

OS HOSPITAIS NA ERA DA CIÊNCIA DOS DADOS: UM ESTUDO A PARTIR DA PRODUÇÃO CIENTÍFICA INDEXADA NA WEB OF SCIENCE

Natanael Vitor Sobral

Doutor em Ciência da Informação
Universidade Federal da Bahia (UFBA)
E-mail: natanvsobral@gmail.com

Gillian Leandro de Queiroga Lima

Doutor em Difusão do Conhecimento
Universidade Federal da Bahia (UFBA)
E-mail: gillianqueiroga@gmail.com

Ana Sara Pereira de Melo Sobral

Mestra em Ciência da Informação
Instituto Federal Baiano (IFBAIANO)
E-mail: anasarap@hotmail.com

INTRODUÇÃO

Ao longo dos anos tem-se notado um aumento na quantidade de dados gerados pelos processos de assistência em saúde. Esses conjuntos de dados, armazenados nos sistemas de registro eletrônico de saúde, são recursos subutilizados, considerando-se as potencialidades dos atuais métodos desenvolvidos no âmbito da Ciência dos Dados, sobretudo, com o propósito de melhorar os serviços hospitalares e de assistência em saúde. (DALIANIS et al., 2015) Numa perspectiva praxeológica, as ações da interlocução entre os domínios da Saúde, Tecnologia e Ciência da Informação centradas em sistemas e serviços de saúde potencializam a concepção e desenvolvimento de ferramentas e métodos capazes de aperfeiçoar os serviços prestados, evitando e solucionando problemas de saúde da população (GALVÃO; VALENTIM, 2019), valendo-se de *data analytics*, *big data* e outras disciplinas. (CAO, 2019)

Motivado pelas possibilidades oferecidas pela Ciência dos Dados aos serviços e sistemas de saúde, essa comunicação objetivou realizar uma análise bibliométrica sobre as aplicações da Ciência dos Dados no âmbito das organizações hospitalares. Tendo em vista a contemporaneidade desse tópico na literatura científica, o recorte temporal selecionado considera 2015 a 2019. A base de dados utilizada foi a Web of Science (WoS),

dada a sua característica pluridisciplinar, sendo referência na indexação de periódicos científicos e demais bibliografias de circulação internacional.

METODOLOGIA

Para o desenvolvimento desta pesquisa, cumpriram-se, respectivamente, as seguintes etapas metodológicas:

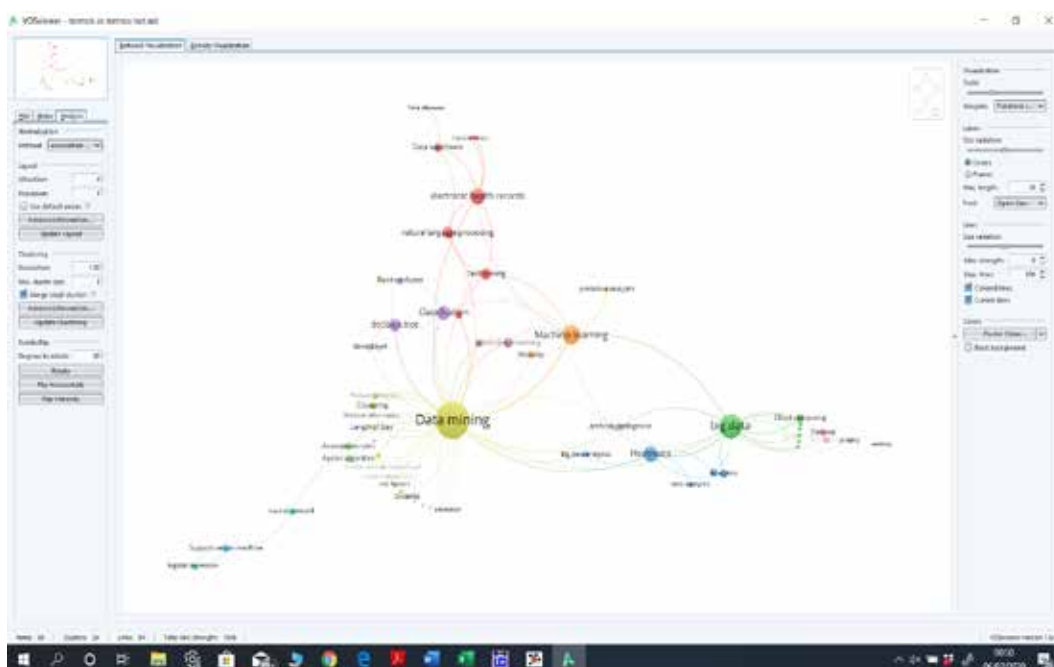
- Recuperação de informações científicas na base de dados WoS: para isto, utilizou-se a expressão: TÓPICO: (((“hospital” OR “hospitals”) AND (“big data” OR “data analytics” OR “business intelligence” OR “data mining” OR “data warehouse” OR “text mining” OR “data science”))). As palavras-chave foram definidas conforme as principais tendências presentes na literatura científica no que tange ao escopo da Ciência de Dados, relacionando-as com os hospitais, objetos de interesse desse estudo. Sendo essa uma pesquisa exploratória, pretende-se, adiante, avançar na identificação de novos termos pertencentes ao domínio com o objetivo de aperfeiçoar a expressão de busca para trabalhos futuros. O período estipulado para o trabalho foi 2015 a 2019, visando obter as publicações mais recentes. Com isto, encontraram-se 1291 registros, sendo este, o *corpus* definitivo. Ressalta-se que o ano 2019 foi considerado até o início de setembro, haja vista que este foi o mês da coleta de dados.
- Processamento dos Dados: os metadados dos trabalhos e artigos encontrados foram compilados e processados na ferramenta de mineração de dados e textos *The Vantage Point*®, aplicando-se instrumentos de limpeza com o propósito de lidar com as possibilidades de dispersão causadas por questões linguísticas comuns aos estudos bibliométricos, e posteriormente, elaboraram-se matrizes, tendo como cerne as palavras-chave, porém, associando-as às áreas de pesquisa definidas pela WoS e às afiliações institucionais dos autores. As fontes também foram levantadas para fins de elaboração de um *ranking* sobre os veículos de publicação.
- Análise dos Dados: para a análise, optou-se pela técnica de Análise de Redes Sociais (ARS), amparada na teoria dos grafos, fazendo-se uso da ferramenta VOSviewer. (VAN ECK; WALTMAN, 2010)

RESULTADOS E DISCUSSÃO

As palavras-chave de maior representatividade presentes nos resultados, por ocorrência, foram: *data mining* (230), *big data* (166), *machine learning* (75), *electronic health records* (59), *healthcare* (44), *text mining* (40), *classification* (39), *decision tree* (30), *electronic medical records* (28), *data analytics* (27), *hospital* (26), *natural language processing* (25), *predictive modeling* (24), *business intelligence* (22), *cloud computing* (22), *data warehouse* (22), *mortality* (22), *artificial intelligence* (20) e *heart failure* (20).

A Figura 1 apresenta a rede de coocorrência de palavras-chave (grau > 5). Os nós foram dispostos no grafo pela centralidade de grau, considerando-se a quantidade de vértices adjacentes (BARABÁSI, 2016), sendo os tamanhos dos nós proporcionais à ocorrência, igualmente às outras figuras apresentadas neste trabalho.

Figura 1 - Rede de coocorrência de palavras-chave (grau > 5)



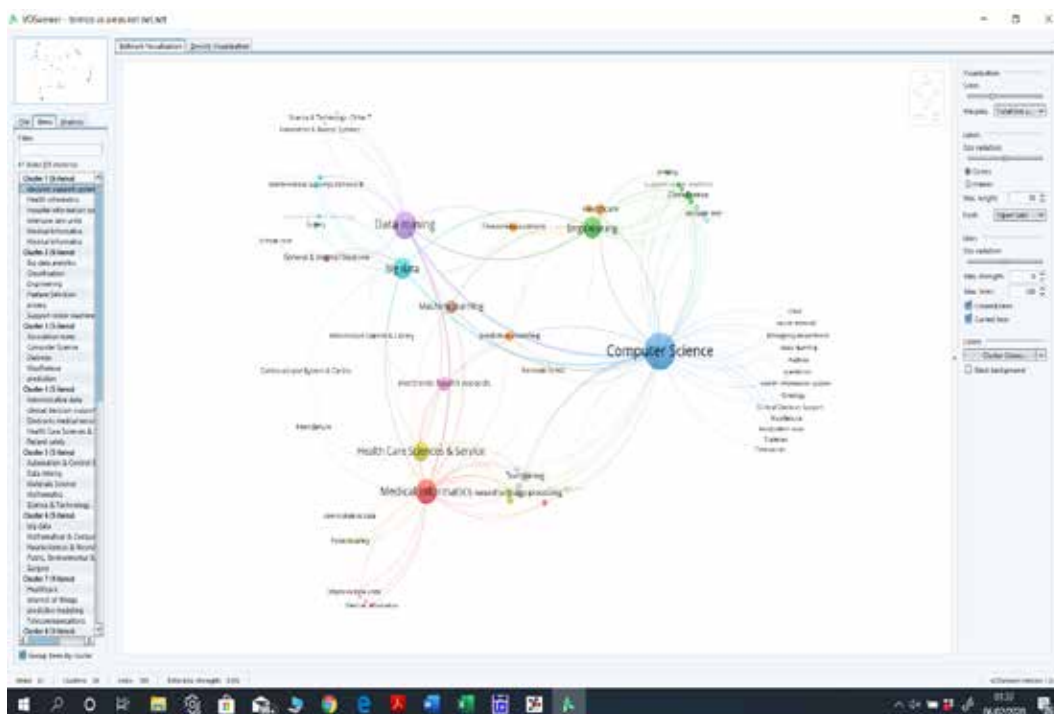
Fonte: dados da pesquisa.

Os pares de vértices (Figura 1) que apresentam maior grau de coocorrência, respectivamente, são: *data mining* & *machine learning* (24), *data mining* & *classification* (23), *data mining* & *decision tree* (19), *big data* & *healthcare* (15), *big data* & *machine learning* (14), *big data* & *cloud computing* (12), *electronic health records* & *data mining* (11), *data mining* & *clustering* (10), *big data* & *analytics* (10), *data mining* & *predictive modeling* (9), *data mining* & *natural language processing* (9), *data mining* & *electronic medical records* (9), *data mining* & *healthcare* (8), *data mining* & *hospital* (8). As associações de termos mais evidentes relacionam-se à discussão trazida por Suchting e demais autores (2018), quando afirmam que a Ciência de Dados é um termo amplo para metodologias pautadas em algoritmos de aprendizado de máquina com o propósito de proporcionar previsões amparadas em conjuntos de dados grandes, complexos e com muitas variáveis. Essas ferramentas auxiliam na solução de problemas de *big data* em uma vasta gama de fenômenos, com aplicações promissoras e multiformes no campo hospitalar, haja vista que neste estudo citado, os autores exploraram a Ciência dos Dados para prever eventos agressivos em um hospital psiquiátrico. (SUCHTING et al., 2018)

As áreas de pesquisa com maior produtividade (Figura 2), associadas às palavras-chave selecionadas foram: *Computer Science* (464), *Engineering* (227), *Medical Informatics* (220), *Health Care Sciences and Services* (195), *General and Internal Medicine* (84), *Telecommunications* (69), *Public, Environmental and Occupational Health* (59), *Mathematical and Computational*

Biology (48), Pharmacology and Pharmacy (42), Science and Technology - Other Topics (37), Surgery (37), Information Science and Library Science (31), Operations Research and Management Science (29), Cardiovascular System and Cardiology (27), Neurosciences and Neurology (27), Nursing (25), Business and Economics (24), Radiology, Nuclear Medicine and Medical Imaging (24).

Figura 2 - Rede de palavras-chave vs áreas de pesquisa (grau > 3)



Fonte: dados da pesquisa.

Conforme apresentado na Figura 2, foram identificados 28 clusters de associações entre palavras-chave e áreas de pesquisa, sendo estes os mais representativos: *Medical Informatics* (área) & *decision support systems, health informatics, hospital information systems, intensive care units, medical informatics* (termos); *Engineering* (área) & *big data analytics, classification, feature selection, privacy, support vector machine* (termos); *Computer Science* (área) & *association rules, diabetes, mapreduce, prediction*; *Health Care Sciences and Service* (área) & *administrative data, clinical decision support systems, electronic medical records, patient safety; data mining* (termo) & *Materials Science, Mathematics, Science and Technology – others, Automation and Control Systems* (áreas); *big data* (termo) & *Mathematical and Computational Biology, Neurosciences and Neurology, Public, Environmental and Occupational Health, Surgery* (áreas); *Telecommunications* (área) & *healthcare, internet of things, predictive modeling* (termos).

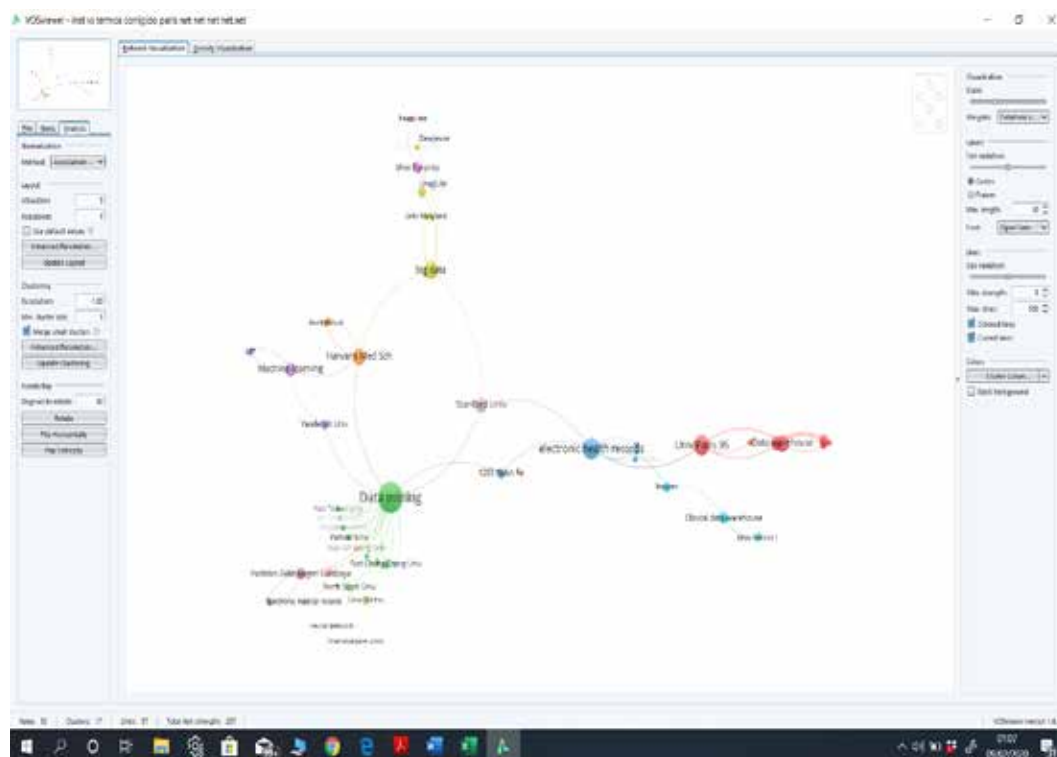
Assim, vê-se que as interações mais expressivas refletem a inclinação da Informática Médica em assuntos relacionados à tomada de decisão, sistemas de informação para

hospitais e unidade de cuidados intensivos. Gómez-Vallejo e demais autores (2016), ao discutirem esse espectro temático, afirmam que a alta disponibilidade de registros eletrônicos em saúde tem estimulado o desenvolvimento de sistemas de tomada de decisão no campo hospitalar, tendo como argumento para o seu desenvolvimento a possibilidade de reduzir custos e a implementação de ações de vigilância.

Sobre a Ciência da Informação, destaca-se sua associação aos *electronic health records* (Figura 2), o que revela a importância de processos organizacionais relacionados à gestão de documentos e à gestão da informação no âmbito dos sistemas e serviços de saúde. Hwang e demais autores (2017) relatam que um dos benefícios mais importantes desses registros é a sua utilização para estudos que envolvem Ciência de Dados, aprendizado de máquina, análise estatística de doenças, previsão personalizada de doenças e análise de doenças com base em coortes.

Concernente aos vínculos entre palavras-chave e instituições, têm-se o panorama apresentado na Figura 3, congregando os principais assuntos face às universidades, empresas e institutos de pesquisa com atuação comprovada através da publicação de artigos em tópicos ligados ao universo da Ciência dos Dados no âmbito dos hospitais. Neste item, também reforça-se o papel dos países conforme a vocação temática constatada nas publicações.

Figura 3 - Rede de palavras-chave vs instituições (Grau > 2)



Fonte: dados da pesquisa.

Os *clusters* mais representativos que apontam as possíveis especialidades das instituições no âmbito deste estudo são: *data warehouse*, *natural language processing* e *rare diseases* (termos) & *Imagine Inst Genet Dis*, *Paris Descartes Sorbonne*, *Univ Paris V*, *Univ Wurzburg* (instituições); *data mining* (termo) & *Florida Polytech Univ*, *Mahidol Univ*, *Natl Changhua Univ*, *Natl Chi Nan Univ*, *Natl Chung Cheng Univ*, *Natl Taiwan Univ*, *Providence Univ* (instituições); *electronic health records* (termo) & *1265 Welch Rd*, *Hop Europeen Georges Pompidou*, *Kings Coll London*, *Univ Florida* (instituições); *big data* e *cloud computing* (termos) & *Cent S Univ*, *Univ Maryland*, *Univ Messina* (instituições); *machine learning* (termo) & *MIT*, *Univ Pittsburgh*, *Vanderblit Univ* (instituições); *clinical data warehouse* (termo) & *CHU Rennes*, *Inserm*, *Univ Rennes* (instituições). Apesar de Estados Unidos (384), China (181) e Índia (82) serem os países mais representativos no cômputo geral, nota-se o destaque das academias francesas de saúde no tocante à especialização temática.

Quanto às fontes, que refletem os veículos de publicação, destacaram-se: *BMC Medical Informatics and Decision Making* (21), *International Journal of Medical Informatics* (21), *IEEE Access* (16), *Journal of Medical Systems* (15), *Journal of Biomedical Informatics* (11), *Journal of the American Medical Informatics Association* (11), *BMJ Open* (10), *Applied Clinical Informatics* (9), *IEEE Journal of Biomedical and Health Informatics* (8), *Medinfo 2017: Precision Healthcare through Informatics* (8), *BMC Health Services Research* (7), *Computer Methods and Programs in Biomedicine* (7), *JMIR Medical Informatics* (7). Os dois periódicos mais representativos integram, respectivamente, os grupos Springer Nature e Elsevier. No mais, nota-se uma diversificada presença de editoras, tais como, IEEE (organização profissional técnica, sem fins lucrativos, dedicada ao avanço da tecnologia), Oxford University Press, BMJ Publishing Group, American Medical Informatics Association, IOS Press e JMIR Publications, evidenciando que o escoamento da produção científica neste campo possui uma significativa diversidade.

CONSIDERAÇÕES FINAIS

O presente trabalho objetivou realizar uma análise bibliométrica sobre as aplicações da Ciência dos Dados, no âmbito dos hospitais, na literatura recente. Com isso, verificou-se a proeminência do *data mining* e *big data* como os principais assuntos, porém, chamando bastante a atenção o posicionamento de *machine learning* e *electronic health records*, em especial, por não serem termos adotados na expressão de busca, o que diminuía suas possibilidades de recuperação e representação nos resultados.

Em relação ao termo *electronic health records* (documentos eletrônicos em saúde), destaca-se sua associação com a área da Ciência da Informação, demonstrando a importância e o potencial dessa área do conhecimento nos estudos sobre as organizações produtoras de serviços de saúde, principalmente, por meio de processos organizacionais

associados à gestão de documentos e à gestão da informação, já inseridos no paradigma contemporâneo da Ciência dos Dados.

A partir desta pesquisa exploratória, foi possível identificar palavras-chave que podem possibilitar, em trabalhos futuros, a realização de um estudo mais aprofundado acerca das tecnologias e métodos informacionais com potencial de utilização nos sistemas e serviços de saúde. Ademais, o trabalho evidenciou a multidisciplinaridade existente em torno do assunto analisado, passando por áreas como a Ciência da Computação, Engenharia, Telecomunicações, Saúde Pública, Matemática, Biologia, Arquivologia, Biblioteconomia e Ciência da Informação, Neurociências, Gestão, dentre outras. Por fim, ressalta-se a necessidade de realização de estudos mais aprofundados acerca da Ciência dos Dados em organizações hospitalares.

REFERÊNCIAS

BARABASI, A. L. *Network science*. Cambridge: University Press, 2016.

CAO, L. Data science and analytics: a new era. *International Journal Of Data Science And Analytics*, Cham, v. 1, p. 1-2, 2019.

DALIANIS, H.; HENRIKSSON, A.; KVIST, M. *et al.* Health bank-A - Workbench for data science applications in healthcare. *Proceedings of the CAiSE-2015 Industry Track*, Estocolmo, SWE, p. 1-18, 2015.

GALVÃO, A. B.; VALENTIM, R. A. M. Desafios para os avanços da análise de big data na Saúde. In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO APLICADA À SAÚDE, 19., 2019, Niterói. *Anais [...]*. Niterói: SBC, 2019. p. 155-160.

GÓMEZ-VALLEJO, H. J.; URIEL-LATORRE, B.; SANDE-MEIJIDE, M. *et al.* Glez-Peña, D. A case-based reasoning system for aiding detection and classification of nosocomial infections. *Decision Support Systems*, Amsterdam, v. 84, p. 104-116, 2016.

HWANG, U.; CHOI, S.; LEE, H. B. *et al.* Disease prediction from electronic health records using generative adversarial networks. *arXiv preprint*, [s. l.], p. 1-10, 2017. Disponível em: <https://arxiv.org/pdf/1711.04126.pdf>. Acesso em: 1 fev. 2019.

SUCHTING, R.; GREEN, C. E.; GLAZIER, S. M. *et al.* A data science approach to predicting patient aggressive events in a psychiatric hospital. *Psychiatry research*, Oxford, v. 268, p. 217-222, 2018.

VAN ECK, N.; WALTMAN, L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, London, v. 84, n. 2, p. 523-538, 2010.

