

Comparando os posicionamentos a favor/contra a vacinação COVID nos Estados Unidos da América e no Brasil

André Mediate de Sousa, Karin Becker

¹Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Porto Alegre – RS – Brazil

{andremediate, karin.becker}@inf.ufrgs.br

Abstract. *The World Health Organization endorses vaccination as the most effective way to fight the COVID-19 pandemic, but hesitation to get vaccinated is a reality in many countries. This article develops a temporal study of stances for/against COVID-19 vaccination in the United States of America (USA), drawing a parallel with Brazil using data from Twitter. We conclude that the anti-vaccination movement in the US is more dominant in relation to Brazil. While there are many common elements between the two countries, the US population is more objective in defending their positions (health and safety versus individual freedom). In Brazil, we observe the population's anxiety about getting vaccinated and a political bias.*

Resumo. *A Organização Mundial da Saúde endossa a vacinação como a forma mais eficaz de combater a pandemia de COVID-19, mas a hesitação em se vacinar é uma realidade em muitos países. Este artigo desenvolve um estudo temporal das posturas pró/contra a vacinação COVID-19 nos Estados Unidos da América (EUA), traçando um paralelo com o Brasil usando dados do Twitter. Concluímos que o movimento antivacinação nos EUA é mais dominante em relação ao Brasil. Embora existam muitos elementos comuns entre os dois países, a população dos EUA é mais objetiva na defesa de suas posições (saúde e segurança versus liberdade individual). No Brasil, observamos a ansiedade da população em se vacinar e um viés político.*

1. Introdução

A vacinação é reconhecida pela Organização Mundial de Saúde (OMS) como uma das medidas de saúde pública de maior sucesso na erradicação de doenças como pólio, sarampo, tuberculose, entre outras. Muitas expectativas foram depositadas na vacinação como instrumento essencial no combate ao coronavírus e grandes esforços foram empreendidos para desenvolver, aprovar e disponibilizar as vacinas COVID-19 no menor prazo possível. Contudo, um número crescente de pessoas vêm percebendo a vacinação como insegura e desnecessária [Hornsey et al. 2018]. Cientistas têm se preocupado com o declínio da vacinação no mundo inteiro, e desejam compreender as razões da hesitação vacinal para estabelecer políticas públicas focadas nestas restrições.

Uma revisão sistemática [Cascini et al. 2021] revelou que as taxas de aceitação da vacina COVID-19 variaram muito entre países, e em diferentes momentos do tempo. Estudos [Omer and *et alli* 2021, Hotez 2021] apontam o prejuízo à saúde pública dos Estados Unidos da América (EUA) causado pelos movimentos antivacina, os quais já estavam organizados para fazer campanha contra medidas consideradas essenciais ao controle da pandemia (e.g., máscaras, isolamento social). O Brasil é tido como um dos países

mais favoráveis à vacinação COVID-19 [Solís Arce and *et alli* 2021, Cascini et al. 2021], mas os movimentos anti-vacina são particularmente motivados por polarização política [Ebeling et al. 2022, de Sousa and Becker 2022].

Diferentes trabalhos têm contribuído à compreensão da percepção de aspectos ligados à COVID expressos em redes sociais, principalmente o Twitter, principalmente com a técnica de modelagem de tópicos LDA [Xue et al. 2020, Tao et al. 2020, Oliveira et al. 2022]. Uma comparação entre as percepções e sentimentos sobre a COVID-19 nos EUA e no Brasil foi desenvolvida em [Garcia and Berton 2021].

Em relação à vacinação COVID, [Huangfu et al. 2022] também explora LDA, mas técnicas mais recentes baseadas em word embeddings têm sido empregadas. Top2Vec [Angelov 2020] foi empregado em [Ma et al. 2020] para investigar hesitação vacinal COVID no mundo. Em nossos trabalhos anteriores [Ebeling et al. 2022, de Sousa and Becker 2022] usamos a variação BERTopic [Grootendorst 2020] para extrair argumentos pró/contra vacinação no contexto brasileiro. Em descompasso com a tradição de vacinação em massa do Brasil, observamos que os brasileiros expressaram resistências à vacinação, mas com um forte viés-político.

Neste trabalho, estendemos nosso estudo anterior [de Sousa and Becker 2022] e verificamos se estas resistências são da mesma natureza observada nos EUA, um país com tradição anti-vacina. Para este fim, desenvolvemos um estudo temporal para identificar e compreender os argumentos usados no Twitter para defender posicionamentos favoráveis e contrários à vacinação contra COVID-19 nos EUA, e compará-los com a situação brasileira. A análise dos posicionamentos relativos aos EUA foi conduzida a partir de 1.195.822 tweets, coletados entre Março de 2020 e Abril de 2022, período que compreende do início da pandemia até o encorajamento para doses de reforço. Nosso objetivo foi responder às seguintes questões de pesquisa:

- *RQ1*: Quais os principais argumentos usados para defender cada posicionamento nos EUA?
- *RQ2*: Como os posicionamentos sobre a campanha de vacinação nos EUA evoluem com o tempo?
- *RQ3*: Como o posicionamento em relação à vacinação nos EUA se compara com o observado no Brasil?

Para responder a estas perguntas, empregamos a técnica de modelagem de tópicos BERTopic, que permite uma compreensão mais detalhada dos mesmos através da identificação dos argumentos utilizados para defender cada posicionamentos. Também estendemos seu uso e o aplicamos na resolução de problemas de qualidade e vieses comuns em análises envolvendo dados coletados de redes sociais. Propomos uma etapa de filtragem que explora recursos de busca semântica e plotagens bi-dimensionais do BERTopic para melhorar a qualidade dos dados.

Nosso trabalho tem duas principais contribuições. A primeira delas é a comparação do comportamento da população do Brasil e dos EUA em relação à vacinação. Observamos um forte movimento pró-vacinação em ambos os países, mas que os movimentos anti-vacinação têm características distintas. Nos EUA o movimento anti-vacina é regular, centrado na liberdade individual, no questionamento da segurança das vacinas. No Brasil ele foi pontual, contra a obrigatoriedade e com motivação política.

Nossa segunda contribuição é a etapa de filtragem para melhorar a qualidade dos dados. Com a filtragem identificamos: a) falsos positivos relacionados com o método de coleta, isto é, quando a hashtag aponta para um posicionamento (e.g., vacinanao), mas, por ser uma resposta, à postagem crítica este posicionamento; b) levantamento artificial de hashtags (e.g., futebol) para promover engajamento; c) identificação de postagens com texto repetitivo, característico de robots.

O restante deste trabalho está estruturado da seguinte forma. A Seção 2 apresenta os trabalhos relacionados. A Seção 3 descreve os dados e o método de análise proposto, e a Seção 4 apresenta uma visão geral do nosso trabalho anterior. Na seção 5 estão os resultados às questões de pesquisa e a Seção 6 apresenta as conclusões e trabalhos futuros.

2. Trabalhos relacionados

A extração da percepção da população sobre a COVID-19 usando dados do Twitter é uma área bastante ativa [Tao et al. 2020, Xue et al. 2020, Oliveira et al. 2022], onde o LDA é a técnica mais popular. Diferenças nas percepções emocionais sobre a COVID-19 entre os EUA e no Brasil é desenvolvida em [Garcia and Berton 2021].

A compreensão dos fatores predominantes para comportamentos anti-vacina é abordada em [Hornsey et al. 2018]. Em relação à COVID-19, uma revisão sistemática [Cascini et al. 2021] revelou que as taxas de aceitação da vacina COVID-19 variou enormemente entre países, sendo que os EUA têm uma menor taxa de aceitação quando comparados ao Brasil. Twitter e LDA também são combinados para estudar a aceitação de vacinas COVID, tais como [Huangfu et al. 2022] que compara vacinas de acordo com o laboratório, e [Liu et al. 2021] que associa tópicos a eventos de desenvolvimento e teste de vacinas. Top2Vec foi empregado sobre dados do Twitter em [Ma et al. 2020] para investigar razões de hesitação à vacinação no mundo (e.g. efeitos colaterais, descrença na ciência e governos). [Ebeling et al. 2022] combina LDA e BERTopic para entender de forma mais detalhada os argumentos associados às posturas pró/anti-vacinação no Brasil, apontando um forte viés político.

Também utilizando o BERTopic, fizemos em [de Sousa and Becker 2022] um estudo temporal dos posicionamentos dos brasileiros no Twitter, confirmando este viés político. Utilizamos este trabalho anterior para comparar os posicionamentos pró/anti-vacinação nos EUA. Resumimos na Seção 4 os principais pontos necessários à comparação.

3. Materiais e métodos

Este trabalho apresenta uma análise usando dados do Twitter para compreender os argumentos usados pelos americanos ao se posicionar em relação a vacinação contra a COVID-19, comparando com o observado no Brasil. O restante desta seção detalha os dados as técnicas usadas para responder às questões de pesquisa listadas na Seção 1. Todo o código foi desenvolvido em Python no ambiente Jupyter Notebook com a ajuda de bibliotecas como NLTK e Pandas, e está disponível em um repositório público¹.

¹<https://github.com/mediote/sbbd>

Tabela 1. Dataset de tweets representando posicionamentos nos EUA

Posicionamento	Hashtags de coleta	Bruto		Pre-processado	
		Tweets	Usuários	Tweets	Usuários
pro-vaxxers	getvaccinated, vaccineswork, vaccinessavelives	842.375	239.681	673.806	207.995
anti-vaxxers	novaccine, novaccineforme, novaccinepassports	353.447	55.342	220.179	44.059
Total		1.195.822	295.023	893.985	252.054

3.1. Coleta e pré-processamento de Tweets

Usamos a TwitterAPI² para coletar os tweets relacionados à vacinação nos EUA no intervalo de 11 de março de 2020 a 06 de abril de 2022. Esse período abrange desde o momento em que a OMS elevou o estado de contaminação à pandemia de Covid-19, até a vacinação de aproximadamente 218 milhões de pessoas nos EUA.

Delimitar o espaço de coleta a um contexto territorial específico é um desafio, pois apenas 3-5% dos tweets são georreferenciados e vários países compartilham o idioma inglês. Definimos um processo de coleta que observa 3 premissas: (1) ser capaz de separar os tweets em datasets diferentes de acordo com a postura expressa em relação a vacinação; (2) os posicionamentos devem ser expressos através de hashtags usadas pela população dos EUA, e (3) a quantidade de tweets deve ser suficiente para entender o panorama geral da vacinação nos EUA. O proposto tem os seguintes passos:

Passo 1) Buscar hashtags representativas em tweets georeferenciados: coletamos tweets georeferenciados dos EUA contendo os termos “vaccination” e/ou “vaccine” (383.248 tweets) a fim de inspecionar como a população dos EUA se expressa. Identificamos manualmente as hashtags que claramente denotem posicionamentos pró e contra a vacinação, selecionando as três prevalentes para cada posicionamento;

Passo 2) Coletar tweets não-georeferenciados usando as hashtags de posicionamento: visando garantir a terceira premissa, coletamos novamente tweets usando as hashtags de posicionamento identificadas, obtendo um total de 1.195.822 tweets. Não é possível garantir que todos os tweets são de usuários dos EUA, mas a modelagem de tópicos nos permitiu eliminar muitos tweets violando esta condição (vide Seção 3.2).

Finalmente, removemos tweets com menos de três termos, retuítes e tweets duplicados. A fim de melhorar a qualidade da modelagem de tópicos, também aplicamos técnicas de pré-processamento usuais (e.g. remoção de pontuação, caracteres especiais, hashtags e URLs). O número resultante de tweets e usuários únicos para cada posicionamento é mostrado na Tabela 1 (Pré-processado).

3.2. Modelagem de Tópico com BERTopic

O BERTopic é um framework que engloba algoritmos que buscam automaticamente clusters densos em uma coleção de documentos, presumindo que clusters de documentos semanticamente semelhantes formam tópicos. Requer como entrada um *corpus* e um modelo de representação de linguagem pré-treinado (e.g. BERT, MPNET) para geração de embeddings. Utilizamos como modelo pré-treinado o MPNET [Song et al. 2020], devido sua maior performance em comparação ao BERT³ em tarefas de agrupamento e busca semântica. Após os embeddings passarem por uma redução de dimensionalidade

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

³https://www.sbert.net/docs/pretrained_models.html

com UMAP, aplica-se um algoritmo de clustering para agrupar documentos semelhantes, formando tópicos. Escolhemos o HDBScan, baseado em densidade. O resultado da modelagem pode gerar um grande número de tópicos, mas o framework oferece recursos de redução e limitação ou junção de tópicos semelhantes.

Para auxiliar na interpretação, o BERTopic rotula cada tópico com n-gramas obtidos através da medida c-TF-IDF aplicada aos termos presentes nos documentos agrupados. Oferece também recursos de busca semântica (i.e. tópicos próximos a um dado termo), identificação de documentos representativos de um dado tópico, e visualização. Como BERTopic permite a modelagem dinâmica, é possível obter a distribuição dos tópicos no tempo, e selecionar os termos c-TF-IDF associados a momentos específicos.

Neste trabalho, a modelagem de tópicos foi resultado de diferentes iterações, onde a entrada foram os tweets pré-processados/filtrados vetorizados com MPNET. Para a garantir a atribuição do maior o maior número possível de tweets a tópicos reais, em vez de rotulá-los como ruídos, usamos os seguintes parâmetros: *num_neighbors* = 15, *n_components* = 5 e *min_cluster_size* = 60. A interpretação dos resultados de cada modelagem foi feita com base nos recursos providos, tais como visualização, termos influentes (c-TF-IDF) e, tweets representativos de cada cluster. O processo de modelagem foi executado em três etapas:

a) Filtragem: esta etapa visou eliminar os problemas decorrentes da coleta que introduziram vieses e ruídos na amostra, conforme mencionado na Seção 1. Fizemos uma modelagem inicial sob o total de tweets pré-processados referente a cada posicionamento conforme a Tabela 1, resultando em 427 tópicos para os pro-vaxxers e 181 tópicos para os anti-vaxxers. Inspecionamos manualmente os tópicos extraídos utilizando recursos de busca semântica, plotagens dos tópicos e amostras de tweets, identificando assim os seguintes problemas: (1) tweets não oriundos de usuários americanos (e.g. tweets semanticamente relacionados a locais como Irlanda, Índia, Inglaterra e África do Sul, por fazerem menções a estas localizações), (2) falsos positivos devido ao uso indevido de hashtags (e.g. xingamentos como “And to that ignorant comment i hope you get the coronavirus” no grupo anti-vax, e vice-versa), (3) tweets agrupados devido ao uso de hashtags artificiais de engajamento (e.g. esporte, música), e (4) tweets muito repetitivos, característicos de robots. Como resultado, excluímos 546 clusters (396 pro-vax e 150 anti-vax), reduzindo nosso datasets a 158.092 tweets (119.775 pro e 38.317 anti).

b) Modelagem de Tópicos: criamos dois novos datasets (pro e anti-vaxxers) a partir dos tweets resultantes da etapa filtragem. Submetemos os datasets uma nova rodada de modelagem, obtendo 64 tópicos para os pro-vaxxers e 19 tópicos para os anti-vaxxers.

c) Redução de Tópicos: A redução de tópicos permite a união de tópicos similares após o modelagem inicial. Esse processo foi necessário devido a dificuldade de interpretar o grande número de tópicos extraídos, mesmo após as exclusões feitas nas etapas anteriores. Para definir um número adequado de tópicos e responder nossas questões de pesquisa, usamos alguns recursos do BERTopic, como plotagens dos tópicos em planos bi-dimensionais e visualizações de como os tópicos se relacionam hierarquicamente em termos de similaridade. Na união, os tweets são deslocados e mesclados em novos tópicos. Após algumas iterações, reduzimos a modelagem final para 10 tópicos (*nr_topic=11*) para ambos os posicionamentos.

A Figura 1 ilustra para os pro-vaxxers os clusters resultantes das etapas descri-

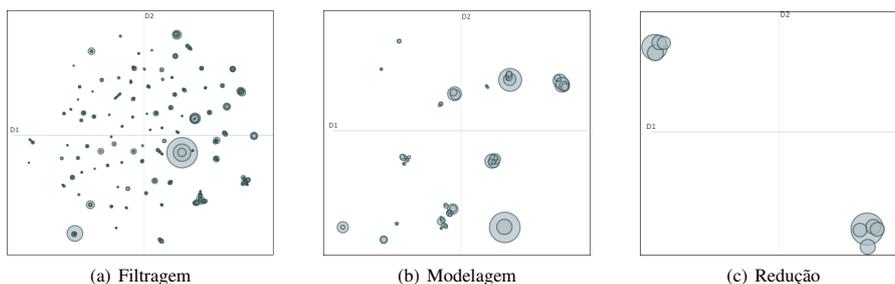


Figura 1. Redução de tópicos: Pro-vaxxers

tas acima. Com a modelagem final, respondemos à questão de pesquisa RQ1 usando os termos significativos para cada tópico (c-TF-IDF), contextualizados com uma amostra central de tweets. Para a questão RQ2, distribuímos os tweets ao longo do tempo, interpretando os principais eventos com termos significativos naqueles pontos temporais.

4. Posturas de vacinação da COVID no Brasil: Visão Geral

Em [de Sousa and Becker 2022] relatamos um estudo temporal do posicionamento brasileiro quanto à vacina COVID utilizando tweets coletados entre 29 de Fevereiro 2020 até 3 de Maio de 2021. Esta data limite foi devido à instauração da CPI da COVID no país, que desviou o debate para negligência e corrupção no manejo da COVID no país, impossibilitando o reconhecimento de posicionamentos pró/anti-vacinação de forma confiável.

Identificamos as hashtags representativas dos posicionamentos de modo semelhante ao explicado na Seção 3.1, porém sem observar as premissas de coleta. O número de tweets brutos/pre-processados para cada posicionamento é mostrado na Tabela 2. Em relação ao uso do BERTopic, utilizamos o modelo pré-treinado *multilingual*, e não aplicamos a etapa de filtragem demonstrado na Seção 3.2. Exploramos principalmente recursos de modelagem dinâmica do BERTopic, onde além de identificarmos os principais tópicos discutidos, também analisamos suas mudanças e como evoluíram no decorrer do tempo.

Tabela 2. Dataset de tweets representando posicionamentos no Brasil

Posicionamento	Hashtags	Bruto		Pre-processado	
		Tweets	Usuários	Tweets	Usuários
pro-vaxxers	todospelasvacinas, vacinaja, vemvacina vacinaparatodos, vacinasim	216.377	76.316	139.131	55.695
anti-vaxxers	eunaovoutomarvacina, vacinaobligatorioanao vacinanao, naovoutomarvacina vacinaobligatorioanunca	39.073	16.675	20.084	10.244
Total		255.450	92.991	139.131	65.939

Tabela 3. Resumo dos argumentos pro/anti-vax no Brasil

	Nro Clusters	Argumentos
Pro-Vaxxers	4	celebração pela aprovação da vacina, alegria por ser vacinado, defesa e celebração da ciência e do SUS
	2	apoio à vacinação COVID (com uso artificial de hashtag)
	2	críticas ao Presidente e ao governo pela indisponibilidade de vacinas
	2	discussões sobre vacinação para retorno às aulas
Anti-Vaxxers	3	contra a vacinação (com uso artificial de hashtags) e vacinação obrigatória
	4	contra a "vacina chinesa"; críticas a João Dória, aos políticos em geral, e ao STF
12.455	3	ironia aos anti-vaxxers (sobra de vacina para pro-vaxxers), críticas ao uso de hashtags para promover movimento anti-vaxxer

A Tabela 3 resume os principais argumentos associados a cada tópico. O movimento pró-vacina expressa a alegria e ansiedade pela oportunidade de se vacinar, elogia a ciência e o SUS, e faz severas críticas ao Presidente e seu governo pela falta de vacinas. A vacinação como condição de retorno às aulas presenciais também é discutida. No movimento anti-vacina há um esforço orquestrado de divulgação de ideias anti-vacina, e insatisfação geral com a possibilidade da obrigatoriedade da mesma. Por isto criticam ou pedem apoio a governadores e políticos, e são contra o STF que declarou a medida constitucional. Há críticas específicas à "vacina chinesa" relacionadas a João Dória. Nota-se a presença significativa de falsos-positivos no movimento anti-vax, com pro-vaxxers criticando e ironizando os anti-vaxxers. Ambos os movimentos usam hashtags artificiais para propagar suas ideias.

O movimento anti-vaxxer foi prevalente em 2020, com dois picos: um em outubro, no dia nacional da vacinação, e outro em dezembro, quando a vacinação iniciou na Europa, tornando-se insignificante em 2021. Já o movimento pró-vaxxer foi prevalente em 2021, com o pico em Janeiro, envolvendo os episódios da falta de oxigênio em Manaus e da vacinação da Enfermeira Mônica. Maiores detalhes sobre o método e os resultados podem ser encontrados em [de Sousa and Becker 2022].

É importante destacar que devido ao método de coleta utilizado, os dados continham diferentes tipos de viés que influenciaram a análise, em particular falso positivos e engajamento artificial devido a hashtags. Estes problemas dificultam a parametrização dos algoritmos para bons resultados. Apesar de termos sido capazes de apontar os tópicos com distorções causadas por esses problemas, os identificamos tardiamente, já na etapa de interpretação dos resultados, tornando o processo custoso e sujeito a falhas.

No presente trabalho, exploramos nosso domínio avançado no uso de BERTopic adicionando uma etapa extra de filtragem à modelagem de tópicos, que permitiu resolver os problemas acima mencionados de forma relativamente fácil, tornando o conjunto de dados relativos à vacinação dos EUA de maior qualidade.

5. Resultados

5.1. RQ1: Quais os principais argumentos usados para defender cada posicionamento?

As tabelas 4 e 5 resumem os tópicos extraídos de acordo com o método descrito na Seção 3.2, ordenados por número de tweets. Cada tópico é caracterizado pela respectiva quantidade de tweets, os quatro termos mais representativos (unigrama ou bigrama) em termos de pontuação c-TF-IDF e um tweet representativo, traduzidos para o português.

a) Pro-vaxxers: como mostra a Tabela 4, este grupo celebra a vacinação, apoia medidas para conter o contágio, e expressa gratidão à ciência. O tópico 0, com o maior número de tweets, revela euforia com a eficácia da vacina, que permite o retorno às atividades com segurança. O tópico 2 celebra a aprovação pelo FDA das vacinas e expressa felicidade pela vacinação da população. Elogios à ciência, alegria e gratidão por se vacinarem ou terem seus entes queridos protegidos, são os argumentos centrais dos tópicos 3 e 6.

No tópico 1, as pessoas apoiam medidas de enfrentamento à pandemia (e.g. máscara, distanciamento social). O tópico 5 demonstra apoio às medidas que restringem certas atividades aos vacinados (e.g. viagens, bares e restaurantes). Os tópicos 4 e

Tabela 4. Representação global dos tópicos dos Pro-vaxxers nos EUA

Tópico	Tweets	Termos Pro-vaxxers	Argumentos representativos
0	43.230	vacina covid, imunização mundial, vacinas covid, totalmente vacinado	“Boa notícia! A vacinação funciona!”
1	18.877	usar máscara, usar máscara, usar máscaras, mandato de máscara	“Use máscara e toma vacina.”
2	11.734	vacina pfizer, vacina pfizer, segunda dose, dose pfizer	“FDA aprova a primeira vacina contra a covid-19!”
3	4.554	braço dolorido, local da injeção, injeção moderna, braço dolorido	“Braço dolorido por dois dias.”
4	4.057	vacina poliomielite, sem poliomielite, vacinação poliomielite erradicação da poliomielite	“Vacinas são ótimas! Você leu sobre a vacina contra a poliomielite?”
5	3.922	vacina obrigatória, profissional saúde cuidados de saúde, funcionários vacinados	“Me chame de radical, mas eu apoio a obrigatoriedade de vacinas para médicos sem exceções!”
6	3.804	segunda dose, tomou dose dose hoje, tomando uma dose	“Tomei minha primeira dose hoje woohoo!”
7	3.523	anti vaxxers, anti vax, axxers anti, anti vaxx	“Para todos os idiotas anti vax f@# em todos os lugares.”
8	3.354	variante delta, delta covid, delta plus, delta vacinado	“A variante covid delta é mortal, por favor cuide de você e dos outros.”
9	3.066	vacina gripe, contra a gripe, pegou gripe, reforço covid	“Acabei de receber minha vacina anual contra a gripe e meu reforço covid.”

Tabela 5. Representação global dos tópicos dos Anti-vaxxers nos EUA

Tópico	Tweets	Termos anti-vaxxers	Argumentos representativos
0	19521	vacinas passaportes, vacinas obrigatória, vacina experimental, não quero	“Vacinação obrigatória é totalitarismo.”
1	3886	apartheid médico, tirania médica, controle de saúde, droga experimental	“Hora de parar este apartheid médico.”
2	3421	máscara, uso obrigatório, nao funciona, distanciamento social	“Não se vacinar e se recusar-se a usar máscara.”
3	3128	junta necessária, extorsão fraude, investigação FBI, corrupção governo	“Necessária investigação conjunta do rcmp fbi sobre corrupção do governo fraude prevenção de extorsão legislação dos direitos dos pacientes”
4	2108	confiar governo, liberdade direitos, direitos humanos liberdades civis	“Os governos não têm o direito de negar a liberdade das pessoas! Não somos animais domésticos não seremos escravizados!”
5	1400	vax passaporte, anti vax, vax obrigatório, mandatoria vax	“Se você é pro vaxpass você é anti liberdade e meu inimigo”
	1288	deixar crianças, abuso infantil, proteger crianças, escola infantil	“Eles não vão tocar em nossos filhos.”
7	935	nova york, florida texas, york estado, nova jersey	“Nyc está acordando da escravidão global que protesta contra vacinação”
8	873	pcr teste, pare testar, falso positivos, teste massa	“Eu não serei testado à força ou injetado à força, eu me recuso a enviar para usar qualquer”
9	816	anti-ciência, ciência real, evidência científica, confiança científica	“Evidências científicas... Me mostre os estudos que você leu!”

9 reforçam a importância da vacinação para outros fins, como o combate à pólio e gripe. No tópico 8 as pessoas mostram preocupação com as novas variantes do vírus.

O tópico 7 é dedicado a críticas aos anti-vaxxers por criarem engajamento usando mensagens contra a vacinação ou qualquer medida adotada no enfrentamento à pandemia como apresentado anteriormente. Esta crítica é encontrada junto a outros tópicos, permeando a maioria das discussões sobre a pandemia de COVID-19 nos EUA.

b) Anti-vaxxers: A Tabela 5 resume os tópicos identificados para o movimento anti-vax. Eles defendem a liberdade individual da vacinação, questionam a segurança da vacina, bem como medidas de enfrentamento em geral.

Os tópicos 0 e 5 são manifestações contra a vacina por ser insegura e/ou obrigatória. São também contra máscaras e medidas de distanciamento social (Tópico 2). Os principais argumentos usados para defender esses posicionamentos são a violação da liberdade de escolha e os direitos civis, que é explicitamente abordada nos tópicos 4, 6 e 8. O tópico 6 é focado em crianças, manifestando que é um direito dos pais escolher se os filhos devem ou não se vacinar, e condenando a exigência de vacinação para retorno às aulas. O tópico 7 apresenta um levante convocando a população de diversos estados para protestar contra a vacinação e sua obrigatoriedade, contrastando posicionamento de locais com maiores restrições (Nova York) ou sem (Flórida, Texas).

Para defender seus argumentos, colocam em dúvida a segurança e a eficácia das medidas de enfrentamento à pandemia e das vacinas disseminando fontes de informação que sustentam seus argumentos. No geral, identificamos postagens relacionadas a teorias da conspiração e de cunho religioso. Existe uma preocupação com um apartheid médico e controle da população por seus governantes. Os argumentos centrais nos tópicos 1 e 9 colocam em cheque os médicos e cientistas, e aventam interesses das companhias farmacêuticas. O tópico 3 aventa corrupções e interesses escusos do governo.

c) Conclusão: os pro-vaxxers estão preocupados com a saúde da população, onde as pessoas expressam alegria ou expectativas sobre a vacinação, apoiam a vacinação e medidas preventivas, e criticam quem não está disposto a se vacinar. Já os anti-vaxxers estão preocupados com a segurança das vacinas e com os direitos individuais que os isentam da obrigatoriedade da vacinação.

5.2. RQ2: Como os posicionamentos sobre a campanha de vacinação nos EUA evoluem com o tempo?

A Figura 2 fornece uma visão geral do volume de atividades desses grupos ao longo do tempo. Até novembro de 2020, quando as vacinas ainda estavam sendo desenvolvidas e testadas, os pro-vaxxers se manifestaram de forma quase isolada. Em dezembro de 2020, com o início da vacinação COVID em nível mundial, observa-se um início tímido nas atividades dos anti-vaxxers, acompanhado por um significativo volume de mensagens postadas pelos pro-vaxxers. A partir daí, as manifestações começaram a crescer, caracterizando de fato a existência de uma corrente a favor e outra contra a vacinação.

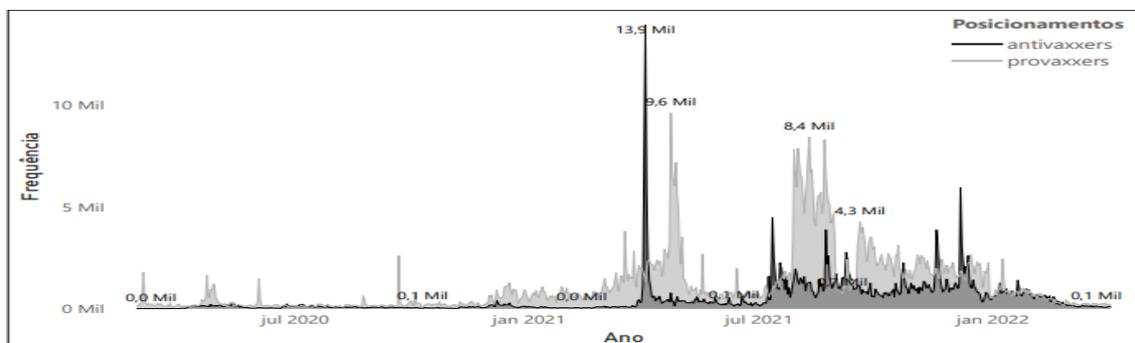


Figura 2. Postagens Pro/Anti-vacina nos EUA ao longo do tempo

Utilizamos notícias de jornais e os termos representativos dos tópicos nos picos de engajamento observados para compreender o comportamento. No movimento anti-vaxxer os picos estão relacionados a fatos questionando a necessidade da vacinação ou sua eficácia. Termos relacionados às variantes ficam evidenciados em tópicos específicos ao longo do tempo. Os tópicos focados no surgimento de novas variantes questionaram a efetividade das vacinas, as taxas de transmissão e de letalidade do vírus. O primeiro e maior pico de atividades é observado em abril de 2021, relacionado a discussões relacionadas à variante Delta. Em 20/08/2021 observa-se reação ao terceiro pico de mortes por COVID-19 nos EUA, em setembro de 2021 (quarta onda) as notícias relatam discussões sobre ações lockdown ao redor do mundo, correlacionados com a aceitação ou hesitação à vacina. O pico associado a dezembro de 2021 está relacionado ao aumento de casos da variante Ômicron.

As manifestações pró-vacina são contínuas e constantes ao longo dos anos de 2021 e 2022, com picos em momentos específicos. Os picos de movimentação dos pro-vaxxers estão relacionados em estabelecer contrapontos sobre eficácia da vacina levantados pelos anti-vaxxers, e a notícias de eficácia da vacina, tais como a de que a vacina pode diminuir a taxa de transmissão e de eficácia de 94% da vacina da Moderna (abril de 2021), ou da segurança da dose de reforço das vacinas Pfizer e AstraZeneca em setembro de 2021. A movimentação a partir de setembro de 2021 também está ligada à mudança de tom do Presidente Biden sobre a necessidade de vacinação da população.

Em conclusão, o movimento pró-vacinação é prevalente, apoiando a vacinação como forma de combate à pandemia e retorno às atividades. Contudo, o movimento anti-vacinação é consistente na defesa do direito individual à vacinação, sendo que as incertezas criadas por novas variantes foram usadas para reforçar as descrenças sobre a eficácia e necessidade de vacinação.

5.3. RQ3: Como o posicionamento em relação à vacinação nos EUA se compara com o observado no Brasil?

5.3.1. Comparação dos Argumentos

Considerando os argumentos da população brasileira (Tabela 3) e dos EUA (Tabelas 4 e 5), a Tabela 6 resume a comparação dos argumentos identificados em ambos os países, junto do número de clusters associados. Observamos a existência de tópicos com argumentos correlatos entre os dois países, destacadas na tabela em cinza. No tocante aos pro-vaxxers, ambos os países celebram a disponibilidade de uma vacina, são gratos à ciência, e apoiam a vacinação como forma de combater a COVID. Os EUA estendem seu apoio à vacinação em geral, outras medidas de combate à COVID (e.g., máscaras), e preocupação quanto a novas variantes. Estes aspectos não são abordados no Brasil, onde o foco é a indisponibilidade de vacinas, a responsabilidade do governo nesta situação, e o impacto na educação. É importante destacar que no período estudado, menos de 6% da população brasileira havia tido acesso à vacina, o que pode explicar este foco. Os brasileiros também usam hashtags artificiais para apoiar a vacinação.

Em ambos os países, os anti-vaxxers são essencialmente contra a obrigatoriedade das vacinas, questionando sua segurança, sendo que o uso artificial de hashtags é frequente para propalar estas ideias. No caso dos EUA, existe uma grande defesa da liberdade individual associada à vacinação, oposição às medidas de combate à COVID em geral, bem como o descrédito da ciência e teorias conspiratórias. Já no Brasil, observa-se um debate mais político, com a população criticando políticos ou o STF em questões relacionadas à obrigatoriedade da vacinação COVID. Nos EUA, este viés político é levemente diferente, quando estados com posicionamentos governamentais distintos são criticados/elogiados.

A crítica dos pro-vaxxers aos anti-vaxxers é observada nos dois países, mas se revela de formas distintas. Parte significativa do movimento anti-vaxxer no Brasil (3 tópicos) são replys de pro-vaxxers contendo críticas e ironias contra a postura anti-vacina. Nos EUA há um tópico nos pro-vaxxers destinado a fazer críticas aos anti-vaxxer

Tabela 6. Comparação de argumentos BR x EUA

Movimento	Qt. Tópicos EUA	Qt. Tópicos BR	Argumentação Central
Pro-vaxxers	3	4	Felicidade e euforia com as vacinas COVID, gratidão a ciência
	2	2	Apoio a vacinação COVID
	1		apoio a medidas de combate a covid
	2		apoio à vacinação em geral
	1		crítica a anti-vaxxers
Anti-vaxxers	1	2	preocupação com novas variantes
	3	2	críticas aos políticos pelo manejo da COVID
	1		volta as aulas
	3	3	contra vacinação COVID (obrigatória, passaporte vacinal)
	3	4	críticas a posicionamentos políticos e jurídicos
Anti-vaxxers	1		contra medidas gerais de enfrentamento COVID
	3		liberdade individual
	3		descrença da ciência, teorias conspiratórias
		3	ironia aos anti-vaxxers (sobra vacina, hashtags)

5.3.2. Comparação dos Posicionamentos ao Longo do Tempo

A Figura 3 permite comparar o comportamento dos posicionamentos ao longo do tempo. No Brasil os anti-vaxxers se manifestaram em eventos específicos ao longo de 2020, mas virtualmente desapareceram após o início da vacinação no mundo, sendo substituídos por um movimento pró-vax crescente e contínuo [de Sousa and Becker 2022]. Já nos EUA observa-se que ao longo de 2020 os pro-vaxxers foram mais ativos, mas que a partir de 2021 existem confrontos recorrentes entre pro-vaxxers e anti-vaxxers. Pode-se fazer a hipótese que as reações anti-vaxxers são devidas a pressões recorrentes incentivando a vacinação principalmente face a novas variantes, as quais são contestadas. Apesar das eleições presidenciais nos EUA disputadas no final de 2020 e a diferença dos partidos no manejo da COVID, não observamos manifestações políticas nos posicionamentos pró/anti-vacinação.

A proporção de postagens dos grupos nos dois países permite concluir que nos EUA existe uma presença mais forte de anti-vaxxers. Nos EUA a relação do pro/anti-vaxxers é de 76/24%, respectivamente, enquanto que no Brasil esta relação é 87/13%.

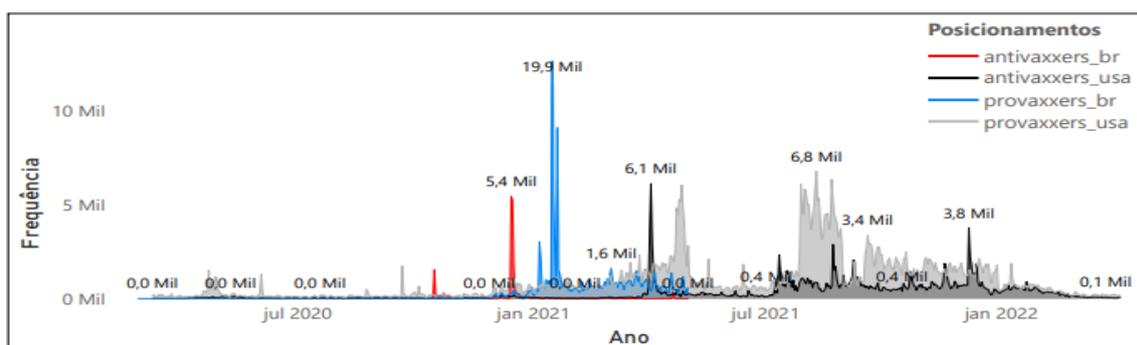


Figura 3. Comparação das postagens Pro/Anti-vacinação nos EUA e no Brasil

6. Conclusões e trabalhos futuros

Neste trabalho, fizemos uma análise temporal dos posicionamentos em relação à vacinação COVID-19 nos EUA, e traçamos um paralelo com o que ocorreu no Brasil. Usamos a técnica de modelagem BERTopic para identificar os principais tópicos e en-

tender com detalhe os argumentos usados para defendê-los. Aprimoramos o emprego do BERTopic para resolver problemas de coleta de dados e de qualidade.

A adição de uma etapa de filtragem nos permitiu mitigar os problemas com postagens oriundas de outros países com língua inglesa, pois o BERTopic agrupou esses tweets, permitindo que pudéssemos removê-los. Esta é uma alternativa à tradicional verificação pela localização fornecida no perfil do usuário, que tem limitações. A filtragem também nos permitiu eliminar problemas identificados em nosso trabalho anterior, a saber, falsos positivos relativos ao reply de tweets contendo uma hashtag de posicionamento oposta ao argumento defendido, levantamento artificial de hashtags e possíveis robôs. Com isto, os tópicos resultantes ficaram bem mais coerentes, facilitando sua interpretação. Além disso, o modelo MPNET, associado à otimização bayesiana de hiperparâmetros, tornou superior a qualidade dos agrupamentos.

Nosso estudo permitiu observar que o grupo dos pro-vaxxers é mais engajado em ambos os países. Contudo, nos EUA, os anti-vaxxers caracterizam um movimento efetivo contra a vacinação, enquanto que no Brasil reagiram a eventos isolados. Concluímos também que os pro-vaxxers e anti-vaxxers dos dois países possuem muitos argumentos em comum. Observa-se uma maior objetividade nas argumentações pró/anti-vacinação nos EUA, alinhadas com as respectivas posições. Assim são a favor/contra a vacinação em geral, sua obrigatoriedade, seus efeitos, e as medidas de combate à COVID. Os anti-vaxxers defendem muito a liberdade individual, e propalam teorias conspiratórias. Já no Brasil, observou-se um componente político forte, enquanto nos EUA a polarização política não é relevante.

As séries temporais comparadas possuem diferenças. Devido à CPI da COVID no Brasil, e incapacidade de identificar os tweets com posicionamentos, não fomos capazes de comparar todo o espectro de eventos observados nos EUA, como reações às variantes (temor ou descrença na vacina) ou a doses de reforço. Ainda, a etapa de filtragem sobre o dataset dos EUA removeu problemas identificados no estudo anterior [de Sousa and Becker 2022], resultando em clusters de melhor qualidade.

Como trabalho futuros, esperamos, entre outros: (i) explorar a filtragem com o BERTopic para ampliar a análise do comportamento brasileiro a fim de abordar outros temas (e.g. doses de reforço) e estender a comparação com o comportamento nos EUA, (ii) desenvolver técnicas para facilitar a interpretação de grande número de tópicos, (iii) empregar a modelagem dinâmica de tópicos para aprofundar a distribuição dos tópicos e argumentos ao longo do tempo, e (iv) identificar e categorizar as teorias da conspiração.

Agradecimentos: Pesquisa financiada pela FAPERGS (19/2551-0001862-2).

Referências

- Angelov, D. (2020). Top2vec: Distributed representations of topics. *arXiv preprint arXiv:2008.09470*.
- Cascini, F., Pantovic, A., Al-Ajlouni, Y., Failla, G., and Ricciardi, W. (2021). Attitudes, acceptance and hesitancy among the general population worldwide to receive the COVID-19 vaccines and their contributing factors: A systematic review. *EClinicalMedicine*, 40.

- de Sousa, A. and Becker, K. (2022). Understanding the COVID vaccination stances in Brazil: a temporal analysis using twitter data. *Journal of Information and Data Management*. Accepted for publication.
- Ebeling, R., Córdova, C., Nobre, J. C., and Becker, K. (2022). Analysis of the influence of political polarization in the vaccination stance: the brazilian covid-19 scenario. In *Proc. of the 15th Intl. Conference on Web and Social Media (ICWSM)*.
- 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*, 101:107057.
- Grootendorst, M. (2020). Bertopic: Leveraging bert and c-tf-idf to create easily interpretable topics. <https://doi.org/10.5281/zenodo.4381785>.
- Hornsey, M. J., Harris, E. A., and Fielding, K. S. (2018). The psychological roots of anti-vaccination attitudes: A 24-nation investigation. *Health Psychology*, 37(4):307–315.
- Hotez, P. (2021). Covid vaccines: time to confront anti-vax aggression. *Nature*, 592(7856):661.
- Huangfu, L., Mo, Y., Zhang, P., Zeng, D. D., and He, S. (2022). Covid-19 vaccine tweets after vaccine rollout: Sentiment-based topic modeling. *J Med Internet Res*, 24(2):e31726.
- Liu, J., Nie, H., Li, S., Chen, X., Cao, H., Ren, J., Lee, I., and Xia, F. (2021). Tracing the pace of covid-19 research: Topic modeling and evolution. *Big Data Research*, 25:100236.
- Ma, P., Zeng-Treitler, Q., and Nelson, S. J. (2020). Use of two topic modeling methods to investigate covid vaccine hesitancy. In *Proc. of the 14th Intl. Conf. on ICT, Society, and Human Beings (ICT2021)*, volume 14, pages 130–140.
- Oliveira, F. B., Haque, A., Mougouei, D., Evans, S., Sichman, J. S., and Singh, M. P. (2022). Investigating the emotional response to covid-19 news on twitter: A topic modeling and emotion classification approaches. *IEEE Access*, 10:16883–16897.
- Omer, S. and *et alli* (2021). Promoting COVID-19 vaccine acceptance: recommendations from the Lancet Commission on Vaccine Refusal, Acceptance, and Demand in the USA. *The Lancet*, 398(10317):2186–2192.
- Solís Arce, J. and *et alli* (2021). COVID-19 vaccine acceptance and hesitancy in low- and middle-income countries. *Nature Medicine*, 27(8):1385–1394.
- Song, K., Tan, X., Qin, T., Lu, J., and Liu, T.-Y. (2020). Mpnet: Masked and permuted pre-training for language understanding. *arXiv preprint arXiv:2004.09297*.
- Tao, G., Miao, Y., and Ng, S. (2020). COVID-19 topic modeling and visualization. In *24th Intl. Conf. on Information Visualisation (IV)*, pages 734–739. IEEE.
- Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., and Zhu, T. (2020). Public discourse and sentiment during the covid 19 pandemic: Using latent dirichlet allocation for topic modeling on twitter. *PLOS ONE*, 15(9):1–12.