

Um Estudo sobre Espalhamento de Desinformação em Comunidades Antivacina no Telegram

Athus Cavalini

athus.cavalini@ifes.edu.br

Instituto Federal do Espírito Santo

Vitória, Espírito Santo, Brasil

Giovanni Comarela

gc@inf.ufes.br

Universidade Federal do Espírito Santo

Vitória, Espírito Santo, Brasil

ABSTRACT

Amid the biggest public health crisis of the century, the COVID-19 pandemic, social platforms amplified the spread of harmful narratives. In this context, this study explores how anti-vaccine communities operated on Telegram. By analyzing nearly 10 million messages from 779 channels and groups, and using machine learning to detect misinformation, harm potential, and toxicity, the results indicate recurring misinformation patterns, high engagement, and links between toxic content and virality. The study provides methodological contributions and insights for understanding and addressing informational disorder on messaging platforms.

KEYWORDS

misinformation, anti-vaccine, telegram, data science

1 INTRODUÇÃO

A plataforma da vida social transformou significativamente os fluxos de informação e os modos de interação entre indivíduos. Ao mesmo tempo em que representam um canal acessível e poderoso para informação, expressão, mobilização e participação pública, as plataformas sociais também se tornaram espaços de propagação de desinformação, discursos de ódio e narrativas extremistas. Esse fenômeno se intensificou durante a pandemia da COVID-19, uma das maiores crises de saúde pública do século, que foi acompanhada por uma *infodemia*, caracterizada por um alto volume de informações, muitas vezes falsas ou enganosas, circulando em alta velocidade [9].

Nesse contexto, o Telegram emergiu como uma plataforma central para a circulação de desinformação, principalmente por sua arquitetura que favorece o anonimato e a ausência de moderação [6]. Em particular, comunidades antivacina encontraram na plataforma um ambiente propício para estruturar redes de desinformação [3, 5], impactando nas políticas públicas de imunização e promovendo narrativas negacionistas. Frente a esse cenário, este trabalho propõe uma análise computacional sobre o funcionamento dessas comunidades no Telegram, buscando entender como a desinformação se espalha, quais padrões discursivos predominam e quais são os impactos em termos de engajamento e toxicidade dos conteúdos.

1.1 Objetivos

O objetivo central desta pesquisa é desenvolver metodologias para identificação e análise de desinformação em comunidades antivacina no Telegram. Entre os objetivos específicos, destacam-se: (i) a construção de um modelo de coleta e modelagem da rede de canais e grupos; (ii) a aplicação de técnicas de aprendizado de máquina para detecção de desinformação; e (iii) a análise das correlações entre desinformação, toxicidade e engajamento.

2 METODOLOGIA

A coleta e análise dos dados foram conduzidas com base em métodos da computação social e da ciência de dados.

Inicialmente, foi construída uma metodologia de amostragem de redes parcialmente observáveis baseada em *Snowball Sampling* enviesado por *Pagerank* [3], com o objetivo de mapear a comunidade antivacina presente no Telegram. A partir dessa estratégia, foram encontrados 779 canais e grupos com participação ativa em narrativas antivacina, das quais foram obtidas 9.904.482 mensagens.

A rede foi modelada como um grafo direcionado e ponderado, em que os nós representam os canais e grupos, enquanto as arestas correspondem ao fluxo de encaminhamento de mensagens entre eles. Para efeito de comparação, também foram coletadas e modeladas duas redes de *baseline*: uma comunidade de conversação comum e outra comunidade extremista [1, 4], permitindo análises comparativas entre os diferentes contextos.

Após a modelagem da rede, foi realizada uma etapa de rotulação manual de 1500 mensagens da rede antivacina com apoio de especialistas da área da saúde e comunicação social, utilizando o *framework* de desordem informacional proposto por [8].

Com esse conjunto rotulado, foram treinados e avaliados modelos de aprendizado de máquina baseados em LLMs [7] para realizar a classificação automática de mensagens a partir de dois eixos: inverdade e potencial de dano [2]. Adicionalmente, foi aplicado um modelo de detecção de toxicidade, além de analisadas métricas de engajamento, a fim de correlacionar os diferentes tipos de conteúdo ao seu desempenho na rede.

3 RESULTADOS

A análise da rede revelou uma comunidade antivacina altamente conectada e coesa, indicando que o Telegram favorece o espalhamento rápido de conteúdo. A Figura 1 ilustra a estrutura da comunidade antivacina, destacando o papel central de alguns canais como *hubs* de disseminação.

Na análise textual, os modelos de aprendizado de máquina demonstraram bom desempenho na classificação de mensagens quanto à inverdade e ao potencial de dano, alcançando F1-score = 0.83 para

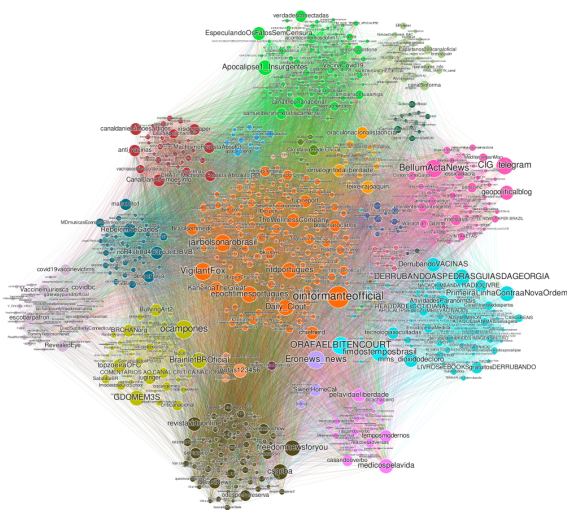


Figura 1: Comunidade antivacina no Telegram.

inverdade e 0.925 para dano, conforme detalhado na Figura 2. Os resultados indicam que uma parcela significativa dos conteúdos apresenta características típicas de desordem informacional, em que 63,49% foi considerada danosa e 53,75% como inverdade, (sendo 45,93% com ambos os rótulos).

Além disso, as análises revelaram a existência de padrões temáticos recorrentes, com destaque para narrativas que envolvem negacionismo científico, teorias conspiratórias sobre controle populacional e supostos efeitos colaterais extremos das vacinas. De maneira geral, mensagens contendo inverdade tiveram, em média, 76,5% mais encaminhamentos, enquanto as potencialmente danosas elevaram os encaminhamentos em cerca de 41,6%.

Outro achado importante refere-se à presença de toxicidade nos discursos analisados. Identificou-se que mensagens classificadas como desinformativas apresentaram maior índice de toxicidade do que o restante da base. Essa sobreposição entre discurso tóxico e desinformação sugere uma estratégia retórica para intensificar o apelo emocional e a capacidade de mobilização desses conteúdos. No entanto, no comparativo com as redes de baseline, a comunidade antivacina exibiu níveis médios de toxicidade inferiores aos da comunidade extremista, indicando que, nesse ecossistema, a viralização está mais associada à inverdade do que à toxicidade.

Esses resultados contribuem para o entendimento das estratégias discursivas em mensageiros, especialmente na comunidade antivacina, e serve de apoio à formulação de estratégias de comunicação e mecanismos de moderação mais adaptados a contextos específicos.

4 CONSIDERAÇÃO FINAIS

Este trabalho apresentou um estudo sobre o espalhamento de desinformação em comunidades antivacina no Telegram, combinando técnicas de ciência de dados, aprendizado de máquina e análise de redes complexas. A partir de uma base inédita de quase 10 milhões de mensagens, foram desenvolvidas metodologias para coleta, modelagem e análise de dados em redes parcialmente observáveis, com destaque para a construção de um classificador de desinformação baseado em LLMs.

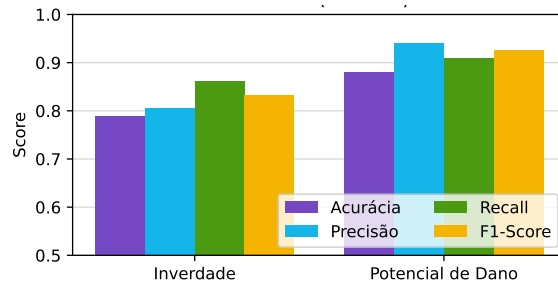


Figura 2: Desempenho do Classificador

Os resultados revelaram que, nesse ecossistema, a viralização é impulsionada principalmente por conteúdos desinformativos. A comparação com outras redes indicou que, ao contrário das comunidades extremistas, onde conteúdos tóxicos apresentam maior engajamento, a dinâmica antivacina é menos dependente desse fator, não havendo influência significativa da toxicidade nas taxas de compartilhamento. Conteúdos potencialmente danosos, por sua vez, atuam de forma distinta conforme o contexto, reduzindo o compartilhamento de conteúdos falsos e ampliando em mensagens verídicas. Esses achados reforçam a importância de compreender a dinâmica dessas redes, a fim de apoiar campanhas de comunicação estratégica e o desenvolvimento de ferramentas de monitoramento adaptadas a plataformas de baixa moderação.

Além da contribuição empírica, o trabalho apresenta um conjunto de ferramentas metodológicas e bases de dados que podem ser aproveitadas por pesquisadores em diferentes contextos e/ou plataformas. Espera-se que essas contribuições inspirem novas pesquisas, ampliando o entendimento sobre a desordem informacional e suas consequências para a saúde pública e para a democracia.

REFERÊNCIAS

- [1] Athus Cavalini, Thamya Donadia, and Giovanni Comarela. 2024. Characterizing the Toxicity of the Brazilian Extremist Communities on Telegram. In *Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'24)*.
- [2] Athus Cavalini, Thamya Donadia, Fabio Malini, and Giovanni Comarela. 2024. Detecting Misinformation on Telegram Anti-vaccine Communities. In *Anais do XXXVIII Simpósio Brasileiro de Bancos de Dados (SBB'D'24)*.
- [3] Athus Cavalini, Fabio Malini, Fabio Gouveia, and Giovanni Comarela. 2023. Politics and disinformation: Analyzing the use of Telegram's information disorder network in Brazil for political mobilization. *First Monday* 28, 5 (mai. 2023). <https://doi.org/10.5210/fm.v28i5.12901>
- [4] Victoria Ferro, Athus Cavalini, Fabio Goveia, and Fabio Malini. 2024. Insights About Radicalism on the Digital Era: A Sociological Approach. In *Proceedings of the 2024 International Conference on Advances in Social Networks Analysis and Mining (Calabria, Italy) (ASONAM'24)*. Springer.
- [5] Fabio Malini, Francis Sodre, Athus Cavalini, Gabriel Herkenhoff, and Fabio Gouveia. 2024. Five Patterns of Vaccine Misinformation on Telegram. In *Proceedings of the 2024 International Conference on Advances in Social Networks Analysis and Mining (Calabria, Italy) (ASONAM'24)*. Springer.
- [6] Richard Rogers. 2020. Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. *Media Studies* (2020). <https://doi.org/10.1177%2F0267323120922066>
- [7] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention Is All You Need. *CoRR abs/1706.03762* (2017). <http://arxiv.org/abs/1706.03762>
- [8] Claire Wardle and Hossein Derakhshan. 2017. Information disorder: Toward an interdisciplinary framework for research and policy making. <https://rm.coe.int/information-disorder-report-november-2017/1680764666>
- [9] World Health Organization. 2022. Health Topics: Infodemic. <https://www.who.int/health-topics/infodemic> Acesso em: 05 set. 2023..