



UNIVERSIDADE FEDERAL DO RIO DE JANEIRO (UFRJ)  
FACULDADE DE ADMINISTRAÇÃO E CIÊNCIAS CONTÁBEIS  
DEPARTAMENTO DE ADMINISTRAÇÃO (FACC)

GUILHERME KOTCHERGENKO BATISTA

**ANÁLISE BIBLIOMETRICA SOBRE A PRESENÇA DO *BIG DATA* E *BUSINESS*  
*ANALYTICS* NO BRASIL**

Rio de Janeiro  
2021

**GUILHERME KOTCHERGENKO BATISTA**

**ANÁLISE BIBLIOMETRICA SOBRE A PRESENÇA DO *BIG DATA* E *BUSINESS*  
*ANALYTICS* NO BRASIL**

Monografia apresentada à Faculdade de Administração e Ciências Contábeis da Universidade Federal do Rio de Janeiro (FACC/UFRJ), como requisito parcial à obtenção do grau de Bacharel em Administração.

Orientador (a): Carlos César Trucíos Maza

Rio de Janeiro

2021

**GUILHERME KOTCHERGENKO BATISTA**

**ANÁLISE BIBLIOMETRICA SOBRE A PRESENÇA DO *BIG DATA* E *BUSINESS*  
*ANALYTICS* NO BRASIL**

Monografia apresentada à Faculdade de Administração e Ciências Contábeis da Universidade Federal do Rio de Janeiro (FACC/UFRJ), como requisito parcial à obtenção do grau em Bacharel em Administração aprovada por:

---

Prof. Carlos César Trucíos Maza (Orientador)

---

Profa. Cristina Pimenta De Mello Spinetti Luz (Leitora)

Rio de Janeiro, 2021

Dedico esse trabalho aos meus pais  
Carlos Eduardo de Souza Batista e Kátia Arruda Kotchergenko Batista que sempre  
batalharam para poder oferecer a mim e meus irmãos a melhor educação possível  
assim como o incentivo para buscá-la.

## **AGRADECIMENTOS**

Agradeço imensamente à minha família pelo suporte, oportunidades oferecidas e ao incentivo de meus estudos.

Agradeço aos meus amigos, em especial ao Guilherme Chamareli, Francesca Scalerandi, Gabriella Paredes, Gabriel Chianca pelo auxílio e palavras de incentivo durante o meu processo acadêmico.

Agradeço aos colegas de trabalho e de estudos, em especial o Armando Falcão e Carlos Trucíos que me apresentaram as possibilidades que existem por meio da programação e análise de dados.

Agradeço aos meus professores na UFRJ e no Anglo Americano pelo ensino humanizado e crítico, em especial à Ricardo Rohm, Thiago Vinícius Vieira e Nina Moreno Campello.

## RESUMO

Em um mundo gradativamente mais globalizado e digital, o *Big Data Analytics* se mostra cada vez mais interessante para as organizações e os tomadores de decisão. Com o início da 4ª Revolução Industrial e a possibilidade da automatização completa da produção nas indústrias, um alto volume de produção científica, relacionado com *Big Data Analytics*, tem sido divulgado na última década. Este trabalho realiza uma análise bibliométrica exaustiva e reproduzível com o objetivo de: (i) mapear a literatura científica existente relacionada com *Big Data Analytics* no Brasil; (ii) identificar quais são os autores, revistas científicas e trabalhos mais influentes; (iii) apresentar os principais temas estudados na literatura e como eles evoluíram e sugerem evoluir. A análise é realizada através do *software* R e utilizando a literatura identificada na base de dados *Web of Science* por meio das palavras-chave *Business Analytics*, *Big Data*, *Business Intelligence*, *Analytics* e *Brazil* como termos de busca, e a partir disto são realizadas as seguintes análises bibliométricas: estatística descritiva, por citação, por cocitação e por co-palavras. Os resultados indicam que desde 2017 há uma alta produção de literatura científica (majoritariamente em língua inglesa) com constantes recordes de produção anual, sugerindo que ao final do ano de 2021 será atingido um novo pico, demonstrando a crescente importância do *Big Data e Analytics* no contexto brasileiro. Entre as revistas científicas identificadas como de maior influência no escopo de pesquisa, são: *Decision Support Systems*, *Scientometrics*, *Environmental Research Letters*, *Ecological Applications* e *International Journal of Digital Earth*. Apesar da maior parte dos artigos identificados estarem em língua inglesa, o trabalho percebe que a maior parte da produção científica encontrada veio do próprio Brasil, e, posteriormente dos países da Europa, EUA e China, sendo a Europa e os EUA grandes parceiros de cooperação científica com o país. É percebido que Chen et al. (2012) e Barney (1991) são os artigos com maior influência sobre os documentos presentes na análise bibliométrica. Quanto à evolução do tema de pesquisa, observou-se uma grande transformação do que vem sendo estudado em *Big Data* e *Business Analytics*, em relação ao Brasil, nos últimos 11 (onze) anos. Os termos de estudo: gerenciamento, formação de modelos e informação foram substituídos por: biodiversidade, impacto, política, saúde e infraestrutura. A análise bibliométrica também indica que o tema de *Big Data* ainda possui muito a crescer e que ainda há espaço para ser desenvolvido na literatura científica relacionada com o Brasil.

Palavras-chave: *Big Data*; *Business Analytics*; *Business Intelligence*; *Analytics*; Brasil; Bibliometria.

## LISTA DE FIGURAS

Figura 1 - Os cinco estágios de extração de conhecimento.....	14
Figura 2 - Redes de colaboração entre países.....	29
Figura 3 - Historiógrafo da rede de citações diretas.....	30
Figura 4 - Redes de cocitações entre artigos.....	34
Figura 5 - Mapa evolutivo das palavras-chaves encontradas nos artigos.....	35
Figura 6 - Mapa temático das palavras-chaves dentro dos artigos entre 2010-2019...37	
Figura 7 - Mapa temático das palavras-chaves dentro dos artigos entre 2020-2021...38	

## LISTA DE GRÁFICOS

Gráfico 1 - Análise temporal dos artigos sobre <i>Big Data</i> na CI brasileira.....	16
Gráfico 2 - Evolução do número de publicações nos últimos 11 anos.....	23
Gráfico 3 - Distribuição do número de citações entre os artigos.....	24

## LISTA DE TABELAS

Tabela 1 - Distribuição por língua e tipo de artigo encontrado.....	21
Tabela 2 – Produção anual de artigos revisados.....	22
Tabela 3 - Distribuição anual de artigos por idioma.....	22
Tabela 4 - Informações sobre os 5 artigos mais citados globalmente.....	26
Tabela 5 - As 10 maiores revistas científicas por número de artigos e citações.....	27
Tabela 6 - As 10 maiores revistas científicas por número de citações por artigo.....	28
Tabela 7 - Os 10 artigos mais citados entre as citações locais.....	32
Tabela 8 - Os 10 artigos com maior índice de centralidade.....	33

## SUMÁRIO

<b>1 O PROBLEMA .....</b>	<b>6</b>
1.1 Introdução.....	6
1.2 Objetivo.....	7
1.3 Objetivos específicos .....	7
1.4 Delimitação.....	8
1.5 Relevância.....	8
<b>2 REFERENCIAL TEÓRICO.....</b>	<b>9</b>
2.1 <i>Big Data</i> .....	9
2.2 Analytics .....	10
2.3 Análise Bibliométrica.....	12
<b>3 METODOLOGIA .....</b>	<b>15</b>
3.1 Tipo de Pesquisa.....	15
3.2 População e Amostra.....	15
3.3 Coleta de Dados .....	15
3.4 Método de Análise .....	16
3.5 Limitações do Método .....	17
<b>4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS.....</b>	<b>18</b>
5.1 Análise por citação; .....	21
5.3 Análise por cocitação;.....	29
5.4 Análise por co-palavras.....	32
<b>6 CONSIDERAÇÕES FINAIS .....</b>	<b>36</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>39</b>

# 1 O PROBLEMA

## 1.1 Introdução

Ao longo dos anos, observa-se que a história econômica, cultural e política do ser humano é fortemente relacionada com a história das transformações tecnológicas (ANNUAL CONFERENCE & EXPOSITION, 2010). Da agricultura às navegações, das lamparinas à invenção da luz, das cartas ao telefone e a criação da internet, observa-se os impactos da tecnologia até hoje.

No século XVIII o poder do vapor na mecanização marca o início do alto volume de produção, com ele também surgem os navios e as locomotivas movidas a vapor que alteraram as dinâmicas existentes. Essa transformação tecnológica é denominada como a 1ª Revolução Industrial. Os séculos seguintes são marcados pelos descobrimentos dos métodos de geração e transmissão da eletricidade e da linha de montagem da produção. Com isso, a produção aumenta significativamente e os custos diminuem, possibilitando a redução de preços e criando o consumo de massa, surgindo assim a 2ª Revolução Industrial. A partir dos anos 70 o mundo é novamente transformado, desta vez, pela automação parcial da produção por meio da computação e robotização, surgindo assim, a 3ª Revolução Industrial (INDUSTRIAL, 2021). No final do século XX, um novo nível de organização e controle sobre toda a cadeia de valor dos produtos é gradativamente desenvolvido (2ND INTERNATIONAL CONFERENCE ON MATERIALS, MANUFACTURING AND DESIGN ENGINEERING, 2017). Esse momento é identificado como o início da 4ª Revolução Industrial, ou Indústria 4.0. Esse novo nível de organização e controle se diferencia da revolução anterior na possibilidade da automação completa da produção por meio da personalização e digitalização de produtos, serviços e processos (INDUSTRIAL, 2021).

Para que essa inovação seja possível é necessário que haja dados e que eles se conectem de forma eficiente (GÖLZER; CATO; AMBERG, 2021) surgindo então o termo *Big Data* que tem sido popularizado nos campos acadêmicos e industriais. Chang e Grady (2019) definem o termo *Big Data*, como: “a grande quantidade de dados encontrada em um mundo orientado pela digitalização, informação, conexão e sensores”, tendo suas características distintivas como o volume, velocidade e variedade (LANEY, 2001), sendo depois também acrescentadas a veracidade e o valor

(RAJARAMAN, 2016). Desta observação, nota-se a participação da *Internet of Things* que contribui para o aumento significativo do fluxo e quantidade dos dados gerados (KHAN; SILVA; JUNG, 2017; GANTZ; REINSEL, 2011). Em um mundo que progressivamente produz dados em escalas cada vez maiores, o *Big Data* traz a maior oportunidade em tecnologia desde o início da internet para as empresas e indivíduos (ZIKOPOULOS; EATON, 2011). No entanto, apenas o armazenamento de dados não é de grande valor sem a capacidade de sua análise e interpretação para assim convertê-los em informação, geração de conhecimento e vantagem competitiva. A análise de dados torna-se então parte essencial do trabalho dos tomadores de decisão (POWER et al., 2018), surgindo assim um novo campo de estudo denominado *Business Analytics*. O termo é entendido neste trabalho como “um processo de pensamento sistemático que aplica ferramentas e métodos computacionais qualitativos, quantitativos e mistos para analisar dados, gerar conhecimento e apoiar a tomada de decisão” (POWER et al., 2018). *Business Analytics* pode ser considerada uma evolução do *Business Intelligence*, entendido como: “uma ferramenta competitiva essencial que fornece informações de inteligência” (GILAD; GILAD, 1988). Frente a estes fatos, ao alto nível de abrangência e produção científica, este trabalho visa aumentar o entendimento sobre o uso destas tecnologias no Brasil por meio de uma análise bibliométrica.

## 1.2 Objetivo

Criar um panorama científico sobre o uso do *Big Data* e *Business Analytics* no Brasil por meio de uma exaustiva revisão bibliográfica dos artigos publicados em revistas científicas indexadas na base de dados *Web of Science* (WoS).

## 1.3 Objetivos específicos

Referente a pesquisa realizada na literatura científica, identificar:

- a) os autores mais influentes;
- b) os trabalhos mais influentes;
- c) as revistas científicas mais influentes;
- d) os principais temas abordados e como eles evoluem.

#### 1.4 Delimitação

O trabalho foca na revisão bibliográfica gerada a partir dos artigos e artigos de revisão, publicados em revistas científicas indexadas na base de dados *Web of Science* publicados no período compreendido entre 01 de janeiro de 2010 até o dia 27 de setembro de 2021. Foram considerados artigos escritos na língua inglesa ou portuguesa, que contenham, no título ou resumo, alguma das seguintes expressões: “*Business Analytics*”, “*Big Data*”, “*Business Intelligence*”, “*Analytics*”, além da palavra “Brazil”.

#### 1.5 Relevância

Ram, Zhangb e Koroniosc (2016) e Mariani et al. (2018) afirmam que há uma demanda ainda não atendida por pesquisas que tratam sobre o uso de *Big Data* e *Business Analytics* assim como das capacidades que elas podem trazer para as organizações. Com isso, este trabalho almeja auxiliar a comunidade dentro e fora da academia em relação ao estado da arte e ao panorama destas tecnologias sob a perspectiva da literatura científica no cenário brasileiro.

## 2 REFERENCIAL TEÓRICO

Através de um levantamento bibliográfico, buscou-se apresentar os principais conceitos deste trabalho de forma a contextualizar os temas que perpassam pela questão apresentada neste estudo. Iniciou-se pelo conceito de *Big Data* e suas possibilidades. Posteriormente, criou-se um entendimento em comum sobre o termo *Analytics*, seus processos e suas vantagens. Por fim, abordou-se o método de análise bibliométrica, seu uso no meio científico e suas vantagens.

### 2.1 *Big Data*

A obtenção de dados nunca foi tão valiosa quanto nos dias de hoje, tornando-se uma das maiores mercadorias comerciais (XIAO, et al.,2014). Isto se deve ao fato de que a grande quantidade de dados gerados tem o potencial de transformar os processos organizacionais, bem como fornecer maior suporte para a tomada de decisão em diferentes contextos, tais como a gestão de negócios e o marketing (GOPALKRISHNAN; STEIER, 2012).

Na literatura científica não há uma única definição para o termo *Big Data*, sendo algumas destas definições, as seguintes:

a) “Coleta, processamento e análise de dados enormes e contínuos” (NOBRE; TAVARES, 2017);

b) “Dados de alto volume, alta velocidade e/ou alta variedade; requer novas tecnologias e técnicas para capturar, armazenar e analisar estes dados, e é usado para melhorar a tomada de decisão, fornecer conhecimento, e otimizar processos.” (Mills et al.,2012);

c) “a grande quantidade de dados encontrada em um mundo orientado pela digitalização, informação, conexão e sensores.” (CHANG; GRADY, 2019);

d) Ambiente no qual diversos tipos de dados são armazenados em grande quantidade, sendo estes dados vindos de várias fontes e, fornecendo informações

rápidas e precisas para a tomada de decisão (*2ND INTERNATIONAL CONFERENCE ON MATERIALS, MANUFACTURING AND DESIGN ENGINEERING*, 2017);

e) “É o conceito de dados, em que, por meio de ferramentas e tecnologias clássicas, é difícil coletar, armazenar, gerenciar e processar.” (*IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS (ICC)*, 2017)

Embora não exista unanimidade sobre a definição de *Big Data* na literatura, há um entendimento entre os diferentes autores referente às suas características intrínsecas, sendo elas definidas por um modelo 5V que consiste no volume, velocidade, variedade, veracidade e valor; sendo as três primeiras características definidas por Laney (2001) e as duas últimas introduzidas pela empresas IBM e Oracle, respectivamente, e incluídas na definição de Rajaraman (2016). De acordo com Gandomi e Haider (2014), o volume indica o tamanho total dos dados, a velocidade refere-se à rápida e constante criação de novos dados para serem armazenados e analisados, a variedade define a diversidade entre as diferentes fontes e tipos de dados (estruturados, não estruturados ou semiestruturados), a veracidade indica a incerteza inerente aos dados volumosos e, por fim, o valor descreve o ganho obtido após o processamentos e análise de dados, com o qual pode-se tomar decisões e iniciar ações.

## 2.2 Analytics

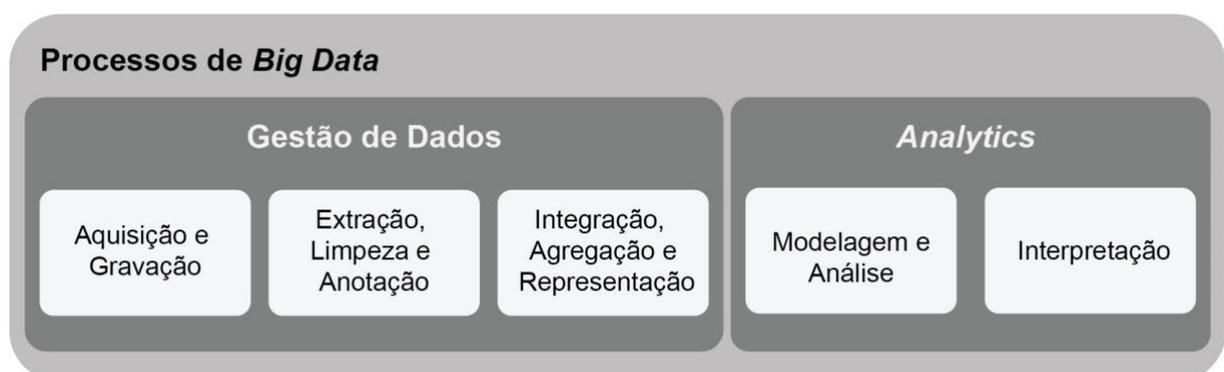
Nas últimas duas décadas, o *Business Intelligence* e *Analytics*, que são campos relacionados de análise de *Big Data*, tem ganhado importância na comunidade acadêmica e empresarial (CHEN et al., 2012). Neste trabalho, entende-se *Business Intelligence* como: “análise dos dados críticos de negócios, com a finalidade de auxiliar a organização a melhor entender suas necessidades e facilitar a tomada de decisão por meio de técnicas, tecnologias, sistemas, práticas, metodologias e aplicativos” (CHEN et al., 2012). De acordo com Ahmad (2015), a literatura de gestão estratégica enfatiza o conceito de *Business Intelligence* como uma ferramenta competitiva essencial na tomada de decisão nos negócios.

Referente ao termo *Analytics*, Armaturo e Aragona (2019) sugerem que a prática não exige necessariamente técnicas computacionais complexas, mas métodos adequados e habilidades específicas. A análise de dados é a etapa mais importante para se obter conhecimento a partir do *Big Data* e obter embasamento para a tomada de decisão (SILVA; DIYAN; HAN, 2019), uma vez que, sua análise inadequada pode levar a conclusões enganosas (RAJARAMAN, 2016). De acordo com Hu et al. (2014), a prática pode ser dividida em seis principais áreas de pesquisa: análise de dados estruturados, análise de texto, análise da Web, análises de rede sociais, análises de multimídia e análises de dados móveis.

O objetivo da análise de dados pode variar em diferentes domínios, mas abordagens tais como: visualização de dados, aprendizado de máquina, redes neurais, análise estatística, reconhecimento de padrões, processamento de sinais e mineração de dados, são benéficas e convenientes para uso em todos os domínios (KHAN et al., 2018).

De acordo com Labrinidis e Jagadish (2012) o processo geral de extração de conhecimento de *Big Data* pode ser dividido em cinco estágios apresentados na Figura 1, sendo eles divididos em dois subprocessos: *Data Management* e *Analytics*. O primeiro corresponde a aquisição de dados, armazenamento, preparação e recuperação para análises. O segundo corresponde aos diversos métodos, técnicas e algoritmos que extraem inteligência do *Big Data*, podendo também ser considerado um subprocesso da tecnologia (SILVA; DIYAN; HAN, 2019).

Figura 1 – Os cinco estágios de extração de conhecimento



Fonte: SILVA; DIYAN; HAN (2019)

O processo de transformação dos dados para a criação de algum valor organizacional é representado pelo termo *Big Data Analytics*, definido como “onde as técnicas analíticas avançadas operam em *Big Data*” (RUSSOM, 2011) ou como “descobrir conhecimento valioso em grandes conjuntos de dados, impondo uma tomada de decisão inteligente” (SILVA, DUYAN e HAN; 2019). Seu uso, traz vantagens competitivas para organizações (KUBINAA et al, 2015), indo além dos negócios, tendo exemplos no combate contra a pandemia de SARS-Cov2 (YANG et al., 2020; BRAGAZZI et al., 2020), na implantação de cidades inteligentes (NUAIMI et al., 2015), melhorias na administração pública (MANYIKA et al., 2011), combate à criminalidade (MINISTÉRIO DA JUSTIÇA E SEGURANÇA PÚBLICA, 2019), entre outros.

Como mencionado, uma grande quantidade de dados é gerada por organizações e comunidades científicas. Desta forma, o uso de *Big Data Analytics* auxilia empresas de pequeno, médio e grande porte na tomada de decisões (INTERNATIONAL CONFERENCE ON COLLABORATION TECHNOLOGIES AND SYSTEMS (CTS), 2013).

Apesar das vantagens e avanços fornecidos pelo *Big Data Analytics*, há controvérsias sobre o uso de dados e como estes são obtidos, uma vez que é possível registrar dados do usuário sem consentimento prévio. A fim de mitigar este problema, órgãos regulamentadores como a GDPR (*General Data Protection Regulation*) na União Europeia e a LGPD (Lei Geral de Proteção de Dados) no Brasil, buscam estabelecer normas para respeitar a privacidade do usuário.

### 2.3 Análise Bibliométrica

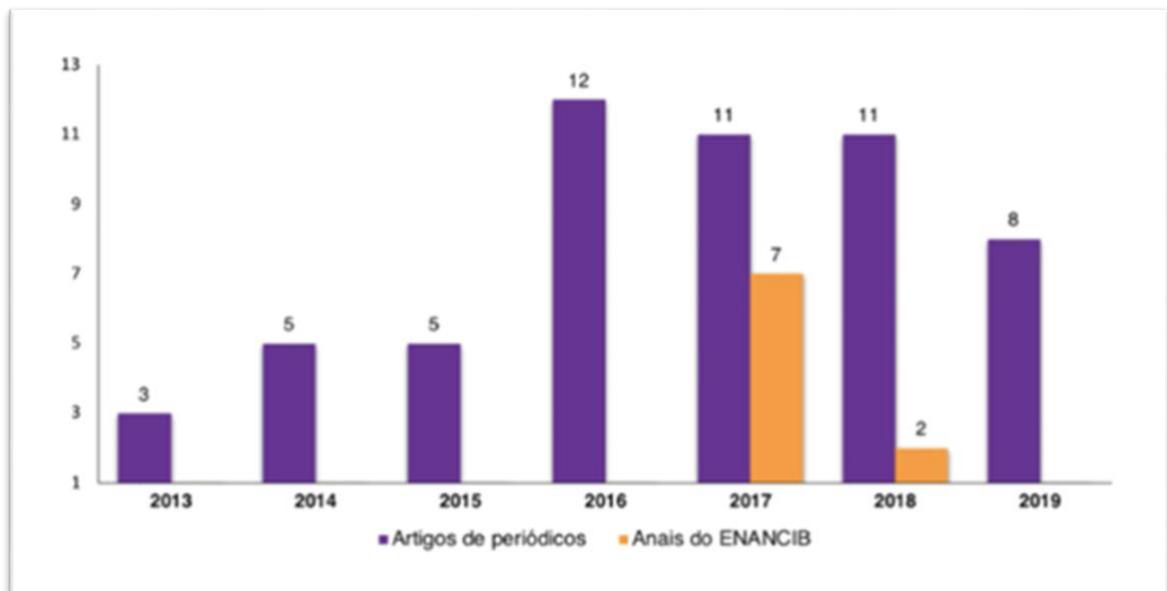
A análise bibliométrica refere-se a um dos métodos para abordar a revisão de literatura a fim de organizar e compreender descobertas anteriores; ela é utilizada quando uma extensa quantidade de materiais científicos é analisada (ARIA; CUCCURULLO, 2017). Este método já se encontra ativamente presente na produção científica, como por exemplo, nos trabalhos de Maçaira et al. (2018), Vargas et al. (2019) e Yoshida (2010).

O potencial da bibliometria é apresentar uma revisão sistemática, transparente e reproduzível, permitindo criar análises mais objetivas e confiáveis (BROADUS, 1987; DIODATO, 1994).

A bibliometria torna-se útil a este trabalho pois, mesmo com a grande quantidade de artigos existentes, permite: realizar de forma mais prática uma análise estruturada; identificar tendências e mudanças nos limites das temáticas; inferir tendências ao longo do tempo; mapear temas pesquisados; identificar os autores e instituições mais prolíficos na área (CRANE, 1973).

Santos-D'Amorim et al. (2020) identifica que somente 64 artigos foram publicados entre os anos 2013 e 2019 sobre o tema de *Big Data*, no campo de Ciências da Informação do Brasil, sendo sua distribuição anual como apresentado pelo gráfico 1:

Gráfico 1 – Análise temporal dos artigos sobre *Big Data* na CI brasileira



Fonte: SANTOS-D'AMORIM et al. (2020)

Uma vez que, a Ciência da Informação é o principal campo de estudo sobre esta tecnologia, nota-se que ainda há espaço de aprofundamentos dentro dos temas. Também se observa que inicialmente não foram encontradas informações que forneçam uma análise bibliométrica sobre o uso dessa tecnologia que se refiram ao Brasil.

Para a criação de um mapa científico sobre os temas estudados neste trabalho é realizado um fluxo de trabalho padrão definido por Zupic e Čater (2014) que consiste em cinco etapas: Design de estudo; Coleta de dados; Análise de dados; Visualização de dados; Interpretação.

O design de estudo define as perguntas a serem respondidas e o método bibliométrico adequado. Na etapa de coleta, os dados desejados são captados e filtrados na sua fonte para a realização de sua posterior análise. Por último, é escolhido um método de visualização para apresentar as interpretações realizadas (ZUPIC; ČATER, 2014).

Dentre tipos de análise mais comuns a serem feitas, encontram-se: análise por acoplamento bibliográfico, análise por cocitação, análise por coautor e análise por co-palavras (ARIA; CUCCURULLO, 2017).

### 3 METODOLOGIA

#### 3.1 Tipo de Pesquisa

Para a classificação da pesquisa utiliza-se a perspectiva apresentada por Vergara (1991) que observa os fins e os meios. A pesquisa pode ser classificada, quanto aos fins, como exploratória, uma vez que, embora o tema seja de grande impacto e abrangência na atualidade, ainda se observa a oportunidade de sistematizar o conhecimento existente em relação ao cenário brasileiro.

Quanto aos meios, realizou-se a pesquisa bibliográfica a partir da extração de artigos científicos na base de dados *Web of Science*, utilizando-se uma pesquisa avançada sobre os temas.

#### 3.2 População e Amostra

Neste trabalho é entendido como população o conjunto de elementos que possuem a característica de serem objetos de estudo (VERGARA, 1998). Assim, neste trabalho, a população objeto de estudo refere-se aos artigos científicos existentes na literatura que contenham as palavras "*Big Data*", "*Business Intelligence*", "*Business Analytics*" "*Analytics*" e "*Brazil*" no título e/ou no resumo, sejam eles escritos na língua inglesa ou portuguesa. Com relação a amostra, composta por 137 artigos, ela se configura como não probabilística por tipicidade, uma vez que foram selecionados unicamente artigos indexados na base de dados *Web of Science*, uma das maiores e mais representativas bases de dados na comunidade científica.

#### 3.3 Coleta de Dados

Foram coletados na *Web of Science* artigos científicos e artigos de revisão escritos em inglês ou português através da seguinte consulta: ("*Business Analytics*" OR "*Big Data*" OR "*Business Intelligence*" OR "*Analytics*") AND "*Brazil*". Ou seja, artigos que contenham uma ou mais das expressões "*Business Analytics*", "*Big Data*", "*Business Intelligence*", ou "*Analytics*" simultaneamente o termo "*Brazil*"; todos presentes no título ou resumo do artigo. Nesta pesquisa, considerou-se todos os artigos publicados desde 01 de janeiro de 2010 até o dia 27 de setembro de 2021.

### 3.4 Método de Análise

Optou-se neste trabalho pela utilização do pacote *bibliometrix* desenvolvido por Aria e Cuccurullo (2017), que é uma ferramenta de código aberto na linguagem R (R Core Team, 2016) para executar uma análise abrangente de mapeamento da literatura científica existente. O seu uso foi programado para ser flexível e facilitar a integração com outros pacotes estatísticos e gráficos, obtendo-se também a flexibilidade de ser rapidamente atualizado e integrado (ARIA; CUCCURULLO, 2017).

Esta ferramenta permite importar dados bibliográficos coletados na Web of Science e, após o filtro do material coletado, é possível realizar análises bibliométricas, com o uso de métodos como matrizes de dados para cocitação, acoplamento, análise de colaboração científica e análise de co-palavras (ARIA; CUCCURULLO, 2017).

O processo de filtragem dos documentos coletados neste trabalho consiste em remover (i) artigos que não sejam escritos em inglês ou português; (ii) documentos que não sejam artigos científicos ou de revisão; (iii) artigos remanescentes que não sejam sobre o tema pesquisado, por meio da leitura das palavras-chaves e resumos.

Para o processo de análise bibliométrica foram utilizados três tipos de métodos, sendo eles: análise por citação; análise por cocitação; análise por co-palavras.

O primeiro método refere-se a esclarecer quais são os autores e revistas científicas mais influentes na literatura sobre o tema pesquisado, isto acontece por meio da apresentação do número de citações e número de artigos publicados por autor.

O segundo método refere-se à análise de cocitação, que esclarece quais são os trabalhos mais influentes na literatura científica sobre o tema pesquisado. A análise por cocitação, ao comparada de acordo com o tempo, tem a vantagem de auxiliar na detecção de mudança de paradigmas e escolas de pensamento. Ela mapeia artigos em diferentes períodos.

O terceiro método tem como objetivo esclarecer quais são os principais temas estudados na literatura e sua evolução. Nele é observado quais são as palavras e palavra-chave mais importantes nos documentos a fim de identificar a estrutura conceitual do campo pesquisado (CALLON et al., 1983).

Após as análises, criou-se formas de visualização que melhor expressassem os resultados.

### 3.5 Limitações do Método

Uma das limitações de pesquisa é a amplitude de artigos científicos existentes na base de dados *Web of Science*. Apesar da base de dados *Web of Science* ser uma das mais representativas da comunidade acadêmica, existem ainda múltiplas bases que também possuem artigos dentro do critério de busca de pesquisa.

Outra limitação é sobre a existência de artigos que abordam o assunto de *Big Data* ou *Big Data Analytics*, porém utilizam outros termos no título e/ou resumo, tais como “*Forecasting*”, “*Prediction*”, “*Machine Learning*”, “*Statistical Learning*” e outros similares. Desta forma, estes não foram considerados no escopo da pesquisa.

## 4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Através do critério de busca "(\"Business Analytics\" OR \"Big Data\" OR \"Business Intelligence\" OR \"Analytics\") AND \"Brazil\" na base de dados *Web of Science*, foram encontrados 252 artigos, dos quais 107 foram removidos por não possuírem características compatíveis com o objetivo do estudo<sup>1</sup>. Os critérios para remover os artigos foram os seguintes: (a) escritos em línguas diferentes da portuguesa ou inglesa (4 artigos), (b) artigos que não são considerados artigos científicos ou de revisão (75 artigos), e (c) artigos que, após uma análise do resumo e palavras-chave, não se enquadram no escopo deste trabalho (36 artigos). Restaram então 137 artigos que constituem a amostra a ser analisada. A distribuição por língua e tipo de artigo encontra-se na Tabela 1:

Tabela 1 – Distribuição por língua e tipo de artigo encontrado

Tipo de Artigo	Idioma		Total
	Inglês	Português	
<b>Científico</b>	113 (82%)	14 (10%)	127 (95%)
<b>De revisão</b>	9 (18%)	1 (1%)	10 (5%)
<b>Total</b>	122 (92%)	15 (8%)	<b>137</b>

Fonte: Elaboração do autor

A Tabela 1 revela que a principal configuração documental encontrada no escopo de pesquisa realizado na base de dados *Web of Science* é de artigos científicos escritos em língua inglesa, correspondendo a 82% do tipo de documento encontrado.

Com a presença de 92% dos artigos em língua inglesa, é observada predominância do idioma entre as publicações científicas encontradas. Também é observado a existência de poucos artigos de revisão publicados entre 01 de janeiro de 2010 e 27 de setembro de 2021 apesar do pequeno aumento nos últimos anos (Tabela 2):

---

<sup>1</sup> Criar um panorama científico sobre o uso do *Big Data* e *Business Analytics* no Brasil ao identificar: quais são os autores e trabalhos mais influentes; quais são os principais temas e como eles evoluem.

Tabela 2 – Produção anual de artigos revisados

Ano	Nº de Artigos de revisão
2017	1
2018	1
2019	2
2020	3
2021	3

Fonte: elaboração do autor

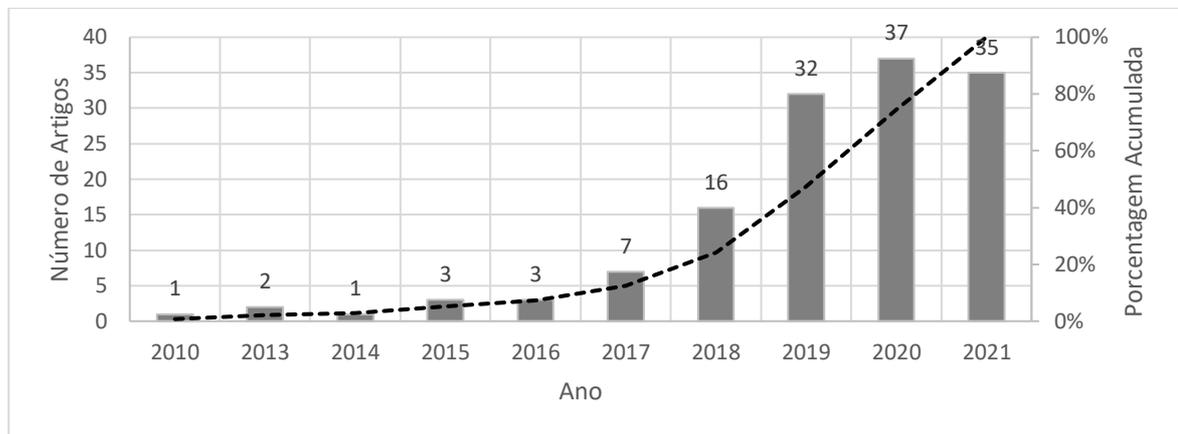
Abaixo, a Tabela 3 apresenta que artigos publicados em português se mantiveram com nível de produção estável, com exceção ao aumento de 2020. Por outro lado, a produção de artigos em inglês aumenta expressivamente desde 2018.

Tabela 3 – Distribuição anual de artigos por idioma

Ano	Inglês	Português	Total Geral
2010	1	0	1
2013	1	1	2
2014	1	0	1
2015	1	2	3
2016	2	1	3
2017	5	2	7
2018	14	2	16
2019	30	2	32
2020	33	4	37
2021	34	1	35

Fonte: elaboração do autor

Gráfico 2 – Evolução do número de publicações nos últimos 11 anos



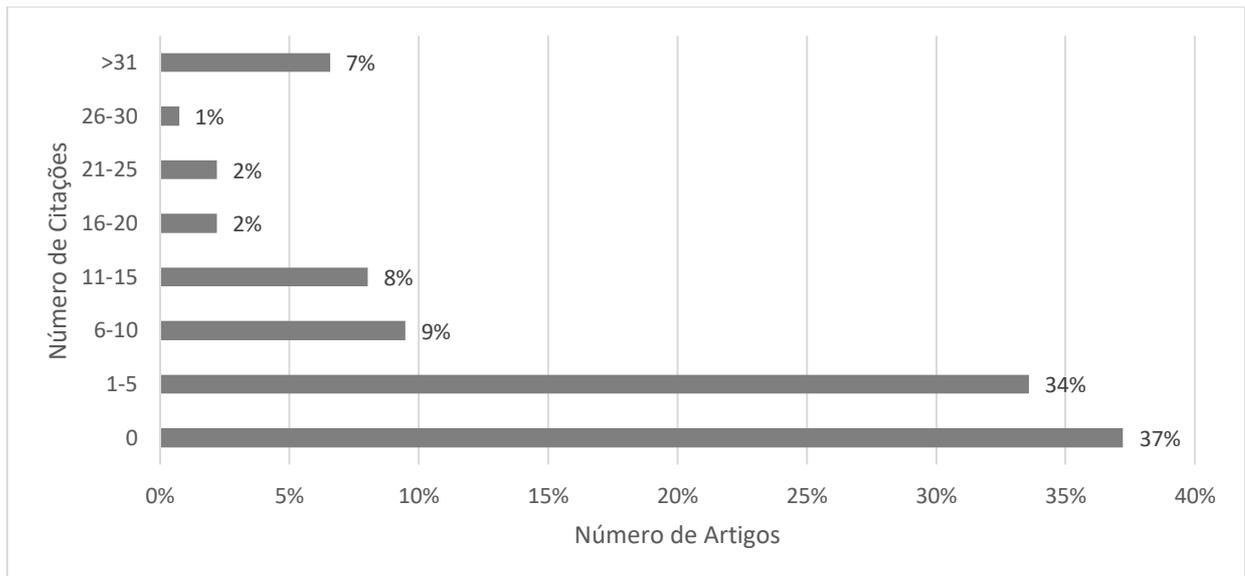
Fonte: Elaboração do autor

A evolução do número de artigos científicos tem acompanhado a popularidade dos termos *Big Data* e *Analytics*, que se tornam cada vez mais populares nas mídias e nas organizações. O Gráfico 2 apresenta a evolução do número de publicações nos últimos 11 anos, sendo observado o aumento significativo dos estudos sobre os termos utilizados neste trabalho. O aumento torna-se considerável a partir de 2017, com crescimento anual da produção e pico registrado em 2020, com 37 artigos publicados.". Os dados de 2021, são referentes às publicações realizadas até o 27 de setembro de 2021, e espera-se que, ao finalizar o ano, o número de publicações supere os valores obtidos em 2020, gerando um novo máximo.

Observou-se que de 2018 a 2021 foram publicados 120 artigos (87,59% dos estudos) e o restante publicado anteriormente a 2018. Isto representa uma média aproximada de 12 artigos publicados por ano. O constante aumento da produção científica sobre os temas relacionados com *Big Data* e *Analytics* ao longo dos anos evidencia a crescente importância percebida tanto por parte dos pesquisadores, quanto nas discussões sobre Big Data e Analytics no contexto brasileiro.

### 5.1 Análise por citação;

Gráfico 3 – Distribuição do número de citações entre os artigos



Fonte: Elaboração do autor

O Gráfico 3 apresenta a distribuição de artigos por número de citações e é possível concluir que 37% dos artigos resultantes da pesquisa não possuem citações, sendo a maior parte deles publicados entre 2020 e 2021. Em sequência, 34% dos artigos possuem entre uma e cinco citações, sendo a maior parte deles artigos publicados entre 2018 e 2020. Também é apresentado que 9% dos artigos têm entre 6 a 10 citações, representado majoritariamente por artigos publicados no ano de 2019. Seguindo, 8% se encontram entre 11 e 15 citações, sendo representados majoritariamente por publicações dos anos de 2019 e 2020. Entre 16 e 20 citações se encontraram 2% dos artigos, sendo eles publicados entre os anos 2019-2021. Observou-se que 2% obtiveram entre 21 e 25 citações, tendo sido publicados entre 2019 e 2020. Entre 26 e 30 citações houve somente 1%, publicado em 2020. Por fim, 7% dos artigos obtiveram mais de 31 citações, estando bem distribuídos entre os anos de 2010 e 2020, tendo destaque somente em 2017.

Os cinco artigos com maior número de citação, são:

- I) Trkman et al. (2010): Este artigo investiga a relação entre as capacidades analíticas na área de planejamento, origem, fabricação e entrega da cadeia de suprimentos e seu desempenho usando o suporte do sistema de informações e a orientação do processo de negócios como moderadores. Os resultados sugerem a existência de uma relação estatisticamente significativa entre capacidades analíticas e desempenho organizacional.
- II) Nobre e Tavares (2017): compreender o *status quo* dos estudos realizados globalmente e ter um panorama geral da análise da literatura científica sobre BD e aplicações da IoT na economia circular. Os resultados indicam que China e EUA são os países mais interessados na área e revelam um contexto com oportunidades significativas para pesquisa. Além disso, grandes produtores de emissões de gases de efeito estufa, como Brasil e Rússia, ainda carecem de estudos na área.
- III) Harris et al. (2017): Apresentam um método que combina ferramentas analíticas de *Big Data* com *Emerging Hot Spot Analysis*<sup>2</sup> para identificar tendências espaço-temporais estatisticamente significativas de perda florestal no Brasil, Indonésia e República Democrática do Congo entre 2000 e 2014. Os resultados indicam que, embora a taxa geral de perda florestal no Brasil diminuiu no período de 14 anos, os padrões espaço-temporais de perda mudaram, com a perda florestal diminuindo significativamente nos estados amazônicos de Mato Grosso e Rondônia e intensificando-se no bioma cerrado. Na Indonésia, a perda de floresta se intensificou na província de Riau e nas regências de Sukamara e West Kotawaringin. Embora os resultados se concentrem na identificação de tendências significativas em escala nacional, também é demonstrado a escalabilidade da abordagem para regiões menores ou maiores.
- IV) Levin, Kark e Crandall (2015): Desenvolveram uma nova abordagem para quantificar a presença humana além de áreas populosas, combinando *Big Data* de mídia social e ferramentas de sensoriamento remoto. Usando essa

---

<sup>2</sup> Tipo de análise conduzida para investigar tendências no espaço ao longo do tempo.

abordagem, os autores identificam locais frequentes de fotografia fora dos centros urbanos, sendo locais, em grande parte, ecologicamente desprotegidos, como o Pantanal do Brasil e o Salar de Uyuni da Bolívia. A abordagem de *Big Data* identifica pressões e lacunas de conservação espacial em tempo real e sua variação espacial e temporal globalmente.

- V) Horita et al. (2017): Apresentam um estudo no contexto da gestão de desastres no Brasil que aplica oDMN+, uma estrutura que conecta a tomada de decisão com fontes de dados por meio de uma notação de modelagem estendida e um processo de modelagem. Os resultados revelam uma abordagem eficaz para melhorar a compreensão de como potencializar o *Big Data* na tomada de decisões da organização.

O número de citações, as instituições de filiação dos autores, bem como a revista científica de onde o artigo foi publicado encontra-se na Tabela 4:

Tabela 4 – Informações sobre os 5 artigos mais citados globalmente

Autores	Filiação	Ano Publicado	Revista	Citações Globais
TRKMAN, Peter; MCCORMACK, Kevin; DE OLIVEIRA, Marcos Paulo Valadares; MADEIRA, Marcelo Bronzo	Univ. Ljubljana; Univ. Fed. Minas Gerais	2010	Decision Support Systems	203
NOBRE, Gustavo Cattelan; TAVARES, Elaine	Univ. Fed. Rio De Janeiro	2017	Scientometrics	113
HARRIS, Nancy et al	World Resources Inst.;Univ. Maryland	2017	Environmental Research Letters	74
LEVIN, Noam; KARK, Salit; CRANDALL, David	Hebrew Univ. Jerusalem; Univ. Queensland; Univ. Queensland; Indiana Univ.	2015	Ecological Applications	63
HORITA, Flávio; DE ALBUQUERQUE, João Porto; MARCHEZINI, Victor; MENDIONDO, Eduardo	Univ. São Paulo; Univ. Warwick; Univ. São Paulo; Brazilian Natl. Ctr. Monitoring and Early Warning Nat. ; Heidelberg Univ.	2017	Decision Support Systems	50

Fonte: Elaboração do autor

Considerando o escopo de artigos encontrados por este trabalho na *Web of Science* até 27 de setembro de 2021, encontrou-se que os artigos foram publicados em 124 diferentes revistas científicas. A Tabela 5 apresenta as 10 maiores revistas por número de artigos e citações identificados no escopo. São apresentadas também suas respectivas classificações CAPES e Scimago.

Tabela 5 - As 10 maiores revistas científicas por número de artigos e citações

Revista Científica	Nº de Artigos	Nº de Citações	Qualis CAPES	Scimago Quarter
JOURNAL OF CLEANER PRODUCTION	3	49	A1	Q1
SUSTAINABILITY	3	4	A1	Q1/Q2
DECISION SUPPORT SYSTEMS	2	253	A1	Q1
EPIDEMIOLOGIA E SERVIÇOS DE SAÚDE	2	11	B2	Q2
THEORETICAL AND APPLIED CLIMATOLOGY	2	11	A3	Q2
RAE-REVISTA DE ADMINISTRAÇÃO DE EMPRESAS	2	6	A2	Q3/Q4
RBGN-REVISTA BRASILEIRA DE GESTÃO DE NEGÓCIOS	2	5	A3	Q3
PERSPECTIVAS EM CIÊNCIA DA INFORMAÇÃO	2	2	A2	Q2/Q3/Q4
REVISTA BRASILEIRA DE MARKETING	2	2	A2	Não encontrado
MATHEMATICS	2	2	Não encontrado	Não encontrado

Fonte: Elaboração do autor

Mesmo que estes jornais representem 16% dos artigos publicados na pesquisa, é observado que não há uma concentração do número de publicações por revista. Uma vez que não há uma diferença significativa na quantidade de artigos publicados, foi também analisado quais são as 10 maiores revistas científicas por número de citações por artigo. A análise é apresentada na Tabela 6:

Tabela 6 - As 10 maiores revistas científicas por número de citações por artigo

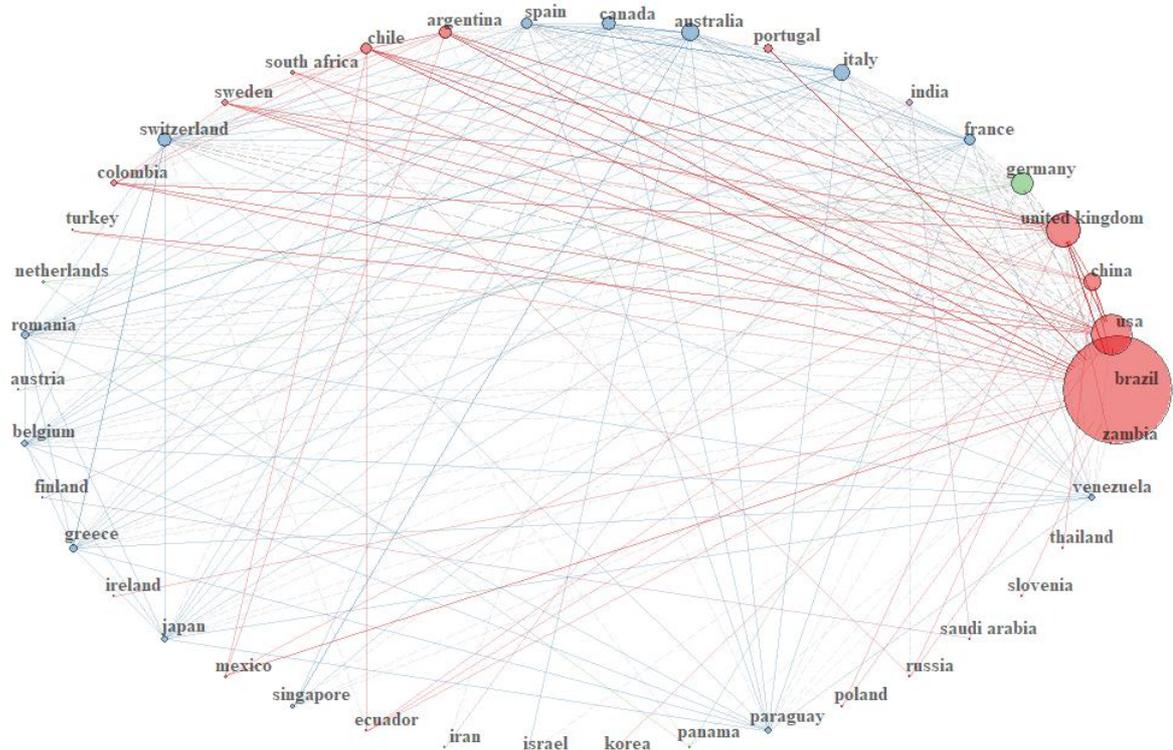
Revista Científica	Citações p/ Artigo
DECISION SUPPORT SYSTEMS	127
SCIENTOMETRICS	113
ENVIRONMENTAL RESEARCH LETTERS	74
ECOLOGICAL APPLICATIONS	63
INTERNATIONAL JOURNAL OF DIGITAL EARTH	42
INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT	41
COMPUTERS IN INDUSTRY	35
JOURNAL OF WEB SEMANTICS	34
NEW MICROBES AND NEW INFECTIONS	29
JOURNAL OF ALLERGY AND CLINICAL IMMUNOLOGY	24

Fonte: Elaboração do autor

Estas análises permitem identificar quais são as revistas mais influentes nestes campos. Uma vez o número de artigos publicados não foi um critério de diferença significativa, considerou-se como principal critério o número de citações por artigo, destacando-se entre as dez maiores, as revistas *Decision Support Systems*, *Scientometrics*, *Environmental Research Letters*. É também interessante mencionar que dois dos artigos mais citados: Trkman et al. (2010) e Horita et al. (2017), possuem 203 e 50 citações respectivamente, e foram publicados na *Decision Support Systems*.

É também observado que a produção científica vinda do Brasil (45%) foi a mais encontrada, sendo seguida pela da Europa (21%), do EUA (14%) e da China (8%), sendo o restante (12%) correspondente a outros países com aproximadamente 1% cada. É natural que a maior parte das publicações sejam oriundas do Brasil, já que uma das palavras-chave de busca foi o nome do próprio país. Em relação aos artigos publicados no exterior, esses números podem ser explicados pelas diferentes frentes de estudo que foram consideradas, sendo exemplos: artigos sobre as tecnologias mencionadas que usam o Brasil como um caso (LEVIN; KARK; CRANDALL, 2015); estudo sobre organizações que se encontram no território brasileiro (TRKMAN ET AL., 2010). Outra explicação para os artigos publicados no exterior é pela rede de cooperação entre os países, apresentada na Figura 2:

Figura 2 - Redes de colaboração entre países

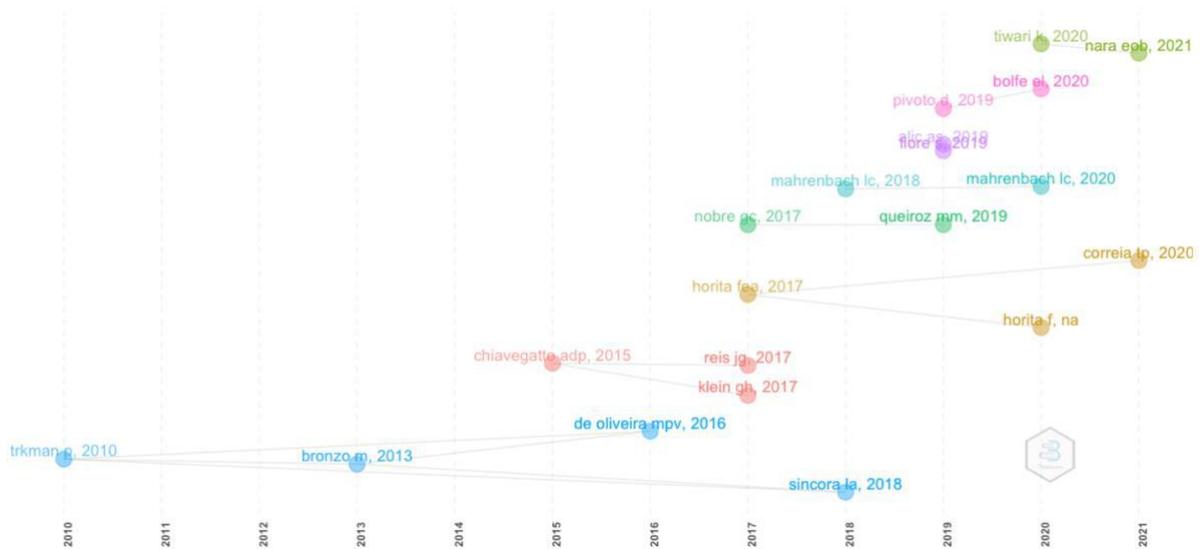


Fonte: Elaboração do autor

A Figura 2 apresenta a rede de colaboração de pesquisa existente entre diversos países, representando o tamanho do círculo a proporção de participação do país.

A partir da Figura 2, é possível observar que os principais países com colaboração da produção científica estudada neste trabalho são: Reino Unido, Estados Unidos, Austrália e Alemanha. Os únicos países que não possuíram colaboração com o Brasil foram: Singapura, Grécia e Turquia.

Figura 3 - Historiografo da rede de citações diretas



Fonte: Elaboração do autor

A análise por citação é apresentada pela Figura 3, onde é construído um historiografo da rede de citações diretas entre os 137 artigos analisados, o que permite observar as conexões intelectuais realizadas entre os artigos. Um historiografo se baseia em citações diretas dos diversos autores. Elas são usadas para reconstruir a estrutura historiográfica do campo, destacando as obras básicas nele contidas e, então, traçando as ligações intelectuais (divididas em cores) em uma ordem cronológica.

Desta forma, a evolução das principais ligações intelectuais é apresentada a seguir, conforme divisão por cor.

- a) Azul – Trkman et al. (2010), Bronzo et al. (2013), De Oliveira et al. (2016) e Sincora et al. (2018):

O primeiro artigo discute a relação estatisticamente significativa entre capacidades analíticas e desempenho organizacional. Sendo continuado pelo segundo artigo em que os resultados mostram que a orientação do processo de negócios, indicadores analíticos e de desempenho podem ser tomados como preditores de desempenho. O terceiro artigo indica que a orientação analítica pode alavancar a inovação do processo, mencionando a importância da qualidade da informação, além de apontar a relevância de aspectos para orientação analítica. Por fim, os resultados do último artigo confirmam as hipóteses teóricas propostas de que

as capacidades analíticas organizacionais e a maturidade da gestão de processos de negócios impactam positivamente a resiliência organizacional.

b) Laranja - Chiavegatto (2015), Klein, Neto e Tezza (2017) e Reis et al. (2017):

O primeiro artigo discute sobre os dilemas, oportunidades e desafios das tecnologias digitais, em especial, *Big Data*, *Internet of Things* e técnicas de análise de dados na área da saúde. O segundo artigo avalia se o monitoramento de mídias sociais pode ser uma ferramenta de previsão de padrões epidemiológicos; seus resultados mostram que há forte correlação entre as variáveis e que o monitoramento pode ser usado como um modelo preditivo por profissionais da saúde e gestores públicos. O terceiro artigo descreve os assuntos abordados e a abrangência geográfica das instituições de vínculo dos autores dos artigos publicados no Informe Epidemiológico do SUS (IESUS) e na Revista Epidemiologia e Serviços de Saúde.

c) Vermelho – Horita (2017) e Correia, Corsi e Quintanilha (2020)

O primeiro artigo apresenta um estudo no contexto da gestão de desastres no Brasil. Ele aplica uma estrutura que conecta a tomada de decisão com fontes de dados por meio de uma notação de modelagem estendida e um processo de modelagem, resultando em uma abordagem eficaz para melhorar a compreensão de como potencializar o *Big Data* na tomada de decisões da organização. O segundo artigo realiza uma revisão bibliográfica sobre a relação de desastres naturais, técnicas de *Big Data* e ferrovias, em nível internacional e nacional.

d) Verde – Nobre e Tavares (2017) e Queiroz e Pereira (2019):

O primeiro artigo visa ter um panorama geral da análise da literatura científica sobre *Big Data* e aplicações da *Internet of Things* na economia circular. Os resultados indicam que China e EUA são os países mais interessados na área e revelam um contexto com oportunidades significativas para pesquisa. Além disso, grandes produtores de emissões de gases de efeito estufa, como Brasil e Rússia, ainda carecem de estudos na área. O segundo artigo avalia as variáveis que influenciam a intenção dos profissionais brasileiros de logística em adotar o *Big Data*. Ele revela que as condições facilitadoras, como infraestrutura de TI, têm uma grande influência na

intenção de adotar *Big Data*. No entanto, a influência social e a expectativa de desempenho não mostraram efeito significativo.

Tabela 7: Os 10 artigos mais citados entre as citações locais.

Rank	Artigo	Revista Científica	Citações Locais	Citações Globais
1	TRKMAN P, 2010	DECISION SUPPORT SYSTEMS	3	203
2	HORITA FEA, 2017	DECISION SUPPORT SYSTEMS	2	50
3	BRONZO M, 2013	INT. JOURNAL OF INFORMATION MANAGEMENT	2	41
4	CHIAVEGATTO ADP, 2015	EPIDEMIOLOGIA E SERVICOS DE SAUDE	2	10
5	NOBRE GC, 2017	SCIENTOMETRICS	1	113
6	TIWARI K, 2020	JOURNAL OF CLEANER PRODUCTION	1	23
7	PIVOTO D, 2019	INT. FOOD AND AGRIBUSINESS MNG. REVIEW	1	10
8	ALIC AS, 2019	FUTURE GENERATION COMPUTER SYSTEMS	1	5
9	MAHRENBACH LC, 2018	THIRD WORLD QUARTERLY	1	4
10	BERSAN SAL, 2013	REVISTA DE SAUDE PUBLICA	0	5

Fonte: Elaboração do autor

A Tabela 7 apresenta os artigos mais citados localmente<sup>3</sup> e globalmente<sup>4</sup>. Desta tabela é possível observar o destaque da *Decision Support Systems*, que além de possuir artigos com maior quantidade de citações locais, também possui a maior quantidade de citações globais.

### 5.3 Análise por cocitação;

Utilizando as referências dos artigos observados na análise bibliográfica, foi observado quais documentos eram citados em comum por artigos em análise deste trabalho. Com as referências dos artigos em análise, é criado um índice de centralidade por proximidade que indica quais artigos mais estão influenciando os artigos em análise bibliométrica. Ou seja, quanto maior o índice de centralidade, mais influente o artigo referenciado é. Desta forma, a Tabela 8 apresenta os dez artigos

<sup>3</sup> citações que uma referência foi recebida de documentos incluídos na análise bibliográfica

<sup>4</sup> citações recebidas por um artigo selecionado "em todo o mundo"

com maior índice de centralidade de proximidade (ICP) entre os artigos analisados nesse estudo.

Tabela 8: Os 10 artigos com maior índice de centralidade.

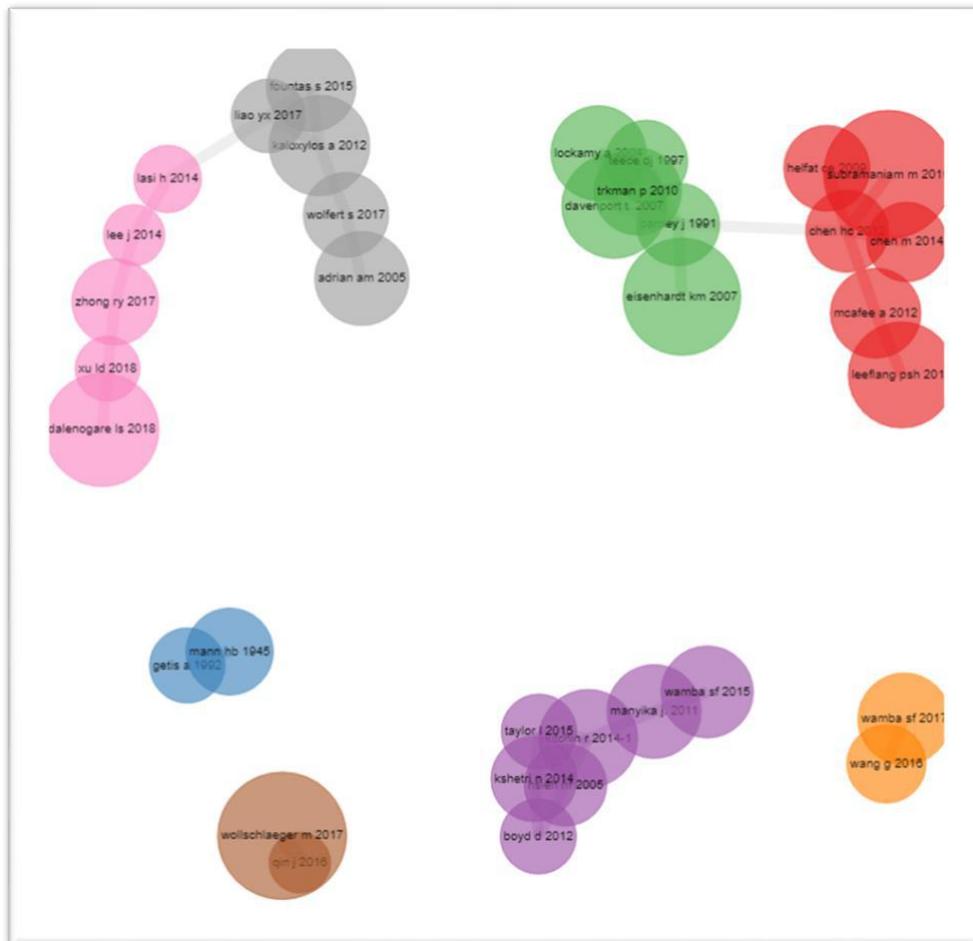
Rank	Artigo	Revista Científica	ICP
1	CHEN; 2012	MIS Quarterly	0,001148
2	BARNEY; 1991	Journal of Management	0,001148
3	DAVENPORT; 2007	Alta Books	0,001110
4	TEEC; 1997	Strategic Management Journal	0,001110
5	TRKMAN; 2010	Decision Support Systems	0,001110
6	MCAFEE; 2012	Harvard Business Review	0,001098
7	CHEN; 2014	Mobile Networks and Applications	0,001086
8	HELFAT; 2009	Strategic Organization	0,001086
9	SUBRAMANIAM; 2019	Business Horizons	0,001086
10	EISENHARDT; 2007	Academy of Management Journal	0,001086

Fonte: Elaboração do autor

Desta forma, é percebido que Chen et al. (2012) e Barney (1991) são os artigos com maior influência sobre o material presente nesta análise bibliográfica. Dos artigos, Chen et al. (2012) fornece uma estrutura que identifica a evolução, aplicações e áreas de pesquisa emergentes além de realizar um estudo bibliométrico de publicações importantes em *Business Intelligence & Analytics*, pesquisadores e tópicos de pesquisa com base em mais de uma década de publicações acadêmicas e industriais relacionadas. Barney (1991) observa a ligação entre os recursos da empresa e a vantagem competitiva sustentável por meio da análise de recursos organizacionais disponíveis. O artigo conclui examinando as implicações desse modelo de recurso para vantagem competitiva sustentada em outras disciplinas de administração.

É interessante apontar que Trkman (2010) também é identificado nesta tabela. Isto se deve ao número de citações locais e o índice de centralidade por proximidade terem uma relação positiva, logo, artigos altamente citados localmente também teriam um alto nível de centralidade.

Figura 4 - Redes de cocitações entre artigos.



Fonte: Elaboração do autor

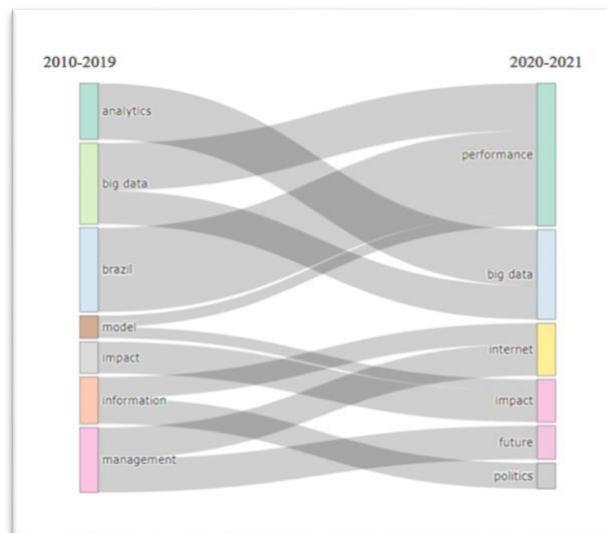
Acima, a Figura 4 apresenta a relação de cocitação realizada entre os artigos. Nela há os 35 artigos que mais foram usados como referência entre os artigos observados na nossa bibliometria. Esses 35 artigos foram divididos em oito grupos diferentes de acordo com as palavras-chave mais utilizadas. A partir desse método é possível indicar a linha de raciocínio utilizada.

É também observado que entre os artigos pertencentes à análise bibliométrica, somente Trkman (2010) aparece.

#### 5.4 Análise por co-palavras

Para a criação de um panorama temático, foi necessário dividir o período estudado em diferentes intervalos de tempo, a fim de comparar as estruturas conceituais e analisar a evolução dos tópicos ao longo do tempo. Para identificar a estrutura conceitual e a evolução dos principais conceitos dos termos pesquisados, separou-se os artigos selecionados para análise bibliométrica em dois períodos: 2010-2019 e 2020-2021, sendo este último encerrado no dia 27 de setembro de 2021. Nota-se que no primeiro período separado (2010-2019) foram publicados 65 artigos, enquanto no segundo (2020-2021) foi de 72. Após a separação dos períodos, foi analisada a ocorrência das palavras chaves presentes dentro dos artigos nos períodos correspondentes.

Figura 5 - Mapa evolutivo das palavras-chaves encontradas nos artigos



Fonte: Elaboração do autor

Na Figura 5 é analisado cada período separadamente pois, assim, é possível rastrear a evolução temporal das pesquisas científicas. As áreas em cinza indicam que os termos compartilham uma associação em comum. Enquanto isso, o tamanho da caixa representa a quantidade de palavras-chaves associadas com cada termo.

Observando a Figura 5, percebe-se que o mapa evolutivo se inicia com 7 termos, diminuindo para 6 no período final. Essa diminuição indica que as áreas pesquisadas neste trabalho estão em um processo de convergência. Um exemplo

disso é o termo “*analytics*” atrelando-se ao “*Big Data*” no período de 2020-2021, mesmo período que o termo “*Big Data Analytics*” começa a se tornar mais comum (Figura 7).

De 2010-2019 para 2020-2021, os termos são reorganizados da seguinte forma: “*analytics*” migra para “*big data*”; “*big data*” e “*Brazil*” se juntam na vertente “*performance*”; “*information*” se divide em “*internet*” e “*politics*”; “*management*” se divide em “*future*” e “*internet*”; “*model*” se divide em “*impact*” e “*performance*”. Desta forma, quase todos se modificam, mantendo-se somente os termos “*big data*” e “*impact*” que ainda são indicados a terem interesse crescente.

Além da convergências dos termos, é interessante observar que eles se modificam, já que somente “*big data*” e “*impact*” se mantem. Essa diferenciação pode ser explicada pelo grande aumento de artigos científicos neste segundo período e a tomada de novos caminhos de estudo. Para apresentar essas modificações foram formuladas as Figuras 6 e 7.

De acordo com Cobo et al. (2010), o mapa temático exemplificado na Figura 6 permite analisar os temas de acordo com quatro quadrantes que estão inseridos:

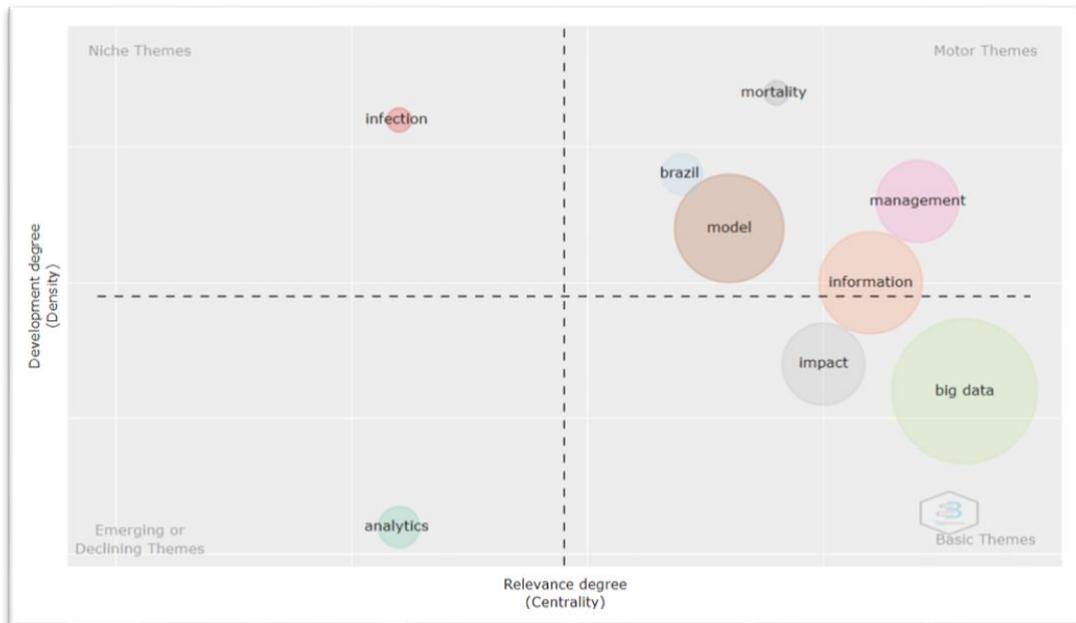
(1) quadrante superior direito: termos motores, que são termos bem desenvolvidos e importantes para a construção do campo científico;

(2) quadrante inferior direito: termos básicos e transversais, termos importantes, mas que não estão ainda bem desenvolvidos;

(3) quadrante inferior esquerdo: termos emergentes ou em desaparecimento, que são termos pouco desenvolvidos e marginais; por ser uma palavra-chave genérica e convergindo para outros termos preferido por outros autores;

(4) quadrante superior esquerdo: termos muito desenvolvidos, mas que se encontram afastados da construção do campo científico.

Figura 6 - Mapa temático das palavras-chaves dentro dos artigos entre 2010-2019



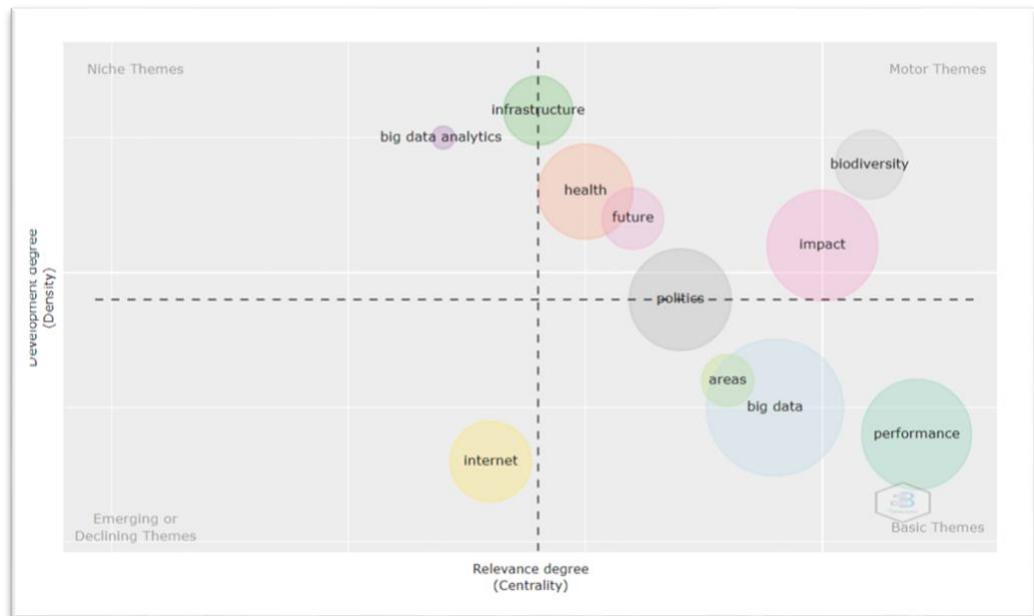
Fonte: Elaboração do autor

Desta forma, é possível identificar na Figura 6 que os termos “Brazil”, “model”, “mortality”, “management” e “information” são tanto termos relevantes para a literatura científica quanto termos em alta na produção científica da época, sendo importantes para a estruturação de um campo de pesquisa.

Também é possível observar que os termos “impact” e “big data” são termos de alta relevância para o campo de pesquisa mas ainda não estão bem desenvolvidos.

Sobre o termo “infection”, é indicado que tem se afastado da literatura científica. Enquanto isso, em relação ao termo “analytics”, é indicado que é um termo que está a desaparecer por falta de uso ou por convergir em termos que os autores preferem, como, por exemplo “big data analytics”, demonstrado na Figura 7.

Figura 7 - Mapa temático das palavras-chaves dentro dos artigos entre 2020-2021



Fonte: Elaboração do autor

Além disso, observa-se como o “*big data*” e “*impact*” ainda são termos de importância na literatura científica desde o início de 2010, sendo o termo “*big data*” com maior necessidade de ser desenvolvido na produção científica.

Na Figura 7 observa-se que os termos “*biodiversity*”, “*impact*”, “*politics*”, “*future*”, “*health*” e “*infrastructure*” são tanto termos relevantes para a literatura científica quanto termos de alta na produção científica da época, sendo importantes para a estruturação de um campo de pesquisa.

Também é possível observar que os termos “*performance*”, “*big data*” e “*áreas*” são termos de alta relevância para o campo de pesquisa, mas ainda não estão bem desenvolvidos. Nota-se que “*big data*” aparece nas duas figuras anteriores, sendo um termo que desde o início de 2010 tem maior necessidade de ser desenvolvido na produção científica.

Sobre o termo “*big data analytics*”, é indicado que tem se afastado da literatura científica. Enquanto isso, em relação ao termo “*internet*”, é indicado que é um termo que está a desaparecer por falta de uso ou por convergir em algum futuro termo que os autores preferem.

## 6 CONSIDERAÇÕES FINAIS

Em um mundo cada vez mais digitalizado e global, a Indústria 4.0 ganha cada vez mais espaço, sendo protagonizada pelo *Big Data* e *Business Analytics*. Com o aumento contínuo do nível de produção e relevância do tema na produção científica, este trabalho investigou o estado da arte e o panorama do tema de *Big Data* e *Business Analytics* na produção científica que envolva o Brasil. Isto, com a finalidade de adquirir uma maior compreensão e diminuir as possíveis lacunas de conhecimento para as organizações, sejam de âmbito privados ou públicas.

Para atender os objetivos pretendidos de buscar esclarecer quais são os autores, revistas científicas e trabalhos mais influentes, além de apresentar quais são os principais temas estudados e como eles evoluem na literatura científica, o estudo realiza um uma exaustiva revisão bibliográfica dos artigos publicados em revistas científicas e uma análise bibliométrica.

A revisão foi realizada tendo em vista artigos científicos e de revisão, escritos na língua portuguesa ou inglesa e que estejam indexados na base de dados *Web of Science* até a data de 27 de setembro de 2021. Os artigos considerados para a revisão bibliográfica incluem os termos *Business Analytics*, *Big Data*, *Business Intelligence*, *Analytics* e o Brasil como tópicos de pesquisa.

A amostra final foi constituída de 137 artigos, os quais foram utilizados para a realização da análise bibliométrica sobre a qual foram conduzidas análises de: estatística descritiva, citações, cocitações e co-palavras, por meio do pacote *bibliometrix*, programado em linguagem de programação R.

A respeito das principais conclusões obtidas, foi observado inicialmente que o tipo de configuração documental mais encontrada no escopo de pesquisa foi de artigos científicos escritos em língua inglesa, correspondendo a 82% do tipo mais encontrado no escopo de pesquisa.

Nota-se o significativo aumento da produção científica dos termos no Brasil desde 2017 e com a tendência de 2021 superar o pico de 37 artigos publicados no ano anterior. Também é observado que 87,5% dos artigos encontrados entre 2010 e 27 de setembro de 2021 foram somente publicados após o ano de 2018, representando uma média aproximada de 12 artigos publicados por ano. Esse constante aumento na produção científica dos temas relacionados com *Big Data* e

*Analytics* nos recentes anos evidencia a crescente importância, recebida por parte dos pesquisadores, quando à discussão sobre *Big Data* e *Analytics* no contexto brasileiro.

Na análise estatística descritiva também foi identificado que 71% dos artigos presentes no escopo de pesquisa possuem menos do que 5 citações, enquanto 7% dos artigos encontrados possuem mais do que 30 citações. Entre os artigos com maior influência devido às citações foram identificados os destaques de Trkman et al. (2010), Nobre e Tavares (2017), Harris et al. (2017), Levin, Kark e Crandall (2015) e Horita et al. (2017). As revistas científicas com maior influência por critério de citações por número de artigos encontrados no escopo de pesquisa são: *Decision Support Systems*, *Scientometrics*, *Environmental Research Letters*, *Ecological Applications* e *International Journal of Digital Earth*, sendo todas, com exceção da última, relacionadas aos artigos mencionados terem maior influência.

Apesar da maioria dos artigos encontrados estarem em língua inglesa, é também observado que a maior parte da produção científica encontrada veio do próprio Brasil, correspondendo à 45% da origem dos artigos publicados, seguido da Europa (21%), dos EUA (14%) e da China (8%). O que vai de acordo com as palavras-chave de busca e os resultados obtidos sobre os países que mais cooperam para a pesquisa científica com o Brasil, sendo eles: Reino Unido, EUA, Austrália e Alemanha.

Na análise por cocitação é percebido que Chen et al. (2012) e Barney (1991) são os artigos com maior influência sobre o material identificado na análise bibliográfica, além de Trkman (2010), mencionado anteriormente, novamente aparecer como um dos com maior influência.

Na análise por co-palavras foi criada o mapa evolutivo sobre os temas mais trabalhados dentro dos artigos, onde foi observado que houve uma grande mudança sobre o que antes era estudado em *Big Data* e *Business Analytics* e o que é estudado hoje. Das mudanças foram observadas principalmente: a alta relevância e aumento de estudos sobre as tecnologias e sua relação com a biodiversidade, impacto, política, saúde e infraestrutura. A análise também sugere que o tema de *Big Data* ainda possui muito a crescer e que ainda não foi bem desenvolvido na literatura científica relacionada com o Brasil.

Das possíveis limitações de pesquisa considera-se a escolha dos termos de busca, cuja delimitação não permitiu a identificação de artigos com outros termos, tais como “*Forecasting*”, “*Prediction*”, “*Machine Learning*”, “*Statistical Learning*” e outros

similares. Outra limitação apontada é a restrição a uma base dados (*Web Of Science*). Para trabalhos futuros, é interessante ocorrer a extensão deste trabalho para expandir o escopo de estudos e enriquecer as análises.

## REFERÊNCIAS BIBLIOGRÁFICAS

AHMAD, Azizah. Business Intelligence for Sustainable Competitive Advantage. **Advances in Business Marketing and Purchasing**, [s. l.], October 2015.

ARCHAMBAULT, É; CAMPBELL, D; GINGRAS, Y; LARIVIÈRE, V. Comparing bibliometric statistics obtained from the Web of Science and Scopus. **Journal of the American Society for Information Science and Technology**, [s. l.], v. 60, ed. 7, p. 1320-1326, 13 abr. 2009.

ARIA, Massimo; CUCCURULLO, Corrado. Bibliometrix: An R-tool for comprehensive science mapping analysis. **Journal of Informetrics**, [s. l.], 27 ago. 2017.

ARMATURO, Enrica; ARAGONA, Biagio. Methods for big data in social sciences. **Mathematical Population Studies**, [S. l.], p. 65-68, 13 maio 2019.

BARNEY, Jay. Firm Resources and Sustained Competitive Advantage. **Journal of Management**, [s. l.], v. 17, n. 1, p. 99-120, 1 mar. 1991.

BRAGAZZI, Nicola Luigi; DAI, Haijiang; DAMIANI, Giovanni; BEHZADIFAR, Masoud; MARTINI, Mariano; WU, Jianhong. How Big Data and Artificial Intelligence Can Help Better Manage the COVID-19 Pandemic. **Int. J. Environ. Res. Public Health 2020**, [s. l.], v. 17, n. 7, 2 maio 2020.

BROADUS, R. Toward a definition of bibliometrics. **Scientometrics**, [S. l.], p. 373-379, 16 jan. 1987.

BRONZO, Marcelo *et al.* Improving performance aligning business analytics with process orientation. **International Journal of Information Management**, [s. l.], v. 33, ed. 4, p. 300-307, 9 jan. 2013.

CALLON, Michel; COURTIAL, Jean-Ierre; TURNER, William A.; BAUIN, Serge. From translations to problematic networks: An introduction to co-word analysis. **Social Science Information**, [s. l.], v. 22, ed. 2, p. 191-235, 1 mar. 1983.

CHANG, Wo L.; GRADY, Nancy. NIST Big Data Interoperability Framework: Volume 1, Definitions. **National Institute of Standards and Technology**, [S. l.], p. 10-53, 21 out. 2019.

CHEN, Hsinchun; CHIANG, Roger H. L.; STOREY, Veda C. Business Intelligence and Analytics: From Big Data to Big Impact. **MIS Quarterly**, [s. l.], v. 36, n. 4, p. 1165-1188, December 2012.

CHIAVEGATTO, Alexandre. Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. **Epidemiologia e Serviços de Saúde**, [s. l.], v. 24, Jun 2015.

COBO, M.J.; LÓPEZ-HERRERA, A.G.; HERRERA-VIDEIRA, E.; HERRERA, F. An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. **Journal of Informetrics**, [s. l.], v. 5, n. 1, p. 146-166, 6 out. 2010.

CORREIA, Thais; CORSI, Alessandra; QUINTANILHA, Jose. Big Data for Natural Disasters in an Urban Railroad Neighborhood: A Systematic Review. **Smart Cities**, [s. l.], p. 202-211, 1 abr. 2020.

CRANE, Diana. INVISIBLE COLLEGES: DIFFUSION OF KNOWLEDGE IN SCIENTIFIC COMMUNITIES. **Social Forces**, [s. l.], v. 52, ed. 1, p. 135-136, 1 set. 1973.

DE OLIVEIRA, Marcos Paulo Valadares; CAVALCANTI, Claudia Xavier; LADEIRA, Marcelo Bronzo; MCCORMACK, Kevin P. Leveraging Process Innovation with Business Analytics. **International Journal of Decision Support System Technology**, [s. l.], v. 8, ed. 3, p. 63-75, 2016.

DIODATO, V. **Dictionary of bibliometrics**. [S. l.: s. n.], 1994.

GANDOMI, Amir; HAIDER, Murtaza. Beyond the hype: Big data concepts, methods, and analytics. **International Journal of Information Management**, [S. l.], v. 35, n. 2, p. 137-144, 3 dez. 2014. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0268401214001066>. Acesso em: 18 ago. 2021.

GANTZ, John; REINSEL, David. Extracting Value from Chaos. **IDC's Digital Universe Study**, [S. l.], p. 1-12, 1 jun. 2011.

GILAD, Benjamin; GILAD, Tamar. **The Business Intelligence System: A New Tool for Competitive Advantage**. [S. l.: s. n.], 1988.

GÖLZER, Phillipp; CATO, Patrick; AMBERG, Michael. Data Processing Requirements of Industry 4.0: Use Cases for Big Data Applications. **ECIS**, [S. l.], p. 1-12, 29 maio 2021.

GOPALKRISHNAN, Vivekanand; STEIER, David. Big data, big busines: bridging the gap. **Association for Computing Machinery**, [s. l.], p. 7-11, 2012.

HARRIS, Nancy *et al.* Using spatial statistics to identify emerging hot spots of forest loss. **Environmental Research Letters**, [s. l.], v. 12, n. 2, 7 fev. 2017.

HORITA, Flávio; DE ALBUQUERQUE, João Porto; MARCHEZINI, Victor; MENDIONDO, Eduardo. Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in Brazil. **Decision Support Systems**, [s. l.], v. 97, p. 12-22, Mai 2017.

HU, H; WEN, Y; CHUA, T S; LI, X. Toward scalable systems for big data analytics: a technology tutorial. **IEEE**, [s. l.], p. 652-687, 2014.

IEEE INTERNATIONAL CONFERENCE ON COMMUNICATIONS (ICC), 2017, Paris, France. **Big data challenges and opportunities in the hype of Industry 4.0 [...]**. [S. l.: s. n.], 2017.

INDUSTRIAL Revolution - From Industry 1.0 to Industry 4.0. [S. l.], 2021. Disponível em: <https://www.desouttertools.com/industry-4-0/news/503/industrial-revolution-from-industry-1-0-to-industry-4-0>. Acesso em: 16 abr. 2021.

KHAN, M; SILVA, BN; JUNG, C. A context-aware smart home control system based on ZigBee sensor network. **KSII Transactions on Internet and Information Systems**, [s. l.], v. 11, 27 fev. 2017.

KHAN, M et al. Efficiently processing big data in real-time employing deep learning algorithms. In: DEEP learning innovations and their convergence with big data. [S. l.: s. n.], 2018. p. 61-78.

KLEIN, Gisiela; NETO, Pedro; TEZZA, Rafael. Big Data e mídias sociais: monitoramento das redes como ferramenta de gestão. **Saúde e Sociedade**, [s. l.], v. 26, p. 208-217, Jan 2017.

KUBINAA, Milan; VARMUS, Michal; KUBINOVA, Irena. Use of big data for competitive advantage of company. **Procedia Economics and Finance**, Ephesus, Turkey, v. 26, p. 561-565, 17 out. 2015.

LABRINIDIS, Alexandros; JAGADISH, H. V. Challenges and opportunities with big data. **Proceedings of the VLDB Endowment**, [S. l.], p. 2032-2033, 1 ago. 2012.

LANEY, D. 3D Data Management: Controlling Data Volume, Velocity, and Variety. **META Group Research Note**, [s. l.], 6 fev. 2001.

LEVIN, Noam; KARK, Salit; CRANDALL, David. Where have all the people gone? Enhancing global conservation using night lights and social media. **Ecological Applications**, [s. l.], Abr 2015.

MAÇAIRA, Paula Medina; THOMÉ, Antônio Marcio Tavares; OLIVEIRA, Fernando Luiz Cyrino; FERRER, Ana Luiza Carvalho. Time series analysis with explanatory variables: systematic literature review. **Environmental Modelling & Software**, [s. l.], 1 jun. 2018.

MANYIKA, James; CHUI, Michael; BROWN, Brad; BUGHIN, Jacques; DOBBS, Richard; ROXBURGH, Charles; BRYERS, Angela Hung. Big data: The next frontier for innovation, competition, and productivity. **McKinsey Global Institute**, [s. l.], 2011.

MARIANI, Marcello; BAGGIO, Rodolfo; FUCHS, Matthias; HÖEPKEN, Wolfram. Business intelligence and big data in hospitality and tourism: A systematic literature review. **International Journal of Contemporary Hospitality Management**, [s. l.], v. 30, ed. 12, p. 3514-3554, 31 out. 2018.

MILLS, S. et al. **Demystifying big data**: a practical guide to transforming the business of government. [s.l.]: TechAmerica Foundation, 2012

MINISTÉRIO entrega aos estados primeiras ferramentas de Big Data e Inteligência Artificial para combater a criminalidade. [S. l.], 20 ago. 2019. Disponível em: <https://www.justica.gov.br/news/collective-nitf-content-1566331890.72>. Acesso em: 17 set. 2021.

NOBRE, Gustavo Cattelan; TAVARES, Elaine. Scientific literature analysis on big data and internet of things applications on circular economy: a bibliometric study. **Scientometrics**, [s. l.], n. 111, p. 463-492, 7 fev. 2017.

NUAIMI, Eiman Al; NEYADI, Hind Al; MOHAMED, Nader; JAROODI, Jamela Al. Applications of big data to smart cities. **Journal of Internet Services and Applications**, [s. l.], ed. 6, 1 dez. 2015.

POWER, D. J.; HEAVIN, C.; MCDERMOTT, J.; DALY, M. Defining business analytics: an empirical approach. **Journal of Business Analytics**, Iowa, USA; Cork, Ireland; Dublin, Ireland, v. 1, p. 40-53, 23 ago. 2018.

QUEIROZ, Maciel; PEREIRA, Susana. INTENTION TO ADOPT BIG DATA IN SUPPLY CHAIN MANAGEMENT: A BRAZILIAN PERSPECTIVE. **Revista de Administração de Empresas**, [s. l.], p. 389-401, Dec 2019.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Disponível em: <https://www.R-project.org>. Acesso em: 27 set. 2021.

RAJARAMAN, V. Big Data Analytics. In: RESONANCE. [S. l.: s. n.], 2016. p. 695-716.

RAM, Jiwat; ZHANGB, Changyu; KORONIOSC, Andy. The implications of Big Data analytics on Business Intelligence: A qualitative study in China. **Procedia Computer Science**, University of South Australia, Adelaide, v. 87, p. 221 - 226, 2016.

REIS, Juliana; DUARTE, Elisete; EBLE, Laeticia; GARCIA, Leila. Epidemiology and Health Services: 25 years in review. **Epidemiologia e Serviços de Saúde**, [s. l.], p. 685-700, Dec 2017.

RUSSOM, Philip. Big data analytics. **TDWI best practices report, fourth quarter**, [S. l.], v. 19, n. 4, p. 1-34, 1 jan. 2011.

SANTOS-D'AMORIM, Karen; SIVLA, Marcela Lino da; CRUZ, Rúbia Wanessa dos Reis; CORREIA, Anna Elizabeth Galvão Coutinho. DOS DADOS AO CONHECIMENTO: TENDÊNCIAS DA PRODUÇÃO CIENTÍFICA SOBRE BIG DATA NA CIÊNCIA DA INFORMAÇÃO NO BRASIL. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, Florianópolis, v. 25, p. 01-23, 2020.

SILVA, Bhagya Nathali; DIYAN, Muhammad; HAN, Kijun. Big Data Analytics. In: DEEP Learning: Convergence to Big Data Analytics. [S. l.: s. n.], 2019. p. 13-30.

SINCORÁ, Larissa Alves; DE OLIVEIRA, Marcos Paulo Valadores; FILHO, Hélio Zanquetto; LADEIRA, Marcelo Bronzo. Business analytics leveraging resilience in organizational processes. **RAUSP Management Journal**, [s. l.], p. 385-403, 18 jul. 2017.

TRKMAN, Peter; MCCORMACK, Kevin; DE OLIVEIRA, Marcos Paulo Valadores; MADEIRA, Marcelo Bronzo. The impact of business analytics on supply chain performance. **Decision Support Systems**, [s. l.], v. 49, ed. 3, p. 318-327, Jun 2010.

VARGAS, Soraida Aguilar; ESTEVES, Gheisa Roberta Telles; MAÇAIRA, Paula Medina; BASTOS, Bruno Quaresma; OLIVEIRA, Fernando Luiz Cyrino; SOUZA, Reinaldo Castro. Wind power generation: **A review and a research agenda**, [s. l.], 2 fev. 2019.

VERGARA, Sylvia Constant. Sugestão para estruturação de um projeto de pesquisa. Caderna de Pesquisa. Rio de Janeiro: EBAP/FGV nº2, 1991.

VERGARA, Sylvia Constant. **Projetos e Relatórios de Pesquisa em Administração**. São Paulo: ATLAS S.A., 1998. 87 p.

XIAO, Yu; LU, Louis Y. Y.; LIU, John; ZHOU, Zhili. Knowledge diffusion path analysis of data quality literature: A main path analysis. **Journal of Informetrics**, [s. l.], v. 8, p. 594-605, Jul 2014.

YANG, Choaowei et al. Taking the pulse of COVID-19: a spatiotemporal perspective. **International Journal of Digital Earth**, [s. l.], v. 13, n. 10, p. 1186-1211, 2020.

YOSHIDA, Nelson D. ANÁLISE BIBLIOMÉTRICA: UM ESTUDO APLICADO À PREVISÃO TECNOLÓGICA. **Future Studies Research Journal: Trends and Strategies**, São Paulo, v. 2, n. 1, p. 52-84, Jun 2010.

ZIKOPOULOS, Paul; EATON, Chris. **Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data**. In: UNDERSTANDING Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. United States: McGraw-Hill Education, 2011. cap. 1, p. 15. ISBN 9780071790543, 0071790543.

ZUPIC, Ivan; ČATER, Tomaž. Bibliometric Methods in Management and Organization. **Organizational Research Methods**, [s. l.], v. 18, ed. 3, p. 429-472, 22 dez. 2014.

2010 ANNUAL CONFERENCE & EXPOSITION, 2010, Louisville, Kentucky. **The Social, Economic, And Political Impact Of Technology: An Historical Perspective** [...]. Louisville, Kentucky: [s. n.], 2010.

2ND INTERNATIONAL CONFERENCE ON MATERIALS, MANUFACTURING AND DESIGN ENGINEERING, iCMMD2017., 2017, Maharashtra, India. **Industry 4.0 – A Glimpse** [...]. Aurangabad, India: [s. n.], 2018. 233-238 p. v. 20.