

Caracterização do debate no Twitter sobre a vacinação contra a COVID-19 no Brasil

Larissa Malagoli¹, Júlia Stancioli¹, Carlos H. G. Ferreira^{1,2}, Marisa Vasconcelos³, Ana Paula Couto da Silva¹, Jussara Almeida¹

¹Universidade Federal de Minas Gerais,

²Universidade Federal de Ouro Preto, ³IBM Research

{larissagomes, juliastancioli, chgferreira}@dcc.ufmg.br,

marisaav@br.ibm.com, {ana.coutosilva, jussara}@dcc.ufmg.br

Abstract. *Twitter is one of the most popular social media applications used by the general public to debate a range of topics, being the COVID-19 pandemic topic one of the most discussed in the last months. Since December 2020, the main focus of the debate has become vaccination, coinciding with the beginning of worldwide vaccination campaigns. We investigate the public perception of this topic, analyzing over 9 million Portuguese-language tweets during two months corresponding to the early stages of vaccination in the world, including in Brazil. Our results provide an initial understanding of the dynamics of the online debate around the vaccination against COVID-19, highlighting how people use the online world to share their impressions and concerns about it.*

Resumo. *O Twitter é uma das redes sociais online mais utilizadas pelo público em geral para o debate de diferentes assuntos, sendo a pandemia da COVID-19 um dos temas mais debatidos nos últimos meses. A partir de dezembro de 2020, o foco do debate passou a ser a vacinação. Neste artigo, investigamos a percepção do público sobre o tema, analisando mais de 9 milhões de tweets em português, em um período de dois meses correspondentes aos estágios iniciais da vacinação no Brasil e no mundo. Nossos resultados fornecem um entendimento inicial sobre a dinâmica do debate online sobre a vacinação contra a COVID-19, evidenciando como as pessoas usam o mundo online para compartilhar suas impressões e preocupações sobre o assunto.*

1. Introdução

Declarada uma pandemia em março de 2020 pela Organização Mundial da Saúde, a COVID-19 ainda causa grandes impactos no mundo. Profissionais de saúde e pesquisadores têm trabalhado 24 horas por dia para conter a propagação do vírus e desenvolver soluções eficazes para combatê-la. No entanto, o pouco conhecimento sobre o novo vírus, os desafios de uma nova vida sob restrições e incertezas e a necessidade de distanciamento social no mundo físico promoveram um aumento acentuado na busca por informações online, notadamente em aplicativos de mídia social [Tsao et al. 2021], como o Twitter.

Após quase um ano de pandemia severa, o debate online passou a ter maior foco na vacinação contra a COVID-19 no final de 2020. O mundo finalmente viu o surgimento de uma série de vacinas para conter a disseminação do vírus, com o início da campanha

de vacinação em vários lugares do mundo, inclusive, mesmo que tardiamente, no Brasil. É senso comum que o sucesso dessa campanha depende principalmente do engajamento massivo da população, uma vez que as vacinas são, até o momento, a solução mais eficiente para conter a disseminação do vírus. No entanto, o surgimento relativamente rápido de novas vacinas e a falta de conhecimento fundamental sobre elas levantam várias preocupações e dúvidas entre as pessoas em geral. Essas preocupações impulsionaram ainda mais os debates no ambiente online e no mundo real.

Neste artigo, investigamos a dinâmica do debate em torno da vacinação contra a COVID-19 no Twitter, com enfoque no contexto brasileiro. Para tal, analisamos mais de 9 milhões de tweets em português que foram coletados entre os meses de dezembro de 2020 e janeiro de 2021, que correspondem aos estágios iniciais de vacinação em vários países, inclusive no Brasil. Nossa investigação explora duas dimensões principais relacionadas à disseminação de informação: (i) engajamento dos usuários e (ii) propriedades dos conteúdos, com foco particular na análise de como as características destas dimensões evoluem ao longo do tempo à medida que eventos externos do mundo real influenciam as discussões no mundo online. Nossos resultados mostram que o volume de tweets está altamente correlacionado com os eventos externos. A partir da análise dos sentimentos expressos pelos tweets, observamos que menções às palavras *vacina* e *vacinação* tendem a ser mais negativas, provavelmente devido à ansiedade e incerteza sobre a aquisição de vacinas e da organização da vacinação em massa no Brasil. Adicionalmente, observamos que palavras relacionadas à *família*, *aprovação*, *ansiedade*, *coletividade* são mais utilizadas nos tweets que incluem as palavras-chave *vacina*, *vacinação*. Textos que possuem as palavras-chave referentes a vacinas específicas (p.ex., *Astrazeneca*, *Coronavac*) abordam menos aspectos relacionados à *morte*, *emoção negativa*, e *negações*.

Este artigo está organizado da seguinte forma. A Seção 2 apresenta os trabalhos relacionados. A Seção 3 descreve nosso conjunto de dados enquanto a Seção 4 apresenta os principais resultados. A Seção 5 conclui o artigo e oferece possíveis direções para trabalhos futuros.

2. Trabalhos Relacionados

Vários estudos na literatura analisaram discussões online sobre vacinação em geral e a pandemia COVID-19. Na sua grande maioria, os dados utilizados para as análises são em língua inglesa. Assim, devido à limitação de espaço, discutimos brevemente a seguir outros trabalhos que consideraram o contexto brasileiro em suas análises. Importante ressaltar que, até o momento da submissão deste artigo, não foram encontradas análises do debate sobre a vacinação contra a COVID-19 em redes sociais online no contexto do Brasil, foco dos resultados apresentados neste artigo.

Discussões sobre vacinação em geral. Os autores em [Araujo et al. 2020] apresentaram uma análise exploratória de como as informações de saúde foram compartilhadas e discutidas no Twitter em termos de tópicos de conscientização e posicionamentos durante surto da Febre Amarela de 2017. O estudo demonstrou que as redes sociais podem auxiliar no monitoramento de alguns aspectos da saúde pública e oferecer dados adicionais aos gestores de saúde de como as pessoas interagem durante um surto. Análises de enquadramento das matérias sobre vacina e vacinação vinculadas em dois perfis de grandes jornais (um americano e um brasileiro) apresentadas em [Santos 2020] mostraram uma

maior presença de uma retórica contra a vacina no público norte-americano em relação ao brasileiro. Os autores em [Pereira da Veiga et al. 2020] avaliaram as postagens publicadas pelo Ministério da Saúde no Facebook, além dos comentários, compartilhamentos e reações em relação à campanha de vacinação contra o HPV entre janeiro e julho de 2018.

Discussões sobre a pandemia da COVID-19. Em [Massarani et al. 2020] os autores analisaram os cem conteúdos de maior engajamento em redes sociais sobre COVID-19 no Brasil (março de 2020). Os resultados mostraram que a desinformação respondeu por 13,5% da amostra e que seu engajamento médio, estimado a partir de métricas como compartilhamentos, comentários e reações, foi maior do que o dos conteúdos que continham informações passíveis de serem checadas em outras fontes e baseadas em evidências científicas. A análise comparativa apresentada em [Garcia and Berton 2021] entre o conteúdo de tweets em inglês e português que abordavam respostas à pandemia COVID-19 no Brasil e nos EUA mostrou que emoções negativas foram dominantes em quase todos os tópicos identificados nos textos compartilhados. O trabalho apresentado em [Brum et al. 2020] caracterizou um conjunto de mais de 56 milhões de tweets escritos em português e coletados ao longo de 70 dias. As análises mostraram que os temas mais discutidos no período foram quarentena, hidroxicloroquina, aglomeração e distância social, sendo as discussões centradas em questões políticas.

3. Base de Dados

Com o objetivo de analisar o debate online sobre a vacinação contra a COVID-19 no Brasil, foi inicialmente definida a seguinte lista de palavras-chave relacionadas ao tema, considerando as principais vacinas desenvolvidas até o momento, bem como termos relacionados ao contexto brasileiro (p.ex., SUS, vacina): *vacinação, vacina, vachina, sputnikv, pfizer, novavax, moderna, coronovac, covaxin, biontech, astrazeneca, bnt162b2¹, SUS*. Foram então coletados tweets em português utilizando a Twitter API Search², restringindo a coleta somente a tweets que mencionam pelo menos uma das palavras-chave nesta lista.

No total, coletamos quase 9 milhões de tweets, durante 9 semanas, no período entre 01 de dezembro de 2020 e 31 de janeiro de 2021.³ O período de coleta inclui vários eventos relevantes relacionados à vacinação contra a COVID-19 no Brasil e no mundo, tais como o início da campanha mundial de vacinação (8 de dezembro de 2020, no Reino Unido)⁴ e a aprovação pela Anvisa⁵ das primeiras vacinas no Brasil (17 de janeiro).⁶

A Tabela 1 mostra uma visão geral da base de dados, apresentando os totais de tweets, retweets e usuários únicos por semana. Nas nossas análises, o conjunto de tweets inclui os tweets e os *quoted tweets*, que são respostas a um tweet postado por outro usuário. De maneira geral, podemos observar que o engajamento dos usuários, estimado pela frequência de postagens (tweets e retweets), na discussão dos tópicos relacionados à vacinação possui uma tendência de crescimento, assim como o total de números de

¹Nome oficial da vacina da Pfizer/Biontech.

²<https://developer.twitter.com/en/docs/twitter-api/v1/tweets/search/api-reference/get-search-tweets>

³https://zenodo.org/record/4818622#.YK_4MJNKhBx

⁴<https://g1.globo.com/bemestar/vacina/noticia/2020/12/08/idoso-de-90-anos-e-a-primeira-a-ser-vacinado-contra-covid-no-reino-unido.ghtml>

⁵<https://www.gov.br/anvisa/pt-br>

⁶<https://www.cnnbrasil.com.br/saude/2021/01/17/votos-anvisa-vacina-coronovac-oxford>.

Tabela 1. Principais estatísticas da base de dados.

Semana	Início	#Tweets	#Retweets	#Usuários Únicos
1	01-12-2020	212.832	453.953	299.583
2	08-12-2020	310.110	746.943	334.204
3	15-12-2020	277.882	803.146	368.051
4	22-12-2020	247.746	680.341	330.496
5	29-12-2020	257.931	612.353	365.331
6	05-01-2021	288.885	928.922	495.354
7	12-01-2021	346.882	1.321.950	530.022
8	19-01-2021	296.886	1.010.262	435.842
9	26-01-2021	195.925	493.325	275.491

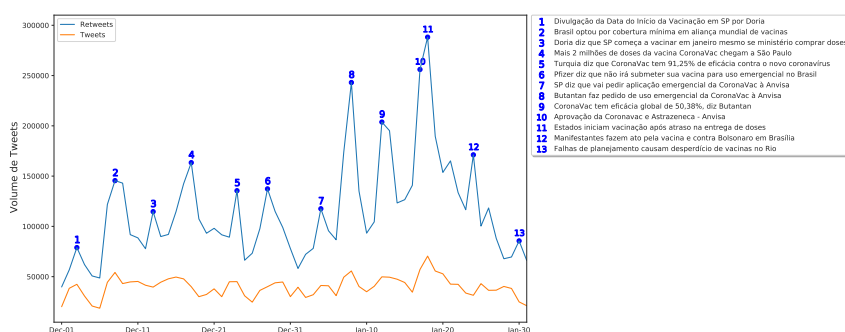


Figura 1. Volume de tweets durante o período coletado.

usuários únicos participantes dessa discussão. Esta tendência geral reflete a importância que o tópico ganha à medida que eventos reais importantes relacionados ao tema ocorrem.

Uma análise mais detalhada da evolução temporal das discussões sobre a vacinação é apresentada na Figura 1, através da série temporal dos números diários de tweets e retweets. A figura também destaca as datas com alguns eventos reais de relevância, tais como divulgação do calendário de vacinação e da eficácia e aprovação do uso de algumas vacinas. Podemos notar picos significativos que coincidem com alguns eventos identificados. Por exemplo, no dia da aprovação pela Anvisa das vacinas Coronavac e Astrazeneca/Oxford (evento 10) o total de tweets e o total de retweets cresceram 63% e 81%, respectivamente. Assim, eventos externos, tais como aprovações de novas vacinas, notícias sobre reações adversas ao uso de uma vacina específica, bem como as constantes mudanças do cronograma da campanha de vacinação no Brasil podem ter impulsionado o debate sobre as vacinas da COVID-19 no Twitter no período selecionado.

A seguir, verificamos o conteúdo dos tweets e retweets em três períodos de tempo cobertos pela coleta realizada. A Figura 2 mostra a nuvem de palavras com os 100 termos mais frequentes (em número de tweets e retweets) durante a primeira, quinta e nona semanas. A palavra-chave *vacina* é a mais frequente nos tweets e retweets em todas as semanas, seguida pela palavra-chave *vacinação*. Menções às palavras-chave relacionadas às vacinas (*coronavac*, *astrazeneca*) aumentam ao longo das semanas. Interessante notar que a discussão em torno do tema da vacinação é permeada fortemente pelo aspecto político, com a presença do nome do governador de São Paulo, João Dória, que liderou as negociações com o Instituto Butantan para a fabricação da Coronavac, além da menção constante ao presidente Jair Bolsonaro. Infelizmente, a presença da palavra *vachina*, termo pejorativo para a Coronavac, não é desprezível na base de dados coletada.

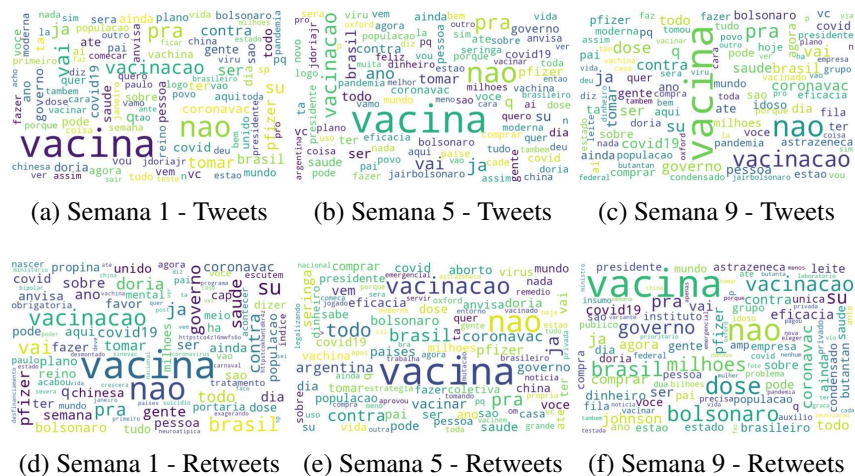


Figura 2. Nuvem de palavras mais populares (top-100) dos tweets e retweets.

Para finalizar esta visão geral dos dados, a Figura 3 mostra a popularidade das palavras-chave usadas para a coleta ao longo do tempo, sendo a popularidade de uma palavra-chave em uma dada semana proporcional ao número total de tweets (ou retweets) que a mencionam compartilhados naquele período. Essa análise é apresentada como mapas de calor, onde a cor de cada célula reflete o total de tweets (retweets) que mencionam cada palavra-chave (eixo-y) em uma dada semana (eixo-x) em escala logarítmica.

Considerando a popularidade estimada em número de tweets, a Figura 3a mostra que as palavras-chave *vacina* e *vacinação* são as mais populares em todas as semanas, sugerindo que discussões mais gerais dominam o debate. Como esperado, a vacina *Coronavac* é a mais mencionada no debate, uma vez que é a vacina de maior disponibilidade de doses no Brasil no período. Menções à *Pfizer* também são frequentes, possivelmente por ter sido a primeira vacina utilizada no mundo, além de ser uma vacina que teve a compra recusada pelo governo.⁷ Menções à vacina *Astrazeneca* são menos frequentes, apesar da mesma estar também disponível para a aplicação no Brasil. O termo pejorativo *vachina* é mencionado recorrentemente nos dados coletados, com um aumento de popularidade na semana 7, período que coincide com a aprovação da *Coronavac* pela Anvisa. Conclusões similares podem ser observadas se considerarmos a popularidade estimada pelo número de retweets, conforme mostrado na Figura 3b.

4. Resultados e Análises

Nesta seção apresentamos e discutimos o conjunto de análises realizadas que visam entender a dinâmica do debate sobre vacinação contra a COVID-19 no Twitter. Para cada análise, primeiramente descrevemos a metodologia utilizada, seguida das principais conclusões alcançadas.

4.1. Perfil dos Usuários

Essa análise tem como objetivo apresentar o perfil dos usuários engajados no debate sobre a vacinação no Twitter. Como em estudos anteriores [Chen et al. 2020,

⁷<https://ultimosegundo.ig.com.br/politica/2021-03-07/governo-negou-3-ofertas-da-pfizer-por-vacina-em-2020-perdendo-milhoes-de-doses.html>

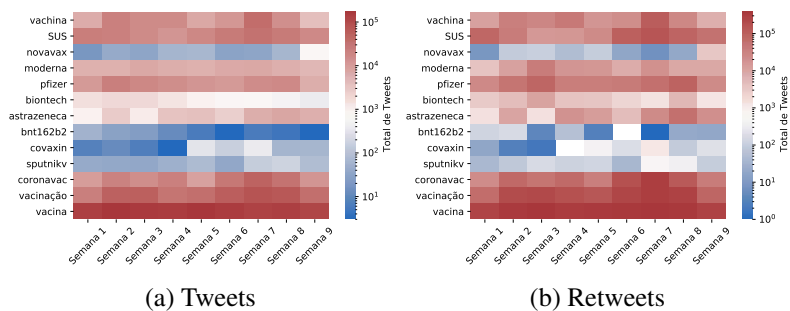


Figura 3. Popularidade das palavras-chave ao longo das semanas.

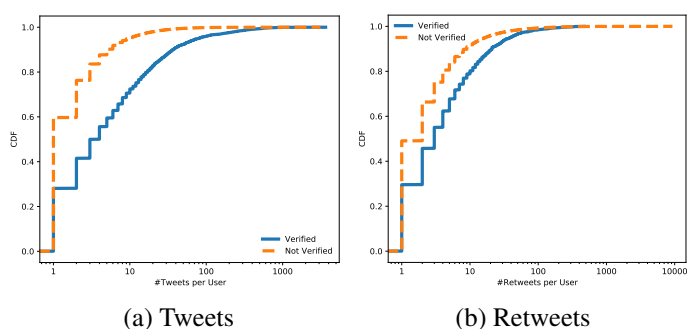


Figura 4. Distribuições dos números de tweets e retweets por tipo de conta de usuário.

Kouzy et al. 2020], separamos os usuários entre aqueles com contas verificadas e aqueles com contas não verificadas. Usuários verificados pelo Twitter são aqueles que despertam maior interesse público e tendem a ser mais ativos nas discussões relacionadas à ocorrência de eventos relevantes para a sociedade [Chen et al. 2020].

As Figuras 4a e 4b apresentam as distribuições de probabilidades acumuladas dos números de tweets e retweets, respectivamente, realizados pelas contas verificadas e não verificadas. Contas verificadas tendem a postar mais tweets e retweets: 80% dos usuários com contas verificadas postam até 37 tweets e 21 retweets, enquanto a mesma porcentagem de usuários com contas não verificadas postam até 5 tweets e 9 retweets. O usuário mais ativo com conta verificada postou 3.654 tweets, enquanto o usuário mais ativo com conta não verificada postou 3.125 tweets. Considerando os retweets, observa-se que as contas não verificadas mais ativas tendem a propagar mais informação, com o máximo de 9.274 retweets para um único usuário, contra somente 471 do usuário mais ativo com conta verificada.

Com o objetivo de explorar as palavras-chave que os usuários com contas (verificadas e não verificadas) mais ativas postam ou repassam, as Tabelas 2 e 3 mostram, para os top-5 usuários mais ativos em cada tipo de conta, os números de seguidores e de ocorrências de cada palavra-chave em seus textos (tweets e retweets). Os usuários são identificados por suas posições no ranking. Apesar destas contas mencionarem mais frequentemente palavras mais gerais (*vacina*, *vacinação*), encontramos também menções a vacinas mais específicas, como, por exemplo, *novavax*. A conta verificada mais ativa, em número de tweets, é a do Instituto Butantan (@butantanoficial). Foi observado que

Tabela 2. Top-5 usuários mais ativos (em # de tweets e retweets): contas verificadas.

Tweets		
Usuário	# Seguidores	Palavras-chave (# ocorrências)
1	140.903	vacina(2.679), coronavac(585), vacinação(310), vachina(55), SUS(25)
2	1.054.898	vacina(521), vacinação(473), coronavac(263), pfizer(212), astrazeneca(73), biontech(54), moderna(48), SUS(10), covaxin(6), novavax(6)
3	32.530	vacina(342), vacinação(204), pfizer(203), coronavac(97), moderna(81), astrazeneca(78), novavax(27), biontech(18), bnt162b2(16), SUS(15), covaxin(2)
4	3.655.485	vacina(333), vacinação(315), coronavac(209), pfizer(71), astrazeneca(32), SUS(17), biontech(16), moderna(16), covaxin(2)
5	1.856.594	vacina(310), vacinação(259), coronavac(184), pfizer(65), astrazeneca(31), SUS(17), biontech(16), moderna(16), covaxin(1), novavax(1)
Retweets		
Usuário	# Seguidores	Palavras-chave (# ocorrências)
1	1.225	vacina(181), vacinação(124), pfizer(59), coronavac(45), SUS(29), astrazeneca(11), moderna(9), vachina(7), biontech(5), novavax(1)
2	50.223	vacina(192), vacinação(147), coronavac (54), pfizer(31), SUS(20), moderna(9), astrazeneca(7), biontech(3)
3	612.725	vacina(131), vacinação(100), coronavac(54), pfizer(30), SUS(13), astrazeneca(10), biontech(4), moderna(3), vachina(1)
4	177.601	vacinação(134), vacina(119), SUS(29), pfizer(26), coronavac(17), astrazeneca(12), biontech(2), moderna(2), vachina(1), novavax(1)
5	10.592	vacinação(134), vacina(123), SUS(44), coronavac(21), pfizer(6), astrazeneca(3), moderna(1)

a palavra-chave *vachina* estava presente em tweets do Instituto para esclarecimento de desinformação e de repúdio a ataques à Coronavac, p.ex.: *Os estudos da fase 3 concluem que a eficácia é superior ao exigido pela OMS e sem efeito adverso grave. É, portanto, segura e eficaz. Só lembrando que não é "vachina", é a vacina do Brasil, ok?* O segundo usuário verificado mais ativo na produção de informação é uma conta de um canal de notícias, seguido por uma biomédica, um portal de notícias e um provedor de conteúdo.

4.2. Análise dos Emojis

Em redes sociais, emojis são utilizados para complementar a comunicação realizada por texto. Esses símbolos são representações visuais de uma emoção, ideia ou simbolismo. Nos nossos dados encontramos uma pequena fração de tweets (10,5%) e retweets (9,9%) com pelo menos um emoji, que foram identificados através da utilização do pacote *emoji*.⁸ A Figura 5 apresenta os emojis mais utilizados pelos usuários que estão debatendo sobre o tema vacinação no Twitter. De uma maneira geral, o emoji mais popular (em termos de tweets e retweets) é o emoji *seringa*, utilizado para representar as vacinas (vacinação) contra a COVID-19. Outros emojis populares refletem estados emocionais como gratidão, oração, súplica (*mãos juntas*), diversão (*carinha chorando de rir*), cognitivos tais como questionamentos (*carinha pensativa*) e para salientar a importância de um conteúdo (*mão com dedo indicador para baixo*). Além desses emojis, é interessante notar a alta frequência do emoji *cara de palhaço*, sugerindo que o usuário se sente bobo ou enganado por causa do conteúdo postado.

Analisando os dados coletados, o emoji *seringa* é usado tanto para expressar a felicidade e esperança das pessoas com o desenvolvimento e descoberta das vacinas quanto para difundir desinformação e notícias falsas. O emoji *mãos juntas* expressa gratidão, seja pelo início da vacinação no mundo ou súplica para que a vacinação comece rapidamente no Brasil. A Tabela 4 apresenta alguns exemplos do uso dos diferentes emojis.

⁸<https://pypi.org/project/emoji/>

Tabela 3. Top-5 usuários mais ativos (em # de tweets e retweets): contas não verificadas.

Tweets		
Usuário	# Seguidores	Palavras-chave(# ocorrências)
1	1.219	vacina(1.619), pfizer(782), vacinação(707), biontech(17)
2	1.651	vacinação(1.091), vacina(367), SUS(253), astrazeneca(6), coronavac(2)
3	815	vacinação(615), vacina(421), pfizer(192), coronavac(105), astrazeneca(98), biontech(75), SUS(47), moderna(20), covaxin(6), sputnikv(5), anti-vacina(3), bnt162b2(1)
4	1.857	vacinação(672), vacina(527), coronavac(204), astrazeneca(74), pfizer(59)
5	25.669	vacina(506), vacinação(428), coronavac(322), pfizer(111), SUS(44), astrazeneca(36), moderna(8), biontech(3), novavax(2), covaxin(1)
Retweets		
Usuário	# Seguidores	Palavras-chave(# ocorrências)
1	2.630	vacinan(3.127), vacinação(2.011), coronavac(1.989), pfizer(916), astrazeneca(320), biontech(319), SUS(284), moderna(178), vachina(59), sputnikv(38), covaxin(31), novavax(1), bnt162b2(1)
2	3.971	vacina(1.253), vacinação(1.070), coronavac(525), pfizer(400), astrazeneca(100), SUS(69), vachina(57), moderna(50), biontech(48), covaxin(5), novavax(4),
3	963	vacina(1.366), vacinação(1.068), coronavac(359), SUS(252), pfizer(212), astrazeneca(100), biontech(35), moderna(34), vachina(18), novavax(5), sputnikv(3), covaxin(2)
4	474	vacina(1.197), vacinação(704), coronavac(539), pfizer(255), astrazeneca(103), SUS(69), vachina(41), moderna(39), biontech (31), novavax(6)
5	6.446	vacina(1.009), vacinação(895), SUS(416), coronavac(229), pfizer(135), astrazeneca(55), vachina(27), moderna(26), biontech(12), sputnikv(1)

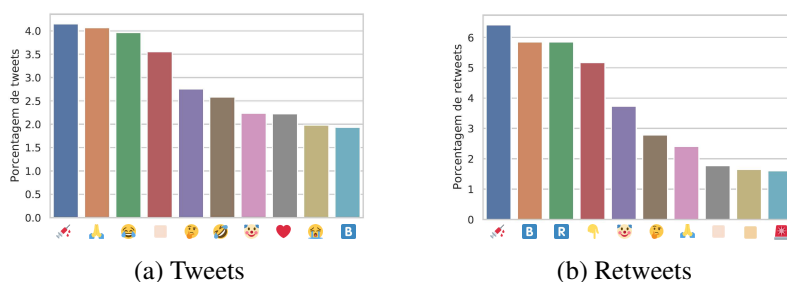


Figura 5. Top-10 emojis mais frequentes em tweets e retweets.

4.3. Análise de Sentimentos

Nesta seção analisamos o sentimento expresso nos tweets e retweets quando são consideradas as diferentes palavras-chave. Para realizar a análise, utilizamos a ferramenta *SentiStrength*⁹ para extrair o sentimento de cada tweet. O SentiStrength fornece uma pontuação inteira variando de -4 (fortemente negativo) a $+4$ (fortemente positivo). A pontuação 0 indica um sentimento neutro. Consideramos aqui como *negativo*, *neutro* e *positivo* tweets e retweets com pontuações menores que 0, iguais a 0 e maiores que 0, respectivamente. O SentiStrength tem sido amplamente aplicado para analisar o sentimento de conteúdo textual postado em redes sociais [Thelwall 2017].

Para entender como o sentimento expresso pelos usuários variam em relação a cada palavra-chave, contrastamos os tweets classificados como positivos e negativos que mencionam cada palavra-chave ao longo das semanas. O mapa de calor da Figura 6 sumariza as diferenças através da *pontuação contrastiva* calculada como a diferença entre a fração de tweets positivos e negativos. As palavras-chave *bnt162b2*, e *covaxin* foram retiradas por não terem sido mencionadas em todas as semanas.

A Figura 6a apresenta os resultados para os tweets. Posts que mencionam as

⁹<http://sentistrength.wlv.ac.uk/index.html>

Tabela 4. Exemplos do uso de emojis.

Emoji	Tweet
🇧🇷	<i>Que dezembro traga o que novembro não trouxe, a vacina e um emprego... Dezembroou.</i>
	<i>Eu acredito no resultado curativo do tratamento nos primeiros dias!! Mas o docinho quer empurrar "para os brasileiros uma vacina que está em fase de testes ainda!!</i>
🙏	<i>Q coisa linda a vacina no Reino Unido, estamos vendo a luz no fim do túnel meu povo</i>
	<i>No Butantan a gente confia! Chega de desinformação e fakenews! Venha logo vacina!</i>
😂	<i>Comprar ultracongeladores pra estocar vacina eficaz de covid: Comprar estoque de cloroquina superfaturado que não serve pra nada.</i>
	<i>@BrazilFight Que tomem ! Essa vacina não é confiável. Querem tomar? Tomem. Mas acho que hierarquicamente era o STF que deveria tomar 1°.</i>
😬	<i>Em março a gente achava a até dezembro teria vacina.</i>
	<i>Foi tão bom saber que não adianta chegar a vacina no Brasil, já que não tem seringas suficientes pra demanda.</i>

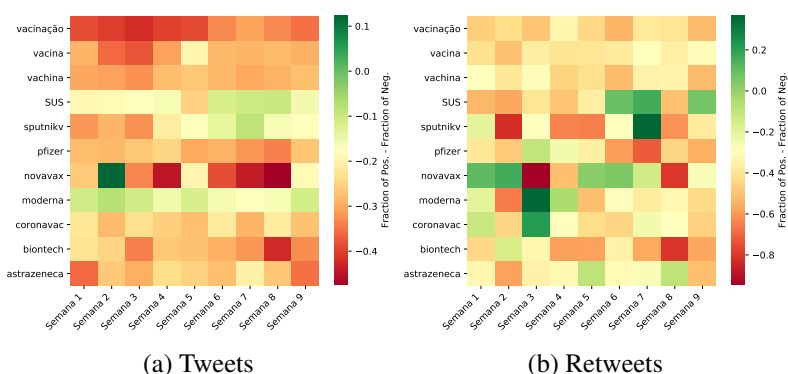


Figura 6. Evolução semanal da diferença entre a fração de sentimentos positivo e negativo.

palavras *vacinação*, *vacina* tendem a ser mais negativos (tons mais vermelhos), principalmente nas primeiras semanas da coleta dos dados. A palavra-chave *SUS* e algumas vacinas tais como *novavax*, *moderna*, *sputnikv* atraem uma fração maior de posts mais positivos. A tendência de uma maior fração de sentimento negativo relacionados aos posts pode não significar unicamente movimentos contra a vacinação, mas também a maneira como a pandemia tem sido conduzida no Brasil, despertando o sentimento de desesperança e ansiedade pela incerteza de quando a população brasileira será massivamente vacinada.

Dado que retweets podem ser considerados como uma medida da intensidade da difusão da informação na rede social, analisamos também a pontuação contrastiva dos mesmos na Figura 6b. De maneira geral, podemos observar que os retweets tendem a ser mais positivos, principalmente em relação às vacinas, sugerindo que, em nossos dados, as pessoas tendem a difundir os tweets menos negativos.

4.4. Redes de Coocorrência

Uma análise complementar de um corpus pode ser feita por meio da observação de redes de co-ocorrência. Para obter essa rede, contamos o número de coocorrências de cada par de palavras-chave. Em seguida, definimos um grafo onde os vértices representam as palavras-chave e as arestas indicam se existe coocorrência nos mesmos tweets e retweets. Em nossa visualização, o tamanho de cada nó é proporcional ao seu grau e a espessura

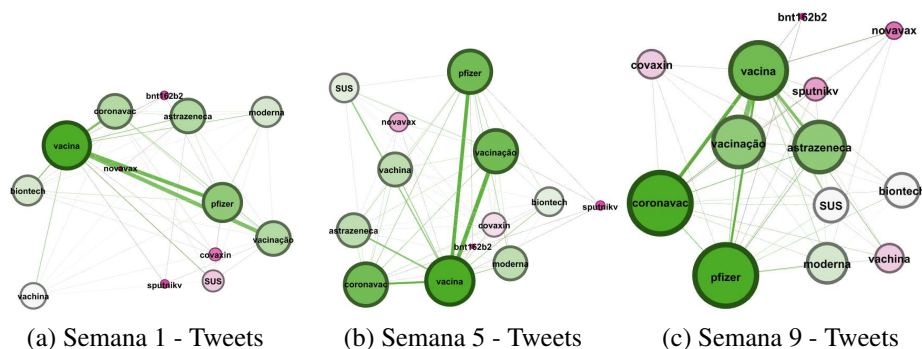


Figura 7. Redes de coocorrências para os tweets.

da aresta é proporcional ao número de vezes que cada par de palavras-chave coocorreram em nossos dados.

As redes obtidas para os tweets para a primeira, quinta e nona semanas são apresentadas na Figura 7. Como esperado, alguns pares de palavras-chave coocorrem com maior frequência, por exemplo, *vacina* e *vacinação* assim como *vacina* e *<nome-vacina>*. Ao longo das semanas, podemos observar que as primeiras vacinas aplicadas no mundo (Pfizer, Astrazeneca) e a vacina de maior disponibilidade no Brasil (Coronavac) passam a ser mais citadas, corroborando o fato de que as discussões no Twitter são, em sua grande maioria, fortemente influenciadas pelos eventos ocorridos no mundo real. Resultados para os retweets seguem as mesmas tendências e não foram incluídos por restrições de espaço. Pode-se observar também que alguns vértices se tornam mais destacados com o passar das semanas como Novavax e Covaxin, devido ao aumento da discussão relacionadas à essas vacinas.

4.5. Análise Pscolinguística

Por fim, esta seção apresenta a análise das propriedades psicolinguísticas dos conteúdos de tweets e retweets, visando encontrar padrões na forma como os usuários se comunicam no debate sobre vacinação. Utilizamos o léxico *Linguistic Inquiry and Word Count* (LIWC)¹⁰ [Tausczik and Pennebaker 2010] para categorizar as palavras em cada tweet/retweet em uma hierarquia de atributos relacionados a estilo linguístico, conceitos afetivos e cognitivos. Para cada palavra-chave, calculamos a frequência média dos vários atributos nos tweets e retweets.

Em nossos dados, identificamos a presença de 61 dos 64 atributos disponíveis na versão em português do LIWC. Em seguida, procuramos diferenças estatísticas entre os debates em torno de diferentes palavras-chave, explorando as frequências médias dos vários atributos nos tweets e retweets associadas à cada palavra-chave. Nós então aplicamos o teste não paramétrico de Kruskal [Kruskal and Wallis 1952] para selecionar apenas atributos para os quais há uma diferença significativa entre as palavras-chave, identificando 23 atributos com diferenças significativas. Em seguida, selecionamos, dentre esses 23 atributos, aqueles mais discriminantes usando o coeficiente de Gini [Yitzhaki 1979].

A Figura 8 mostra o mapa de calor para os atributos mais discriminativos para

¹⁰A versão em português utilizada foi a de 2007.

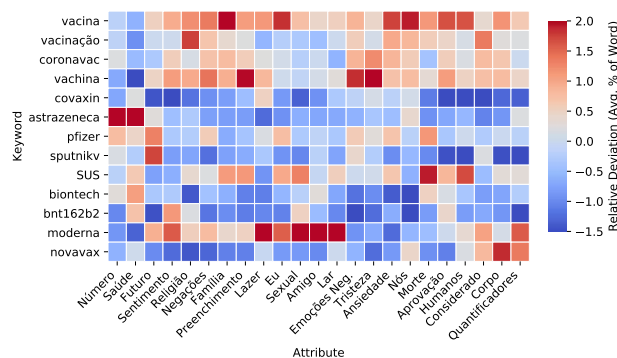


Figura 8. Top LIWC atributos extraídos dos tweets e retweets coletados.

cada palavra-chave, considerando todos os tweets e retweets. As células do mapa de calor em uma coluna indicam o desvio relativo de um determinado atributo para uma determinada palavra-chave em relação às outras palavras-chave. As células possuem cor vermelha (resp. azul) quando o atributo está mais (resp. menos) presente do que a média. Para este fim, cada coluna (atributo) é normalizada utilizando a métrica z -score - ou seja, $z = (x - \text{média}) / \text{std}$, ou seja, o valor da coluna é subtraído da média da coluna e dividido pelo desvio padrão da coluna.

Os resultados apresentados na Figura 8 mostram que postagens relacionadas às palavras-chave são diferentes em relação aos atributos selecionados. Por exemplo, postagens com *vacinação*, *vacina* frequentemente utilizam palavras relacionadas à *família*, *aprovação*, *ansiedade*, além da noção de coletividade (*nós*). Postagens com a palavra-chave *vachina* tendem a utilizar palavras relacionadas a *sentimentos negativos* e *tristeza*. Já os textos que possuem as palavras-chave referentes a vacinas específicas (exceto *pfizer*) abordam menos aspectos relacionados à *morte*, *emoção negativa*, *negações*. Em resumo, LIWC é uma ferramenta útil para analisar o conteúdo do debate em curso no Twitter, fornecendo uma boa imagem das narrativas dominantes em relação às vacinas contra a COVID-19 em nosso conjunto de dados. Por fim, é importante ressaltar que este trabalho é um estudo preliminar sobre o debate sobre vacinação no Twitter. Análises mais profundas sobre o conteúdo do texto devem ser realizadas para identificar usuários que estão contra ou a favor da campanha de vacinação contra a COVID-19.

5. Conclusões e Trabalhos Futuros

Neste artigo, investigamos a dinâmica do debate online no Twitter em torno da vacinação contra a COVID-19. Nossas análises foram realizadas a partir da coleta de mais de 9 milhões de tweets em português entre os meses de dezembro de 2020 e janeiro de 2021 que englobam os estágios iniciais da vacinação no mundo, inclusive no Brasil.

Nossos resultados mostram que o volume de tweets está correlacionado com eventos externos, como a aprovação pela Anvisa das primeiras vacinas a serem aplicadas no Brasil. A partir da análise dos sentimentos expressos pelos tweets, observamos que menções às palavras *vacina* e *vacinação* tendem a ser mais negativas, podendo significar a atuação de movimentos anti-vacinação no Twitter juntamente com o sentimento de ansiedade devido à falta de políticas públicas que deixam claras quando a população brasileira será massivamente vacinada. Adicionalmente, observamos que palavras relac-

onadas à *família, aprovação, ansiedade, coletividade* são mais utilizadas nos tweets que incluem as palavras-chave *vacina, vacinação*. Textos com as palavras-chave referentes à vacinas específicas abordam menos aspectos relacionados à *morte, emoção negativa, negações*.

Este artigo apresenta um estudo preliminar sobre o debate no Twitter da vacinação contra a COVID-19 no contexto brasileiro. Sendo um dos primeiros estudos, optamos por uma análise mais simples e geral. Como trabalhos futuros, pretendemos focar nas diferentes narrativas envolvidas no debate (perfis a favor e contra a vacinação), além de uma análise da propagação de notícias falsas relacionadas à este tópico.

Agradecimentos. Este trabalho foi realizado com apoio financeiro do CNPq, FAPEMIG, CAPES e do projeto Covid Data Analytics (PrPq/UFMG/SESU/MEC).

Referências

- [Araujo et al. 2020] Araujo, G., Moraes, F., and Pisa, I. (2020). Análise exploratória de dados do twitter: compreendendo as conexões da informação de saúde durante o surto da febre amarela em 2017. *Brazilian Journal of Information Science: Research Trends*, 14(3).
- [Brum et al. 2020] Brum, P. V., Teixeira, M. C., Miranda, R., Vimieiro, R., Jr, W. M., and Pappa, G. L. (2020). A characterization of portuguese tweets regarding the covid-19 pandemic. In *Anais do VIII Kdmine*.
- [Chen et al. 2020] Chen, E., Lerman, K., and Ferrara, E. (2020). Tracking social media discourse about the covid-19 pandemic: Development of a public coronavirus twitter data set. *JMIR Public Health Surveill.*
- [Garcia and Berton 2021] Garcia, K. and Berton, L. (2021). Topic detection and sentiment analysis in twitter content related to covid-19 from brazil and the usa. *Applied soft computing*, 101.
- [Kouzy et al. 2020] Kouzy, R., Abi Jaoude, J., Kraitem, A., El Alam, M. B., Karam, B., Adib, E., Zarka, J., Traboulsi, C., Akl, E. W., and Baddour, K. (2020). Coronavirus goes viral: Quantifying the covid-19 misinformation epidemic on twitter. 12(3).
- [Kruskal and Wallis 1952] Kruskal, W. H. and Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association*, 47(260).
- [Massarani et al. 2020] Massarani, L., Waltz, I., and Leal, T. (2020). Covid-19 in brazil: an analysis about the consumption of information on social networks. *Journal of Science Communication*, 19.
- [Pereira da Veiga et al. 2020] Pereira da Veiga, C. R., Semprebon, E., da Silva, J. L., Lins Ferreira, V., and Pereira da Veiga, C. (2020). Facebook hpv vaccine campaign: insights from brazil. *Human vaccines & immunotherapeutics*, 16(8):1824–1834.
- [Santos 2020] Santos, C. C. (2020). Análise do enquadramento do debate sobre vacina no twitter: Uma comparação entre brasil e estados unidos. In *10º Encontro Nacional de Jovens Pesquisadores em Jornalismo (JPJOR)*.
- [Tausczik and Pennebaker 2010] Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.
- [Thelwall 2017] Thelwall, M. (2017). The heart and soul of the web? sentiment strength detection in the social web with sentistrength. In *Cyberemotions*. Springer.
- [Tsao et al. 2021] Tsao, S.-F., Chen, H., Tisseverasinghe, T., Yang, Y., Li, L., and Butt, Z. (2021). What social media told us in the time of covid-19: a scoping review. *The Lancet Digital Health*, 3(3):e175–e194.
- [Yitzhaki 1979] Yitzhaki, S. (1979). Relative deprivation and the gini coefficient. *The quarterly journal of economics*, pages 321–324.