



## (DATA SCIENCE APLICADA A EXCELÊNCIA OPERACIONAL: UMA REVISÃO BIBLIOMÉTRICA) Contribuições da Engenharia de Produção para a Gestão de Operações Energéticas Sustentáveis

**Juliano Endrigo Sordan (UFSCar)**  
[julianosordan@yahoo.com.br](mailto:julianosordan@yahoo.com.br)

**Clesio Aparecido Marinho (UFSCar)**  
[clesio.marinho@yahoo.com.br](mailto:clesio.marinho@yahoo.com.br)

**Pedro Carlos Oprime (UFSCar)**  
[pedro@dep.ufscar.br](mailto:pedro@dep.ufscar.br)

**Márcio Lopes Pimenta (UFU)**  
[pimenta@ufu.com.br](mailto:pimenta@ufu.com.br)

*O desenvolvimento científico de temas relacionados à quarta revolução industrial, incluindo Big Data Analytics, inteligência artificial, machine learning e Data Science, tem ampliado as possibilidades de otimização dos processos empresariais e criado novas demandas por competências relacionadas às áreas de tecnologia da informação e estatística. O presente artigo tem como objetivo apresentar uma revisão bibliométrica da produção científica sobre Data Science aplicada a excelência operacional disponível na base de dados Scopus no período 2000-2019. Para alcançar esse objetivo, uma amostra composta por 147 artigos foi analisada quanto ao número de publicações registradas em três subperíodos distintos e revelou um significativo crescimento na pesquisa dedicada o tema, com expressivo destaque no último quadriênio. A análise de indicadores bibliométricos, redes de associação entre autores, assim como análise de palavras-chave e co-word revelaram os trabalhos mais citados, as fontes mais prolíficas, os países mais produtivos e as palavras mais empregadas nesta nova área de conhecimento. Este estudo pode oferecer insights a respeito da pesquisa sobre Data Science aplicada a excelência operacional.*

*Palavras-chave: Data Science, Big Data Analytics, Excelência Operacional, Bibliometria.*

## 1. Introdução

O paradigma da quarta revolução industrial, também conhecido como indústria 4.0 (I4.0) e suas tecnologias habilitadoras, tais como Internet das Coisas (IoT), computação em nuvem e *Big Data Analytics* têm ampliado as possibilidades para a utilização de dados digitais provenientes do chão de fábrica na tomada de decisão e otimização dos processos organizacionais. Neste cenário, um conjunto de ferramentas e técnicas específicas para a coleta, tratamento e análise de grandes volumes de dados tornam-se uma alternativa às tradicionais abordagens voltadas à excelência operacional, incluindo Lean Seis Sigma (LSS), Teoria das Restrições, Sistemas de Medição de Desempenho, entre outras.

De acordo com Snee e Hoerl (2017) a abordagem LSS necessita de um novo paradigma capaz de gerar melhorias contínuas em um cenário que incorpora o gerenciamento de riscos e *Big Data Analytics*. Na visão de Pysdek e Keller (2011), a integração entre as tecnologias da informação e projetos LSS pode ser realizada por meio de três formas: (i) dados armazenados pela empresa (*data warehousing*), disponíveis para serem usados em iniciativas de melhoria; (ii) Processamento Analítico Online (*Online Analytic Processing – OLAP*); e (iii) Mineração de Dados (*Data Mining*), que permite a análise retrospectiva dos dados, combinando avançadas técnicas e ferramentas matemáticas e estatísticas

Davenport e Kim (2015), enfatizam que a extração de valor dos dados requer análise matemática e estatística por meio de uma ciência analítica. De acordo com esses autores, o termo “analítica” se refere ao uso amplo de dados empregando: (i) **análise estatística descritiva**, que envolve coleta, organização, tabulação e apresentação de dados para a exposição das características do que está sendo estudado; (ii) **análise preditiva**, que utiliza dados do passado para prever o futuro por meio de associações entre as variáveis e estimativas de probabilidades de um fenômeno; (iii) **análise prescritiva**, que inclui métodos para fazer inferências causais entre as variáveis de interesse.

Para Provost e Fawcett (2018), *Data Science (DS)* é “*um conjunto de princípios que norteiam a extração de conhecimento a partir de dados*”. Tal definição permite compreender que essa área de conhecimento tem ampliado as possibilidades de melhoria das operações a partir de dados relacionados à excelência operacional, incluindo o monitoramento estatístico dos processos de manufatura, dados recuperados em tempo real provenientes de máquinas e recursos produtivos, informações e registros recuperados dos clientes e fornecedores, entre outros.

A implementação de estudos envolvendo DS requer do praticante um conjunto de habilidades multidisciplinares, sendo comumente descritas por meio de um diagrama de Venn, onde a DS encontra-se em uma região de intersecção entre ciência da computação, matemática e estatística, e o conhecimento sobre o negócio. Desta forma, a habilidade em utilizar ferramentas e técnicas estatísticas para melhorar a qualidade e o desempenho dos processos de negócios é uma necessidade emergente imposta aos engenheiros de produção e gerentes no século XXI (JOHN *et al.*, 2001).

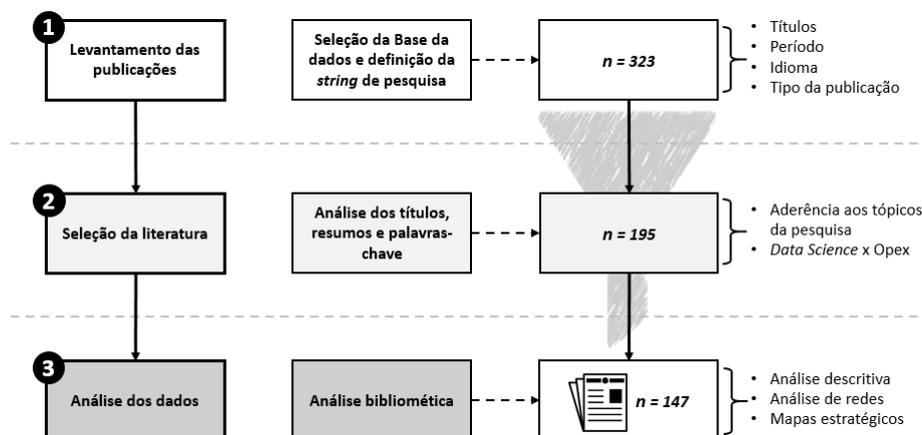
A Bibliometria pode ser compreendida como um estudo dos aspectos quantitativos da produção científica sobre determinado tema, disseminação e uso de informações registradas usando métodos matemáticos e estatísticos (SPINAK, 1996). Esse tipo de pesquisa baseia-se na contagem de artigos científicos, patentes e citações e pode considerar como dados de análise o conteúdo do texto de uma publicação ou os elementos presentes em registros extraídos de bases de dados, tais como títulos, nomes dos autores, nacionalidades, palavras-chave e citações (ZHU *et al.*, 1999).

Nos últimos anos, estudos bibliométricos têm sido utilizados para medir o progresso científico em diversas áreas da engenharia, representando um instrumento de pesquisa comum para análises sistemáticas (VAN RAAN, 2005). Contudo, revisões bibliométricas sobre DS aplicada a excelência operacional ainda representam uma lacuna de pesquisa. Este artigo tem como objetivo apresentar um estudo bibliométrico da produção científica sobre DS aplicada a excelência operacional disponível na base de dados Scopus no período 2000-2019.

## 2. Método de pesquisa

Para atingir o objetivo geral desta pesquisa, os autores estruturaram um procedimento metodológico em três partes: (i) levantamento das publicações; (ii) seleção da literatura; e (iii) análise bibliométrica dos dados, conforme ilustrado na Figura 1. O levantamento das publicações científicas foi executado no período de maio de 2020 por meio de consultas na base Scopus, que foi selecionada para o estudo considerando a relevância de periódicos com fator de impacto, bem como o banco de dados indexado que torna possível a extração de metadados. A busca foi executada por meio da seguinte *string* de pesquisa: (((*"Data Science" OR "Analytics"*) AND (*"Lean" OR "Six Sigma" OR "Theory of Constraints" OR "Quality Control" OR "Quality management" OR "Operational excellence"*))). Após a recuperação dos registros, os artigos foram filtrados pelos critérios: período selecionado no estudo, trabalhos publicados no idioma inglês e publicações na forma de artigos de periódicos e conferências.

Figura 1 – Método de pesquisa



Fonte: Elaborado pelos autores

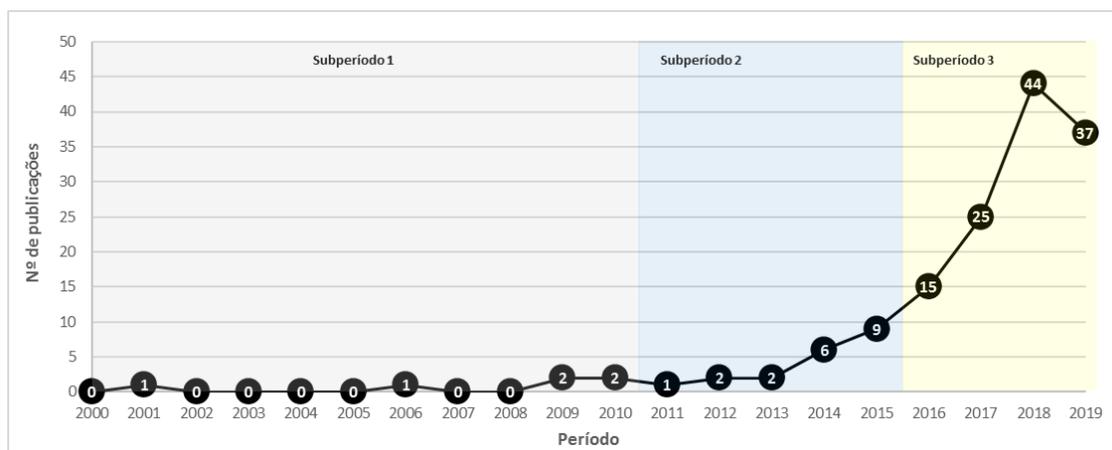
O levantamento inicial recuperou 323 trabalhos publicados no período 2000-2019. Um arquivo com os metadados extraídos da base de dados foi analisado com o auxílio do *software* SciMat (versão 1.1.04), onde foi possível aplicar um segundo filtro por meio da leitura completa dos títulos, resumos e palavras-chave de cada publicação, de acordo com a aderência ao tema desta pesquisa (DS aplicado à excelência operacional). Esse procedimento reduziu a amostra para 195 publicações. Por fim, aplicou-se um terceiro filtro, considerando a presença de todos os dados necessários à análise bibliométrica (tais como afiliação, número de citações, ano de publicação etc.), onde foram excluídos 48 trabalhos por não apresentarem algum dado relevante para a análise. Dessa forma, foram selecionados 147 artigos para compor a amostra final do trabalho.

### 3. Resultados e discussões

#### 3.1 Análise descritiva inicial

A Figura 2 mostra o evolutivo da produção científica sobre DS no período 2000-2019 (20 anos) destacando três subperíodos de análise. A produção científica no subperíodo 1 (2000-2010) representa 4,08% do número total de publicações, com apenas 6 artigos publicados. Essa baixa produtividade pode ser compreendida considerando que o tema de pesquisa pode estar vinculado ao quarto paradigma industrial, que se tornou popular a partir da realização do evento “Hannover fair” no ano de 2011, assim como a disseminação do conceito sobre *Big Data Analytics*. Portanto, observa-se a partir desse período, um número crescente nas publicações sobre DS em periódicos voltados para a gestão da produção e operações.

Figura 2 – Evolução da produção científica sobre DS aplicada a excelência operacional



Fonte: Dados da pesquisa

O segundo subperíodo (2011-2015) reúne 20 artigos (13,60% da amostra) e pode ser compreendido com um período de reconhecimento acadêmico sobre DS e crescimento no volume de pesquisas sobre o tema. Contudo, é no terceiro subperíodo (2016-2019) que o número de publicações cresce significativamente abrangendo 121 trabalhos (82,31% da amostra), que corresponde a uma média aproximada de 30 artigos/ano. Além disso, enquanto o número de citações soma apenas 31 registros no primeiro subperíodo e 789 registros no subperíodo 2, os trabalhos publicados no subperíodo 3 atingem 981 citações (54,47%).

Para compreender as abordagens de pesquisa e os temas de maior impacto no estudo sobre DS aplicada à excelência operacional foram mapeadas as publicações com maior número de citações dentro da amostra pesquisada. Esses valores foram extraídos da base de dados Scopus, conforme procedimento descrito na seção anterior. O Quadro 1 apresenta os dez artigos mais citados, assim como seus autores, as abordagens de pesquisa e a contribuição de cada trabalho. Os dez trabalhos mais citados correspondem a 48,47% do total de citações registradas na amostra. Dentre as principais contribuições reveladas nesses artigos, destacam-se as pesquisas direcionadas para *Big Data Analytics* (CAI; ZHU, 2015; CLARKE, 2016; LI *et al.*, 2016; MERINO *et al.*, 2016), técnicas para tratamento e visualização de uma grande quantidade de dados (RAGAN *et al.*, 2016; HAN; GOLPARVAR-FARD, 2017), estudos sobre a arquitetura de tecnologia da informação necessária as iniciativas de DS (AL-FUQAHA *et al.*, 2015; AKHAVIAN; BEHZADAN, 2015), assim como a otimização de processos por meio de modelagem orientada a dados (TROUP *et al.*, 2013).

Quadro 1 – Principais publicações sobre DS aplicada à excelência operacional

Ordem	Autor(es) /Ano	Nº de Citações	Abordagem de Pesquisa	Contribuição
1	Cai; Zhu, 2015	164	Estudo Teórico	O artigo apresenta uma breve revisão sobre <i>data quality</i> e fornece uma estrutura de análise para a avaliação da qualidade dos dados no contexto do <i>Big Data</i> .
2	Li <i>et al.</i> , 2016	116	Survey	Explora o desenvolvimento de técnicas para <i>crowdsourcing</i> e discute algumas técnicas de gerenciamento de dados, como controle da qualidade dos dados, controle de custos e latência (velocidade).
3	Ali <i>et al.</i> , 2012	110	Survey	O artigo mostra os resultados de dois estudos qualitativos para avaliação de um <i>software (Learning Object Context Ontology)</i> voltado ao aprendizado de técnicas analíticas.
4	Al-Fuqaha <i>et al.</i> , 2015	108	Experimento	Soluções em Internet das Coisas (IoT) e protocolos de comunicação são analisados quanto à operação em diversos cenários operacionais envolvendo máquina/máquina (M2M), máquina/servidor (M2S) e servidor/servidor (S2S).
5	Han; Golparvar-Fard, 2017	68	Estudo de caso	O artigo discute a visualização de grandes quantidades de dados, utilizando sensores visuais e analíticos, assim como aplicativos voltados para a solução de problemas de comunicação entre <i>stakeholders</i> na construção civil.
6	Clarke, 2016	65	Estudo teórico	Sob a ótica do <i>Big Data</i> , o artigo aborda algumas dimensões associadas à qualidade dos dados e destaca algumas responsabilidades de caráter moral e legal para os profissionais e pesquisadores na área de computação.
7	Ragan <i>et al.</i> , 2016	64	Estudo teórico	Os autores apresentam uma estrutura de análise para a seleção e visualização de diferentes tipos de dados ( <i>provenance types</i> ) e discutem as relações entre esses fatores e os métodos usados em DS e <i>visual analytics</i> .
8	Akhavian; Behzadan, 2015	63	Experimento	O artigo apresenta um experimento para avaliar a utilização de sensores embutidos em <i>smartphones</i> , a fim de monitorar equipamentos e atividades na construção civil, bem como utilizar esses dados em processos de simulação.
9	Troup <i>et al.</i> , 2013	60	Survey	Os autores investigam práticas de excelência operacional na indústria farmacêutica e descrevem uma abordagem de modelagem orientada a dados ( <i>Design of Experiences Dinâmicas</i> ) para a otimização de processos em lotes.
10	Merino <i>et al.</i> , 2016	55	Experimento	O artigo propõe um modelo de avaliação da qualidade para um grande volume de dados denominado “3As”, composto por três características: a adequação contextual, a adequação operacional e a adequação temporal.

Fonte: Dados da pesquisa

### 3.2 Análise das fontes e país de origem

O Quadro 2 mostra as principais fontes das publicações em conferências e periódicos relacionados às áreas de gestão da produção, com três ou mais publicações no período analisado, assim como o fator de impacto de cada fonte. No total foram identificadas 112 fontes, sendo 91 artigos de periódicos (81,25%) e 21 artigos de conferências (18,75%). As fontes *IFAC-Papers OnLine* e *IEEE Access* representam juntas 11,61% do total das publicações.

Cinco periódicos com 3 publicações cada (2,68%) também se destacam dentre as fontes mais produtivas observadas na amostra: *International Journal of Production Research*; *Journal of Manufacturing Systems*; *Quality Engineering*; *Computers and Industrial Engineering*; e *International Journal of Lean Six Sigma*. As sete fontes ilustradas abaixo representam juntas 25,01% das publicações sobre o tema disponíveis na base Scopus.

Quadro 2 – Periódicos/conferências mais prolíficos

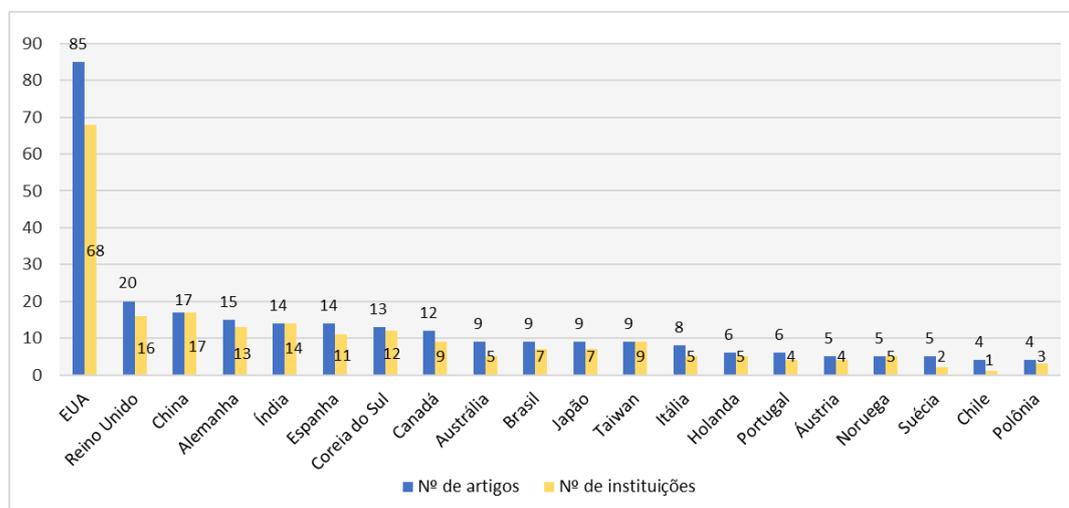
Fontes	Nº de publicações	%	Fator de Impacto*
<i>IFAC-Papers OnLine</i>	7	4,76	0,99
<i>IEEE Access</i>	6	4,08	4,09
<i>International Journal of Production Research</i>	3	2,04	4,34
<i>Journal of Manufacturing Systems</i>	3	2,04	5,45
<i>Quality Engineering</i>	3	2,04	1,24
<i>Computers and Industrial Engineering</i>	3	2,04	6,64
<i>International Journal of Lean Six Sigma</i>	3	2,04	4,89
Outros	119	80,95	-

\*CiteScore 2018

Fonte: Dados da pesquisa

Os 147 artigos analisados tiveram a participação de 51 países e 263 instituições de pesquisa e ensino. A Figura 3 mostra a distribuição do volume de publicações cobrindo os 20 países com mais de três artigos publicados no período. As barras azuis representam o número de artigos produzidos em cada país e as barras amarelas o número de instituições que participaram dessas publicações. Destacam-se neste cenário, a produtividade dos EUA (57,82% das publicações) e do Reino Unido (34,01% das publicações). Contudo, dentre os países com maior produtividade na proporção artigos/instituições destacam-se: Chile, Suécia, Austrália e Itália.

Figura 3 – Contribuição dos países e instituições

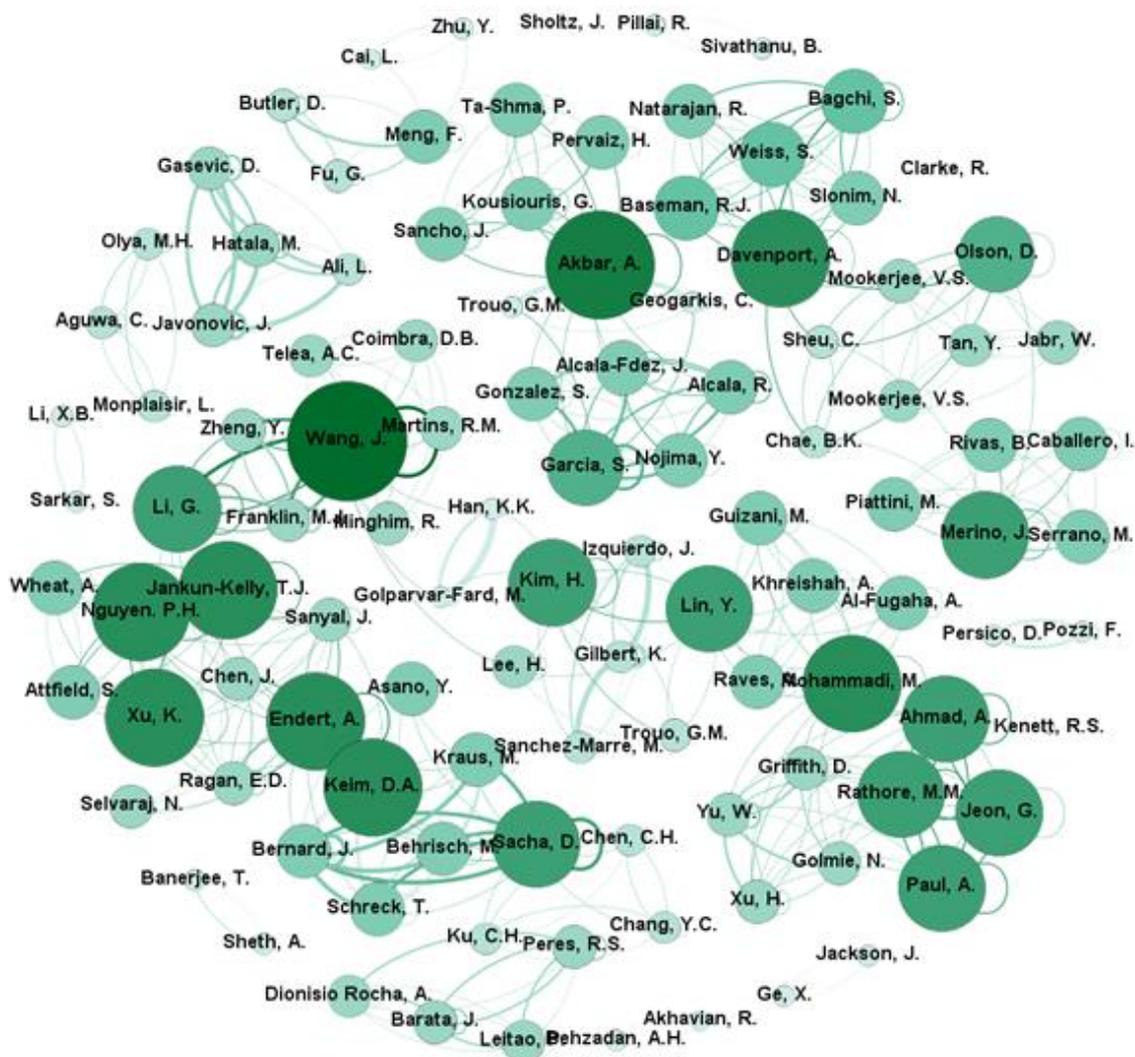


Fonte: Dados da pesquisa

### 3.3 Análise de redes e cocitação

Para analisar as relações entre os autores mais citados na amostra (cocitação) foi elaborada uma rede social com o auxílio do *software* Gephi v.0.9.2. Uma rede é um tipo de grafo que associa nomes aos círculos (nós), às arestas (linhas), ou a ambos, de modo a reduzir o nível de abstração e facilitar o entendimento do grafo (MUELLER; MASSARON, 2018). A Figura 4 ilustra a rede de citação e cocitação entre 113 autores cujos trabalhos possuem mais de dez citações observadas na amostra. Essa rede possui 113 nós e 1.131 arestas, onde cada nó representa um autor e cada aresta representa a citação de um ou mais trabalhos ou artigos em parcerias com os autores conectados. O volume e a intensidade de coloração em cada nó, assim como a espessura de cada aresta são proporcionais ao número de citações recebidas.

Figura 4 – Rede de associação entre autores



Fonte: Dados da pesquisa

A observação das interações entre os diversos autores representados na Figura 4 revela os agrupamentos caracterizados por meio de parcerias nas publicações. Ao todo, foram identificados 532 autores. Os autores de maior grau de intermediação incluem Ali, L.; Hatala, M.; e Gašević, D. (grau 9), seguidos por Javonović, J.; Kenett, R.S.; Sacha, D.; Kraus, M.; Bernard, J.; Behrisch, M.; e Schreck, T. (ambos com grau 8). A rede revela também que a prática de autocitação é comum para vários autores.

### 3.4 Análise das palavras-chave e *co-word*

A análise das palavras-chave contidas nas publicações científicas pode ser auxiliada pela aplicação das leis de Zipf, onde a primeira lei aborda a frequência na qual as palavras aparecem em um texto e estabelece que o produto da ordem de uma série de palavras ( $r$ ) pela sua frequência ( $f$ ) organizada em um *ranking* é aproximadamente constante ( $c$ ), ou seja,  $r \times f = c$ . A segunda lei Zipf propõe uma separação entre as palavras de alta frequência (núcleo), das palavras de baixa frequência (dispersão), de modo a identificar as palavras mais significativas, representadas pelo núcleo. Para facilitar a separação entre essas duas categorias de palavras, Goffman criou um ponto de transição, denominado “Ponto T de Goffman”, de acordo com a seguinte expressão (GUEDES; SANTOS, 2013):

$$n = \frac{-1 + \sqrt{1 + 8I_1}}{2}$$

Onde  $n$  representa o ponto de transição T,  $I_1$  é o número de palavras com frequência igual a 1 e 8 é uma constante derivada da língua inglesa. Após o ajuste das palavras com o mesmo significado, foram identificadas 1.641 palavras-chave informadas pelos autores, sendo 1.380 palavras com frequência igual a 1. Assim, o Ponto de transição foi atribuído à 18ª palavra contida na lista ordenada. O Quadro 3 apresenta as 18 palavras contidas no núcleo de palavras-chave, onde é possível perceber a influência das tecnologias habilitadoras da indústria 4.0 nos estudos sobre DS aplicada a excelência operacional, tais como *Big Data Analytics*, *Artificial Intelligence*, *Industry 4.0*, *Machine-learning* e *Internet of Things*. Além disso, observa-se a aplicação dessas tecnologias nas atividades de controle da produção “*Decision-making*”, “*Process-Control*” e “*Manufacturing*”, assim como nas atividades de controle da qualidade “*Quality Control*” e “*Predictive analytics*”.

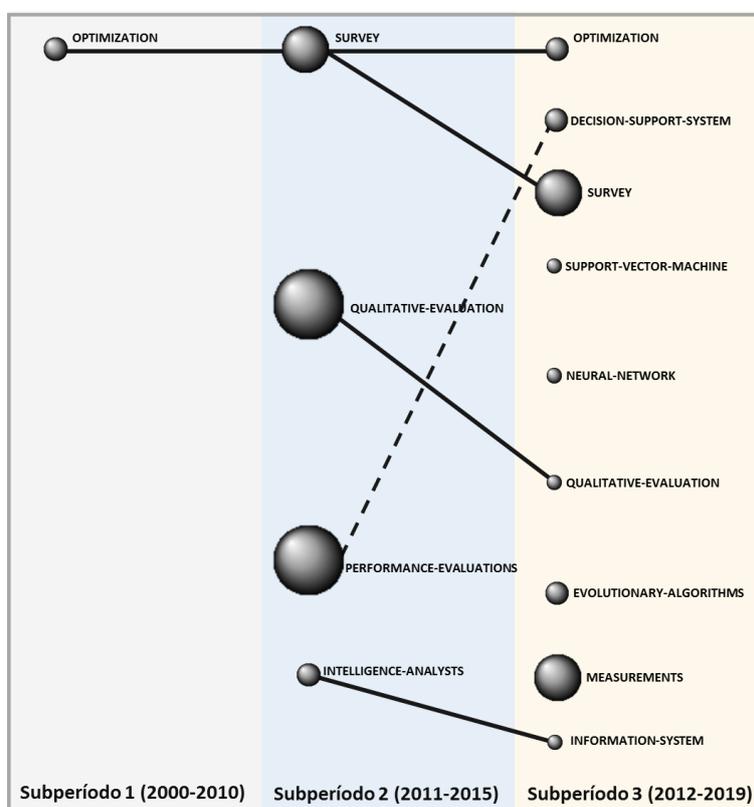
Quadro 3 – Palavras-chave contidas no núcleo da amostra

Palavra	Frequência	Palavra	Frequência
QUALITY CONTROL	104	VISUALIZATION	13
BIG DATA ANALYTICS	68	ARTIFICIAL INTELLIGENCE	11
DATA MINING	20	INDUSTRY 4.0	11
LEARNING SYSTEMS	18	PREDICTIVE ANALYTICS	11
DECISION-MAKING	17	MACHINE-LEARNING	10
DATA HANDLING	16	MANUFACTURING	10
DATA REDUCTION	14	INTERNET OF THINGS	9
INFORMATION ANALYSIS	14	QUALITY OF SERVICE	9
PROCESS CONTROL	13	DATA-QUALITY	8

Fonte: dados da pesquisa

A análise baseada na co-ocorrência das palavras (*co-word analysis*) permite estabelecer a intensidade de ligação entre termos a partir da verificação do número de artigos em que esses termos aparecem. De acordo com Cobo *et al.*, (2011) essa análise pode ser usada para mapear a ciência, agrupamentos de palavras-chave e suas interconexões e um campo de pesquisa pode ser entendido como um conjunto de temas, mapeados em um espaço bidimensional caracterizado por dois parâmetros (densidade e centralidade). A Figura 5 mostra a evolução dos termos que se destacam em cada subperíodo por meio de um mapa de evolução temático.

Figura 5 – Mapa de evolução temático



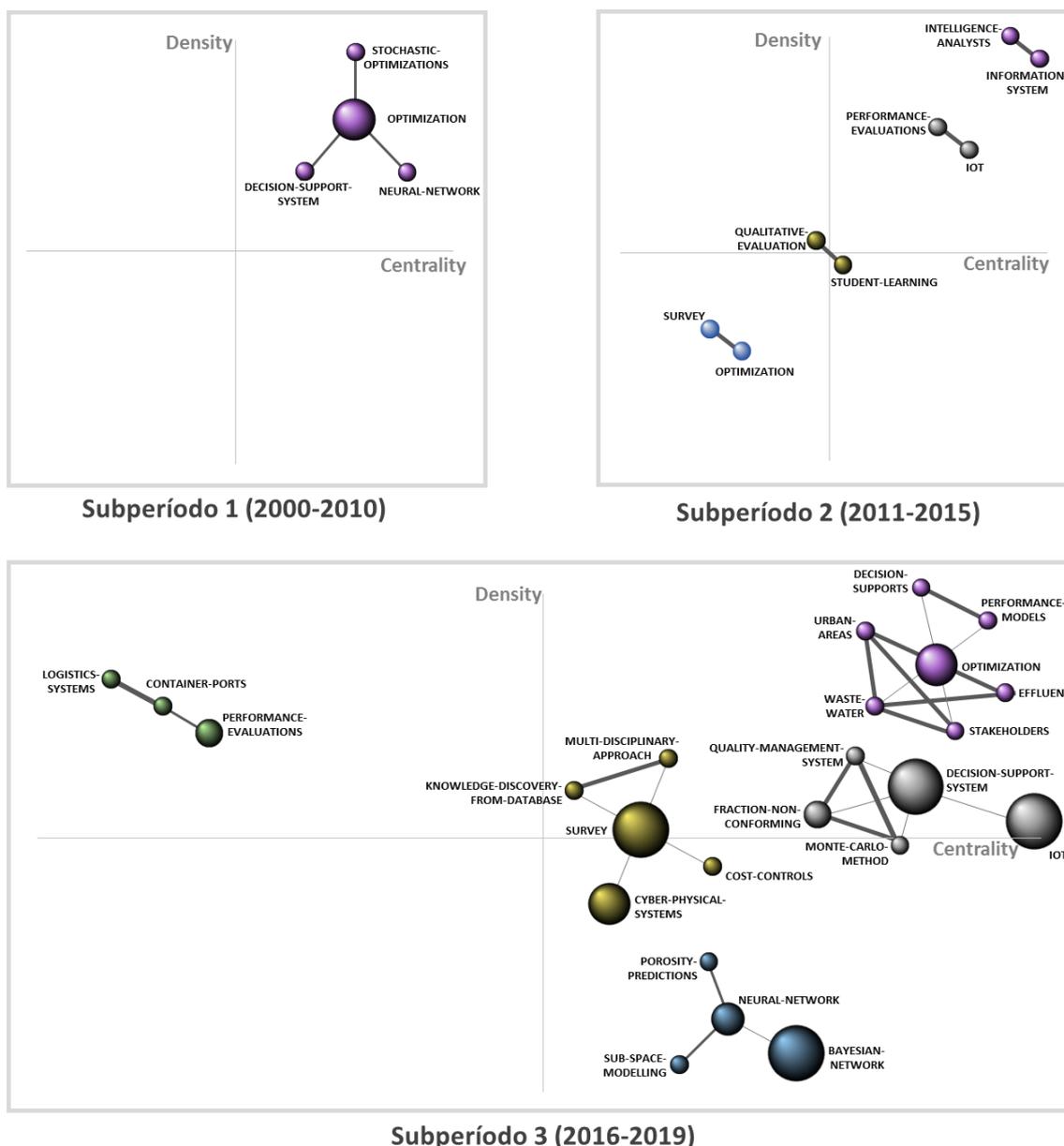
Fonte: dados da pesquisa

Esse mapa foi elaborado com o auxílio do *software* SciMAT utilizando os critérios de não redução de dados, índice de equivalência e algoritmo de centros simples. Os nós (círculos) representam as palavras ou temas com co-ocorrência e o seu volume é proporcional ao número de artigos publicados, enquanto a espessura das linhas é proporcional ao índice de inclusão, calculado a partir do número de elementos associados a cada termo. Além disso, a linha contínua indica os termos com o mesmo significado, como por exemplo a palavra “*survey*” que aparece nos subperíodos 2 e 3. A linha tracejada indica que diferentes temas compartilham elementos comuns, como é o caso do termo “*Qualitative-evaluation*” (subperíodo 2) relacionado com o termo “*Decision-support-system*” (subperíodo 3). A ausência de linhas conectoras entre as palavras indica a descontinuidade dos termos na transição entre os períodos de análise.

Os mapas estratégicos envolvendo *co-words* representam uma alternativa para a compreensão da evolução dos termos empregados em uma determinada área de conhecimento. Os diagramas estratégicos ilustrados na Figura 6 foram elaborados com o *software* SciMAT. Os termos são agrupados por similaridade e plotados em diagramas bidimensionais posicionados nos valores de densidade (eixo y) e centralidade (eixo x), resultando em quatro quadrantes: **a) temas motores**, localizados no quadrante superior direito indicando conceitos aplicáveis a outros temas intimamente relacionados a área de conhecimento; **b) temas isolados**, localizados no quadrante superior esquerdo abrangendo temas muito especializados, com importância marginal; **c) temas básicos e transversais**, localizados no quadrante inferior direito revelando aspectos gerais ainda não desenvolvidos na área de pesquisa; e **d) temas emergentes ou em declínio**, localizados no quadrante inferior esquerdo, representando temas incipientes ou em desaparecimento, visto que possuem baixa centralidade e baixa densidade.

A elaboração dos mapas estratégicos seguiu o mesmo procedimento adotado na construção do mapa de evolução temático. O tamanho dos nós corresponde à contagem de documentos que apresentam o termo e a espessura das linhas é proporcional ao índice de inclusão. Deste modo, dentre os seis artigos publicados no subperíodo 1, três apresentam a palavra “*Optimization*”, que aparece relacionada com as palavras “*Stochastic-optimization*”, “*Neural-network*” e “*Decision-support-system*”, ambas representadas como temas motores e, portanto, relacionadas com outras áreas de conhecimento.

Figura 6 – Mapa estratégico de palavras



O subperíodo 2 reúne 20 artigos que destacam quatro grupos de palavras (cada palavra aparece em apenas um artigo). No quadrante de temas motores destacam-se os termos “*Intelligence-analysts*”, “*Information-system*”, “*performance-evaluations*” e “*IoT*”. No centro do gráfico aparece um grupo conectando as palavras “*Qualitative-evaluation*” e “*Student-learning*”. Por fim, no quadrante de temas emergentes ou em declínio, a palavra “*Optimization*” aparece conectadas com o termo “*Survey*”. A conexão entre essas duas palavras também foi destacada no mapa de evolução temático na transição entre os subperíodos 2 e 3.

Devido à grande produtividade científica observada no subperíodo 3, quatorze grupos de palavras destacam-se no mapa estratégico deste subperíodo. Para facilitar a visualização das palavras-chave foram priorizados os grupos com maior densidade. Destacam-se como temas motores as palavras: “*Decision-support-system*” e “*IoT*” (presentes em 5 artigos); “*Optimization*” (4 artigos); e “*Fraction-non-conforming*” (2 artigos). Um terceiro grupo surge entre os quadrantes de termos motores e transversais contendo as palavras “*Survey*” (5 artigos); “*Cyber-physical-systems*” (4 artigos); e “*Multi-disciplinary-approach*”, “*Knowledge-discovery-from-database*” e “*Cost-controls*”, ambas registradas em 1 artigo.

O quadrante de temas básicos e transversais do subperíodo 3 também destaca um grupo de palavras, cujo núcleo é o termo “*Neural-network*” (3 artigos) que está conectado com as palavras “*Bayesian-network*” (5 artigos), “*Porosity-predictions*” e “*sub-space-modelling*” (1 artigo cada). Por fim, um último grupo aparece no quarto quadrante de temas isolados destacando três palavras presentes em apenas 1 artigo: “*Performance-evaluations*” (2 artigos), “*Container-ports*” e “*Logistics-systems*”.

#### 4. Conclusão

Este artigo apresentou um estudo bibliométrico da produção científica sobre a DS aplicada a excelência operacional disponível na base de dados Scopus no período 2000-2019. Ao todo, foram analisados 147 trabalhos na forma de artigos de periódicos e conferências publicados em 112 fontes por 532 autores, abrangendo 51 países e 263 instituições de pesquisa e ensino. A análise da evolução científica sobre o tema, quando estratificada em três subperíodos, revela um expressivo aumento no volume de artigos no último período (2016-2019), abrangendo 121 trabalhos (82,31% das publicações).

Dentre as publicações mais citadas, destacam-se as pesquisas direcionadas para o tratamento e visualização de grandes quantidades de dados (*Big Data Analytics*), diferentes técnicas de visualização desses dados, arquiteturas de tecnologia da informação necessária as iniciativas de DS, assim como a otimização de processos por meio de modelagem orientada a dados. A rede de associação entre os autores destaca os autores mais citados e evidencia diversos aglomerados que se caracterizam pela prática da cocitação e autocitação.

A análise das palavras-chave salientou a importância do *Big Data* e das tecnologias da informação como suporte às atividades de controle de qualidade, análise preditiva e monitoramento de processos. A análise de *co-words* usando o mapa de evolução temático e mapas estratégicos, destacou a evolução de termos motores e transversais incluindo, otimização, sistemas de suporte à decisão, redes neurais, algoritmos, *machine learning*, IoT, entre outros. A principal limitação deste estudo está na amostra obtida em uma única base de dados. Contudo, os resultados discutidos neste artigo podem oferecer *insights* a respeito da pesquisa científica sobre DS aplicada a excelência operacional.

## REFERÊNCIAS

AKHAVIAN, R.; BEHZADAN, A.H. Construction equipment activity recognition for simulation input modeling using mobile sensors and machine learning classifiers. **Advanced Engineering Informatics**, v. 29, n. 4, p. 867-877, 2015.

ALI, L.; HATALA, M.; GAŠEVIĆ, D.; JOVANOVIĆ, J. A qualitative evaluation of evolution of a learning analytics tool. **Computers & Education**, v. 58, n. 1, p. 470-489, 2012.

AL-FUQAHA, A.; KHREISHAH, A.; GUIZANI, M.; RAYES, A.; MOHAMMADI, M. Toward better horizontal integration among IoT services. **IEEE Communications Magazine**, v. 53, n. 9, p. 72-79, 2015.

CAI, L.; ZHU, Y. The challenges of data quality and data quality assessment in the big data era. **Data science journal**, v. 14, 2015.

CLARKE, R. Big data, big risks. **Information Systems Journal**, v. 26, n. 1, p. 77-90, 2016.

COBO, M.J.; LÓPEZ-HERRERA, A.G.; HERRERA-VIEDMA, 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**, v. 5, n. 1, p. 146-166, 2011.

DAVENPORT, T.; KIM, J. **Dados demais!:** como desenvolver habilidades analíticas para resolver problemas complexos, reduzir riscos e decidir melhor. Rio de Janeiro: Elsevier, 2015.

GUEDES, V.L.S; SANTOS, M.J.V.C. Recorrência de nominalizações deverbais em resumos de cartas científicas em língua portuguesa e a indexação temática. **Linguística**, v. 29, n. 1, p. 37-57, 2013.

HAN, K.K.; GOLPARVAR-FARD, M. Potential of big visual data and building information modeling for construction performance analytics: An exploratory study. **Automation in Construction**, v. 73, p. 184-198, 2017.

JOHN, J.; WHITAKER, D.; JOHNSON, D. **Statistical Thinking for Managers**, Boca Raton: Chapman & Hall/CRC, 2001.

LI, G.; WANG, J.; ZHENG, Y.; FRANKLIN, M. J. Crowdsourced data management: A survey. **IEEE Transactions on Knowledge and Data Engineering**, v. 28, n. 9, p. 2296-2319, 2016.

MERINO, J.; CABALLERO, I.; RIVAS, B.; SERRANO, M.; PIATTINI, M. A data quality in use model for big data. **Future Generation Computer Systems**, v. 63, p. 123-130, 2016.

MUELLER, J.P.; MASSARON, L. **Algoritmos para leigos**. Rio de Janeiro: Alta Books, 2018.

PROVOST, F.; FAWCETT, T. **Data Science para Negócios: O que você precisa saber sobre mineração de dados e pensamento analítico de dados**. Alta Books Editora, 2018.

PYSDEK, T.; KELLER, P. A. **Seis Sigma guia do profissional: um guia completo para green belts, black belts e gerentes em todos os níveis**. Rio de Janeiro: Alta Books, 2011.

RAGAN, E.D.; ENDERT, A.; SANYAL, J.; CHEN, J. Characterizing provenance in visualization and data analysis: an organizational framework of provenance types and purposes. **IEEE transactions on visualization and computer graphics**, v. 22, n. 1, p. 31-40, 2015.

SNEE, R. D.; HOERL, R. W. Time for Lean Six Sigma 2.0? **Quality Progress**, v. 50, n. 5, 2017.

SPINAK, E. **Diccionario enciclopédico de bibliometría, cienciométrica e informetría**. Montevideo: UNESCO, 1996.

TROUP, G. M.; GEORGAKIS, C. Process systems engineering tools in the pharmaceutical industry. **Computers & Chemical Engineering**, v. 51, p. 157-171, 2013.

VAN RAAN, A.F. For your citations only? Hot topics in bibliometric analysis. **Measurement: interdisciplinary research and perspectives**, v.3, n. 1, p. 50-62, 2005.

ZHU, D.; PORTER, A.; CUNNINGHAM, S.; CARLISIE, J.; NAYAK, A. A process for mining science & technology documents databases, illustrated for the case of " knowledge discovery and data mining". **Ciência da Informação**, v. 28, n. 1, p. 07-14, 1999.