e bibliométricos. Entretanto, não são capazes de explicar os *por quês* dessa evidência. Como o próprio nome diz, são apenas indicadores, é importante que sejam qualificados e investigado, por exemplo, os atores que compartilham e contexto desses compartilhamentos (NEYLON, 2015; HAUSTEIN, 2016; SUGIMOTO et al., 2017; ARAUJO, 2020).

Nesse sentido, a exploração qualitativa dos conteúdos, fontes de informação e contextos sociopolíticos das publicações citadas e mencionadas são necessários para um entendimento mais explicativo mais aprofundado desses fenômenos, cujo caráter envolve, sem dúvidas, uma dimensão social importante. Os números alométricos e bibliométricos trazem as evidências indicativas, mas é na investigação qualitativa criteriosa que suas explicações podem se desvelar suficientemente.

## AGRADECIMENTOS

Os autores expressam agradecimento à Dimensions pelo acesso e uso não comercial dos dados da produção científica e de citação utilizados no estudo. De igual modo, agradecem à Altmetric.com por fornecer os dados altmétricos deste estudo gratuitamente para fins de pesquisa.

## REFERÊNCIAS

ABBAGNANO, N. **Dicionário de filosofia**. 5. ed. São Paulo: Martins Fontes, 2007. p. 225.

AGUIAR, E. M. d. **Aplicação do Word2vec e do Gradiente descendente estocástico em**

**tradução automática**. Dissertação (mestrado) – Fundação Getúlio Vargas, Escola de Matemática Aplicada, Rio de Janeiro, 2016.

ARAUJO, R. F. Da altmetria à análise de citações: uma análise da revista Datagramazero. **Datagramazero** (Rio de Janeiro), v. 16, p. 1-20, 2015. Disponível em: <https://doi.org/10.6084/m9.figshare.3409003.v1>. Acesso em: 07 jan. 2021.

ARAUJO, R. F.; CARAN, G. M.; SOUZA, I. V. P. Orientação temática e coeficiente de correlação para análise comparativa entre altmetrics e citações. **Em Questão**, 22(3), 184-200, 2016. Disponível em: <http://10.19132/1808-5245223.184-200>. Acesso em: 10 jan. 2021.

ARAUJO, R.F. Communities of attention networks: introducing qualitative and conversational perspectives for altmetrics. **Scientometrics**, v.124, p.1793-1809, set., 2020. Disponível em: <https://doi.org/10.1007/s11192-020-03566-7>. Acesso em: 13 fev. 2021.

BARBETTA, P. A.. **Estatística aplicada às Ciências Sociais**. 5. ed. Florianópolis: Editora da UFSC, 2013.

BERNERS-Lee T.; LASSILA, O.; HENDLER, J.. **The semantic web**. Scientific American, New York, v. 5, 2001.

BODE, C.; HERZOG, C.; HOOK, D.; MCGRATH, R.. **A guide to the dimensions data approach: a collaborative approach to creating a modern infrastructure for data describing research: Where we are and where we want to take it**. London: Digital Science. 2018

BORNMANN L.. Alternative metrics in scientometrics: A meta-analysis of research on three altmetrics. **Scientometrics**, v.103, n.3, p.1123-1144, 2015. Disponível em: <https://doi.org/10.1007/s11192-015-1565-y>

BORNMANN, L.. Field classification of publications in dimensions: A first case study testing its reliability and validity. **Scientometrics**, v.117, n.1, 2018. Disponível em: <https://doi.org/10.1007/s11192-018-2855-y>. Acesso em: 22 nov. 2020.

CHAMBERS, T.; MILOJEVIC, S; DING, Y. Female semantic web researchers: does collaboration with male researchers influence their network status? In 2014 ACM conference on Web science (WebSci '14). **Proceedings...** Association for Computing Machinery, New York, NY, USA, 301–302. 2018. Disponível em: <https://doi.org/10.1145/2615569.2615659>. Acesso em: 04 mar. 2021.

DING, Y. Semantic web : who is who in the field – a bibliometric analysis. **Journal of Information Science**, 36(3), 335-336, 2010. Disponível em: <https://doi.org/10.1177/0165551510365295>. Acesso em: 17 fev. 2021.

FIGUEIREDO F. de C.; ALMEIDA F. G. Ontologias em ciência da informação: um estudo bibliométrico no Brasil. **Ciência da Informação**, 46(1), 2017. Disponível em: <http://revista.ibict.br/ciinf/article/view/4011>. Acesso em: 03 fev. 2021.

GANDON, F. **A Survey of the First 20 Years of Research on Semantic Web and Linked Data**: Revue des Sciences et Technologies de l'Information - Série ISI : Ingénierie des Systèmes 'Information, Lavoisier, 2018. Disponível em: <https://hal.inria.fr/hal-01935898/document>. Acesso em: 07 jan. 2021.

GONTIJO, M. C. A. **A produção científica sobre inteligência artificial e seus impactos: análise de indicadores bibliométricos e altmétricos**. Dissertação (Mestrado em Gestão e Organização do Conhecimento) – Escola de Ciência da Informação. Universidade Federal de Minas Gerais, Belo Horizonte, p.151, 2021.

HAMMARFELT, B.. Using altmetrics for assessing research impact in the humanities. **Scientometrics**, 101(2), 2014. Disponível em: <https://10.1007/s11192-014-1261-3>. Acesso em: 07 jan. 2021.

HAUSTEIN, S. Grand challenges in altmetrics: heterogeneity, data quality and dependencies. **Scientometrics**, 108, p.413-423, 2016. Disponível em: <https://doi.org/10.1007/s11192-016-1910-9>. Acesso em: 03 fev. 2021.

HITZLER, P.; JANOWICZ, K. Semantic Web – Interoperability, Usability, Applicability. **Semantic Web**, v.2, n.1, mai., 2011. Disponível em: <http://www.semantic-web-journal.net/sites/default/files/swj206.pdf> . Acesso em 12 dez., 2020.

MACHADO, L.; SIMÕES, M. G.; SOUZA, R. R.; ALMEIDA, M. B.. Ciência da informação e Web Semântica: uma relação efetiva ou apócrifa? In: III Congresso ISKO Espanha-Portugal - XII Congresso ISKO Espanha. **Actas...** Coimbra - Portugal, p. 453-466, 2017.

MICROSOFT. (2019). Suporte office: descrição, sintaxe, comentários e exemplo da função CORREL do Microsoft Excel. Disponível em: <https://support.office.com/pt-br/article/correl-fun%C3%A7%C3%A3o-correl-995dcef7-0c0a-4bed-a3fb-239d7b68ca92> . Acesso em: 24 jan. 2020.

NEYLON, C. O caminho menos trilhado: otimizando para os impactos desconhecidos e inesperados da pesquisa. In: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Org.). **Ciência aberta, questões abertas**. Brasília: Ibict; Rio de Janeiro: Unirio. 2015

PASTOR-SÁNCHEZ, J-A. “Quince años de web semántica: de las tecnologías a las buenas prácticas”. **Anuario ThinkEPI**, v. 10, p. 264-268, 2016. Disponível em: <http://dx.doi.org/10.3145/thinkepi.2016.58>. Acesso em: 03 fev. 2021.

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

PINHEIRO, C. B. F. **A construção do conhecimento científico: a web semântica como objeto de estudo**. 2008. f. Dissertação (Mestrado em Ciência da Informação). Faculdade de Filosofia e Ciências – Universidade Estadual Paulista, Marília, 2008.

REISS, T.; VIGNOLA-GAGNE, E.; KUKK, P.; GLANZEL, W.; THIJS, B.. **ERACEP - Emerging research topics and their coverage by ERC-supported projects**. European Research Council. 2013. Disponível em: [http://www.scienceonthenet.eu/files/eracep\\_final\\_report.pdf](http://www.scienceonthenet.eu/files/eracep_final_report.pdf). Acesso em: 03 fev. 2021.

ROBREDO, J. Ciência da informação e Web semântica: Linhas convergentes ou linhas paralelas? In: ROBREDO, J.; BRÄSCHER, M. (Org). **Passeios pelo bosque da informação: estudos sobre representação e organização da informação e do conhecimento**, p. 12–47. Brasília: IBICT. 2010. Disponível em: <http://www.ibict.br/publicacoes/eroic.pdf>. Acesso em: 20 maio 2020.

SÁ, E. G. L.; LEITÃO, C. R. S. Produção científica brasileira sobre conservadorismo contábil: uma análise do período de 2005 a 2014. **Revista Unemat de Contabilidade**, v. 6, n. 12, p. 76-91, 2018. Disponível em: <https://periodicos.unemat.br/index.php/ruc/article/view/1779>. Acesso em: 12 fev. 2020.

SANTAREM SEGUNDO, J. E.. Web Semântica, dados ligados e dados abertos: uma visão dos desafios do Brasil frente as iniciativas internacionais. **Tendências da Pesquisa Brasileira em Ciência da Informação**, João Pessoa, v. 8, p. 219–239, 2015. Disponível em: <https://revistas.ancib.org/index.php/tpbci/article/view/359>. Acesso em: 20 maio 2020.

ST-GERMAIN, M. & MONGEON, P. The Contribution of Information Science in the Semantic Web Research Landscape. In L. Freund (Ed.) Association for Information Science and Technology (p. 470– 477). **Proceedings...** Hoboken, NJ: Wiley, 2018. Disponível em: <https://doi.org/10.1002/pra2.2018.14505501051>. Acesso em: 20 jan. 2021.

SUGIMOTO, C. R.; WORK, S.; LARIVIÈRE, V.; HAUSTEIN, S. Scholarly use of social media and altmetrics: a review of the literature. **Journal of the Association for Information Science and Technology**, v.68, n.9, set., 2017. Disponível em: <https://doi.org/10.1002/asi.23833>. Acesso em: 03 fev. 2021.

TRAVIESO, C.; ARAÚJO, R. F.. Indicadores altmétricos y de citación en la producción científica en ScienceOpen: estudio descriptivo para Brasil, España y Portugal. **Bibliotecas. Anales de Investigación**, 14(2), p. 124-137, 2018. Disponível em: <http://revistas.bnjm.cu/index.php/anales/article/view/4244/3897>. Acesso em: 20 maio 2020.

VISSER, M.; JAN VAN ECK, N.; WALTMAN, L. Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. **Quantitative Science Studies**, 2021. Disponível em: [https://doi.org/10.1162/qss\\_a\\_00112](https://doi.org/10.1162/qss_a_00112). Acesso em: 03 fev. 2021.

ZHANG, Y., HUA, W.; YUAN, S. Mapping the scientific research on open data: A bibliometric review. **Learned Publishing**, 31: 95-106, 2018. Disponível em: <https://doi.org/10.1002/leap.1110> Acesso em: 20 maio 2020.