

Uma análise do programa Saúde sem Fake News através de uma abordagem baseada em análise de dados

An analysis of the Saúde sem Fake News program based on data analytics

Jully Porto Lopes Melo  

Anelise Souza Rocha  

Larissa Machado Vieira   

Douglas Farias Cordeiro   

Resumo

A circulação de *fake news* tem se tornado um dos maiores desafios no âmbito da comunicação, sob os mais variados aspectos, estando isso diretamente ligado ao advento e acessibilidade às tecnologias de informação e comunicação. Diante disso, autoridades têm empenhado esforços em ações que buscam o enfrentamento deste problema, através, por exemplo, do desenvolvimento de projetos que promovam a conscientização social. Neste contexto, se levando em conta o cenário da pandemia da Covid-19, a circulação de *fake news* se torna um problema que demanda uma maior atenção. Este artigo se propõe a realizar um estudo exploratório, através da aplicação de análise de dados, contribuindo com análise das *fake news* na área da saúde, a partir da aplicação de técnicas de processamento de linguagem natural em dados obtidos de publicações veiculadas pelo projeto do Saúde sem *Fake News*, do Ministério da Saúde.

Palavras-chave: Covid-19; Fake news; Análise de dados; Mineração de textos.

Abstract

The circulation of fake news has become one of the biggest challenges in the field of communication, under the most varied aspects, this being directly linked to the advent and accessibility to information and communication technologies. Authorities have made efforts in actions that seek to face this problem, through, for example, the development of projects that promote social awareness. At the Covid-19 pandemic scenario, the circulation of fake news becomes a problem that demands greater attention. This article proposes to carry out an analytical study, through the application of data analysis, to understand the panorama of fakes news in the health area, based on analyzes carried out with the publications published by the Saúde sem Fake News project, from the Brazilian Ministry of Health.

Keywords: Covid-19; Fake news; Data analytics; Text mining.



folha de rosto

Revista de Biblioteconomia e Ciência da Informação

Juazeiro do Norte, v. 7, n. 2, p. 3-19, maio/ago. 2021. ISSN 2447-0120. DOI 10.46902/2021n2p3-19.

1 Introdução

Os avanços tecnológicos alcançados a partir da segunda metade do século XX proporcionaram uma transformação no contexto da geração e disseminação da informação. De acordo com Lévy (1993), existe uma forte relação entre a inteligência coletiva, conceituada como “[...] uma inteligência distribuída por toda parte, incessantemente valorizada, coordenada em tempo real, que resulta em uma mobilização efetiva das competências” (LÉVY, 1993, p. 28) e os referidos avanços tecnológicos. Tal ligação está presente na oralidade, na escrita, e até mesmo através de soluções computacionais, as quais ultrapassam as características de simples ferramentas para solucionar problemas para um panorama onde se tornam ampliadores das habilidades humanas, tais como: calcular, inferir, organizar e armazenar (CONFORTO; CAVEDINI; MIRANDA; CAETANO, 2018).

Neste contexto, junto ao advento da Internet, as fronteiras da comunicação e informação se expandiram de tal forma que o indivíduo, enquanto usuário das tecnologias, passou de mero consumidor de conteúdo para um ator ativo, com habilidades para geração e transmissão de conteúdo, em um ambiente conhecido como Web 2.0 (O'REILLY, 2005). É importante salientar que ainda existem os conceitos de Web 3.0 e Web 4.0, conforme demonstrado por Ariño (2018), porém para as análises e reflexões realizadas neste estudo, os conceitos descritos por O'Reilly (2005) já demonstram as potencialidades alcançadas pela Internet.

Em março de 2020, a Organização Mundial de Saúde (OMS) reconheceu o estado de pandemia da Covid-19, situação a partir da qual surgiu a necessidade de medidas de contenção do contágio e manejo da doença cujas nuances ainda eram desconhecidas. Especialmente diante do cenário pontuado, a presente pesquisa se faz necessária por se compreender que a desinformação na área da saúde é potencialmente letal. Observa-se que sua circulação tem sido mais evidente no contexto da Covid-19, que teve sua origem em dezembro de 2019, quando a doença mencionada foi identificada pela primeira vez em Wuhan, na China, espalhando-se rapidamente pelo mundo.

A pandemia da Covid-19 evidenciou a desordem informativa, cenário que acontece há muitas décadas, em torno de temáticas da saúde. Tal situação é potencializada pelas facilidades de acesso às tecnologias da informação e comunicação, fator este característico da Web 2.0, na qual usuários deixaram de ser apenas consumidores ou espectadores, e passaram a ser geradores ou

disseminadores de (des)informação, alcançando um cenário que se configura enquanto uma revolução participativa democrática (SERRANO, 2013).

Este panorama é também acentuado por alguns fatores relacionados a questões de cunho político e comunicacional. Um dos pontos de destaque se refere à confusão informativa advinda de declarações de órgãos oficiais quanto ao manejo da doença, assim como a abordagem midiática a partir dessas contradições, visto que os diversos meios de comunicação se encontram repletos de incertezas no que tange a noticiar uma doença recente, que apresenta peculiaridades ainda não abarcadas pelos estudos científicos. Além disso, cabe ressaltar ainda o considerável volume de *fake news* que circulam comumente em torno de temas da saúde, abrangendo o coronavírus como mais uma de suas pautas, repleta de promessas não comprovadas de possíveis curas naturais e teorias conspiracionistas. Todo este cenário está envolto pelas incertezas naturais que rondam uma doença pouco conhecida, como é o caso da Covid-19.

Envolto neste ambiente que propicia a circulação de *fake news*, o Ministério da Saúde criou, ainda antes do surgimento da pandemia da Covid-19, em 2018, um projeto de enfrentamento à desinformação batizado como Saúde sem *Fake News*, voltado ao esclarecimento e verificação de notícias falsas que circulam na Internet. De acordo com relatório apresentado, durante o primeiro ano de atuação, o projeto conseguiu realizar 11,5 mil atendimentos, contemplando um total de 104 registros classificados como *fake news* (AGÊNCIA BRASIL, 2019), com destaque para temas como: vacinação, alimentação e câncer (VIEIRA; SILVA; CORDEIRO, 2019).

A importância característica de ações de enfrentamento à desinformação é algo fundamental e esperado, principalmente no âmbito da saúde e em situações que demonstram uma maior sensibilidade, como o caso de uma pandemia. Diante disso, para além das ações de combate, é interessante ainda compreender as especificidades que permeiam as *fake news* e sua circulação, dentro das quais, os conteúdos que são veiculados e compartilhados. Neste sentido, este artigo tem como proposta realizar um estudo exploratório e análise de padrões sobre o conjunto de publicações disponibilizadas pelo projeto Saúde sem *Fake News*. Os resultados alcançados apresentam as variações quantitativas temporais, assim como os indicativos de termos e temas abordados nos conteúdos tratados pelo projeto.

2 Desinformação, Fake News e Saúde

Na atualidade, um fenômeno denominado desinformação tem se alastrado e ganhado força de forma efetiva, merecendo atenção e exploração para compreensão e enfrentamento de suas causas e efeitos, os quais têm se mostrado, de certa maneira, de grande complexidade de controle. A partir disso, uma série de estudos relacionados a este tema emergem nos centros de pesquisa ao redor de todo o mundo, revelando curiosidade e preocupação quanto a este fenômeno, em especial, nas áreas de informação e comunicação.

Face a esta problemática, é fundamental destacar que um de seus principais terrenos se refere às redes sociais virtuais, as quais se caracterizam enquanto um ambiente de relação interpessoal com dinâmica própria, que, conforme abordado por Martino (2014, p. 55) “a ideia de rede ganhou mais força quando a tecnologia auxiliou a construção de redes sociais conectadas pela Internet, definidas pela interação via mídias digitais”. Em outras palavras, no âmbito das mídias digitais, se vê a transformação de um modelo tradicional de comunicação, onde se denota um fluxo de um para muitos, para um modelo comunicacional de muitos para muitos (LIMA; SANCHEZ-TARRAGÓ; MORAES; GRINGS; MAIA, 2020).

É nesta circunstância que a desinformação acaba por se alastrar de forma exponencial, através de um processo que é acelerado pela velocidade dos compartilhamentos de conteúdo que ocorrem nas redes sociais, onde se observa a conexão entre usuários que possuem opiniões semelhantes, consumindo notícias personalizadas que reforçam suas crenças e ideias preconcebidas. Essa característica promove a formação de ambientes denominados de bolhas, nos quais convivem aqueles que possuem similaridade de pensamento, opinião e ideias (KAKUTANI, 2018).

Cabe ressaltar que as redes sociais virtuais possuem um alto poder de promover a interatividade e o encontro entre os indivíduos que as utilizam, conectando pessoas de todo o mundo, independente da distância física que se encontrem umas das outras. Nesses ambientes virtuais, os conteúdos circulam em uma velocidade vertiginosa, gerando nos usuários uma necessidade de consumo rápido de tudo que lhes é oferecido, corroborando para um cenário de insegurança informativa (RAMONET, 2013), já que essas informações necessitam ser consumidas de forma imediata, pois, provavelmente, perderão sua importância no momento seguinte.

Nesse sentido, as redes sociais virtuais são um dos terrenos mais férteis para a disseminação de desinformação, que se configura como um conteúdo

intencionalmente enganoso, gerado com o intuito de prejudicar pessoas, grupos ou organizações (WARDLE, 2017). No âmbito da desinformação, depara-se com um de seus principais produtos, as *fake news*, que, de maneira literal, significam notícias falsas, possuindo um caráter deliberadamente fraudulento, cuja falsidade pode ser verificada, e, além disso, imitam uma fonte noticiosa legítima para aparentar credibilidade (ALLCOTT; GENTZKOW, 2017). As *fake news* também envolvem uma economia de emoções que as fomentam. Na Internet, os sites que as hospedam são financiados, muitas vezes, com o tráfego que tais conteúdos geram (BAKIR; MCSTAY, 2017).

O consumo e replicação de conteúdos desinformativos, como as *fake news*, provoca, diversas vezes, a eclosão de certas reações características de um cenário da pós-verdade, caracterizado como uma legitimação de procedimentos que se remetem a uma esfera de poder do que diretamente com a própria verdade (CARRERA, 2018), onde emoções como entusiasmo, desconfiança exacerbada e medo fortalecem crenças pré-existentes e incitam muitos sujeitos a terem uma visão romântica de processos políticos e sociais complexos (D'ANCONA, 2018).

Um estudo realizado pela *Science*, em 2018, apontou que uma pessoa tem 70% mais chance de compartilhar uma *fake news* em suas redes sociais do que uma notícia verdadeira. Isto acontece em virtude do apelo que a notícia falsa provoca nos indivíduos que se comprazem em seu consumo, estimulando sensações como surpresa e medo, em detrimento do caráter muitas vezes repetitivo das notícias verdadeiras que, em determinadas ocasiões, vão na contramão do que certos indivíduos e grupos acreditam (VOSOUGHI; ROY; ARAL, 2018).

Neste enquadramento, podem ser analisadas diversas questões acerca da temática da saúde, que se mal compreendida, podem levar a comportamentos geradores de risco, como a tomada de decisão em torno da não vacinação, a ingestão de medicamentos sem acompanhamento médico e a adoção de tratamentos alternativos, sem comprovação científica, em substituição aos protocolos baseados nas melhores evidências. No contexto da Covid-19, os comportamentos geradores de risco giram em torno da abstenção, por parte de certos indivíduos, da obediência às medidas para evitar o contágio, como o uso da máscara, do álcool em gel e a desobediência do isolamento social recomendado.

Nas próximas seções, será descrita a metodologia aplicada neste artigo, bem como os resultados obtidos e a interpretação dos mesmos, os quais retratam o fenômeno nocivo da circulação de *fake news*, que são fomentadoras de um

cenário pós-verdadeiro, derivando, deste contexto, parte dos comportamentos geradores de risco pontuados anteriormente.

3 Procedimentos Metodológicos

O objetivo principal deste artigo é realizar uma análise que possibilite a compreensão do panorama das *fake news* no contexto da saúde a partir da aplicação de soluções de análise exploratória e mineração de textos, sobre o conjunto de publicações do projeto Saúde Sem *Fake News*, do Ministério da Saúde. Para tanto, os procedimentos metodológicos adotados estão baseados no processo denominado de Descoberta do Conhecimento em Bases de Dados (do inglês, *Knowledge Discovery in Databases* - KDD).

O KDD, proposto por Fayyad, Piatetsky-Shapiro e Smyth (1996) é composto de cinco atividades, as quais possuem como principal característica o fato de serem iterativas e interativas, ou seja, uma atividade anterior pode ser retomada, caso necessário, e alterações na amostra de dados ou nos métodos de processamento e análise podem ser realizadas. As atividades que compreendem o KDD são: seleção, pré-processamento, transformação, mineração de dados e interpretação (RISTOSKI; PAULHEIM, 2016). A seleção de dados diz respeito à definição do conjunto de dados a ser considerado para as análises, assim como os procedimentos necessários para a obtenção dos mesmos (PROVOST; FAWCETT, 2016). Neste sentido, em consonância com o objetivo do artigo, a amostra de dados a ser trabalhada se refere ao conjunto de dados das publicações veiculadas no portal do projeto Saúde Sem *Fake News*, do Ministério da Saúde. É importante destacar que os sites do Governo Federal passaram por uma reformulação e reestruturação, nesse sentido, o novo portal do Ministério da Saúde não possui um link direto para as análises publicadas no âmbito do referido projeto, sendo que elas se encontram disponíveis no espelho do portal anterior, ainda acessível.

As publicações foram extraídas através da utilização de uma solução baseada em *Web Scraping* (MITCHELL, 2015), desenvolvida na linguagem de programação Python, a qual permitiu, inicialmente, a listagem íntegra de todas as publicações do projeto Saúde sem *Fake news* e, posteriormente, a extração dos atributos de interesse: título, data de publicação e texto. Neste sentido, foram obtidas 231 publicações, as quais foram persistidas em um arquivo estruturado em formato CSV (do inglês, *comma separated values*).

A etapa de pré-processamento se refere à aplicação de soluções que possibilitem a limpeza, filtragem e identificação de dados secundários (PROVOST;

FAWCETT, 2016). Neste sentido, foram utilizadas rotinas baseadas em expressões regulares, as quais permitem a identificação de sequências textuais estruturadas e posterior manipulação delas. Desta maneira, uma vez que o conteúdo textual apresentado se trata de análises realizadas por profissionais vinculados ao projeto, utilizando norma culta da escrita, foram realizados apenas procedimentos de limpeza no sentido de remover eventuais links presentes nos textos, uma vez que estes podem comprometer a eficiência dos algoritmos de mineração de texto utilizados. De semelhante forma, foram utilizadas expressões regulares para classificação dos conteúdos nas seguintes classes: Covid-19 *Fake News*, Covid-19 Verdade, Outros *Fake News*, Outros Verdade. A identificação dos grupos foi feita com base na ocorrência de termos como: covid, coronavírus, “é falso”, “é *fake news*”, “é verdade”, “não compartilhe”.

Após o pré-processamento dos dados, na etapa de transformação, responsável pela adequação das estruturas de dados aos formatos necessários para os procedimentos algorítmicos de análise (PROVOST; FAWCETT, 2016), foram geradas duas saídas, uma primeira seguindo o formato supracitado, CSV, constando dos atributos de identificação de data e de classes, e um segundo arquivo em formato textual, utilizado como entrada para aplicação das soluções de mineração de texto.

A etapa de mineração de dados se refere a um conjunto de processos e procedimentos que são utilizados, de forma algorítmica, para análise de bases de dados, de forma a encontrar padrões, tendências, associações, anomalias, entre outros, podendo tanto ter seu foco em análises descritivas ou exploratórias (AMARAL, 2016). No presente trabalho, é realizada inicialmente uma análise exploratória do conjunto de dados, de modo a evidenciar a evolução temporal das publicações realizadas junto ao projeto do Ministério da Saúde, em face das classes identificadas, com destaque para o período que sucede a determinação de pandemia por parte da OMS.

Além disso, os dados textuais obtidos foram analisados através do software Iramuteq, possibilitando a geração dos grafos de similitude para todas as classes identificadas, os quais apresentam os termos de maior frequência e as relações entre estes, obtidas através de conexões semânticas entre eles. De forma geral, a técnica de análise de similitude se refere à identificação dos termos que possuam maior frequência textual, conectados através de uma estrutura de um grafo no qual as arestas representam as ligações semânticas entre estes (VERGÈS; BOURICHE, 2001). Os resultados obtidos são interpretados em face das questões abordadas previamente sobre desinformação e *fake news*, assim

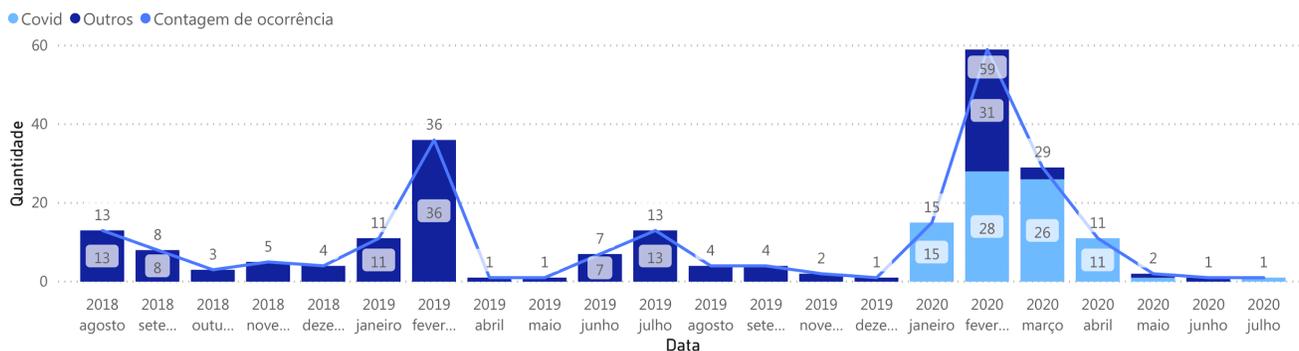
como fatores relacionados a políticas públicas de saúde no âmbito do Brasil e do mundo, os quais podem influenciar a potencialização deste tipo de fenômeno.

4 Resultados e Discussão

Em face do objetivo proposto neste artigo, e a partir da aplicação dos procedimentos metodológicos definidos, foi extraída a totalidade de publicações veiculadas no projeto Saúde sem *Fake News*, contemplando um total de 231 registros, entre o dia 24 de agosto de 2018 e 17 de julho de 2020. Como destacado anteriormente, não foram identificadas inserções após a data da última publicação obtida. Neste sentido, os primeiros resultados se referem às análises das séries temporais que demonstram a variação quantitativa do total de postagens, assim como a variação classificada, de acordo com as classes pré-definidas e identificadas através da solução algorítmica proposta.

Da totalidade de registros obtidos, 160 (69,2%) dizem respeito ao compartilhamento de conteúdos classificados como *fake news*, enquanto 71 (30,8%) são identificados como conteúdos verídicos ou informativos. A Figura 1 apresenta a série temporal para essa distribuição, onde é possível observar que no mês de fevereiro de 2020 houve um crescimento considerável no número de registros publicados, o que pode estar associado a um crescimento no número de compartilhamento de conteúdos por usuários de redes sociais e aplicativos de troca mensagens, assim como o crescimento no uso da Internet, motivado pelo isolamento social e as medidas de restrição.

Figura 1 - Série temporal quantitativa classificada por tipo (“Covid” e “Outros”)

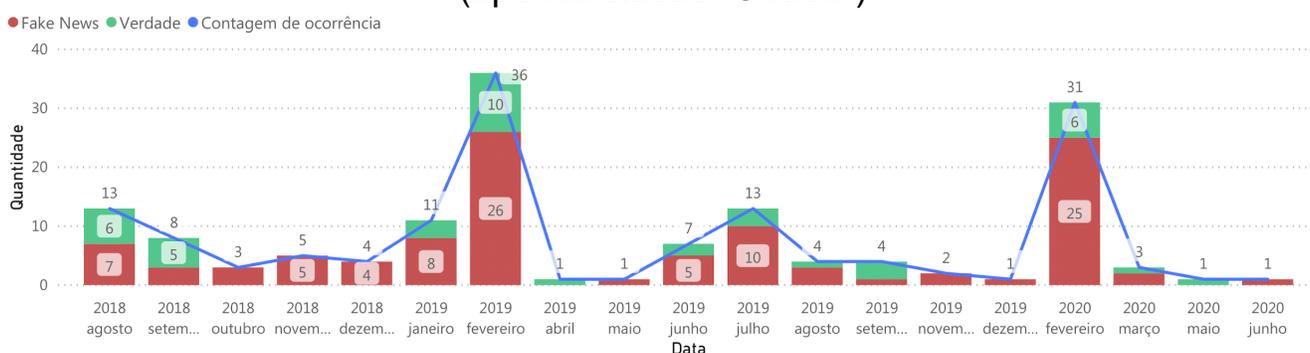


Fonte: dados de pesquisa.

As publicações que não estão relacionadas à Covid-19 representam um total de 149 registros, o que equivale a um percentual de 64,6% da amostra de dados (Figura 2). Destes registros, 107 (71,8%) foram classificados como *fake news*, enquanto 42 (28,2%) como informações verdadeiras. Por outro lado, para as

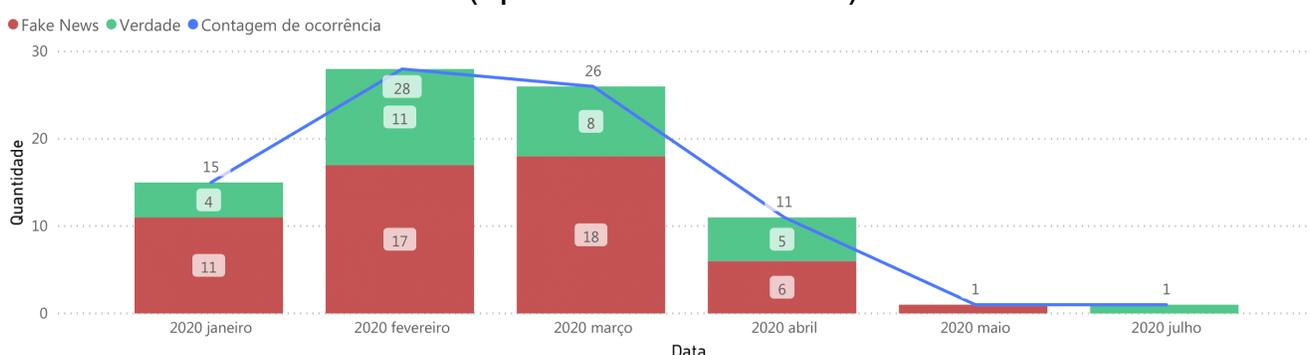
publicações que se referem à Covid-19, o número de registros aferidos foi de 82, totalizando 35,4% da base (Figura 3). Entretanto, é importante destacar que em termos proporcionais, levando-se em conta o período efetivo da pandemia da Covid-19, o número médio de publicações da classe “Outros” é de aproximadamente 7,8 registros por mês, enquanto para aquelas da classe “Covid-19”, esse valor é de aproximadamente 13,6%, o que vai ao encontro dos fatores previamente destacados da maior circulação de dados e informações na rede mundial de computadores. Outro fator de destaque está ligado ao fato de que para os registros relacionados à Covid-19, 53 (64,6%) foram classificados como *fake news*, enquanto 29 (35,4%) como verdade.

Figura 2 - Série temporal quantitativa categorizada por “*fake news*” e “verdade” (apenas classe “Outros”)



Fonte: dados de pesquisa.

Figura 3 - Série temporal quantitativa categorizada por “*fake news*” e “verdade” (apenas classe “Covid”)



Fonte: dados de pesquisa.

A partir da separação dos registros nas classes identificadas como “Covid-19” (*fake news* e verdade) e “Outros” (*fake news* e verdade), foram gerados grafos de similitude, através do uso do software de mineração de textos Iramuteq (Figura 4). É possível observar, a partir do grafo da Figura 4-a, a existência de

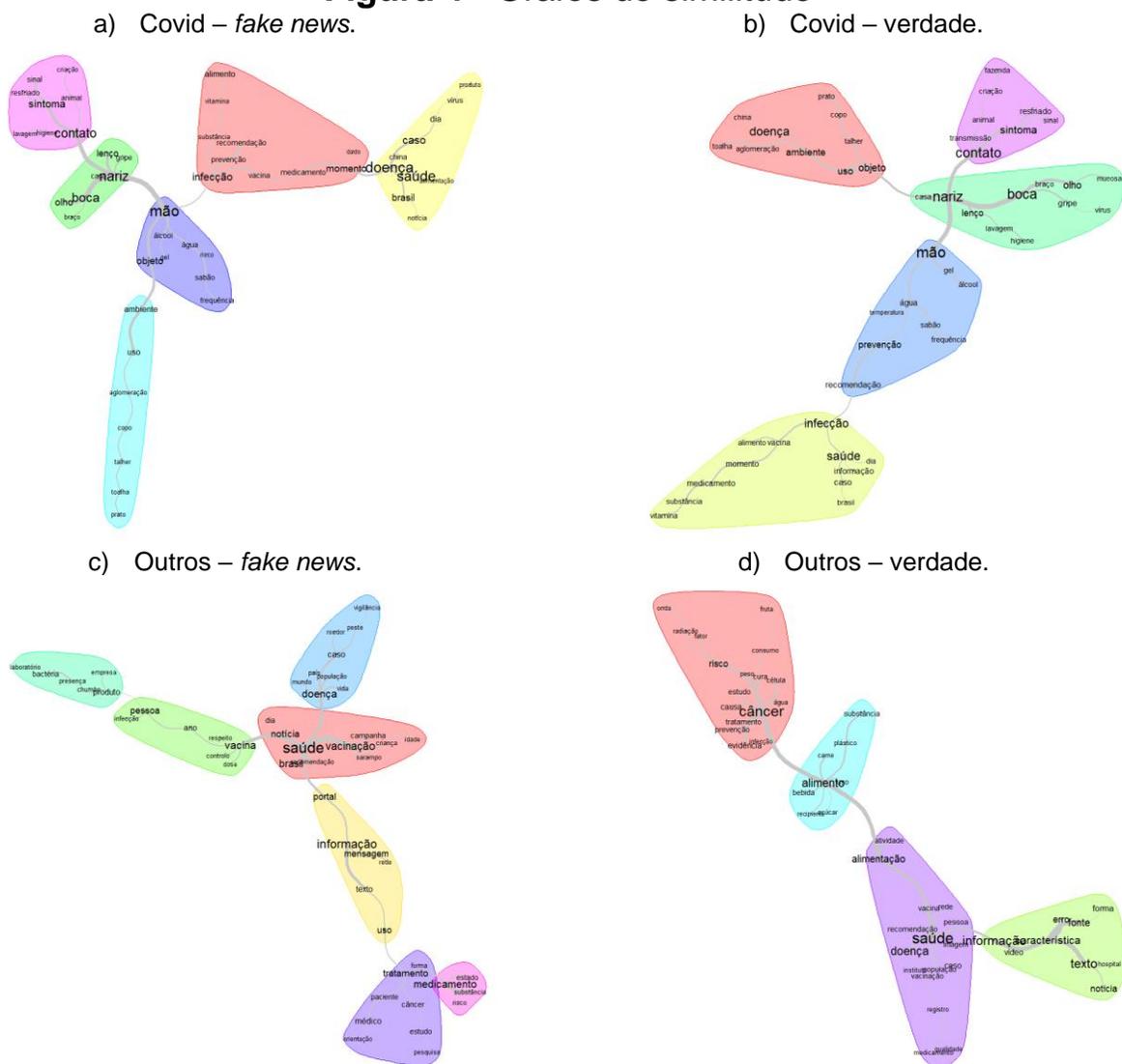
seis grupos contendo os principais termos relacionados às *fake news* sobre a Covid-19.

Destacam-se, relativos à Figura 4-a, os grupos verde e azul, onde se localizam as palavras mais usadas no âmbito pontuado, cujos vocábulos principais são “nariz” e “mão”, respectivamente. O grupo verde aponta para uma grande quantidade de *fake news* acerca das vias de contágio da Covid-19, que são as mucosas do nariz, da boca e até mesmo dos olhos, as quais absorvem as gotículas respiratórias expelidas por alguém infectado (ZHOU; ZHENHUA; CASTIGLIONE; SOIBERMAN; EBERHART; DUH, 2020). Já o grupo azul contém a temática da higiene para diminuir riscos de infecção, contendo os termos como “mão”, “álcool”, “gel”, “risco”, “água”, “sabão”, “frequência”, “objeto”, palavras que fazem parte do rol de recomendações da Organização Mundial de Saúde para a higienização das pessoas e dos objetos. O tipo de *fake news* contida nestes dois grupos pode incitar comportamentos de risco por parte de determinados indivíduos, quando estes, ao se depararem com a desinformação circulante, tomam a decisão de não adotarem as medidas de higiene e isolamento social necessárias para controle do contágio da Covid-19.

Analisando-se o grafo da Figura 4-b, observa-se que, dentre os cinco grupos de termos principais, os dois grupos em destaque coincidem com aqueles mencionados no grafo da Figura 4-a, quais sejam, os grupos verde e azul, com seus respectivos termos principais também coincidindo. É possível inferir, a partir disso, que da mesma forma que chegaram *fake news* para o Ministério da Saúde quanto às vias de contágio da Covid-19, têm circulado também informações confiáveis sobre o mesmo aspecto da doença.

Esses conteúdos apontados como verdadeiros pelo órgão mencionado demonstram que há uma difusão de informações baseadas em evidências científicas confiáveis e, a partir deste cenário, é possível inferir também que essa circulação se deve às iniciativas das agências de *fact-checking*. Teixeira e Martins (2020) demonstrou, em pesquisa recente, que a influência deste modelo de agência digital no combate às *fake news* foi decisiva desde o início da pandemia da Covid-19, onde se observa que o número de verificações feitas pelas empresas Lupa e Fato ou *Fake* foi três vezes maior a partir de março de 2020, sendo o pico de verificações em abril do mesmo ano.

Figura 4 - Grafos de similitude



Fonte: dados de pesquisa.

O grafo da Figura 4-c aponta para as *fake news* com outro foco, sendo possível inferir que o grupo base deste Grafo, de cor vermelha, tem como uma das palavras centrais a “vacinação”, evidenciando a campanha de combate ao sarampo, doença com surto no ano de 2019, bem como a campanha de vacinação promovida pelo Ministério da Saúde em 2020, mesmo após ser considerada uma doença erradicada em 2016 (ESCOLA NACIONAL DE SAÚDE PÚBLICA SERGIO AROUCA, 2016). Outro grupo importante deste grafo, o sinalizado em verde claro, aponta as palavras “presença”, “chumbo”, “bactéria” e “produto”, ligadas ao grupo anterior, verde escuro, em que aparece o termo “vacina” em evidência. Esses grupos evidenciam *fake news* que circulam, comumente, acerca dos possíveis malefícios das vacinas. A palavra “chumbo”,

nesse contexto, pode estar relacionada à propagação da ideia de que as vacinas são contaminadas com metais pesados, e que este fato poderia ocasionar o desenvolvimento de autismo em crianças. Entretanto, não foram demonstradas relações causais entre os metais contidos nas vacinas e o autismo em crianças, visto que essas substâncias são usadas em doses comprovadamente seguras, apenas como conservantes das vacinas (APS; PEREIRA; PEREIRA; CASTRO; SANTOS; FERREIRA, 2018). O grupo de cor amarela contém palavras como “Informação”, “Mensagem”, “rede”, as quais incitam uma possível disseminação dessas *fake news* por meio de redes de contatos, dando aos usuários que as recebem um certo grau de confiança ao notar a presença da palavra “informação”, como forma de comprovar uma mensagem fidedigna.

Finalmente, o grafo da Figura 4-d, onde os conteúdos sobre outros assuntos foram classificados como verdadeiros, evidenciam informações detectadas como verdadeiras ou informativas acerca do câncer (grupo vermelho), destacando também as palavras “consumo”, “risco” e “evidências”, elucidando os fatores que podem levar à referida doença. O grupo vermelho está diretamente ligado ao grupo azul, que contém termos comumente encontrados em notícias sobre o desenvolvimento do câncer, como fatores alimentares. O grupo roxo, que destaca os termos “saúde” e “doença”, volta a falar sobre vacinação, entretanto, a presença da palavra “recomendação” mostra que houve informações positivas sobre a temática em questão, provavelmente suscitadas pelas campanhas de vacinação ocorridas no período, que visam estimular o uso das vacinas como um ato de promoção da saúde.

5 Considerações finais

A desinformação na área da saúde não é uma novidade dos dias atuais. Entretanto, conforme pontuado na presente pesquisa, trata-se de um fenômeno que se ampliou devido à facilidade de interação entre os usuários das redes sociais virtuais, tudo isso fruto de um cenário propiciado pela Web 2.0, onde os indivíduos são consumidores, geradores e propagadores de todo tipo de (des)informação.

Nesse sentido, visando combater as *fake news* na área da saúde, visto que seu consumo apresenta um potencial letal para a vida humana, o Ministério da Saúde do Brasil criou, em 2018, o programa Saúde sem *Fake News*. Para empreender a análise proposta no artigo, utilizou-se os procedimentos metodológicos baseados no processo KDD, a partir do qual foram extraídas todas as publicações veiculadas no referido projeto, totalizando 231 registros compreendidos entre o dia 24 de agosto de 2018 e 17 de julho de 2020.

Os resultados evidenciaram que, dentro do número total de registros, 149 (64,6%) deles não estão relacionados à Covid-19. Por conseguinte, foram levantados 82 registros de publicações que dizem respeito à Covid-19, totalizando 35,4% da base de dados coletada. A partir disso, foram gerados quatro grafos de similitude que apontaram para os principais grupos de conteúdos identificados: o grafo contido na Figura 4-a, que abrange os principais termos ligados às *fake news* sobre a Covid-19; o grafo presente na Figura 4-b, que contém os termos de notícias verdadeiras acerca da Covid-19; o grafo evidenciado na Figura 4-c, abarcando os termos ligados às *fake news* sobre outros assuntos e, finalmente, o grafo da Figura 4-d, contendo as palavras ligadas às informações verdadeiras de outros assuntos.

Os dois primeiros grafos apontam para a temática da Covid-19, onde é possível notar a circulação de *fake news* relativas às vias de contágio da doença e sobre as medidas de higienização necessárias para frear a contaminação. Foi possível observar que, ao mesmo tempo em que as *fake news* se alastram, as ações informativas baseadas em evidências científicas também têm circulado, especialmente aquelas ligadas às iniciativas das agências de *fact-checking*, que têm atuado mais fortemente no combate à desinformação neste período de pandemia.

O terceiro grafo, por sua vez, mostra as *fake news* relacionadas principalmente aos malefícios da vacinação e, finalmente, o quarto grafo aponta para os conteúdos verídicos em torno de assuntos como câncer, alimentação e manutenção da saúde, incluindo ainda o tema da vacinação, provavelmente derivado das campanhas de incentivo ao uso das vacinas ocorridas no período pesquisado.

Conclui-se que o cenário evidenciado na presente pesquisa deixa patente a atualidade e a pertinência do assunto, o qual deve motivar investigações futuras no sentido de compreender o papel da informação baseada em evidências científicas no combate à desinformação circulante nas diversas ambiências sociais.

Referências

AGÊNCIA BRASIL. Em um ano, ministério respondeu a 11,5 mil dúvidas sobre saúde. **Agência Brasil**, Brasília, 27 ago. 2019. Disponível em: <https://agenciabrasil.ebc.com.br/saude/noticia/2019-08/ministerio-respondeu-por-aplicativo-12-mil-duvidas-sobre-saude>. Acesso em: 13 nov. 2020.

ALLCOTT, Hunt; GENTZKOW, Matthew. Social media and fake news in the 2016 election. **Journal of Economic Perspectives**, Nashville, v. 31, n. 2, p. 211–236, 2017. Disponível em: <https://www.aeaweb.org/articles?id=10.1257/jep.31.2.211>. Acesso em: 5 set. 2021.

AMARAL, Fernando. **Aprenda Mineração de Dados: teoria e prática**. Rio de Janeiro: Alta Books, 2016.

APS, Luana Raposo de Melo Moraes; PIANTOLA, Marco Aurélio Floriano; PEREIRA, Sara Araújo; CASTRO, Júlia Tavares de; SANTOS, Fernanda Ayane de Oliveira; FERREIRA, Luís Carlos de Souza. Eventos adversos de vacinas e as consequências da não vacinação: uma análise crítica. **Revista Saúde Pública**, São Paulo, v. 52, n. 40, p. 1-13, 2018. DOI: <https://doi.org/10.11606/S1518-8787.2018052000384>. Disponível em: <https://www.scielo.br/j/rsp/a/6T6JH8wZHMgqVsVkjZ85xLm/?lang=pt>. Acesso em: 5 set. 2021.

ARIÑO, Marino Latorre. **Historia de las Web, 1.0, 2.0, 3.0 y 4.0**. Santiago de Surco, marzo 2018. Disponível em: https://umch.edu.pe/arch/hnomarino/74_Historia%20de%20la%20Web.pdf. Acesso em: 5 set. 2021.

BAKIR, Vian; McSTAY, Andrew. Fake news and the economy of emotions: Problems, causes, solutions. **Digital Journalism**, [S.l.], v. 6, n. 2, p. 154–175, 2017. DOI: <https://doi.org/10.1080/21670811.2017.1345645>. Disponível em: <https://www.tandfonline.com/doi/abs/10.1080/21670811.2017.1345645>. Acesso em: 5 set. 2021.

CARRERA, Pilar. Estratagemas de la posverdad. **Revista Latina de Comunicación Social**, Islas Canarias, v. 73, p. 1469-1482, 2018. Disponível em: <http://www.revistalatinacs.org/073paper/1317/76es.html>. Acesso em: 05 set. 2021.

CONFORTO, Débora; CAVEDINI, Patrícia; MIRANDA, Roxane; CAETANO, Saulo. Pensamento computacional na educação básica: interface tecnológica na construção de competências do século XXI. **Revista Brasileira de Ensino de Ciências e Matemática**, Passo Fundo, v. 1, n. 1, p. 99-112, 2018. DOI: <https://doi.org/10.5335/rbecm.v1i1.8481>. Disponível em: <http://seer.upf.br/index.php/rbecm/article/view/8481>. Acesso em: 5 set. 2021.

D'ANCONA, Matthew. **Pós-verdade: a nova guerra contra os fatos em tempos de fake news**. Barueri: Faro Editorial, 2018.

ESCOLA NACIONAL DE SAÚDE PÚBLICA SERGIO AROUCA (ENSP). **Brasil recebe certificado de eliminação do sarampo**. Rio de Janeiro, 30 set. 2016. Disponível em: <http://informe.ensp.fiocruz.br/noticias/40569>. Acesso em: 14 nov. 2020.

FAYYAD, Usama; PIATETSKY-SHAPIRO, Gregory; SMYTH, Padhraic. From data mining to knowledge discovery in databases. **AI Magazine**, Palo Alto, v. 17, n. 3, p. 37-54, 1996. DOI: <https://doi.org/10.1609/aimag.v17i3.1230>. Disponível em: <https://ojs.aaai.org/index.php/aimagazine/article/view/1230>. Acesso em: 5 set. 2021.

KAKUTANI, Michiko. **A morte da verdade: notas sobre a mentira na era Trump**. Rio de Janeiro: Intrínseca, 2018.

LÉVY, Pierre. **As tecnologias da inteligência: o futuro do pensamento na era da informática**. São Paulo: Editora 34, 1993.

LIMA, Clóvis Ricardo Montenegro de; SÁNCHEZ-TARRAGÓ, Nancy; MORAES, Danielle; GRINGS, Luciana; MAIA, Mariângela Rabelo. Emergência de saúde pública global por pandemia de Covid-19: desinformação, assimetria de informações e validação discursiva.

Folha de Rosto, Juazeiro do Norte, v. 6, n. 2, p. 5-21, 2020. DOI:

<https://doi.org/10.46902/2020n2p5-21>. Disponível em:

<https://periodicos.ufca.edu.br/ojs/index.php/folhaderosto/article/view/490>. Acesso em: 5 set. 2021.

MARTINO, Luís Mauro Sá. **Teoria das mídias digitais**: linguagens, ambientes, redes. Rio de Janeiro: Vozes, 2014.

MITCHELL, Ryan. **Web Scraping com Python**. São Paulo: Novatec, 2015.

O'REILLY, Tim. **What Is Web 2.0** - Design Patterns and Business Models for the Next Generation of Software. [S.l.]: O'Reilly Publishing, 2005.

PROVOST, Foster; FAWCETT, Tom. **Data Science para Negócios**. Rio de Janeiro: Alta Books, 2016.

RAMONET, Ignácio. A explosão do jornalismo na era digital. *In*: MORAES, Denis; RAMONET, Ignácio; SERRANO, Pascual. (org.). **Mídia, poder e contrapoder**: da concentração monopólica à democratização da informação. São Paulo: Boitempo, 2013, p. 85-102.

RISTOSKI, Petar; PAULHEIM, Heiko. Semantic web in data mining and knowledge discovery: a comprehensive survey. **Journal of Web Semantics**, Alemanha, v. 36, p. 1-22, Jan. 2016.

DOI: <https://doi.org/10.1016/j.websem.2016.01.001>. Disponível em:

<https://www.sciencedirect.com/science/article/abs/pii/S1570826816000020>. Acesso em: 05 set. 2021.

SERRANO, Pascual. Outro jornalismo possível na Internet. *In*: MORAES, Denis; RAMONET, Ignácio; SERRANO, Pascual. (org.). **Mídia, poder e contrapoder**: da concentração monopólica à democratização da informação. São Paulo: Boitempo, 2013. p. 145-182.

TEIXEIRA, Juliana Fernandes; MARTINS, Allysson Viana. *Fact-checking* no combate às *fake news* sobre a COVID-19: um estudo exploratório das agências digitais de checagem de fatos contra a desinformação da pandemia. **Comunicação e Inovação**, São Caetano do Sul, v. 21 n. 47 p. 63-81, 2020. Disponível em:

https://seer.uscs.edu.br/index.php/revista_comunicacao_inovacao/article/view/7178. Acesso em: 5 set. 2021.

VERGÈS, Pierre; BOURICHE, Boumedienne. L'analyse des données par les graphes des similitude. **Sciences Humaines**, França, p. 1-90, juin 2001. Disponível em:

<https://www.scienceshumaines.com/textesInedits/Bouriche.pdf>. Acesso em: 5 set. 2021.

VIEIRA, Larissa Machado; SILVA, Núbia Rosa da; CORDEIRO, Douglas Farias. Análise descritiva das *fake news* da saúde através de mineração de textos no Portal da Saúde. *In*: CONGRESSO DE CIÊNCIAS DA COMUNICAÇÃO NA REGIÃO CENTRO OESTE, 21., Goiânia, 2019. **Anais [...]**. Goiânia: Intercom, 2019. Disponível em:

<https://portalintercom.org.br/anais/centrooeste2019/resumos/R66-0230-1.pdf>. Acesso em: 5 set. 2021.

VOSOUGHI, Soroush; ROY, Deb; ARAL; Sinan. The spread of true and false news online. **Science**, Estados Unidos da América, v. 359, n. 6380, p. 1146–1151, Mar. 2018. DOI: 10.1126/science.aap9559. Disponível em:

<https://www.science.org/doi/abs/10.1126/science.aap9559>. Acesso em: 5 set. 2021.

WARDLE, Claire. **Information disorder**: toward an interdisciplinary framework for research and policy making. Strasbourg: Council of Europe, 2017. Disponível em:

<https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-research/168076277c>. Acesso em: 5 set. 2021.

ZHOU, Lingli; ZHENHUA, Xu; CASTIGLIONE, Gianni M.; SOIBERMAN, Uri S.; EBERHART, Charles G.; DUH, Elia J. ACE2 and TMPRSS2 are expressed on the human ocular surface, suggesting susceptibility to SARS-CoV-2 infection. **BioRxiv**, 2020. DOI:

<https://doi.org/10.1101/2020.05.09.086165>. Disponível em:

<https://www.biorxiv.org/content/10.1101/2020.05.09.086165v1>. Acesso em: 13 nov. 2020.

Sobre a autoria

Jully Porto Lopes Melo

Graduanda em Gestão da Informação, pela Universidade Federal de Goiás (UFG). Bolsista de Iniciação Científica do CNPq.

jullyporto15@gmail.com

Anelise Souza Rocha

Mestranda em Comunicação pela Universidade Federal de Goiás (UFG). Graduada em Gestão da Informação (UFG). Pesquisadora do Grupo de Pesquisa Tecnologia e Computação Aplicada à Informação e Comunicação da FIC/UFG.

anelisesrocha@gmail.com

Larissa Machado Vieira

Doutoranda em Comunicação pela Universidade Federal de Goiás (UFG). Mestre em Comunicação e Graduada em Jornalismo, pela UFG.

vieira.mlarissa@gmail.com

Douglas Farias Cordeiro

Doutor em Ciência da Computação e Matemática Computacional, pela Universidade de São Paulo (USP). Professor da Faculdade de Informação e Comunicação (FIC) da Universidade Federal de Goiás (UFG). Docente do Programa de Pós-Graduação em Comunicação da UFG e do Programa de Pós-Graduação em Engenharia de Produção, da Universidade Federal de Catalão (UFCAT).

cordeiro@ufg.br

Artigo submetido em: 18 fev. 2021.
Aceito em: 16 jun. 2021.



Esta obra está licenciada com uma Licença [Creative Commons Atribuição-NãoComercial-Compartilhalgal 4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/).

UFCA UNIVERSIDADE
FEDERAL DO CARIRI

Centro de Ciências Sociais Aplicadas
Mestrado Profissional em Biblioteconomia
Revista Folha de Rosto



✉ folhaderosto@ufca.edu.br

📷 [@revistafolhaderosto](https://www.instagram.com/revistafolhaderosto)

🐦 [@revfolhaderosto](https://twitter.com/revfolhaderosto)

Este periódico é uma publicação do [Programa de Pós-Graduação em Biblioteconomia da Universidade Federal do Cariri](#) em formato digital e periodicidade quadrimestral.