

ANÁLISE BIBLIOMÉTRICA DE PUBLICAÇÕES NA TEMÁTICA DO *BIG DATA* UTILIZANDO O *VOSviewer*

Bibliometric analysis of *big data* publications using *vosviewer*

Sara Rodrigues Correa¹
Ricardo Luiz Machado, Dr.²

¹ Mestranda em Engenharia de Produção e Sistemas pela Pontifícia Universidade Católica de Goiás. sara_02rodrigues@hotmail.com

² Professor Dr. Pontifícia Universidade Católica de Goiás, Programa de Mestrado em Engenharia de Produção e Sistemas e UniEVANGÉLICA – Centro Universitário de Anápolis. drmachado@gmail.com

RESUMO

Este trabalho teve como objetivo identificar a ocorrência do tema *Big data* associado aos negócios em pesquisas de natureza científica. Para atingir seus objetivos foi realizada uma pesquisa bibliométrica, utilizando o *software VOSviewer*. Os dados foram extraídos da plataforma *Web of Science*. Os resultados obtidos permitiram identificar as palavras com maior índice de ocorrência nos artigos, as datas em que foram publicados e os autores com maior número de pesquisas realizadas na área. Como principal conclusão do trabalho foi constatado que ainda é baixo o índice de publicação de trabalhos científicos associando o termo *big data* aos negócios.

Palavras-chave: *Big data*; *Web of science*; análise bibliométrica.

ABSTRACT

This work aimed to identify the occurrence of the Big Data theme associated with business in scientific research. To achieve the objectives of this research we carried out a bibliometric survey using the *VOSviewer* software. We extracted the data from the *Web of Science* platform. The results obtained allowed to identify the words with the highest index of occurrence in the articles, the dates in which they were published and the authors with the greatest number of studies in the area. As the main conclusion of the work, we found that the number of publication of scientific papers is still low associating the term big data with the business.

Keywords: Big Data; Web of science; bibliometric analysis.

1. INTRODUÇÃO COM REFERENCIAL TEÓRICO

O termo *Big Data* é comumente descrito por suas características, que incluem três dimensões (conhecidas como os 3V's) propostas para caracterizá-lo: volume, velocidade e variedade. O volume está relacionado à grande quantidade de dados disponíveis; a velocidade está relacionada à rapidez com que os dados são criados e a agilidade com que são processados; e a variedade refere-se aos vários tipos de dados (MCAFEE & BRYNJOLFSSON, 2012; ISASI *et al.*, 2015; HILBERT, 2016). Posteriormente, outras duas características foram incorporadas ao tema *Big data*: a veracidade, que se refere à confiabilidade dos dados e o valor, que está associado à utilidade dos dados (NELSON & OLOVSSON, 2016; LAKSHEN *et al.*, 2016).

O *Big Data* tem atraído a atenção tanto da comunidade acadêmica quanto empresarial, devido à extensa possibilidade de estudos e às grandes oportunidades de negócios que podem ser obtidos por meio da análise de dados. McAfee & Brynjolfsson (2012) afirmam que a quantidade e a variedade de dados dispostos no mundo cresce exponencialmente e saber analisá-los é o segredo para aproveitar as oportunidades geradas.

Encontra-se nos bancos de dados científicos uma grande diversidade de estudos sobre o *Big Data*. São identificadas aplicações nas áreas da saúde, segurança, manufatura, finanças, economia, logística, inteligência de negócios, mídias sociais, *e-commerce*, inovação tecnológica, turismo, dentre outras (ADRIAN *et al.*, 2017; HUANG *et al.*, 2016; AKOKA *et al.*, 2017).

No que concerne ao meio empresarial, Huang *et al.*, (2016) afirmam que, ao analisar e tratar dados pode-se deparar com tendências e condições de negócios, técnicas de gerenciamento, estratégias e práticas de gestão, além de prever caminhos para inovação. Sabe-se que, com a ascensão do *Big Data*, torna-se necessário que as empresas mudem sua forma de gerenciamento, e tenham profissionais qualificados para analisar os dados estruturados e não estruturados que estão dispostos, para assim prever cenários futuros e obter vantagens competitivas (MCAFEE & BRYNJOLFSSON, 2012).

É importante ressaltar que o *Big Data* não implica apenas em benefícios para as empresas como perspectivas mais precisas de negócios, melhor

segmentação de clientes, conhecimento de vendas e oportunidades de mercado, detecção de fraudes e melhor planejamento. A era do *Big Data* traz consigo limitações que as empresas precisam lidar como recursos humanos desqualificados para analisar dados, alto custo do processo de análise, incapacidade de bancos de dados em realizar análises e lentidão no resgate dos dados, dentre outros (RUSSOM, 2011).

A qualificação do pessoal é tão importante quanto um *software* de análise de dados. Inicialmente, as ferramentas dispostas para análise de dados não eram capazes de processá-los com eficiência, afirmam Lakshen *et al.* (2016). O Instituto Mckinsey (MCKINSEY, 2011) realizou um levantamento das tecnologias que tem sido usadas e desenvolvidas nos últimos anos, afim de agregar, manipular e gerenciar dados, e identificou as seguintes opções existentes: *Big Table*, *Cloud Computing*, *Data mart*, *Google File System*, *Hadoop*, *Map Reduce*, *HBase* e *Metadata*, dentre outras.

No contexto descrito nos parágrafos anteriores, foi realizada uma pesquisa bibliométrica com o intuito de acompanhar / mapear a evolução das produções científicas de diversas áreas do conhecimento, apoiando o desenvolvimento de novos trabalhos e indicando suas tendências (ARAÚJO, 2006; OLIVEIRA ET AL., 2013; CAFÉ & BRÄSCHER, 2008). Esse tipo de pesquisa também assessora pesquisadores com informações sobre a influência e o alcance de periódicos e pesquisadores da área.

A delimitação do trabalho consistiu na seleção de cinco periódicos internacionais que abordassem os temas *Big Data*, negócios e *softwares* para a análise de dados. Por meio desses periódicos buscou-se identificar a crescente quantidade de pesquisas sobre o tema e, a partir dos artigos publicados, identificar os principais autores e as principais palavras-chaves utilizadas para identificar as pesquisas realizadas.

2. METODOLOGIA

A etapa inicial da pesquisa foi a escolha do *software VOSviewer* para condução do levantamento e análise de dados. A partir das opções que o *software* oferecia, foi realizada a escolha da base de dados a ser utilizada. O *software VOSviewer* cria mapas para análise de redes bibliométricas com base

em dados extraídos das bases *Scopus*, *PubMed* e *Web of Science*. A base de dados escolhida neste trabalho foi a *Web of Science*.

A segunda etapa foi a seleção dos periódicos. Nesta etapa foram selecionados: *Big data analytics*; *Big data research*; *Journal of big data*; *Journal of business research* e *Computational statistics and data analysis*. Os critérios de seleção desses periódicos foram: selecionar revistas que abordassem o tema *Big Data* (isto foi feito através de uma análise da política editorial de cada periódico); escolher um periódico que abordasse o tema negócios, além do *Big Data*; e acessar uma revista que tratasse da utilização de *softwares* para análise de dados do *Big Data*. Outro critério adotado foi o fator de impacto das revistas e sua relevância para o meio acadêmico.

A etapa subsequente consistiu em acessar a base de dados *Web of Science* e coletar os dados necessários para análise. Realizou-se um levantamento de todos os artigos publicados nos periódicos selecionados, com a finalidade de analisá-los por meio do *software VOSviewer* e identificar as principais palavras e autores.

A quarta etapa consistiu em analisar os dados da etapa anterior. Para isto, foram criados mapas das principais palavras presentes nos títulos e resumos da amostra. As palavras e os autores que foram consideradas mais relevantes foram listadas através de um algoritmo existente no *software VOSviewer*. Os resultados obtidos com a análise proporcionaram o entendimento das correntes teóricas apresentadas em cada *cluster* identificado na pesquisa, como será abordado nas seções seguintes.

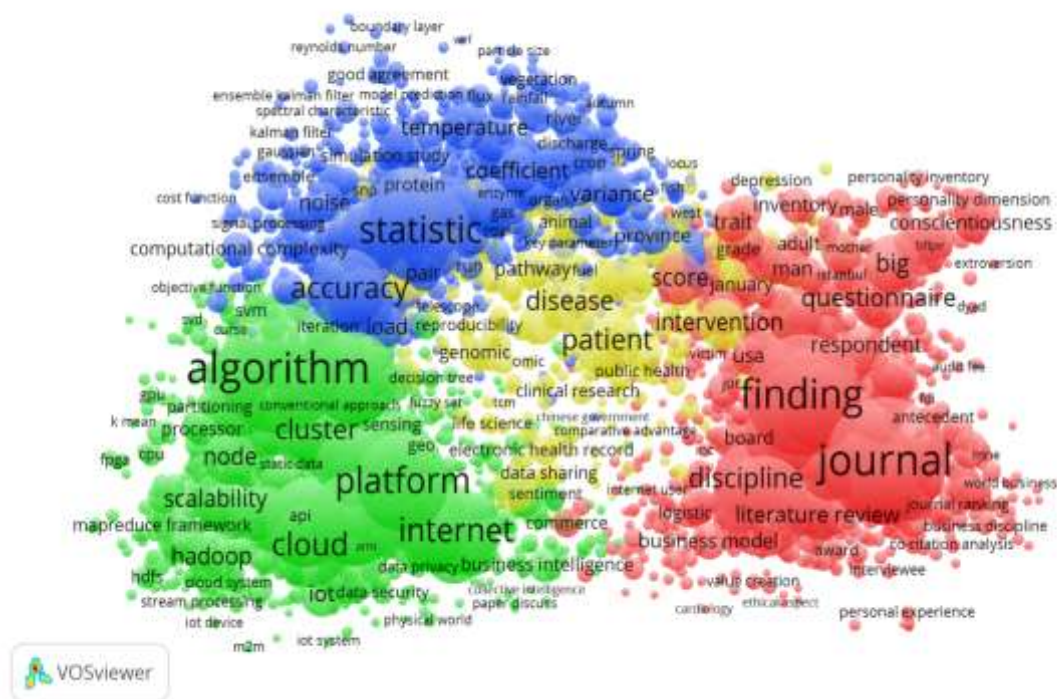
3. RESULTADOS

Com o auxílio do *software VOSviewer* foi possível detectar as principais palavras existentes nos títulos e nos resumos dos artigos das revistas selecionadas. As palavras foram detectadas com base em uma rede de co-ocorrência, como se pode observar na Figura 1. Por meio da técnica de agrupamento do *VOSviewer*, observa-se que quatro grupos foram formados a partir de 3.725 palavras identificadas algoritmicamente pelo *software*. Constatou-se ainda que as palavras apresentadas em cada *cluster* estão

relacionadas às diferentes áreas de pesquisa e uma parte tem forte relação com o tema *Big Data*.

As cores apresentadas na Figura 1 são aleatórias, fazendo apenas a separação dos grupos de palavra. Cada círculo representa um termo e apenas parte deles tem seu nome apresentado, pois, o *software*, para evitar sobreposições, identifica apenas alguns. O tamanho do círculo é referente a quantidade de aparições do termo. As palavras presentes nos *clusters* tem relação direta entre si, correspondendo ao fator de separação destes. O tamanho de cada palavra no *cluster* está relacionado ao seu peso, ou seja, a sua co-ocorrência nas publicações.

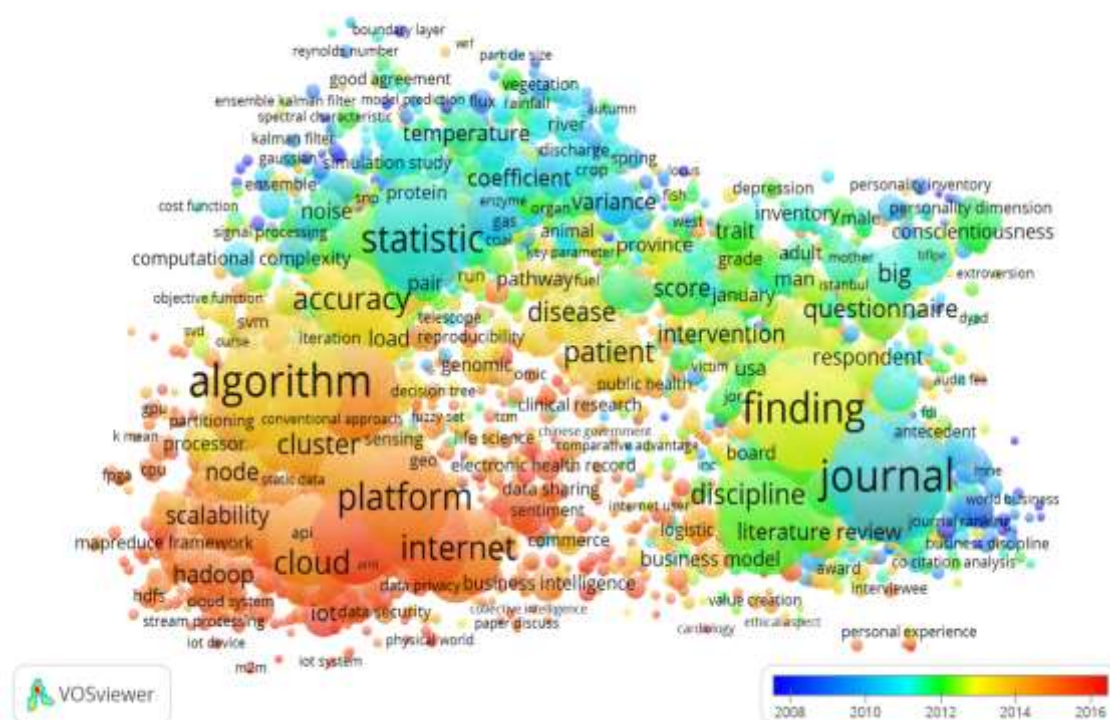
Figura 1: Mapa das principais palavras presente nos títulos e resumos da amostra



Fonte: VOSviewer (2017)

As palavras e termos com maior relação com o *Big Data* estão inseridas no grupo verde. No total, 53 termos *Big Data* foram listados pelo *software*. A busca foi realizada colocando o termo *Big Data* no filtro. Muitos termos são visíveis apenas quando a imagem é ampliada no *software*, como se pode observar nas Figuras 2 e 3.

ANÁLISE BIBLIOMÉTRICA DE PUBLICAÇÕES NA TEMÁTICA DO BIG DATA UTILIZANDO O VOSviewer

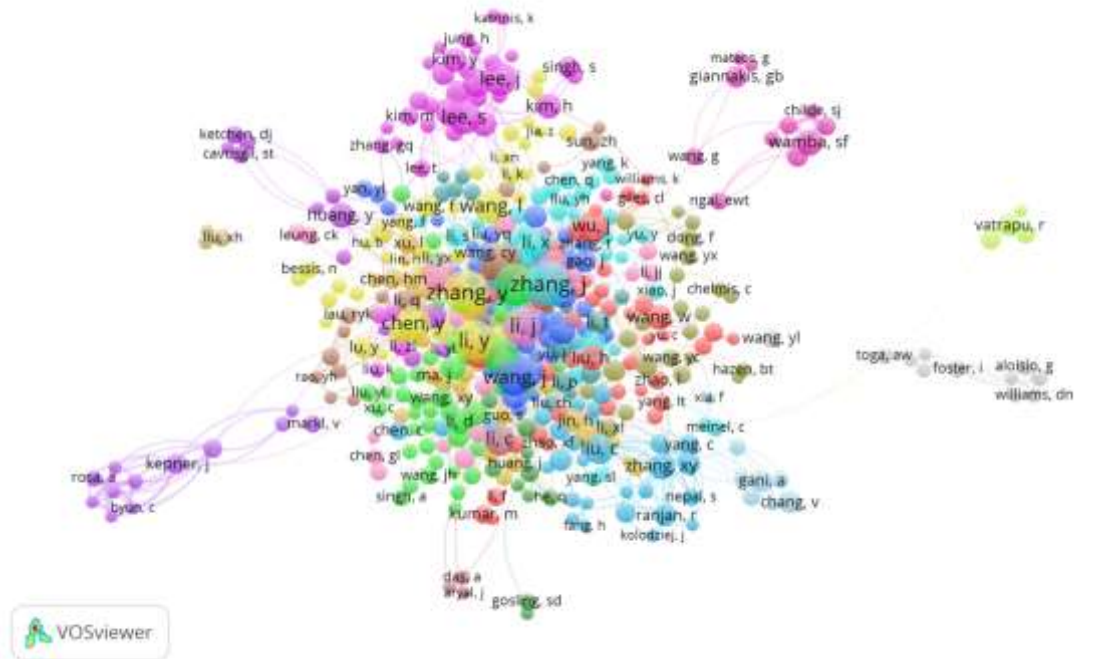


Fonte: VOSviewer (2017)

O mapa de densidade apresentado na Figura 5, indica que a proximidade dos termos e os pontos de cada termo, que são calculados com base na quantidade de vezes que o termo aparece é determinante na identificação da cor. Quanto maior a proximidade dos termos mais próximos da cor vermelha os itens estarão. Em contrapartida, quanto menor a proximidade dos termos mais próximos da cor azul os itens estarão.

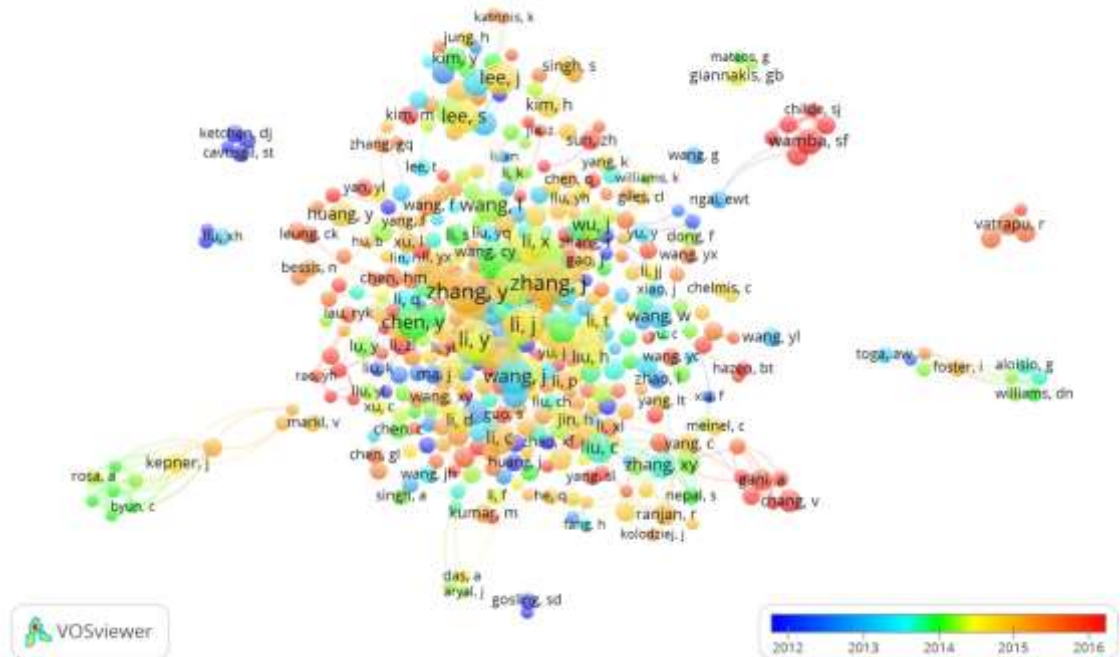
Figura 5: Mapa de densidade

ANÁLISE BIBLIOMÉTRICA DE PUBLICAÇÕES NA TEMÁTICA DO BIG DATA UTILIZANDO O VOSviewer



Fonte: VOSviewer (2017)

Figura 7: Mapa dos principais autores classificados por ano de publicação



Fonte: VOSviewer (2017)

4. CONCLUSÃO

O termo *Big Data* tem recebido bastante atenção nos últimos anos. Consequentemente é crescente o número de artigos que vem sendo publicados. Nesse contexto, o presente trabalho se propôs a analisar a ocorrência do termo *Big Data* nos artigos e identificar as palavras que vem sendo associadas a ele, além de identificar a relação entre *Big Data*, negócios e *softwares* de análise de dados.

A análise realizada com base nos periódicos selecionados, mostrou que há grande ocorrência de termos relacionados ao *Big Data* nas publicações dos últimos anos. A análise também mostrou que nesses periódicos pouco tem sido publicado sobre *Big Data* associado a negócios, que pode ser outra realidade se as buscas forem realizadas em outros periódicos.

Analizou-se também a que os principais autores não são aqueles com publicações mais recentes, mas os que possuem mais trabalhos, maior número de citações e força de ligação com outros autores.

Conclui-se que é crescente e recente a aparição do assunto nos periódicos e que o tema tem grande potencial para estudos e aplicações. Espera-se que o presente trabalho apresente alguma contribuição ao meio acadêmico e que mais estudos sejam realizados envolvendo o *Big Data* a área de negócios.

5. REFERÊNCIAS

- ADRIAN, C.; ABDULLAH, R.; ATAN, R.; JUSOH, Y. Y. Factors Influencing to the Implementation Success of Big Data Analytics: A Systematic Literature Review. **IEEE**, 2017.
- AKOKA, J.; COMYN-WATTIAU, I.; LAOUFI, N. Research on Big Data - A systematic mapping study. **Computer Standards & Interfaces**, v. 54, p. 105–115, 2017.
- ARAÚJO, C. A. Bibliometria: evolução histórica e questões atuais. **Em Questão, Porto Alegre**, v. 12, p. 11–32, 2006.
- CAFÉ, L.; BRÄSCHER, M. Organização da informação e bibliometria. **Revista Eletrônica de Biblioteconomia e Ciência da Informação**, p. 54–75, 2008.
- HILBERT, M. Big Data for Development: A Review of Promises and

- Challenges. **Development Policy Review**, v. 34, n. 1, p. 135–174, 2016.
- HUANG, Y.; ROBINSON, D. K. R.; YOUTIE, J.; et al. Big Data and Business : Tech mining to capture business interests and activities around Big Data. **IEEE**, p. 145–150, 2016.
- ISASI, N. K. G.; FRAZZON, E. M.; URIONA, M. Big Data and Business Analytics in the Supply Chain : A Review of the Literature. **IEEE**, v. 13, n. 10, p. 3382–3391, 2015.
- LAKSHEN, G. A.; VRANEŠ, S.; JANEV, V. Big Data and Quality : A Literature Review. **IEEE**, 2016.
- MCAFEE, A.; BRYNJOLFSSON, E. Big Data. The management revolution. **Harvard Business Review**, v. 90, n. 10, p. 61–68, 2012.
- MCKINSEY. Big data: The next frontier for innovation, competition, and productivity. **McKinsey Global Institute**, , n. June, p. 156, 2011.
- NELSON, B.; OLOVSSON, T. Security and Privacy for Big Data : A Systematic Literature Review. **IEEE**, p. 3693–3702, 2016.
- OLIVEIRA, S. C. M.; BARBOSA, E. S.; RESENDE, I. C. C.; SILVA, R. P. A.; ALBUQUERQUE, L. S. Bibliometria em artigos de contabilidade aplicada ao setor público. **Congresso Brasileiro de Custos**, 2013.
- RUSSOM, P. Big data analytics . **TWDI Best Practices Report**, , n. Fourth Quarter, p. 1–34, 2011.
- VOSviewer. www.vosviewer.com . Acesso em 10/10/2017