

**Estudo comparativo das produções científicas e tecnológicas dos  
especialistas seniores em covid-19 após  
um ano de pandemia no Brasil**

*Comparative study of the scientific and technological productions  
of senior specialists in covid-19 after  
a year of pandemic in Brazil*

*Estudio comparativo de las producciones científicas y tecnológicas  
de especialistas senior en covid-19 después de  
un año de pandemia en Brasil*

Henrique Koch Chaves<sup>58</sup>

Júlia Maria Brandão da Silva<sup>59</sup>

Marcos Emiliano Lima Alves Hir<sup>60</sup>

Carla Cristina de Freitas da Silveira<sup>61</sup>

Jorge Lima de Magalhães<sup>62</sup>

Jesús Pascual Mena-Chalco<sup>63</sup>

**RESUMO**

A pandemia da covid-19 mostrou-se como um dos maiores desafios sanitários de todos os tempos. Em tempos de crises e pandemias, o Big Data surge como uma ferramenta benéfica e essencial tanto do ponto de vista econômico quanto científico, especialmente no cenário no qual nos encontramos: ausência de medicamentos com eficácia clinicamente comprovada contra o novo vírus. Nesse contexto, torna-se cada vez mais importante o trabalho em rede, o compartilhamento aberto de dados em pesquisa, assim como a rápida identificação de especialistas em áreas específicas da ciência. A Plataforma Lattes é o principal repositório curricular de profissionais e especialistas no Brasil, sendo possível extrair diversas informações dos mesmos tais como

<sup>58</sup>Pesquisa em Gestão e Inovação Científica (Fiocruz), Email: henrique.chaves@fiocruz.br.

<sup>59</sup>Fundação Oswaldo Cruz (Fiocruz), Email: brandaojulia093@gmail.com.

<sup>60</sup>Fundação Oswaldo Cruz (Fiocruz), Email: marcos.hir@far.fiocruz.br.

<sup>61</sup>Fundação Oswaldo Cruz (Fiocruz), Email: carla.silveira@fiocruz.br.

<sup>62</sup>Fundação Oswaldo Cruz (Fiocruz), Email: jorge.magalhaes@fiocruz.br.

<sup>63</sup>Pós-doutorado em Ciência da Computação, Universidade de São Paulo, Email: jesus.mena@ufabc.edu.br.

instituição onde trabalham, projetos, linhas de pesquisa, produções científicas e tecnológicas, dentre outras. Baseado no mapeamento desses currículos, foi realizado um estudo comparativo em dois momentos (2020 e 2021) visando analisar a evolução do cenário científico no país dos especialistas em covid-19. Para facilitar a identificação, extração e tratamento dessas informações, utilizou-se a ferramenta computacional ScriptLattes. Foi observado um aumento substancial na quantidade de currículos na plataforma após um ano de pandemia, assim como a quantidade de projetos iniciados, artigos completos publicados em periódicos, orientações. Em ambos os casos, observou-se também a predominância dos especialistas na região sudeste do Brasil. Os resultados mostraram-se eficientes pelo tratamento dos dados e o procedimento poderá ser replicado em diversas áreas da ciência.

**Palavras-chave:** Covid-19; Plataforma Lattes; ScriptLattes; big data; gestão da informação.

#### **ABSTRACT**

The covid-19 pandemic has proved to be one of the greatest health challenges of all time. In times of crises and pandemics, Big Data emerges as a beneficial and essential tool both from an economic and scientific point of view, especially in the scenario we find ourselves in: absence of drugs with clinically proven effectiveness against the new virus. In this context, networking, open sharing of research data, as well as the rapid identification of specialists in specific areas of science become increasingly important. The Lattes Platform is the main curriculum repository for professionals and specialists in Brazil, making it possible to extract various information from them, such as the institution where they work, projects, research lines, scientific and technological productions, among others. Based on the mapping of these curricula, a comparative study was carried out in two moments (2020 and 2021) in order to analyze the evolution of the scientific scenario in the country of specialists in covid-19. To facilitate the identification, extraction and treatment of this information, the computational tool ScriptLattes was used. A substantial increase in the number of CVs on the platform was observed after a year of the pandemic, as well as the number of projects started, complete articles published in journals, guidelines. In both cases, there was also a predominance of specialists in the southeast region of Brazil. The results proved to be efficient due to data processing and the procedure could be replicated in several areas of science.

**Keywords:** Covid-19, Lattes Platform, ScriptLattes, big data, information management.

## RESUMEN

La pandemia de covid-19 ha demostrado ser uno de los mayores desafíos de salud de todos los tiempos. En tiempos de crisis y pandemias, el Big Data surge como una herramienta beneficiosa e imprescindible tanto desde el punto de vista económico como científico, especialmente en el escenario en el que nos encontramos: ausencia de fármacos con eficacia clínicamente probada frente al nuevo virus. En este contexto, la creación de redes, el intercambio abierto de datos de investigación, así como la identificación rápida de especialistas en áreas específicas de la ciencia, se vuelven cada vez más importantes. La Plataforma Lattes es el principal repositorio de currículos de profesionales y especialistas de Brasil, siendo posible extraer de ellos diversas informaciones, como la institución donde actúan, proyectos, líneas de investigación, producciones científicas y tecnológicas, entre otras. A partir del mapeo de estos planes de estudios, se realizó un estudio comparativo en dos momentos (2020 y 2021) con el fin de analizar la evolución del escenario científico en el país de los especialistas en covid-19. Para facilitar la identificación, extracción y tratamiento de esta información se utilizó la herramienta computacional ScriptLattes. Se observó un aumento sustancial en la cantidad de CV en la plataforma después de un año de la pandemia, así como en la cantidad de proyectos iniciados, artículos completos publicados en revistas, directrices. En ambos casos, también hubo predominio de especialistas en la región sureste de Brasil. Los resultados demostraron ser eficientes debido a la minería de datos y el procedimiento podría replicarse en varias áreas de la ciencia.

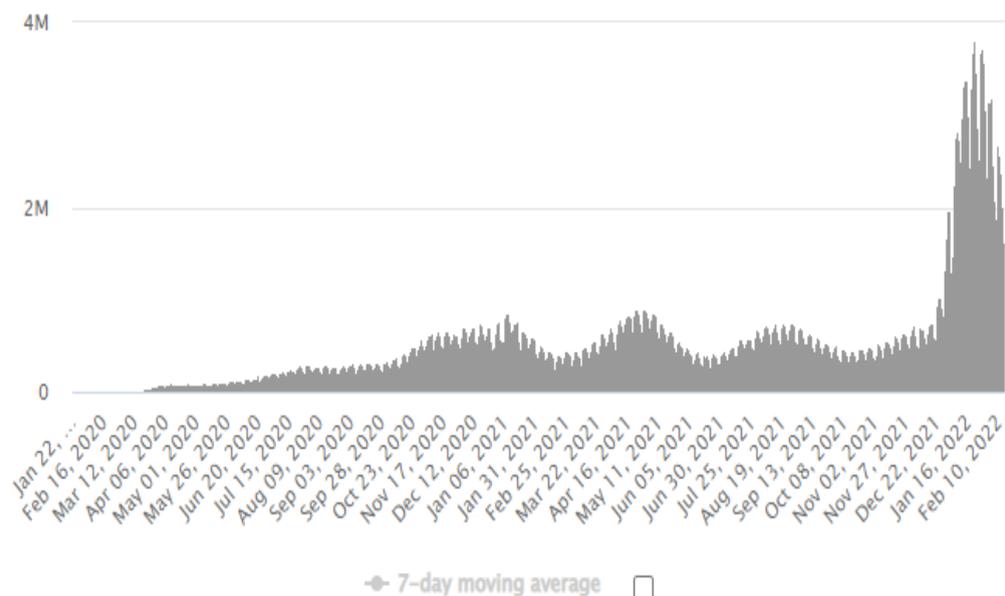
**Palabras clave:** Covid-19, Lattes Platform, ScriptLattes, big data, gestión de la información.

## 1 INTRODUÇÃO

A pandemia de Covid-19 foi relatada pela primeira vez em dezembro de 2019 em Wuhan, na província chinesa de Hubei<sup>1,2</sup>. Mas somente em 31 de janeiro de 2020, a Organização Mundial da Saúde (OMS) declarou a infecção pelo novo coronavírus como uma emergência global<sup>3</sup>. Inicialmente foi identificado como um novo tipo de coronavírus (síndrome respiratória aguda grave coronavírus 2 ou SARS-CoV-2), e posteriormente renomeado como Doença de Coronavírus-19 ou covid-19<sup>4,5</sup>. A facilidade de transmissão do vírus resultou em uma tragédia humana em nível global resultando em impactos sociais, políticos, culturais, históricos e econômicos como nunca visto antes em outras pandemias. Até a data de submissão do presente artigo, a pandemia tinha

atingido 223 países, apresentando um total de 415.244.367 de casos confirmados, 337.582.642 casos recuperados, e 5.849.986 de mortes (Figura 1)<sup>6</sup>.

**Figura 1.** Evolução na quantidade de novos casos de Covid-19 no mundo.



**Fonte:** Adaptado do site <https://www.worldometers.info/coronavirus/> (2021).

No Brasil, a população convive com a pandemia do novo coronavírus desde 11 de março de 2020<sup>3</sup>. O conhecimento científico limitado sobre a nova doença levou a questionamentos sobre quais seriam as melhores estratégias a serem utilizadas para combater a epidemia, que no caso do Brasil, apresentam outros agravantes, tais como a extrema desigualdade social, com populações de baixa renda vivendo em condições precárias de saneamento e habitação e em situação de aglomeração<sup>7</sup>. Diferente de outros países que conseguiram controlar de maneira eficaz a pandemia, no Brasil ocorreu ainda o agravante da desqualificação dos riscos de propagação da doença, assim como a adoção de medidas de prevenção por parte dos representantes do governo brasileiro, em especial, na figura do Presidente da República (Jair Bolsonaro)<sup>8</sup>. Ao invés do governo incentivar medidas de proteção e tratamentos baseados em evidências cientificamente comprovadas, houve a defesa exacerbada da denominada

“imunidade de rebanho”<sup>8</sup>. Assim como em outros países que mais sofreram os efeitos da pandemia, com a justificativa de defender o emprego e a economia, foi realizada uma oposição ferrenha ao isolamento social, resultando em um rápido e crescente número de contágios e mortes<sup>9</sup>. Até a data de submissão do presente artigo, a pandemia no Brasil apresentava um total de 27.538.503 de casos confirmados e 638.835 de mortes<sup>10</sup>.

Com os sistemas hospitalares de todos os países em colapso, devido aos crescentes casos de infecção pela doença, grande parte da sociedade teve que se mobilizar para impedir o avanço do vírus. O surto pandêmico levou a comunidade científica a repensar sobre uma possível resposta tecnológica que poderia auxiliar a comunidade médica, dentre elas, o uso de ferramentas para tratar o Big Data em saúde. No século 21 o Big Data é definido por muitos autores pelos “7 V’s”: velocidade, volume, variedade, veracidade, versatilidade e viabilidade, no qual a junção das seis primeiras dimensões levaria ao sétimo V: valor<sup>11-13</sup>. O *Big Data* em saúde permite a análise de grandes volumes de informação provenientes das mais diversas fontes (redes sociais, artigos científicos, patentes, etc), permitindo à comunidade médica, gestores de saúde e formuladores de políticas públicas estabelecerem padrões e antecipar ações frente às doenças. Dessa maneira, através do uso sistemático dos dados e das tecnologias do *Big Data*, insights e novas soluções são geradas, auxiliando na resolução dos problemas tanto da saúde pública quanto privada.

Diariamente são gerados quintilhões de dados na web em praticamente todos os lugares, tais como universidades, empresas e residências, no qual esses dados são oriundos das mais diversas fontes, tais como redes sociais, internet das coisas, dispositivos móveis, transações bancárias e comerciais, satélites, sistemas de monitoramento, sensores (dados de localização e dados meteorológicos), registro de softwares, dentre outros<sup>14</sup>. Assim, novas arquiteturas computacionais devem ser desenvolvidas e adotadas para manipular, armazenar e extrair valores desses dados. No que tange a saúde, as ferramentas do *Big Data*, otimizariam tanto o tempo quanto a administração

dos recursos financeiros, evitando desperdícios e melhorando a prestação de serviços à comunidade<sup>15</sup>.

Uma vez que a comunidade médica ainda está aprendendo a lidar com a nova pandemia, um dos fatores que podem contribuir para o sucesso do *Big Data* em covid-19 é a integração e disponibilização dos dados médicos dos pacientes (histórico de doenças, idade, tamanho, tratamentos realizados etc.), pois fornecem insumos para as autoridades médicas estabelecerem padrões de atendimento baseados em procedimentos com eficácia anteriormente comprovada. A pandemia de covid-19 nos mostrou claramente a importância no compartilhamento rápido dos dados, sejam eles políticos, epidemiológicos, moleculares (a partir das sequências aos alvos de medicamentos) ou, ainda, estratégias de intervenção<sup>16</sup>.

As análises dos dados fornecidos pelo *Big Data* tornam-se componentes cruciais para que as atividades de tomada de decisão sejam mais assertivas nos mais diversos. Por exemplo, através do uso das tecnologias de inteligência artificial (IA) e informações do Big Data, outras pandemias como a da covid-19 poderiam ser mais facilmente contidas ou evitadas, pois baseado em experiências passadas, seria possível prever as ações necessárias para combatê-las, assim como encontrar a cura em um prazo menor. Dessa maneira, IA aliada a medicina em rede, nos mostram como as informações do *Big Data* corroboram na compreensão das pandemias, seja pela definição de novas doenças, medicamentos, terapias, ou ainda, pela identificação de alvos com o mínimo de erro<sup>17</sup>. Como exemplo se pode citar o estudo realizado por Hu *et al.* (2020), no qual utilizaram um conjunto de dados do vírus Covid-19 integrado a um modelo de *Deep Learning*. Foram analisadas 4.985 drogas disponíveis comercialmente que poderiam atuar contra as proteínas alvo do SARS-CoV-2, no qual o modelo listou 10 drogas como potenciais inibidores, como os estimulantes respiratórios (mesilato de almitrina e roflumilaste) e a droga já utilizadas no tratamento do HIV (abacavir)<sup>18</sup>.

Para se ter uma ideia da dimensão do Big Data no cenário de pandemia, até a data de submissão do presente artigo, ao utilizar o termo de busca “covid-19” no portal brasileiro de periódicos Scielo (*Scientific Electronic Library Online*), são encontrados um total de 7.131 artigos científicos publicados. Já no *Espacenet*, site de busca de patentes desenvolvido pelo EPO (*European Patent Office*), são encontradas um total de 12.218 famílias de patentes utilizando o mesmo termo de busca.

Como exemplo de *Big Data* no Brasil, é possível citar a Plataforma Lattes do CNPq. Nessa base de dados são disponibilizados os currículos acadêmicos de profissionais de todas as áreas do conhecimento, tanto de pesquisadores brasileiros quanto de pesquisadores estrangeiros que trabalham no Brasil ou desenvolveram projetos em parceria com brasileiros<sup>19</sup>. Para facilitar a extração simultânea dos dados dos pesquisadores cadastrados nessa Plataforma, foi desenvolvida a ferramenta computacional de código aberto ScriptLattes, que apresenta, organiza e disponibiliza as informações desses pesquisadores (produções científicas, orientações, colaborações, geolocalização, dentre outras)<sup>19-22</sup>. As informações geradas pelo programa fornecem uma visão global do cenário científico no país, facilitando, dessa maneira, as tomadas de decisão das agências governamentais, empresas e comunidades científicas. Um outro ponto positivo desse programa é que, através do mapa de geolocalização gerado pelo mesmo, torna-se possível mapear a produção científica no país, evidenciando as regiões que possuem capacidade científica instalada, assim como aquelas que precisam ser desenvolvidas pelos ministérios e agências governamentais.

Dessa maneira, este estudo avalia o uso do programa ScriptLattes na gestão das tecnologias do *Big Data* na Plataforma Lattes do CNPq, com o objetivo de identificar os principais cientistas seniores no Brasil na área do Covid-19 no início da pandemia e após um ano da mesma, assim como analisar suas principais produções científicas e tecnológicas.

## 2 METODOLOGIA

A pesquisa foi baseada em cientometria e bibliometria em dados abertos da ciência. Utilizou-se software *open source* ScriptLattes, em ambiente Linux para extração dos dados abertos na plataforma Lattes brasileira (veja <http://scriptlattes.sourceforge.net/>), visando identificar, em território brasileiro, os especialistas de maiores relevâncias atuando em pesquisas de covid-19. Dentre os filtros disponíveis na plataforma Lattes para extração dos dados, foram selecionados os abaixo listados a fim de identificar os pesquisadores com maior experiência acadêmica:

- a) Nível de formação (doutorado completo);
- b) Nacionalidade (brasileira ou estrangeira);
- c) Bolsista de produtividade em Pesquisa do CNPq (todas as categorias);
- d) Presença/registro no Diretório de Grupos de Pesquisa do CNPq.

No sentido de identificar, preliminarmente, o número de CVs disponíveis em cada estratégia de pesquisa, foram realizadas buscas no campo “avançado” da Plataforma Lattes utilizando os seguintes termos de busca: “Coronavírus\* or MERS-CoV\* or SARS-CoV\* or COVID\* or HCoV or SARS or MERS”. Também se aplicaram filtros associados à categoria “Bolsistas de Produtividade do CNPq” e “Presença no Diretório de Grupos de pesquisa”. A identificação e extração dos currículos foi realizada em duas etapas: a primeira em março de 2020 e a segunda em março de 2021. Os currículos foram categorizados pelo ID Lattes de cada especialista, um código de 16 dígitos de identificação dos currículos disponibilizados na plataforma<sup>23</sup>. Para ser possível utilizar o Scriptlattes, foi necessário criar uma lista desses CV, utilizando o software gratuito note++. Foi elencado uma lista com o número de ID de cada pesquisador na plataforma Lattes. Os dados foram agrupados, organizados, processados e, posteriormente, disponibilizados em ambiente na *web* em formato HTML.

### **3 RESULTADOS E DISCUSSÃO**

Na Tabela 1, pode-se observar os resultados das buscas realizadas em março dos anos de 2020 e 2021. Utilizando os termos de busca citados na metodologia, o termo, "COVID\*" foi o que apresentou uma maior quantidade de currículos para todos os filtros utilizados. Um dos fatores limitantes nas buscas foi o filtro “Bolsistas de Produtividade do CNPq”, no qual a quantidade de currículos diminuiu drasticamente, reduzindo em alguns casos para quase 1/6 do encontrado inicialmente. Dessa maneira decidiu-se analisar os currículos dos doutores brasileiros ou estrangeiros com qualquer nível de bolsa de produtividade e presença no diretório de GP, resultando em 818 e 2585 currículos, nos anos de 2020 e 2021, respectivamente.

**Tabela 1.** Número de pesquisadores identificados na base de dados Lattes contendo os termos de busca listados no método

<b>Número de currículos identificados na Plataforma Lattes utilizando os termos de busca e operadores listados no método</b>																
<b>Critérios selecionados para busca na base de dados Lattes</b>	<b>"Coronavír**"</b>		<b>"MERS-CoV**"</b>		<b>"SARS-CoV**"</b>		<b>"COVID**"</b>		<b>"HCoV"</b>		<b>"SARS"</b>		<b>"MERS"</b>		<b>((Coronavír*) or (MERS-CoV*) or (SARS-CoV*) or (COVID*) or (HCoV) or (SARS) or (MERS))</b>	
<b>Ano da extração dos currículos</b>	<b>2020</b>	<b>2021</b>	<b>2020</b>	<b>2021</b>	<b>2020</b>	<b>2021</b>	<b>2020</b>	<b>2021</b>	<b>2020</b>	<b>2021</b>	<b>2020</b>	<b>2021</b>	<b>2020</b>	<b>2021</b>	<b>2020</b>	<b>2021</b>
<b>Todos (doutores e demais pesquisadores, brasileiros ou estrangeiros)</b>	54 23	18 23 2	23	55	10 54	73 32	12 65 8	70 54 7	9	13	15 59	82 02	69	11 3	142 88	7358 2
<b>Doutores brasileiros ou estrangeiros</b>	27 35	72 65	15	32	57 1	32 33	64 87	26 15 6	8	9	88 6	88 6	46	68	705 9	2744 7
<b>Doutores brasileiros</b>	26 73	26 73	14	31	55 7	31 70	63 48	25 62 6	7	7	84 9	84 9	38	58	688 4	2684 7
<b>Doutores brasileiros ou estrangeiros com qualquer nível de bolsa de produtividade</b>	31 1	71 0	3	3	10 7	51 7	80 3	25 83	0	0	15 7	57 3	10	12	888	2775
<b>Doutores brasileiros com qualquer nível de bolsa de produtividade</b>	30 2	68 8	3	6	10 5	50 4	77 4	25 10	0	0	15 1	55 5	10	12	856	2694

<b>Doutores brasileiros ou estrangeiros com qualquer nível de bolsa de produtividade e presença no diretório de GP</b>	28 9	62 2	9	6	10 5	48 2	74 2	24 04	0	0	14 1	53 3	9	11	<b>792</b>	<b>2615</b>
<b>Doutores brasileiros com qualquer nível de bolsa de produtividade e presença no diretório de GP</b>	28 2	64 8	9	6	96	47 1	72 2	23 55	0	0	13 6	51 8	9	11	790	2529

**Fonte:** elaborado pelos autores com base na Plataforma Lattes (2020 e 2021)

Os resultados extraídos, tratados e, posteriormente, plotados em informações essenciais, estão disponibilizados e podem ser acessadas para cada ano em questão, nos seguintes endereços eletrônicos:

- Ano de 2020: <https://pesquisa.ufabc.edu.br/cientometria/covid/2020>
- Ano de 2021: <https://pesquisa.ufabc.edu.br/cientometria/covid/2021>

A Figura 2 apresenta a página principal gerada pelo ScipLattes da busca realizada em março de 2020 em formato HTML. O cabeçalho da página mostra os links onde pode-se extrair várias informações, tais como, produção bibliográfica, a produção técnica, orientações, projetos, dentre outras informações dos especialistas seniores analisados pelo software. Todos os resultados nos links da página principal apresentam informações desde o ano de 2010.

**Figura 2.** Menu Principal gerado pelo ScipLattes.

**Especialistas Seniores em Covid-19 com atuação no Brasil**

[ [Membros](#) | [Produção bibliográfica](#) | [Produção técnica](#) | [Produção artística](#) | [Orientações](#) | [Projetos](#) | [Prêmios](#) | [Eventos](#) | [Gráfico de colaborações](#) | [Mapa de geolocalização](#) | [Métricas](#) ]

**Produção bibliográfica**

- Artigos completos publicados em periódicos (51864)
- Livros publicados/organizados ou edições (2523)
- Capítulos de livros publicados (9740)
- Textos em jornais de notícias/revistas (5509)
- Trabalhos completos publicados em anais de congressos (8802)
- Resumos expandidos publicados em anais de congressos (3389)
- Resumos publicados em anais de congressos (23335)
- Artigos aceitos para publicação (1044)
- Apresentações de trabalho (25285)
- Demais tipos de produção bibliográfica (3116)
- Total de produção bibliográfica (134607)

**Produção técnica**

- Produtos tecnológicos (693)
- Processos ou técnicas (154)
- Trabalhos técnicos (23293)
- Demais tipos de produção técnica (9593)
- Total de produção técnica (33733)

**Produção artística**

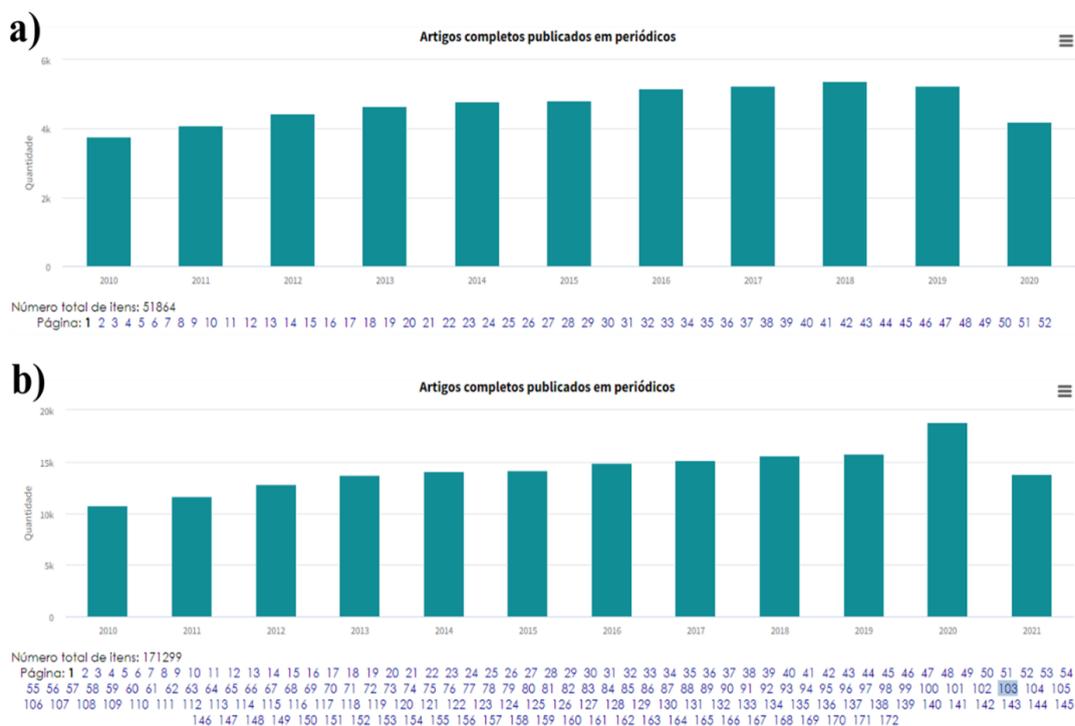
- Total de produção artística (189)

**Fonte:** ScipLattes (2021).

Clicando qualquer um dos itens desse menu, é possível observar a evolução da produção científica ao longo dos anos, sendo possível acessar na página a produção por ano e obter a lista de publicações ordenadas

alfabeticamente pelo título do artigo publicado. A Figura 3 exibe, respectivamente, a quantidade de artigos publicados em periódicos pelos pesquisadores até a data da primeira (Figura 3a) e segunda extração (Figura 3b) dos currículos pelo software ScriptLattes. Em ambos os casos é observado um aumento na quantidade de publicações ao longo dos anos. Na primeira extração das informações dos currículos, tinham sido publicados um total de 51.864 artigos em periódicos, aumentando mais de 03 (três) vezes em apenas 01 (um) ano, especificamente, 171.299 manuscritos na segunda extração. Concomitantemente ao início da pandemia de covid-19 no Brasil, observa-se no ano de 2020 uma quantidade maior de publicações (18.864) até a data em que os dados foram analisados (Figura 3b).

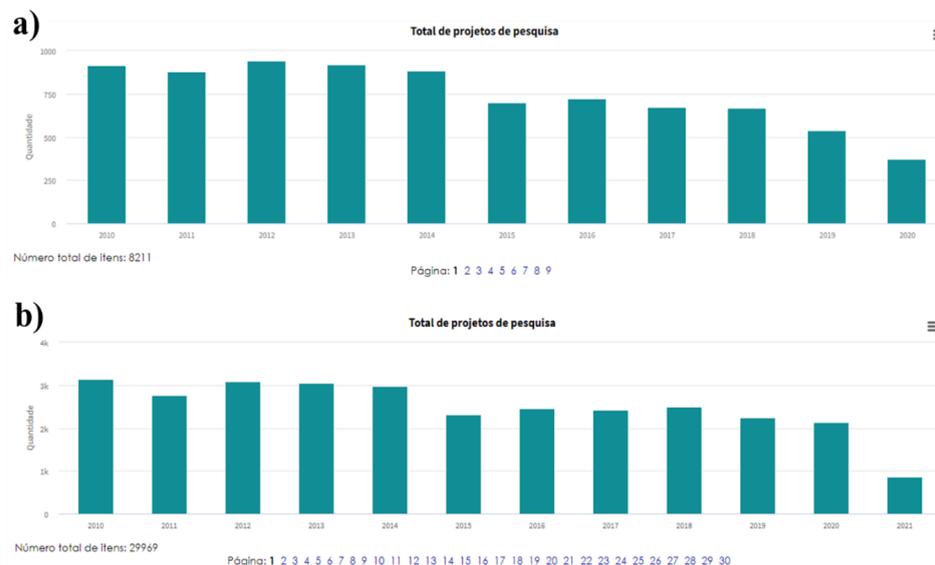
**Figura 3:** Exemplos de resultados que podem ser visualizados a partir do índice listado na Figura 2.



Fonte: ScriptLattes (2021).

A Figura 4 apresenta, respectivamente, a quantidade de projetos implementados pelos pesquisadores ao longo dos anos na primeira (Figura 4a) e segunda extração (Figura 4b) dos currículos. Em ambos os casos se observou uma diminuição na quantidade de novos projetos ao longo dos anos. Até a data da primeira extração, tinham sido implementados 8.211 projetos pelos especialistas analisados. Com o aumento na quantidade de currículos na segunda extração, foi observado também um aumento significativo na quantidade de novos projetos implementados (29.969).

**Figura 4:** Quantidade de projetos de pesquisa iniciados ao longo dos anos no Brasil.



Fonte: ScriptLattes (2021).

Na Figura 5 são exibidos, respectivamente, eventos organizados pelos especialistas seniores analisados no estudo na primeira (Figura 5a) e segunda extração (Figura 5b) dos currículos. Na primeira extração dos currículos tinham sido organizados 9.712 eventos acadêmicos pelos analistas analisados, aumentando para 34.662 na segunda extração. Em ambos os casos pode-se observar uma flutuação na quantidade de eventos organizados ao longo dos anos. Em fevereiro de 2020 foram registrados os primeiros casos de covid-19 no Brasil, o que levou a comunidade científica a repensar os meios de realizar

os seus eventos de divulgação científica, como congressos, seminários, simpósios e workshops. Dessa maneira, houve uma diminuição na quantidade de eventos presenciais nas diversas áreas do conhecimento. Por medida de segurança, a maioria dos eventos acadêmicos têm sido realizados, tanto no Brasil quanto em outros países, de maneira remota através de plataformas de videoconferência tais como o Google Meet<sup>®</sup>, Microsoft Teams<sup>®</sup>, Zoom<sup>®</sup>, Cisco Webex<sup>®</sup> e GoToMeeting<sup>®</sup>.

**Figura 5.** Quantidade de eventos organizados no Brasil ao longo dos anos.

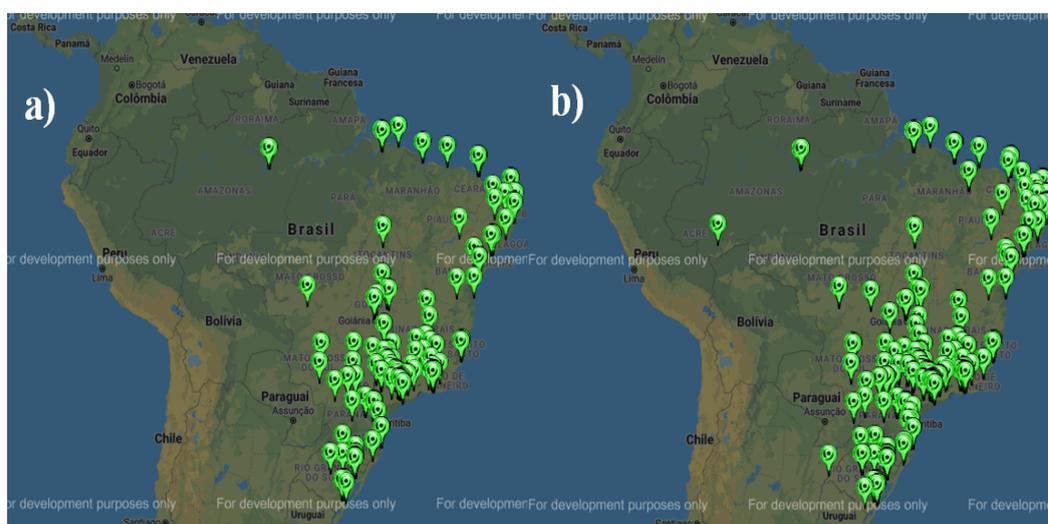


**Fonte:** ScriptLattes (2021).

Na Figura 6 são exibidos, respectivamente, os mapas de geolocalização dos especialistas seniores obtidos na primeira (Figura 6a) e segunda extração (Figura 6b) dos currículos. Os símbolos em verde identificam onde cada pesquisador sênior atua, ou seja, seu endereço profissional informado em seu currículo<sup>21</sup>. Ao clicar nesse símbolo, é possível ver algumas informações do pesquisador, como o seu nome, endereço profissional (universidade ou instituto) e o endereço para acessar o currículo Lattes. Em ambos os casos, com os termos de busca utilizados na pesquisa, observa-se um aumento na

quantidade de especialistas na área de estudo, principalmente nas regiões sul e sudeste do Brasil. Embora a presença dos pesquisadores, situam-se, praticamente, nas mesmas bases de outrora, a produção científica e tecnológica aumentaram progressivamente. Da mesma forma, após 01 (um) ano observou-se o aparecimento de um pesquisador no estado do Acre.

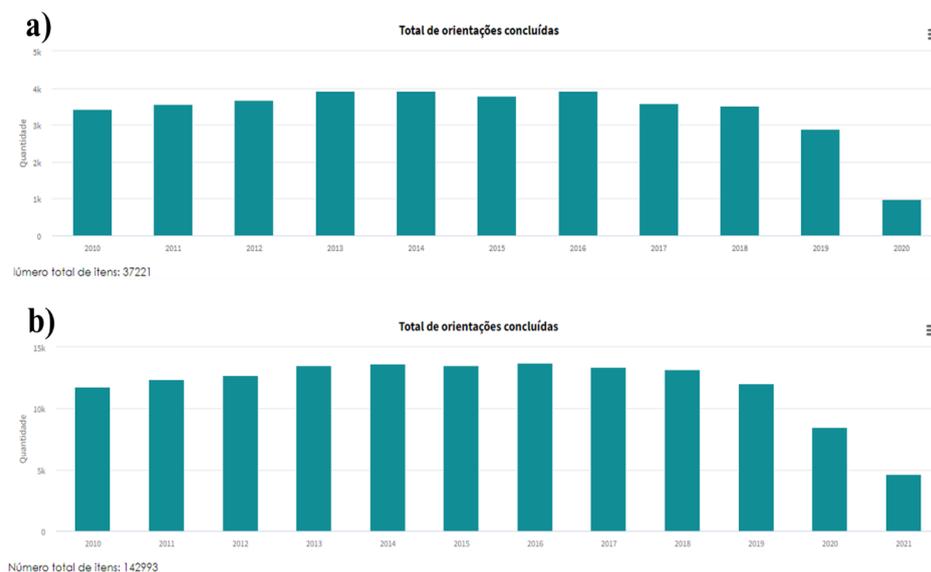
**Figura 6:** Mapa localização de habilidades essenciais em covid-19 no Brasil.



**Fonte:** ScriptLattes (2020).

No menu “Orientações”, podem ser observados diversos tipos de orientação acadêmica, desde alunos de iniciação científica até pós-doutorado. Na Figura 7 são exibidos, respectivamente, o número total de orientações concluídas pelos especialistas seniores obtidos na primeira (Figura 7a) e segunda extração (Figura 7b) pelo software ScriptLattes. A soma total de orientações concluídas pelos pesquisadores analisados até a data da primeira extração foi de 37.221. Já na segunda extração dos currículos, foi identificado um total de 142.993 orientações concluídas. Em ambos os casos, observou-se: i) uma maior quantidade de orientações concluídas pelos alunos de mestrado e ii) uma diminuição na quantidade de orientações concluídas a partir de 2016.

**Figura 7:** Quantidade de orientações realizadas pelos pesquisadores seniores entre 1982 e 2021.



**Fonte:** ScriptLattes (2021).

Uma das vantagens da ferramenta Scriptlattes é a possibilidade realizar a análise de redes sociais (ARS) através de outras ferramentas computacionais como Treecloud<sup>®</sup>, Gephi<sup>®</sup>, Cowo<sup>®</sup> e VOSviewer<sup>®</sup>, já acopladas ao Scriptlattes<sup>24</sup>. No presente estudo, os dados da Tabela 2 e o gráfico das Figura 8 foram produzidos com o auxílio do software gratuito e de código aberto Gephi<sup>®</sup>, que permite a visualização e exploração para todos os tipos de gráficos e redes<sup>25</sup>. A colaboração científica, no que tange aos termos de coautoria, pode ser observado na Tabela 2, no qual os vértices representam os pesquisadores e as arestas representam as ligações entre os vertice<sup>26</sup>. Nela são apresentadas as características de cada rede de coautoria. As redes foram construídas seguindo períodos sobrepostos/aumentados, isto é, 2010-2020 e 2010-2021. É possível notar o grande aumento tanto no número de pessoas quanto no número e grau de colaborações, evidenciando o maior interesse na temática. É importante destacar a característica de diâmetro da rede. Esta se manteve invariante, ou seja, igual a 13. Este atributo representa a maior distância entre quaisquer par de pesquisadores. A rede se tornou mais coesa no segundo período.

**Tabela 2.** Principais características das redes de colaboração

Características	2010-2020	2010-2021
Pessoas (vértices)	792	2615
Colaborações (arestas)	1.636	13.932
Pessoas com pelo menos um colaborador	628 (79,3%)	2.446 (93,5%)
Grau médio (número médio de colaboradores)	4,1	10,7
Diâmetro da rede (maior distância entre 2 pessoas)	13	13
Tamanho da maior componente conexa	573 (72,4%)	2398 (91,7%)

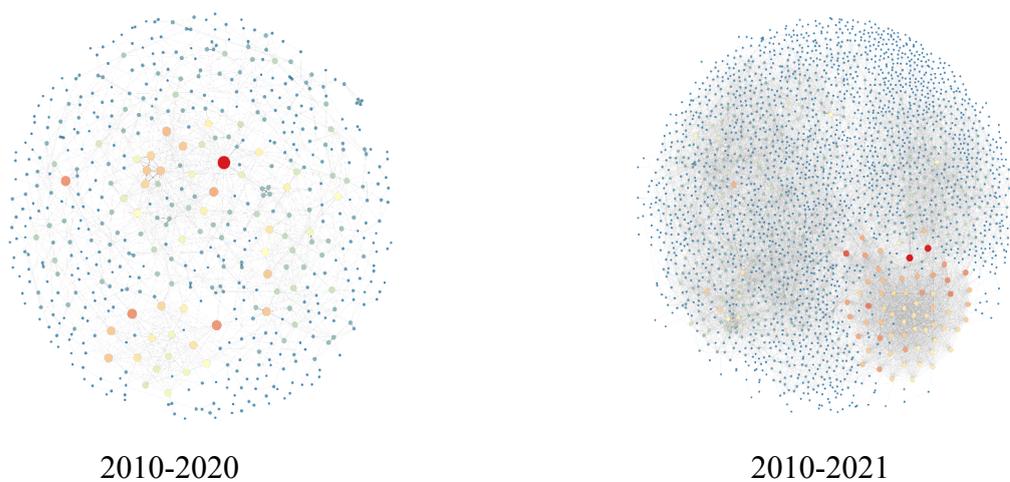
**Fonte:** Elaboração própria dos autores.

A Figura 8 apresenta a maior componente conexa de ambas as redes de colaboração. Uma versão interativa das redes está disponível em <https://pesquisa.ufabc.edu.br/cientometria/covid/rede/>. A colaboração acadêmica entre os pesquisadores selecionados foi identificada baseada nas publicações realizadas em coautoria entre eles. Nessa figura, os vértices representam os pesquisadores e as arestas representam as ligações entre os vértices. Tanto os diâmetros dos vértices quanto a sua cor estão associados ao número de conexões de cada pesquisador. Um pesquisador bem conectado terá um vértice de tamanho maior e a cor vermelha. Quanto mais próximo do vermelho, mais conexões esse pesquisador tem na sua rede. Já um pesquisador menos conectado terá um tamanho de vértice pequeno e uma cor azul clara. Na primeira figura, 2010-2020, é observado apenas um vértice na cor vermelha, referente ao pesquisador Mauricio Lima Barreto, pesquisador (especialista) da Fiocruz-Bahia e Bolsista de Produtividade em Pesquisa do CNPq - Nível 1A. Até a data em que foi processada a pesquisa, esse pesquisador possuía 280 artigos periódicos, 1 livro, 23 capítulos de livros e 304 publicações bibliográficas. Esse vértice apresenta uma centralidade de grau igual a 36, isto é, ele colabora de forma direta com 36 pesquisadores identificados no contexto desta pesquisa. Com esta mesma analogia, pode-se identificar e analisar qualquer pesquisador e suas respectivas interações.

Já na segunda figura, que compreende o período de 2010-2021, são observados dois vértices na cor vermelha. Um desses vértices é referente ao pesquisador Roberto Kant de Lima, Professor Titular aposentado da

Universidade Federal Fluminense (UFF) e Bolsista de Produtividade em Pesquisa do CNPq - Nível 1A. Até a data em que foi processada a pesquisa, este pesquisador possuía 22 artigos em periódicos, 16 livros, 15 capítulos de livros e 54 publicações bibliográficas. Esse vértice apresenta uma centralidade de grau igual a 115. O outro vértice é referente ao pesquisador Camilo Albuquerque de Braz, Professor Associado II de Antropologia da Universidade Federal de Goiás (UFG) e bolsista de Produtividade em Pesquisa do CNPq – Nível 2. Até a data em que foi processada a pesquisa, esse pesquisador possuía 25 artigos periódicos, 5 livros, 17 capítulos de livros, 113 produções bibliográficas e um vértice com centralidade de grau igual a 113.

**Figura 8:** Redes de coautoria (maior componente conexa) entre os pesquisadores.



**Fonte:** Visualização gerada usando o software livre Gephi.

Convém ressaltar que esta rede de pesquisadores é coesa, uma vez que há uma grande colaboração entre eles. Estas redes evidenciam graficamente o aumento dos pesquisadores e o aumento na sua colaboração. No primeiro período analisado, a maior componente conexa da rede está composta por 573

pesquisadores (72,4%) com 13.932 conexões. Já no segundo período analisado, a maior componente conexa da rede está composta por 2.398 pesquisadores (91,7%) com 13.932 conexões. Evidencia-se, portanto, o exponencial trabalho em sinergia, pelas equipes formadas e expressas no avanço das produções científicas e tecnológicas.

#### **4 CONCLUSÕES**

Surto pandêmicos repetem-se pelos séculos, apresentando semelhanças tanto nas suas formas de propagação quanto de contenção. Essa não foi a primeira vez que presenciamos uma pandemia de proporções continentais e mundiais no século XXI. Como exemplo podemos citar as epidemias de Ebola na África, a gripe aviária (H5N1) e as duas epidemias de coronavírus (pelo SARS-CoV e a síndrome respiratória do Oriente Médio – MERS), que em conjunto provocaram menos mortes do que a covid-19<sup>7</sup>. A principal diferença da pandemia de covid-19 em relação às outras reside no fato de todas as outras pandemias terem sido disseminadas geograficamente em menor escala e, também, terem sido mais facilmente controladas.

Mesmo com origens distintas, os surtos pandêmicos apresentam algumas características em comum, como o caos social, paralisação da vida cotidiana, mudanças no comportamento humano, propagação de informações falsas, deixando transparecer antigos problemas sociais e de saúde pública. Ao colocar sociedades inteiras sob ameaça, as pandemias acabam se tornando catalisadores de transformações sociais e econômicas, mobilizando a sociedade a repensar valores e atitudes. Devido ao temor frente à uma nova doença e a falta de conhecimento dela, a população acaba se apegando a crenças populares ou informações falsas de cura ou tratamento precoce. Como exemplo pode-se citar o uso da cloroquina e hidroxicloroquina incentivado pelo governo brasileiro para o tratamento precoce do Covid-19, sem nenhuma comprovação científica. A própria Organização Mundial da Saúde (OMS), afirmou que essa

medicação não reduzia a mortalidade e poderia causar ainda vários efeitos colaterais, como dor de cabeça e náuseas<sup>27</sup>.

E justamente em um momento tão delicado, cortes dos investimentos em saúde e ciência no Brasil foram realizados pelo governo<sup>7</sup>. Levando em consideração que o Brasil é referência em pesquisa de doenças negligenciadas (dengue, zika, malária, chikungunya, leishmaniose), demonstra a importância dos investimentos e da valorização da pesquisa no país, assim como a manutenção de um Sistema Único de Saúde robusto capaz de garantir o direito universal à saúde.

Na era informacional de dados, o Big Data em saúde não é diferente, pois grandes volumes de dados são adicionados constantemente à Web. Assim, novas ferramentas para a identificação, extração e processamento desses dados são prementes. Softwares como o *open source* ScriptLattes, são cada vez mais indispensáveis, a fim de subsidiar os gestores na obtenção da informação essencial de forma gráfica e prática.

As informações obtidas com o uso do programa, serviram como importantes indicadores sobre alguns aspectos da ciência no Brasil na área do presente estudo. Tanto as análises micro (pesquisador individual e grupos de pesquisadores) quanto macro (redes de colaboração ou representatividade da área), retrataram o que é desenvolvido e publicado no Brasil nos setores de Pesquisa, Desenvolvimento e Inovação (P,D&I). Os resultados comparativos, no intervalo de 01 ano da pandemia de covid-19, demonstraram que a quantidade de produções e pesquisas em covid-19, foram intensificadas no Brasil. Estes foram evidenciados pela quantidade de novos especialistas que surgiram um após da primeira extração com os mesmos termos de busca utilizados na Plataforma Lattes.

Cabe destacar, algumas limitações tanto na Plataforma Lattes quanto nas informações apresentadas pelo software ScriptLattes. Através dos operadores booleanos é possível realizar buscas simples na Plataforma Lattes. Já buscas relacionais mais complexas, que era o desejado nesse estudo, ainda

não são possíveis na plataforma. O motor de busca disponível da plataforma não permite extração e recuperação das informações de forma gerencial, como por exemplo, identificar os termos de busca numa mesma sentença ou parágrafo, resumo ou título da produção científica ou tecnológica dos pesquisadores analisados. Em relação ao programa ScriptLattes, ainda não é possível saber se as informações apresentadas na página gerada pelo mesmo (orientações, artigos científicos, eventos realizados, dentre outras), são relativas aos termos de busca utilizados.

Apesar da delimitação da pesquisa, a principal contribuição do trabalho recai na possibilidade de se obter na Plataforma Lattes informações essenciais e confiáveis dos especialistas em covid-19 no Brasil. Por ser uma doença nova, a gestão do conhecimento torna-se um fator imprescindível para avanços na área da saúde, política e economia, possibilitando ações mais rápidas e assertivas pelos gestores e tomadores de decisão. Ademais, as informações obtidas pelo programa permitiram uma visão geral sobre o engajamento desses pesquisadores com a atual pandemia, demonstrado através das contribuições (artigos, eventos, orientações) de cada um deles na sua área de conhecimento. Novas metodologias deverão ser aprimoradas e desenvolvidas para que o tratamento desses dados seja mais assertivo.

A ferramenta computacional ScriptLattes têm se mostrado importante na gestão da informação e, a cada nova atualização, tenta minimizar os problemas encontrados tanto nesse como em outros estudos já reportados na literatura<sup>21,22,28</sup>. A indexação das informações na Plataforma Lattes é o principal problema enfrentado pelo ScriptLattes na extração dos dados. Dessa maneira, para melhorar o processo de extração e recuperação dos dados pelo ScriptLattes, torna-se necessário um maior empenho também por parte dos pesquisadores e instituições em indexar de maneira padronizada as informações na Plataforma Lattes.

Concernente a outros surtos pandêmicos em que a humanidade já passou, se observa a premente necessidade de se investir e valorizar as

pesquisas científicas em todas as áreas do conhecimento. Apesar dos progressos obtidos pela ciência ao longo dos séculos, há muitas lacunas ainda a serem preenchidas, para evitar que esse tipo de tragédia assole novamente a humanidade. Dessa maneira, motivado pelos avanços e aplicações recentes do *Big Data* nas diversas áreas do conhecimento, este artigo procurou enfatizar sua importância na resposta ao surto de covid-19. Espera-se que este artigo possibilite aos pesquisadores e comunidades científicas novos *insights* sobre como o Big Data em saúde pode ajudar na compreensão da atual pandemia e, assim, impulsionar novos estudos visando controlar tanto o surto de covid-19, quanto outros que vierem a surgir.

## REFERÊNCIAS

1. Agbehadji IE, Awuzie BO, Ngowi AB, et al. Review of Big Data Analytics, Artificial Intelligence and Nature-Inspired Computing Models towards Accurate Detection of COVID-19 Pandemic Cases and Contact Tacing. *Int J Environ Res Public Health* 2020; 17: 5330.
  2. Esakandari H, Nabi-Afjadi M, Fakkari-Afjadi J, et al. A comprehensive review of COVID-19 characteristics. *Biol Proced Online* 2020; 22: 19.
  3. Strabelli TMV, Uip DE. COVID-19 e o Coração. *Arq Bras Cardiol* 2020; 114: 598–600.
  4. Brodeur A, Gray D, Islam A, et al. A literature review of the economics of COVID-19. *J Econ Surv* 2021; 35: 1007–1044.
  5. Qiu Y, Chen X, Shi W. Impacts of social and economic factors on the transmission of coronavirus disease 2019 (COVID-19) in China. *J Popul Econ* 2020; 33: 1127–1172.
  6. COVID Live - Coronavirus Statistics - Worldometer, <https://www.worldometers.info/coronavirus/> (2022, accessed 15 February 2022).
  7. Werneck GL, Carvalho MS. A pandemia de COVID-19 no Brasil: crônica de uma crise sanitária anunciada. *Cad Saúde Pública* 2020; 36: e00068820.
- Campos GW de S. O pesadelo macabro da Covid-19 no Brasil: entre negacionismos e desvarios. *Trab Educ E Saúde* 2020; 18: e00279111.

9. Burki T. COVID-19 in Latin America. *Lancet Infect Dis* 2020; 20: 547–548.
10. MonitoraCovid-19, <https://bigdata-covid19.icict.fiocruz.br/> (2022, accessed 15 February 2022).
11. Dias JA, Duarte P. BIG DATA OPPORTUNITIES IN HEALTHCARE. HOW CAN MEDICAL AFFAIRS CONTRIBUTE? 2015; 7.
12. Gandomi A, Haider M. Beyond the hype: Big data concepts, methods, and analytics. *Int J Inf Manag* 2015; 35: 137–144.
13. White S. A review of big data in health care: challenges and opportunities. *Open Access Bioinforma* 2014; 13.
14. Martino BD, Aversa R, Cretella G, et al. Big data (lost) in the cloud. *Int J Big Data Intell* 2014; 1: 3.
15. Galvão AB, Valentim RA de M. Desafios para os Avanços da Análise de Big Data na Saúde. In: *Anais Estendidos do Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*. SBC, pp. 155–160.
16. Dagliati A, Malovini A, Tibollo V, et al. Health informatics and EHR to support clinical research in the COVID-19 pandemic: an overview. *Brief Bioinform* 2021; 22: 812–822.
17. Zhou Y, Wang F, Tang J, et al. Artificial intelligence in COVID-19 drug repurposing. *Lancet Digit Health* 2020; 2: e667–e676.
18. Hu F, Jiang J, Yin P. Prediction of potential commercial inhibitors against SARS-CoV-2 by multi-task deep model. 2020; 9.
19. Brito AGC de, Quoniam L, Mena-Chalco JP. Exploração da Plataforma Lattes por assunto: proposta de metodologia. *Transinformação* 2016; 28: 77–86.
20. Ferraz RRN, Barnabé AS, Quoniam L, et al. Aspectos históricos da criação dos grupos de pesquisa em dengue no Brasil com a utilização da ferramenta computacional ScriptGP. *Ciênc Saúde Coletiva* 2018; 23: 837–848.
21. Magalhães J, Hir M, Quoniam L, et al. A MANAGEMENT TOOL TO AID IN THE TROPICAL OUTBREAK OF THE 21st CENTURY: SENIOR SCIENTISTS AND THEIR KNOWLEDGE OF THE TRIPLE THREAT DENGUE, ZIKA AND CHIKUNGUNYA. *Probl Manag 21st Century* 2020; 15: 40–55.
22. Mena-Chalco JP, Junior RMC. scriptLattes: an open-source knowledge extraction system from the Lattes platform. *J Braz Comput Soc* 2009; 10.

23. Mena-Chalco JP, Digiampietri LA, Lopes FM, et al. Brazilian bibliometric coauthorship networks. *J Assoc Inf Sci Technol* 2014; 65: 1424–1445.
24. Quoniam L, Ferraz RRN, Alvares LMA de R. Avaliação de redes multidisciplinares com a ferramenta scriptlattes: os casos da nanotecnologia, da dengue e de um programa de pós-graduação Stricto Sensu em Administração. *Encontros Bibli Rev Eletrônica Bibl E Ciênc Informação* 2014; 19: 67–97.
25. Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proc Int AAAI Conf Web Soc Media* 2009; 3: 361–362.
26. Giordano DM, Bruning E, Bordin AS. Uso do scriptLattes e Gephi na Análise da Colaboração Científica. *An Comput Beach* 2015; 239–248.
27. CNN BRASIL. OMS, Europa e Anvisa não recomendam cloroquina para tratar a Covid-19. CNN Brasil, <https://www.cnnbrasil.com.br/saude/entenda-as-recomendacoes-das-instituicoes-de-saude-contra-o-uso-de-cloroquina/> (2021, accessed 18 November 2021).
28. Magalhães JL, Quoniam L, Mena-Chalco JP, et al. Extração e tratamento de dados na base lattes para identificação de core competencies em dengue. *Informação Informação* 2014; 19: 30