

**SEMANTOMETRIA: MÉTRICA SEMÂNTICA PARA CALCULAR O VALOR DE UMA
PUBLICAÇÃO POR MEIO DE SEU TEXTO COMPLETO**

***SEMANTOMETRICS: SEMANTIC METRICS TO CALCULATE THE VALUE OF A
PUBLICATION BY ITS FULL TEXT***

 Nivaldo Calixto Ribeiro¹
 Dalgiza Andrade Oliveira²
 Jonas Aron Cardoso Diniz³

¹ Doutor em Gestão e Organização do Conhecimento pela Universidade Federal de Minas Gerais. Bibliotecário/Documentalista na Universidade Federal de Lavras.

E-mail: zoopas@gmail.com

² Professora Associada da Escola de Ciência da Informação da Universidade Federal de Minas Gerais. Doutora em Ciência da Informação pelo Programa de Pós-Graduação em Ciência da Informação da UFMG.

E-mail: dalgizamg@gmail.com

³ Mestrando do Programa de Pós-Graduação em Gestão e Organização do Conhecimento da Universidade Federal de Minas Gerais.

E-mail: jonasacd@gmail.com



ACESSO ABERTO

Copyright: Esta obra está licenciada com uma Licença Creative Commons Atribuição 4.0 Internacional. 

Conflito de interesses: Os autores declaram que não há conflito de interesses.

Financiamento: Não há.

Declaração de Disponibilidade dos dados: Todos os dados relevantes estão disponíveis neste artigo.

Recebido em: 11 jan. 2023.

Aceito em: 24 ago. 2023.

Publicado em: 12 out. 2023.

Como citar este artigo:

RIBEIRO, Nivaldo Calixto; OLIVEIRA, Dalgiza Andrade; DINIZ, Jonas Aron Cardoso. Semantometria: métrica semântica para calcular o valor de uma publicação por meio de seu texto completo. **Informação em Pauta**, Fortaleza, v. 8, p. 1-16, 2023. DOI: 10.36517/2525-3468.ip.v8i0.2023.83267.1-16.

RESUMO

Introdução: A aplicação e o uso das métricas alternativas podem ser considerados como estratégias a serem adotadas para complementar os métodos convencionais de avaliação da produção científica que consideram apenas a contagem de citações. Distinta de outras métricas que se baseiam em medir o número de interações, na rede de comunicação científica ou na web social, a semantometria, objeto de estudo deste artigo, parte do pressuposto de que é necessário o texto completo para avaliar o valor de uma publicação. Diante dessa abordagem, este estudo tem como objetivo discorrer sobre essa classe de métrica, expondo, sua origem em Herrmannova e Knoth (2014), sua metodologia de análise e sua fórmula de aplicação, por meio de investigação bibliográfica. Método: trata-se de uma pesquisa exploratória, em que, no levantamento

bibliográfico preliminar, foi conduzida, adotando sete dos dez passos iniciais de busca de representação Query, propostos por Cendón (2018). Foi empregada a revisão narrativa de literatura com análise dos documentos recuperados no todo. Resultados: considerou-se relevante a pesquisa sobre o tema, visto que há uma grande preocupação com a demonstração da qualidade de resultados de pesquisa à sociedade e constatou-se que há poucas publicações sobre o tema, materializada pelos resultados em buscas nas bases de dados: Dimensions, Gale, Scopus e Web of Science. Conclusão: embora a semantometria disponha de vantagens de sua aplicação, considera-se como necessária a continuidade de estudos nesse campo para entender melhor quais facetas da qualidade da pesquisa podem ser capturadas e como podem ser melhor aplicadas. Observou-se, ainda, a sua associação com a mineração, análise de dados textuais e relação com técnicas e ferramentas de similaridades textuais, bem como com pesquisas envolvendo a altmetria.

Palavras-chave: semantometria; métricas alternativas; publicações científicas - qualidade; publicações científicas - texto completo.

ABSTRACT

Introduction: The application and use of alternative metrics can be considered as a strategy to be adopted to complement the conventional methods of evaluating scientific production that consider only the citation count. Unlike other metrics that are based on

measuring the number of interactions in the scientific communication network, semantometry, the object of study of this research, assumes that the full text is necessary to assess the value of a publication. Given this approach, this article aims to discuss this class of metric, exposing its origin in Herrmannova and Knoth (2014), its analysis methodology and its application formula by means of literature research. Method: this is an exploratory research, in which the preliminary bibliographic survey was conducted adopting seven of the ten initial steps of Query Representation, proposed by Cendón (2018). A narrative literature review was used, with analysis of the recovered documents as a whole. Results: the research on the subject was considered relevant, since there is a great concern with demonstrating the quality of research results to society and it was found that there are few publications on the subject, materialized by the results of searches in databases: Dimensions, Gale, Scopus and Web of Science. Conclusions: it is concluded that, although semantometry has advantages in its application, it is necessary to continue studying this field to better understand which facets of research quality they can capture and how they can best be applied. It was also observed its association with the mining and analysis of textual data and its relationship with techniques and tools of textual similarities, as well as with research involving altmetrics.

Keywords: semantometry; alternative metrics; scientific publications - quality; scientific publications - fulltext.

1 INTRODUÇÃO

Em tempos, em que são obtidos inúmeros dados a cada instante, por meio de sensores, experimentos científicos e observações, tais como dados biológicos, dados neurais, dados astronômicos e dados de sensoriamento remoto, oriundos de redes sociais, Internet, programas de saúde, de finanças, de economia, de transporte e de diversos outros recursos ou aparatos tecnológicos, observa-se também a necessidade de atenção para métricas que possam avaliar a qualidade desses dados, principalmente, ao se tratar de documentos científicos.

As metodologias, para analisar a produção da ciência e tecnológica, têm surgido na mesma via, por meio de indicadores e dispositivos desenvolvidos, com a finalidade de analisar e compreender o universo da pesquisa por meio de menções ou outros aspectos. O investimento de fundos públicos em pesquisa exige a capacidade de demonstrar claramente retornos benéficos, responsáveis e bem gerenciados.

A questão de como avaliar a qualidade dos resultados da pesquisa é muito complexa de responder e, apesar de anos de pesquisa e debates sobre essa agenda, ainda não existe uma solução definitiva para esse problema. Dada a crescente necessidade de avaliação de pesquisas, é cada vez mais necessário entender como devem ser avaliadas e se os métodos existentes atendem a essa demanda. As soluções, predominantemente, baseadas na contagem do número de interações e menções, nas redes de comunicação acadêmica e tradicionais, são insuficientes (Herrmannova, 2018).

É salutar expor que as menções a uma pesquisa podem refletir o uso dos produtos científicos e, nesse aspecto, evidencia-se que eles são a materialidade da produção científica (Ferreira, 2017). As citações podem ser entendidas como “o conjunto de uma ou mais referências, que, incluídas em uma publicação, evidenciam elos entre indivíduos, instituições e áreas de pesquisa, visto que mostram o relacionamento de uma publicação com outra” (Rodrigues, 1982, p. 36). Contudo, elas representam apenas um pequeno nicho do ambiente científico, apesar de serem úteis para a avaliação de pesquisadores e publicações (Barros, 2015).

Ressalta-se que é necessário identificar caminhos para legitimar o processo de reconhecimento para garantir as funções basilares de investimento e certificação da científica (Barros, 2015). Para esse autor, a contagem de citações, quando aplicada de forma exclusiva, pode ignorar diversos atributos que contribuem para o impacto de uma publicação científica. Segundo Souza (2015), por meio das métricas alternativas às convencionais, há a possibilidade de se obter informações, a respeito do alcance e do uso de trabalhos científicos complexos de conseguir com processos tradicionais, como a exclusiva análise de citação.

Registra-se que existem diversas métricas, como altmetria, que analisam o fluxo da informação e do conhecimento nas redes e mídias sociais (Vanti; Sanz-Casado, 2016); a bibliometria, que tem como objetos de estudo os livros ou as revistas científicas por meio de métodos estatísticos e matemáticos; a cientometria, que observa a dinâmica da ciência como atividade social, com foco na análise da produção, da circulação e do

consumo da produção científica; a infometria, que utiliza diversos recursos para medir e analisar os aspectos da informação técnica e científica (Santos; Kobashi, 2009); a webometria, que é a aplicação da infometria à *Word Wide Web* (Vanti, 2015), no intuito de obter conhecimentos a respeito de acessos, *hiperlink*, estrutura e padrões de uso e a semantometria, de interesse desta pesquisa, que vem se expandindo lentamente na academia. Embora as métricas citadas tenham como objetivo demonstrar sobremaneira o impacto das publicações científicas, cada uma delas tem suas peculiaridades.

A semantometria, uma das métricas citadas na taxonomia da Ciência Aberta em Pontika *et al.* (2015) e Silveira *et al.* (2021), pode ser considerada como uma extensão da bibliometria, webometria e altmetria. A ideia da semantometria surgiu, pela primeira vez, em Herrmannova e Knoth (2014). De acordo com o estudo desses autores, baseia-se no pressuposto de que é necessário o texto completo para avaliar o valor de uma publicação, utilizando-se da representação semântica.

Considerando a relevância que as métricas têm alcançado para a ciência e os esforços de se utilizar métodos mais completos, para avaliar o impacto e a qualidade das publicações científicas, este artigo tem como objetivo discorrer sobre a semantometria, expondo sua origem em Herrmannova e Knoth (2014), sua metodologia de análise e sua fórmula de aplicação. A proposição do estudo é a investigação bibliográfica da literatura relacionada à semantometria e sua aplicabilidade como complemento aos atuais indicadores quantitativos de publicações.

Diante do exposto, justifica-se esta pesquisa por entender que se trata de uma proposta original e inovadora, trazendo um assunto, semantometria, ainda não abordado na literatura da Ciência da Informação em contexto nacional, podendo incidir em possíveis novos *insights* e novas perguntas de pesquisa envolvendo e impactando em estudos métricos da informação.

2 OPÇÕES METODOLÓGICAS

O estudo tem características exploratórias, uma vez que visa proporcionar mais familiaridade sobre a métrica semantometria, levantando seus conceitos e práticas alinhados à definição de Herrmannova e Knoth (2014). Segundo Raupp e Beuren (2013), uma das características da pesquisa bibliográfica consiste no aprofundamento de determinados conceitos de um tema ainda não apresentado satisfatoriamente.

Da mesma forma, a pesquisa caracteriza-se como qualitativa, à medida que utiliza coleta de dados, baseada em levantamentos bibliográficos, sem o uso de dados quantitativos ou técnicas estatísticas. Destaca-se que, para o levantamento preliminar da pesquisa bibliográfica, foram adotados sete dos dez passos iniciais de buscas, propostos por Cendón (2018).

Essa técnica consiste no processo em que a expressão da necessidade de informação é representada, por meio de uma pergunta de busca em linguagem natural e, a seguir, transformada numa consulta, usando a sintaxe do sistema e o vocabulário controlado, quando for o caso (Cendón, 2018). Normalmente é composta por palavras-chave, operadores booleanos ou algum outro modificador para realizar consultas em bancos de dados e sistemas de informação, Quadro 1.

Quadro 1 - Passos da busca

Passos		Aplicação na pesquisa
1	Formular e compreender a pergunta da pesquisa.	Que é a semantometria, sua metodologia e sua aplicação?
2	Selecionar as bases de dados - justificar.	Dimensions - oferece uma coleção abrangente de dados vinculados em uma única plataforma; de subsídios, publicações, conjuntos de dados e ensaios clínicos, patentes e documentos de políticas. Gale - disponibiliza bancos de dados e fontes primárias, bem como diversos recursos de aprendizagem. A Scopus e Web of Science - foram pré-selecionadas pela grande aceitação pela comunidade acadêmica.
3	Fazer a análise conceitual e desmembrar a pergunta em conceitos.	Semantometria, Semântica, Métricas. Avaliação científica.
4	Identificar termos alternativos.	Métrica alternativa. Mineração de dados textuais. Análise textual.
5	Definir a relação entre termos.	Não aplicado.
6	Definir a estratégia de busca.	A busca foi realizada em inglês, visto que a literatura técnico-científica sobre o tema se apresenta concentrada nesse idioma, com o argumento da pesquisa: "semantometrics".
7	Executar a busca nas bases de dados informando o número de documentos encontrados para cada conceito e combinação tentada. Utilizou-se apenas o termo "semantometrics".	Dimensions - 32 Gale - 3 Scopus - 8 Web of Science - 1

8	Verificar alguns documentos para determinar a necessidade de usar outros termos ou outras combinações de buscas.	Em razão da especificidade do tema e número reduzido de publicações, essa etapa foi desconsiderada.
9	Refinar/reformular a busca.	Com o baixo número de documentos recuperados, essa etapa não foi aplicada.
10	Apresentar os resultados.	Disponível na seção 3 - optou-se por apresentar os textos em ordem cronológica, facilitando a compreensão da métrica e a sua repercussão até o momento.

Fonte: Adaptado de Cendón (2018).

Com os passos estabelecidos, o resultado de busca possibilitou a recuperação de documentos completos para análise, evidenciando os conceitos e aplicações relacionados com a semantometria, podendo verificar o seu surgimento e a sua aplicação em alguns contextos. Foram recuperados 44 documentos e, após a análise dos títulos e da eliminação dos duplicados, foram selecionados 18 estudos para a leitura do resumo e para a redação da fundamentação. Entretanto, após a leitura dos textos, para a identificação do alinhamento ao escopo deste trabalho e reconhecimento dos tipos de abordagens, em torno do tema em estudo, foram eliminados mais seis documentos que citavam o termo apenas nas referências utilizadas, sem ser mencionado no texto e aqueles que o citavam, em prefácios de publicações, compondo o *corpus* da pesquisa com 12 publicações, aproximadamente, 27% do total encontrado na busca inicial.

3 SEMANTOMETRIA: CÁLCULO DA CONTRIBUIÇÃO SEMÂNTICA

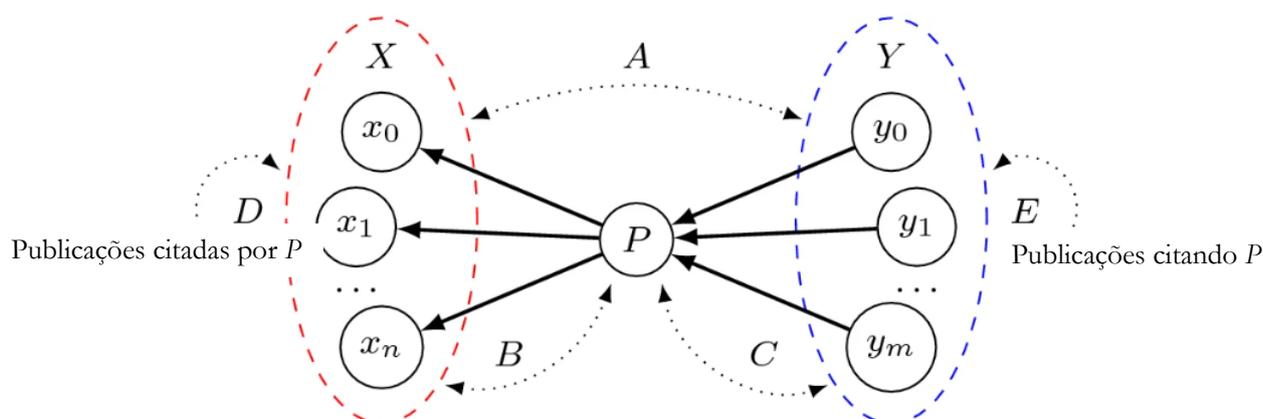
Na visão de Knoth e Herrmannova (2014), o uso atual das métricas de desempenho das publicações de pesquisa, bibliometria, altmetria, webometria, entre outras, está ancorado em um falso silogismo de que o impacto, ou até mesmo a qualidade de um trabalho de pesquisa, pode ser avaliado puramente com base em dados externos, sem considerar o manuscrito da publicação em si. Para os autores, essa suposição se assemelha à ideia de julgar uma ação judicial sem que o suspeito tenha a oportunidade de estar no tribunal e, conseqüentemente, falha da mesma maneira. Em seus estudos, demonstram que novas medidas de impacto, levando-se em consideração o manuscrito da publicação, poderiam ser desenvolvidas, as quais denominaram de semantometria.

Na mesma esteira, o Knowledge Media Institute, The Open University, ancorado no Projeto Sematometrics, expõe que a semantometria, por meio de métricas semânticas, defende que o texto completo de uma publicação é elemento essencial, para avaliar o valor da publicação, indo além da simples medição do número de interações na rede de comunicação acadêmica ou somente pelo número de citações (Semantometrics, 2017).

No artigo de Knoth e Herrmannova (2014), foi apresentada a primeira medida semantométrica que estimou a contribuição da pesquisa. Os pesquisadores mediram a semelhança semântica das publicações conectadas a uma rede de citações, usando uma fórmula básica para avaliar sua contribuição. O cálculo da similaridade semântica, valor numérico que reflete a proximidade em significado entre termos, necessita da aplicação de técnicas apropriadas, conforme o objetivo prospectado, como “contagem de nós” ou métodos baseados em índices (Silva, 2008).

Para Knoth e Herrmannova (2014), o valor agregado da publicação p pode ser estimado, com base na distância semântica das publicações citadas por p às publicações citando p , Figura 1. A distância semântica não está relacionada com o espaço entre termos, palavras ou expressões, pois tem a ver com completude, relevância, natureza e níveis de estruturação da informação, envolvendo a quantidade de processamento necessária para a sua compreensão. Quanto mais esforço cognitivo exigido, maior a distância semântica (Garcia, 2016). Em síntese, pode ser entendida como a forma em que as palavras se relacionam, como se conectam em frases, parágrafos e outros elementos.

Figura 1 - Explicação do cálculo da contribuição (p)



Fonte: Kreutz, Sahitaj e Schenkel (2020), adaptado de Knoth e Herrmannova (2014, 2015).

Na Figura 1, os nós simbolizam as publicações, as setas retas entre os artigos representam as citações. As variáveis $X=\{x_0, \dots, x_n\}$ são artigos referenciados por P e as variáveis $Y=\{y_0, \dots, y_n\}$ são artigos que citam P . As setas com linhas pontilhadas simbolizam relações observadas entre publicações. O grupo A contém distâncias entre pares de artigos referenciados (X) e citados (Y). O grupo B contém as distâncias entre os artigos referenciados (X) e P . O grupo C contém as distâncias entre P e os artigos que citam (Y). O grupo D contém distâncias entre pares de artigos referenciados (X). O grupo E contém distâncias entre pares de documentos citados (Y) (Kreutz; Sahitaj; Schenkel, 2020).

A hipótese projetada na semantometria foi baseada, em um processo desenhado da mesma forma em que a pesquisa se baseia no conhecimento já existente, a fim de criar novos conhecimentos sobre os quais outros possam construir, como um elo de ligação entre o estado da arte e trabalhos futuros. Na fórmula desenvolvida por Herrmannova e Knoth (2014), B é o conjunto de publicações que citam a publicação p e A é o conjunto citado por p , Fórmula 1.

Fórmula 1 - Distância semântica das publicações

$$Contribution(p) = \frac{\bar{B}}{\bar{A}} \cdot \frac{1}{|B| \cdot |A|} \sum_{a \in A, b \in B, a \neq b} dist(a, b)$$

Fonte: Knoth e Herrmannova (2014, 2015).

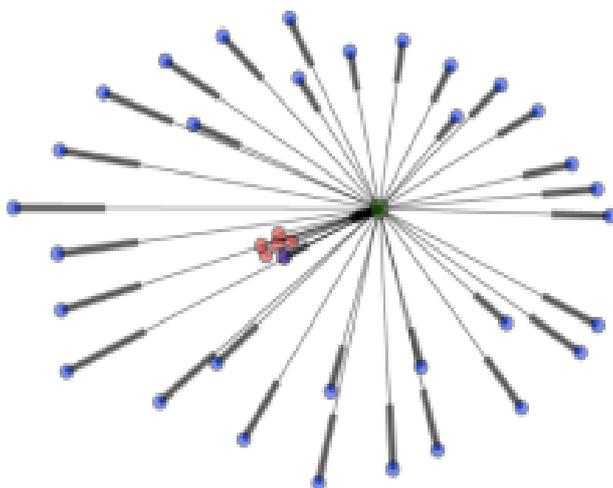
A equação é utilizada para calcular a distância total entre todas as combinações de publicações nos conjuntos A e B . A expectativa é de que a distância seja estimada, usando medidas de similaridade semântica, no texto completo das publicações, como a similaridade de cosseno nos vetores de documentos tf-idf, técnica utilizada para comparação entre documentos, com finalidades, como detecção de plágio, sistemas de recomendação, busca de informação, identificação de semelhanças entre textos, entre outras (Jurafsky; Martin, 2019).

A segunda fração na equação é um fator de normalização que se ajusta a todas as combinações entre membros dos conjuntos A e B , o que resulta em uma distância média entre os membros dos dois conjuntos. A primeira fração da equação acima é outro fator de normalização, responsável por ajustar o valor da contribuição a um domínio e tipo de

publicação específica. Baseia-se na medição da distância interior média de publicações nos conjuntos A e B (Knoth; Herrmannova, 2014).

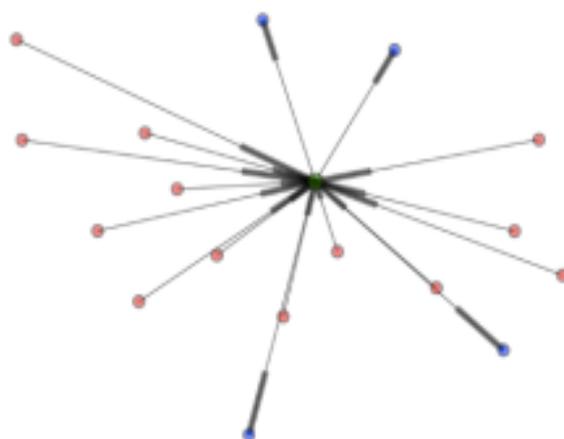
Para demonstrar como o cálculo da medida de contribuição funciona na prática, os pesquisadores escolheram algumas publicações, em um conjunto de dados com pontuação de citação semelhante, mas com contribuição diferente, expostas nas Figuras 2 e 3¹. Em ambas as figuras, o ponto verde, situado na parte central, representa as publicações para as quais foi calculada a contribuição de impacto dentro do conjunto de publicações utilizado pelos autores. Os pontos vermelhos são publicações citadas pelas publicações do ponto verde e os pontos azuis são as publicações que citam as publicações dos pontos verdes.

Figura 2 - Contribuição semântica $A = 0.8452$



Fonte: Semantometrics (2017).

¹ É importante mencionar que é necessário o uso do *Python*, uma linguagem de programação, para executar o *script* que calcula a contribuição da publicação. O *script* para calcular a contribuição de publicação de pesquisa, de acordo com Knoth e Herrmannova (2014), está disponível em: <https://github.com/robodasha/semantometrics>. Acesso em: 22 ago. 2022.

Figura 3 - Contribuição semântica $B = 0,9220$ 

Fonte: Semantometrics (2017).

Ao comparar a contribuição dos exemplos expostos nas Figuras 2 e 3, pode-se observar que a medida de contribuição de B é mais alta que A , embora A tenha recebido mais citações que B . Isso pode indicar que B é uma publicação que tem mais possibilidade de impacto na ciência, em função de ser um texto com menos semelhança a outros textos do conjunto de dados selecionados para o experimento da pesquisa de Knoth e Herrmannova (2014).

4 RESULTADOS DA INVESTIGAÇÃO BIBLIOGRÁFICA DA LITERATURA

Com a pesquisa, identificou-se que os veículos de comunicação científica que mais publicaram sobre o tema foram a *D-Lib Magazine* e a *Semantometrics*. Além disso, constatou-se que Drahomira Herrmannova foi a pesquisadora que desenvolveu a métrica, sendo, naturalmente, a autora com mais publicações sobre o tema, em colaboração com o orientador da sua pesquisa Petr Knoth.

De acordo com seu currículo, Herrmannova (2021) é pesquisadora no Learning Systems Group no Oak Ridge National Laboratory, nos Estados Unidos. Atua com pesquisas sobre inteligência artificial, para permitir o acesso inteligente ao conteúdo das publicações de pesquisa, concentrando-se no desenvolvimento de modelos para triagem de literatura e extração de informações de publicações científicas, extração de dados de tabelas em documentos científicos, entre outros. Antes, desenvolveu sua pesquisa de pós-doutorado no Knowledge Media Institute, The Open University, Reino Unido, onde desenvolveu a semantometria.

Com relação às demais publicações sobre o tema, em uma reflexão sobre o futuro da altmetria, Roemer e Borchardt (2015) pontuam que é importante reconhecer que, quando tudo estiver dito e feito, a altmetria de amanhã pode parecer muito diferente da altmetria discutida e debatida, assumindo que, futuramente, pode de fato ser algo muito menos "alternativo" e, em vez disso, parecer mais próximo da abordagem formal. Citam a semantometria, como proposta da análise semântica de texto completo de publicações, para determinar seu nível de contribuição em uma rede de citações. Concluem que a próxima fase da altmetria depende em grande parte das ações, esforços e práticas dos defensores e inovadores de hoje.

Quanto à aplicação da métrica, Herrmannova e Knoth (2015) buscaram compreender como a semantometria pode ajudar a caracterizar os tipos de colaboração de pesquisa, em redes de publicação acadêmica e a natureza dos laços entre comunidades e como essas informações podem ser utilizadas, para auxiliar na avaliação de pesquisas. Usando o conjunto de dados CORE, um framework nativo utilizado para manipular dados, foi analisada a relação entre a distância semântica dos autores e seu valor de endogamia de pesquisa. Concluíram que essa classificação pode ser útil em estudos e análises de avaliação de pesquisa, por exemplo, para identificar colaborações de pesquisa emergentes ou grupos de especialistas já estabelecidos. Embora as pontes tenham sido a preocupação de muitos estudos de pesquisa, sua identificação tem sido limitada à estrutura das redes de interação. Em contraste com essas abordagens, este estudo levou em consideração tanto a rede de interação (coautoria, citações) quanto a distância semântica entre trabalhos de pesquisa ou comunidades. Foram identificados quatro tipos potenciais de colaboração em uma rede de coautoria. Os resultados sugerem que medidas semelhantes podem ser usadas, para fornecer informações significativas sobre a natureza da colaboração, em redes de publicações acadêmicas.

Em outro estudo, a pesquisa de Herrmannova e Knoth (2016) enfocou a avaliação de publicações de pesquisa e a criação de novos métodos que utilizam o conteúdo da publicação, apresentando uma nova classe de métricas baseadas em manuscritos. Ao fazer a ponte de campos de avaliação de pesquisa e mineração de texto e dados, foi possível fornecer ferramentas para analisar os resultados da pesquisa e para aliviar a sobrecarga de informações em publicações acadêmicas. Destaca-se que mineração de textos está associada ao procedimento de captura de informações relevantes de um texto.

Hill (2016), ao discutir o futuro da comunicação científica à época, destacou o uso de mineração de texto e análise semântica com o desenvolvimento de algoritmos apropriados e a disponibilidade de poder de processamento suficiente, favorecendo a geração de insights reais a partir de grandes corpos de texto. Cita o exemplo da semantometria como um modelo interessante e promissor.

Em sua pesquisa, Erdt *et al.* (2016) retratam a altmetria e apontam o crescimento do interesse em suas práticas, gerando muitas dúvidas sobre seus potenciais benefícios e desafios. Diante disso, por meio dessa pesquisa, os autores buscaram responder a algumas dessas questões e identificaram que problemas de processamento e desambiguação de dados altmétricos são baseados no conceito de rastreamento de menções de saída de pesquisa, para os objetos de investigação e que solucionar essas questões relacionadas aos links, para identificadores exclusivos, pode ser muito desafiador. Essa constatação, da possibilidade de existirem várias versões do mesmo documento, em distintos sites, usando identificadores diferentes, ou mesmo a ausência dos links, pode revelar a semantometria como uma possível solução como métrica de avaliação da produção científica, por meio de diferentes maneiras de mapear menções aos artigos, considerando a semântica dos textos envolvidos.

O trabalho de Burland e Grout (2017) descreveu o papel que a *Joint Information Systems Committee* (JISC), organização sem fins lucrativos do Reino Unido, que fornece serviços de rede, tecnologia e recursos digitais, em apoio a instituições de ensino superior e pesquisa, desempenhou. Expuseram que as práticas de pesquisa aberta estão oferecendo oportunidades, para explorar métricas de download e novas formas de dados de citação, para uso no estabelecimento de indicadores de atividade de pesquisa aberta. Entre as ações da Jisc, registra-se apoio ao experimento, em torno de um conjunto de dados aberto e semântico, que visa fornecer transparência no nível do artigo, com indicadores de citação de Herrmannova e Knoth (2015).

Kreutz, Sahitaj e Schenkel (2019) reavaliaram as ideias de semantometria apresentadas por Herrmannova e Knoth (2014), para aprender padrões de recursos extraídos das distâncias de publicação, em suas redes de citações, com o objetivo de distinguir documentos seminais e de pesquisa na área de Ciência da Computação. Para a avaliação, os pesquisadores utilizaram o conjunto de dados SeminalSurveyDBLP. Ao usarem diferentes representações de conteúdo de documentos, a incorporação de medidas de distância semântica, bem como vários algoritmos de aprendizado de

máquina para a classificação, alcançou uma precisão de até 0,8015 em seu conjunto de dados.

Parinov, Bakarov e Vodolazcky (2020) contextualizam os métodos de aprendizado de máquina e processamento de linguagem natural, elucidando que também pertence o método de modelagem de tópicos, sinalizando que, na área de análise de conteúdo de citações, eles foram implementados para várias tarefas, entre elas semantometria, como medição de semelhanças de texto, de citação e artigos neles citados. No estudo de Parinov (2021), que teve como objetivo a proposição de uma abordagem geral, para a construção dos indicadores de consumo acadêmico e apresentação dos resultados das experiências de avaliação de uma estrutura temática, o mesmo contexto de medição de semelhanças de texto referindo-se à semantometria foi mencionado.

A publicação de Zhang, Tam e Cox (2021) discute o crescimento de pesquisas em inteligência artificial e big data, na última década, o que levou a um aumento significativo de pesquisas baseadas em dados publicados em Biblioteconomia e Ciência da Informação. Segundo os autores, a implicação dessa situação é que se torna extremamente desafiadora para pesquisadores desenvolverem e manterem uma visão atualizada dos métodos de pesquisa usados no campo. Acrescentam que há uma necessidade crescente de métodos automatizados que possam ajudar na análise de métodos de pesquisa na área, pois o número de publicações e de métodos de pesquisa aumenta rapidamente. No entanto, para Zhang, Tam e Cox (2021), não foram encontrados trabalhos nesse sentido até o momento, embora já tenha sido tentado em outras disciplinas. Apontam a necessidade de fornecer acesso estruturado ao conteúdo da literatura científica, articulada por meio do conceito de “semantometria”, apontando a urgência de realizar pesquisas semelhantes na área. Ainda, citam que, em razão da complexidade de definir e concordar com uma classificação dos métodos de pesquisa, a tarefa de análise automatizada pode enfrentar muitos obstáculos.

Por fim, o estudo de Brierley *et al.* (2022) sobre o acompanhamento das mudanças entre a publicação de preprint e a publicação em periódico, durante a pandemia do covid-19, cita a semantometria como uma das variedades de ferramentas e técnicas para medir a similaridade de textos, com foco na “semelhança textual semântica”.

5 CONSIDERAÇÕES FINAIS

Em tempos que se discutem os avanços das práticas do movimento da Ciência Aberta, do uso de métricas responsáveis, este estudo se propôs a discorrer sobre o que é a métrica semantometria, que se refere ao uso da distância semântica, para calcular o valor de uma publicação, por meio de seu texto completo, suas metodologias e sua aplicação.

Quanto às suas metodologias e sua aplicação, verifica-se que há *locus* para um cenário promissor para análise de publicações. Em boa parte das publicações recuperadas, a semantometria foi associada pelos autores consultados à mineração de dados e análise textual, lançando olhares para pesquisas interdisciplinares abordando esses temas.

Ainda, constatou-se que a métrica em estudo pode atuar com possível continuação e avanço de estudos alométricos, uma vez que foi citada como uma possível solução de seus problemas, em virtude das diferentes maneiras de mapear menções aos artigos, considerando a semântica dos textos envolvidos.

Há de se destacar ainda a preocupação dos autores com relação à necessidade de fornecer acesso estruturado ao conteúdo da literatura científica, por meio da semantometria e a referência às suas ferramentas e técnicas para medir a similaridade de textos.

Por fim, foi observado que, apesar de a semantometria dispor de vantagens em sua aplicação, entende-se como necessária a continuidade de estudos nesse campo para entender melhor quais facetas da qualidade da pesquisa podem capturar e como podem ser melhor aplicadas.

Com relação à literatura sobre o tema, identificou-se que há poucas publicações relacionadas ao tema, o que é comprovado pelo baixo número de textos recuperados na etapa de levantamento bibliográfico.

Como proposta de pesquisas futuras, recomenda-se a utilização de banco de dados temáticos, voltados para a área da Ciência da Informação e aplicação do algoritmo criado por Knoth e Herrmannova (2014), para verificar as publicações com mais potencial de inovação, não considerando, exclusivamente, os dados formais de citações ou de dados alométricos.

REFERÊNCIAS

- BARROS, M. Altmetrics: métricas alternativas de impacto científico com base em redes sociais. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 20, n. 2, p. 19-37, jun. 2015. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/1782>. Acesso em: 30 jun. 2022.
- BRIERLEY, L. *et al.* Tracking changes between preprint posting and journal publication during a pandemic. **PLoS Biology**, San Francisco, v. 20, n. 2, e3001285, 2022. Disponível em: <https://doi.org/10.1371/journal.pbio.3001285>. Acesso em: 30 jun. 2022.
- BURLAND, T.; GROUT, C. Standards and Interoperability: how Jisc's Work Supports Reporting, Communicating and Measuring Research in the UK. **Procedia Computer Science**, Amsterdam, v. 106, p. 276-282, 2017. Disponível em: <https://doi.org/10.1016/j.procs.2017.03.026>. Acesso em: 2 jul. 2022.
- CENDON, B. V. **Exercícios: passos na busca.** Disciplina Metodologia da Pesquisa. Escola de Ciência da Informação da Universidade Federal de Minas Gerais, 2018.
- ERDT, M. *et al.* Altmetrics: an analysis of the state-of-the-art in measuring research impact on social media. **Scientometrics**, Amsterdam, v. 109, p. 1117-1166, 10 aug. 2016. Disponível em: <https://doi.org/10.1007/s11192-016-2077-0>. Acesso em: 30 jun. 2022.
- FERREIRA, S. A. **Produção Científica sobre Biblioteca Pública nos Programa de Pós-Graduação em Ciência da Informação no Brasil.** 2017. Dissertação (Mestrado em Ciência da Informação) – Escola de Ciência da Informação, Universidade Federal de Minas Gerais, Belo Horizonte, 2017.
- GARCIA, L. S. **Interação humano-computador: espaços cognitivos no uso de sistemas computacionais.** Curitiba: UFPR, 2016. Disponível em: <https://www.inf.ufpr.br/laura/IHC-2016-2/Material%20anterior/IHC-Engenharia-Cognitiva-26-09-16.pdf>. Acesso em: 30 jun. 2022.
- HERRMANNOVA, D. [Curriculum vitae]. Bethel Valley Road, 10 set. 2021. Disponível em: https://dasha.tech/documents/dasha_herrmannova_cv.pdf. Acesso em: 30 jun. 2022.
- HERRMANNOVA, D. **Mining scholarly publications for research evaluation.** 2018. 428p. Thesis (PhD) - The Open University. Disponível em: http://oro.open.ac.uk/55421/7/dissertation_final.pdf. Acesso em: 30 jun. 2022.
- HERRMANNOVA, D.; KNOTH, P. An analysis of the microsoft academic graph. **D-Lib Magazine**, Reston, v. 22, n. 9/10, 2016. Disponível em: <https://www.dlib.org/dlib/september16/herrmannova/09herrmannova.html>. Acesso em: 30 jun. 2022.
- HERRMANNOVA, D.; KNOTH, P. Semantometrics in coauthorship networks: Fulltext-based approach for analysing patterns of research collaboration. **D-Lib Magazine**, Reston, v. 21, n. 11-12, 2015. Disponível em: <https://doi.org/10.1045/november2015-herrmannova>. Acesso em: 30 jun. 2022.
- HILL, S. A. Making the future of scholarly communications. **Learned Publishing**, Hatfield, v. 29, n. 1, p. 366-370, oct. 2016. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1002/leap.1052>. Acesso em: 30 jun. 2022.
- JURAFSKY, D.; MARTIN, J. H. **Speech and language processing: an introduction tonatural language processing, computational linguistics, and speech recognition.** New Jersey: Prentice Hall. 2019.
- KNOTH, P.; HERRMANNOVA, D. Semantometrics: fulltext-based measures for analyzing research collaboration. In: INTERNATIONAL CONFERENCE OF THE INTERNATIONAL SOCIETY FOR SCIENTOMETRICS AND INFORMETRICS, 15, 2015, Istanbul, TURKEY. **Anais...** Istanbul: ISSIS, 2015. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.698.159&rep=rep1&type=pdf>. Acesso em: 30 jun. 2022.
- KNOTH, P.; HERRMANNOVA, D. Towards semantometrics: a new semantic similarity based measure for assessing a research publication's contribution. **D-Lib Magazine**, Reston, v. 20, n. 11/12, 2014. Disponível em: <http://www.dlib.org/dlib/november14/knoth/11knoth.html>. Acesso em: 30 jun. 2022.

KREUTZ, C. K.; SAHITAJ, P.; SCHENKEL, R. Evaluating semantometrics from computer science publications. **Scientometrics**, Budapest, v. 125, p. 2915-2954, Dec. 2020. Disponível em: <https://doi.org/10.1007/s11192-020-03409-5>. Acesso em: 21 ago. 2023.

KREUTZ, C. K.; SAHITAJ, P.; SCHENKEL, R. Revaluating Semantometrics from Computer Science Publications. *In: INTERNATIONAL CONFERENCE ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL (SIGIR) 42., 2019, Paris, France. Anais...* Paris: SIGIR, 2019. p. 42-55. Disponível em: <http://ceur-ws.org/Vol-2414/paper5.pdf>. Acesso em: 30 jun. 2022.

PARINOV, S. Citation contexts as a data source for evaluation of scholarly consumption. **Scientometrics**, Amsterdam, v. 126, n. 1, p. 9249-9265, Sep. 2021. <https://doi.org/10.1007/s11192-021-04165-w>. Acesso em: 30 jun. 2022.

PARINOV, S.; BAKAROV, A.; VODOLAZSKY, D. Layout logical labelling and finding the semantic relationships between citing and cited paper content. **International Journal of Metadata Semantics and Ontologies**, London, v. 14, n. 1, p. 54-62, jun. 2020. Disponível em: <https://www.inderscienceonline.com/doi/abs/10.1504/IJMSO.2020.107796>. Acesso em: 30 jun. 2022.

PONTIKA, N. *et al.* Fostering open science to research using taxonomy and an elearning portal. *In: INTERNATIONAL CONFERENCE ON KNOWLEDGE TECHNOLOGIES AND DATA-DRIVEN BUSINESS, 15., 2015. Proceedings...* Graz, Áustria: Association for Computing Machinery, 2015. Disponível em: <http://oro.open.ac.uk/44719/>. Acesso em: 30 jun. 2022.

RAUPP, F. M.; BEUREN, I. M. Metodologia da Pesquisa Aplicável às Ciências Sociais. *In: BEUREN, I. M. (Org.). Como elaborar trabalhos monográficos em contabilidade: teoria e prática.* 3. ed. São Paulo: Atlas, 2013. p. 76-97.

RODRIGUES, M. P. L. Citações nas dissertações de mestrado em Ciência da Informação. **Ciência da Informação**, Brasília, v. 11, n. 1, p. 35-61, 1982. Disponível em: <https://revista.ibict.br/ciinf/article/view/177>. Acesso em: 08 out. 2022.

ROEMER, R. C.; BORCHARDT, R. Altmetrics. **Library Technology Reports**, Chicago, v. 51, n. 5, jul. 2015. Disponível em: link.gale.com/apps/doc/A433686573/AONE?u=cap&sid=bookmark-AONE&xid=97f13838. Acesso em: 30 jun. 2022.

SANTOS, R. N. M. D.; KOBASHI, N. Y. Bibliometria, cientometria, infometria: conceitos e aplicações. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 2, n. 1, 2009. Disponível em: <https://repositorio.ufpe.br/handle/123456789/10089>. Acesso em: 08 out. 2022.

SEMANTOMETRICS. **Contribution measure**. Milton Keynes: Knowledge Media Institute, The Open University, 2017. Disponível em: <http://semantometrics.org/contact.html>. Acesso em: 30 jun. 2022.

SILVA, D. F. **Estudo de funções de similaridade semântica de termos aplicadas a um domínio**. 2008. 45p. Trabalho de Conclusão de Curso (Graduação) - Ciência da Computação do Centro de Informática, Universidade Federal de Pernambuco, Recife, 2008.

SILVEIRA, L. *et al.* Ciência aberta na perspectiva de especialistas brasileiros: proposta de taxonomia. **Encontros Bibli: Revista eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, v. 26, n. 1, p. 1-27, 2021. DOI 10.5007/1518-2924.2021.e79646. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/79646>. Acesso em: 30 jun. 2022.

SOUZA, I. V. P. Altméria ou métricas alternativas: conceitos e principais características. **AtoZ: novas práticas em informação e conhecimento**, Curitiba, v. 4, n. 2, p. 58-60, 2015. Disponível em: <http://dx.doi.org/10.5380/atoz.v4i2.44554>. Acesso em: 30 jun. 2022.

VANTI, N. Os links e os estudos webométricos. **Ciência da Informação**, Brasília, v. 34, n. 1, 2005. Disponível em: <https://revista.ibict.br/ciinf/article/view/1104>. Acesso em: 08 out. 2022.

VANTI, N.; SANZ-CASADO, E. Altméria: a métrica social a serviço de uma ciência mais democrática. **Transinformação**, Campinas, v. 28, n. 3, p. 349-358, 2016. Disponível em: <https://www.scielo.br/j/tinf/a/KsKpZZrfp3nZWF3pLVy7cP/abstract/?lang=pt>. Acesso em: 08 out. 2022.

ZHANG, Z.; TAM, W.; COX, A.; Towards automated analysis of research methods in library and information science. **Quantitative Science Studies**, Cambridge, v. 2, n. 2, p. 698-732, jul. 2021. Disponível em: https://doi.org/10.1162/qss_a_00123. Acesso em: 30 jul. 2022.