

Samuel Assis Vieira

Análise de dados de conteúdos do TikTok relacionados ao TEA

Belo Horizonte, Minas Gerais

2023

Samuel Assis Vieira

Análise de dados de conteúdos do TikTok relacionados ao TEA

Relatório de Pesquisa de Análise Quanti-
tativa de Dados

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Orientador: Ana Paula Couto da Silva

Belo Horizonte, Minas Gerais
2023

Sumário

1	INTRODUÇÃO	3
1.1	Objetivos Gerais	4
1.2	Objetivos Específicos	4
2	REFERENCIAL TEÓRICO	5
3	METODOLOGIA	6
3.1	Tratamento e Limpeza dos dados	7
4	CARACTERIZAÇÃO E ANÁLISES	9
4.1	Vídeos	9
4.1.1	Taxa de Engajamento	10
4.1.2	Hashtags	12
4.2	Autores	14
4.3	Comentários	16
4.3.1	Linguistic Inquiry and Word Count (LIWC)	19
4.3.2	Análise de Sentimentos	21
4.3.3	BERTopic	23
5	CONSIDERAÇÕES FINAIS E TRABALHO FUTURO	26
	REFERÊNCIAS	27

1 Introdução

O TikTok, com milhões de usuários em todo o mundo, emergiu como um fenômeno significativo na esfera da internet. Fundado em 2017, o aplicativo chinês agora reúne mais de um bilhão de usuários ativos diariamente em todo o mundo, com cerca de 80 milhões desses usuários situados no Brasil (Revista Exame, 2023). Este trabalho analisa o papel do TikTok como plataforma de disseminação de informações sobre o Transtorno do Espectro Autista (TEA) e examina as implicações dessa disseminação para a compreensão pública e a percepção do TEA.

O TikTok, amplamente adotado por pessoas de todas as idades, concentra discussões sobre uma variedade de tópicos relevantes para a sociedade, frequentemente trazidos à tona pelos próprios usuários da plataforma. No entanto, muitas vezes, a abordagem desses tópicos carece do cuidado técnico e validação necessários por parte dos produtores de conteúdo. Grande parte dos conteúdos consiste em opiniões e experiências pessoais de indivíduos comuns, alinhando-se com o propósito fundamental da plataforma de criar conexões interpessoais.

Dentro desse ecossistema, destacam-se as discussões sobre questões técnicas e complexas da área da saúde, incluindo o Transtorno do Espectro Autista. Observa-se a existência de uma comunidade no TikTok dedicada à disseminação de informações e experiências relacionadas ao TEA, identificada pelo uso de hashtags como #autismo, #tea, #autism, entre outras. No entanto, essa abundância de informações, frequentemente desprovidas de validação por especialistas no campo da saúde, suscita preocupações. A disseminação de informações potencialmente imprecisas pode causar confusão entre aqueles que consomem esse conteúdo, levando a uma interpretação incorreta do TEA. Isso, por sua vez, pode fomentar a prática do autodiagnóstico com base em informações incorretas, ampliando a ilusão de compreensão profunda sobre o assunto.

1.1 Objetivos Gerais

Este estudo está inserido nas atividades desenvolvidas pela equipe de Educação e Difusão do Conhecimento do Centro de Inteligência Artificial para Saúde (CIA Saúde). O objetivo principal deste projeto é avaliar, compreender e mapear conteúdos relacionados ao Transtorno do Espectro Autista (TEA) presentes na plataforma TikTok, com ênfase na temática do autodiagnóstico. Para alcançar esse objetivo geral, foram realizadas coletas extensivas de vídeos e comentários relacionados ao TEA na referida plataforma, utilizando como ponto de partida a identificação de hashtags pertinentes ao tema. A partir dessa rica base de dados, o estudo visa conduzir uma análise abrangente e sistemática, com o propósito de identificar e compreender os padrões de comportamento e as dinâmicas subjacentes à disseminação de informações sobre o TEA no TikTok.

1.2 Objetivos Específicos

O objetivo do trabalho é entregar o suporte de análise e entendimento quantitativo e filtragem das informações coletadas utilizando técnicas de mineração de dados e análise de dados em conjunto com especialistas de domínio que irão classificar o conteúdo.

2 Referencial Teórico

O TikTok tem ganhado crescente destaque como objeto de estudo devido à sua significativa relevância e influência sociocultural. Vários trabalhos de pesquisa têm se dedicado à análise e compreensão dos conteúdos presentes nessa plataforma, com um foco particular na área da psicologia.

[Yeung, Ng e Abi-Jaoude 2022] et. al., por exemplo, empreenderam esforços para classificar conteúdos relacionados ao Transtorno de Déficit de Atenção e Hiperatividade (TDAH) no TikTok. Sua abordagem consistiu em avaliar a relevância informativa desses conteúdos, distinguir se foram produzidos por instituições médicas ou por usuários comuns, e categorizá-los como informativos, desinformação ou relatos de experiências pessoais. Os resultados revelaram que aproximadamente um terço dos conteúdos avaliados consistia em desinformação.

Seguindo uma abordagem semelhante, mas agora aplicada ao contexto do Transtorno do Espectro Autista (TEA),[Jensen e Borgen 2023], em sua tese de mestrado, obteve resultados consistentes, demonstrando que a maioria dos conteúdos relacionados ao TEA no TikTok também era considerada desinformação em uma proporção comparável.

Por outro lado,[Pineda et al. 2023] et. al., em um trabalho recente, validou a percepção da disseminação de autodiagnósticos por meio de comentários e conteúdo de vídeos na plataforma TikTok. O estudo envolveu a coleta de comentários em vídeos relacionados ao TDAH no TikTok, seguida da rotulação desses comentários de acordo com categorias específicas, que incluíam TDAH, Autodiagnóstico e Sintomas. Especialistas no domínio foram responsáveis por essa rotulação, e os dados resultantes foram utilizados para treinar um modelo de classificação multiclasse baseado em BERT. Os resultados do modelo indicaram que a combinação mais prevalente nos comentários coletados foi a tupla (TDAH, Autodiagnóstico), representando 18% dos comentários rotulados com essa categoria.

Esses estudos oferecem uma base teórica sólida para nossa pesquisa, que se concentra na análise e compreensão dos conteúdos relacionados ao TEA no TikTok, com especial atenção ao fenômeno do autodiagnóstico na plataforma.

3 Metodologia

Para atingir os objetivos propostos neste estudo, será adotada uma abordagem metodológica abrangente e estruturada, composta por várias etapas distintas. O fluxo de trabalho delineado a seguir descreve as principais fases da metodologia a serem seguidas:

1. **Entendimento e Caracterização dos Dados:** A primeira etapa consistirá em adquirir um entendimento completo dos dados coletados da plataforma TikTok. Isso envolverá a análise preliminar dos dados para compreender sua natureza, estrutura e conteúdo. Além disso, será realizada uma caracterização inicial da base de dados para identificar as variáveis-chave e seus atributos.
2. **Filtragem e Limpeza dos Dados:** Após o entendimento e caracterização dos dados, será iniciada a fase de filtragem e limpeza. Nessa etapa, serão aplicados critérios de seleção para reter apenas os dados que sejam relevantes e adequados para a análise. A limpeza dos dados envolverá a correção de erros, a padronização de formatos e a eliminação de informações redundantes ou inconsistentes.
3. **Desenvolvimento de Análises Quantitativas e Qualitativas:** Com a base de dados refinada e preparada, serão conduzidas análises quantitativas e qualitativas. As análises quantitativas visam a identificar tendências, padrões e métricas relacionadas ao Transtorno do Espectro Autista (TEA) no TikTok. Por outro lado, as análises qualitativas explorarão o conteúdo de maneira mais aprofundada, buscando insights qualitativos relevantes.
4. **Validação por Especialistas de Domínio:** As análises quantitativas e qualitativas serão submetidas à validação por especialistas no campo do TEA. Esses especialistas serão responsáveis por avaliar a validade e a interpretação das análises, garantindo que os resultados estejam alinhados com o conhecimento do domínio e sejam confiáveis.

3.1 Tratamento e Limpeza dos dados

Após a coleta baseada nas 13 hashtags definidas por especialistas da área da saúde tínhamos 3 bases principais:

Bases	Atributos	Instâncias
Vídeos	Autor, Duração, Música, Data de Publicação, Hashtags, Qnt. Interações, Qnt. Visualizações, Descrição	7.448
Autores	Nome de Usuário, Qnt. Seguidores, Qnt. Seguidos, Qnt. Vídeos	2.065
Comentários	Texto, Qnt. Curtidas, Qnt. Respostas, Fixado, Curtido pelo Autor, Linguagem	1.735.544

A análise dos dados coletados envolveu a aplicação de diversas tratativas para garantir a qualidade e relevância das informações. Inicialmente, a base de vídeos continha 7.448 registros provenientes de 13 *hashtags* relacionadas ao Transtorno do Espectro Autista (TEA). Contudo, identificou-se que a similaridade entre o português e o espanhol resultou em uma presença significativa de vídeos de criadores de língua espanhola.

Além dessa questão linguística, observou-se uma particularidade em relação à *hashtag* *#2deabril*, que remete ao Dia Mundial da Conscientização do Autismo, instituído pela ONU em 2007. No entanto, essa *hashtag* também era utilizada em vídeos sobre a Guerra das Malvinas, que iniciou em 2 de abril de 1982. Para resolver essa ambiguidade, aplicou-se um filtro específico, analisando as *hashtags* utilizadas e a descrição dos vídeos. Dessa forma, foram removidos os vídeos que mencionavam o termo 'malvina' e que não continham trechos relacionados ao TEA.

Outra etapa do tratamento de dados foi a análise das *hashtags* contendo o termo 'tea'. Um filtro foi aplicado para excluir vídeos não relacionados ao autismo, mas sim a chá. Similarmente ao filtro anterior, buscou-se termos específicos relacionados ao autismo, e, por fim, criadores de língua espanhola foram excluídos da base.

Durante o desenvolvimento do trabalho, procedeu-se à visualização de alguns vídeos, permitindo a anotação dos *nicknames* de alguns autores de língua estrangeira. Esses autores foram posteriormente removidos da base de dados.

Para os comentários, o único critério de filtragem foi a linguagem, sendo selecionados apenas os comentários em língua portuguesa, identificados com a

abreviação 'pt'. Em relação aos autores, realizou-se um cruzamento com a base de vídeos, considerando apenas aqueles com pelo menos um vídeo na base.

Ao término desse processo, a base de dados resultante contou com 6.750 vídeos, 1.700 autores e 931.133 comentários.

4 Caracterização e Análises

A coleta de dados foi conduzida com base em 13 hashtags principais, cuidadosamente selecionadas por especialistas da área da saúde. As hashtags e sua representatividade em termos quantitativos na base de dados podem ser visualizadas na Tabela 1.

Quantidade de videos por # de coleta

challenge	Quantidade
autismoinfantil	867
transtornoespectroautista	813
autismobrasil	803
autismobr	778
autista	748
espectroautista	668
souautista	591
maeutista	442
2deabril	374
transtornoespectroautista	348
paiautista	174
teabrasil	134
teabr	10

Figura 1 – Tabela de Vídeos por Hashtag.

Todas as hashtags consideradas apresentam uma contagem mínima de centenas de vídeos, exceto a #teabrasil, que é menos frequente entre os criadores de conteúdo dedicados ao Transtorno do Espectro Autista (TEA). Ao excluir a #teabrasil, a média de vídeos por hashtag na coleta é de aproximadamente 561.

4.1 Vídeos

Para a base de vídeos, foi crucial investigar as distribuições fundamentais dos atributos numéricos, destacando-se as interações (curtidas, comentários e compartilhamentos) e as visualizações por vídeo.

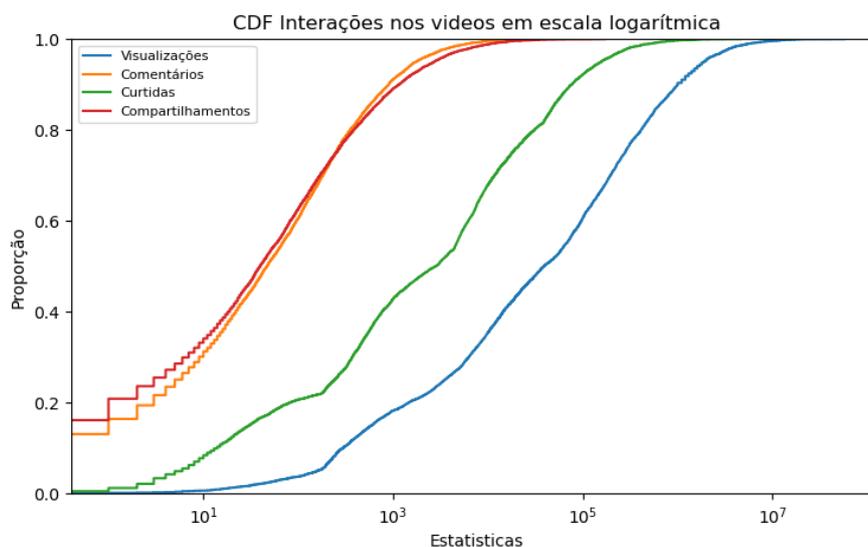


Figura 2 – CDF Interações nos Vídeos.

Ao analisar a função cumulativa de distribuição (CDF) das interações, conforme representado na Figura 2, em uma escala logarítmica, torna-se evidente que a quantidade de visualizações se destaca notavelmente das demais, compreendendo aproximadamente 40% dos vídeos na faixa das dezenas de milhares (10^5). Em seguida, o número de curtidas também apresenta valores expressivos para os vídeos, indicando interações frequentes, enquanto comentários e compartilhamentos são menos comuns. Essas interações demandam um esforço mais significativo por parte do usuário, refletindo um comportamento mais ativo na interação com o conteúdo e a plataforma. Isso é evidenciado pelos valores menores de quantidade de comentários, com uma mediana situada entre dezenas (10^1) e centenas (10^2), e menos de 20% da base de dados apresenta mais do que centenas de comentários.

4.1.1 Taxa de Engajamento

Além das interações, uma métrica de extrema importância, amplamente utilizada como indicador de sucesso para uma conta ou conteúdo, é a taxa de engajamento. Essa métrica representa uma simples proporção entre a quantidade de interações e o número de visualizações, conforme expresso pela seguinte fórmula:

$$\text{Taxa de Engajamento} = \frac{\text{Número de Interações}}{\text{Número de Visualizações}} \times 100$$

De acordo com blogs especializados em mídias sociais, o valor médio considerado adequado para uma taxa de engajamento é de 10% para um vídeo ou conta. Quanto maior essa taxa, maior a probabilidade de a plataforma promover o conteúdo. Contudo, a distribuição e as recomendações exatas dessas plataformas ainda permanecem obscuros.

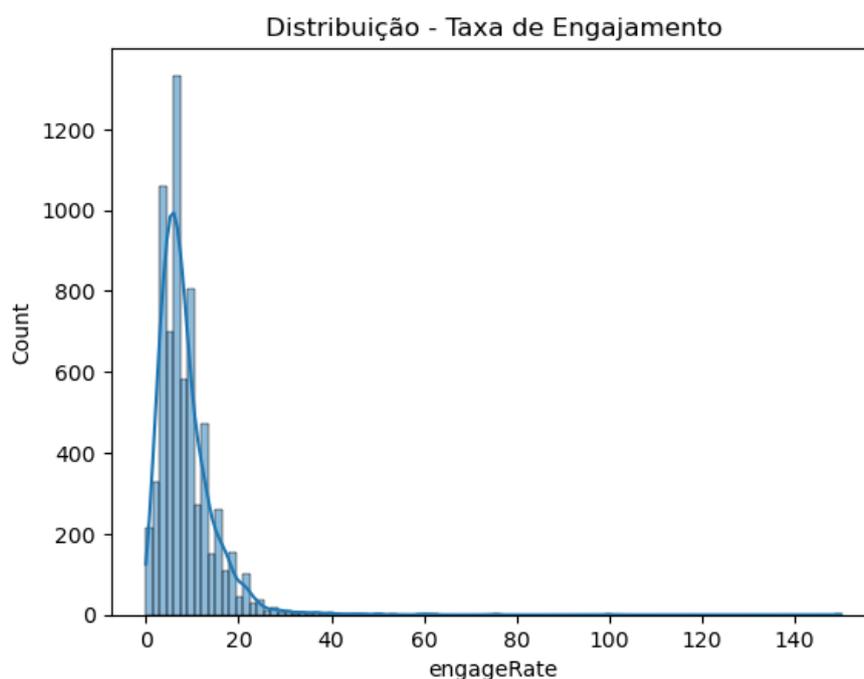


Figura 3 – Histograma e KDE da Taxa de Engajamento.

Ao examinar a distribuição, é perceptível que os valores tendem a se concentrar em torno da média esperada de 10%. Para a base de dados, a média da taxa de engajamento dos vídeos é de 8,4, com um desvio padrão de 6,2.

A maioria dos vídeos apresenta taxas de engajamento no intervalo de [0%, 20%]. No futuro, a análise dos vídeos mais engajados pode ser de interesse para identificar elementos de viralização. Outra abordagem possível consiste em comparar essa métrica entre vídeos classificados como informativos e os de desinformação. Ao segmentar essa métrica com base nas hashtags da coleta, conforme ilustrado na

Figura 4, os vídeos não apresentaram diferenças significativas. A exceção ocorre nos vídeos da hashtag #autista, que atingiram uma mediana de engajamento ligeiramente superior aos demais, possivelmente devido à quantidade de comentários e curtidas entre os vídeos coletados dessa hashtag.

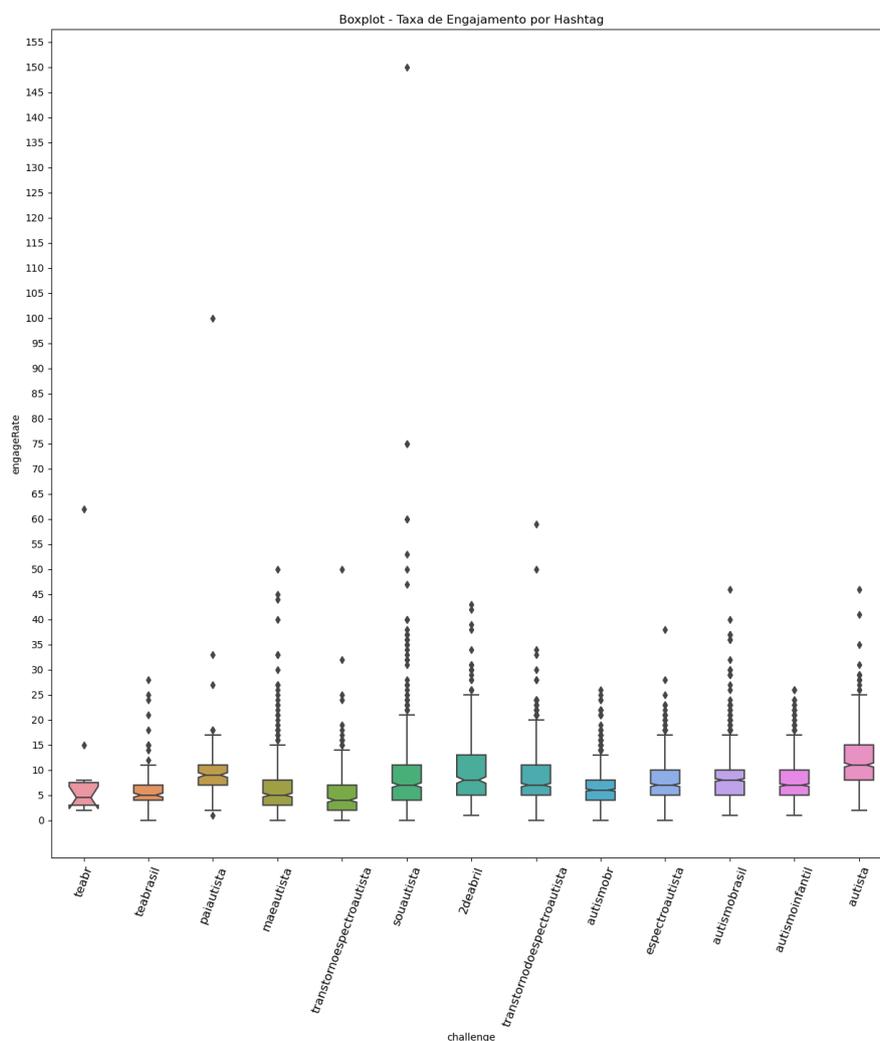


Figura 4 – Boxplot da Taxa de Engajamento por hashtag de coleta.

4.1.2 Hashtags

Para cada vídeo, há sempre um conjunto de hashtags escolhido pelos autores para indexar os conteúdos, servindo como palavras-chave na recomendação e

4.2 Autores

Esta seção dedica-se à exploração da base de autores, os criadores de conteúdos responsáveis pela produção dos vídeos. Ao todo, temos 1700 criadores, sendo que a grande maioria, 75%, contribui com até 2 vídeos na base. A distribuição da quantidade de vídeos por autor pode ser visualizada na Figura 6.



Figura 6 – CDF Quantidade de Video por Autores.

Apenas 20% dos autores possuem dezenas de vídeos, enquanto a parcela que atinge a casa das centenas representa um percentual ainda menor, cerca de 0,5%. O tamanho das contas, ou seja, a quantidade de seguidores e curtidas que essas contas possuem, pode ser considerado como uma medida de influência, visto que teoricamente alcançarão mais pessoas.

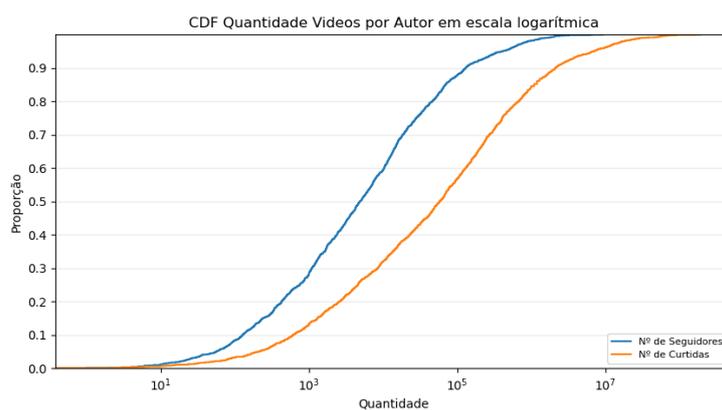


Figura 7 – CDF Quantidade de Video por Autores.

referente ao nome da filha de Shirley, da conta *@shirleybolzan*, que é a conta com a maior quantidade de vídeos.



(a) #autista



(b) #teabr



(c) #espectroautista



(d) #transtornoespectroautista

Figura 10 – 200 termos mais utilizados por hashtag de coleta

Além disso, é interessante analisar como se distribui a quantidade de comentários, considerando o recorte por hashtags de coleta. Na Figura 11, é possível observar que, para a grande maioria das hashtags, metade dos vídeos apresentam menos de 20 comentários. Com exceção das hashtags #autista, #autismobrasil, #autismoinfantil e #espectroautista, que exibem uma porcentagem maior de vídeos com centenas de comentários. Destaca-se a #autista, com cerca de 20% dos vídeos registrando mais de cem comentários.

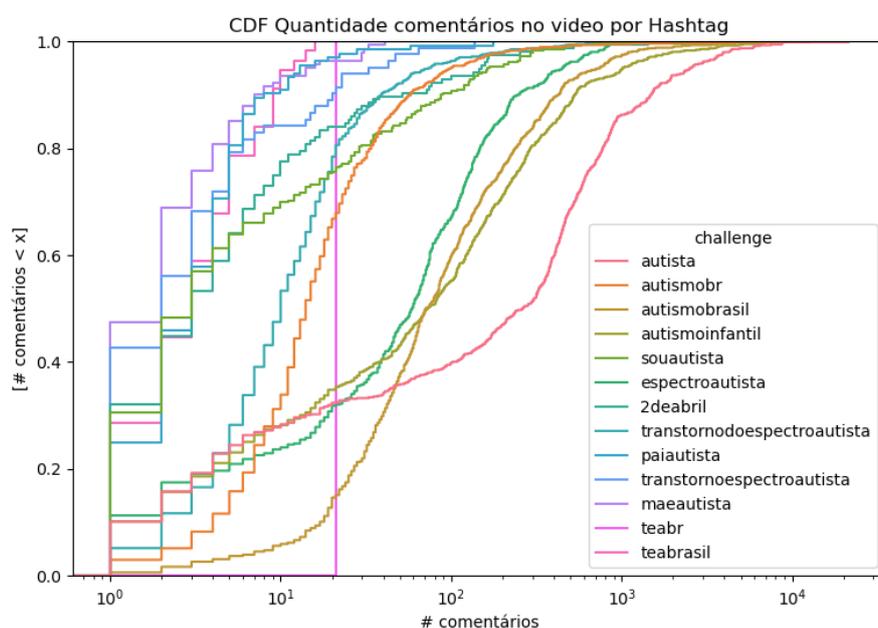


Figura 11 – CDF Quantidade de Comentários no video por hashtag.

Em seguida, foram realizados recortes na base para focar a análise, considerando as limitações computacionais diante dos 930 mil comentários. O primeiro recorte consistiu em capturar os comentários dos 100 vídeos mais comentados da base, resultando em uma base de 300 mil comentários.

Ao observarmos as distribuições de interações para esse grupo de vídeos na Figura 12, é interessante notar como a média de comentários aumenta substancialmente, atingindo 2700 comentários, distribuindo-se entre 1000 e 7500 comentários por vídeo. As visualizações desses vídeos, por sua vez, passam a variar na casa dos milhões (10^6).

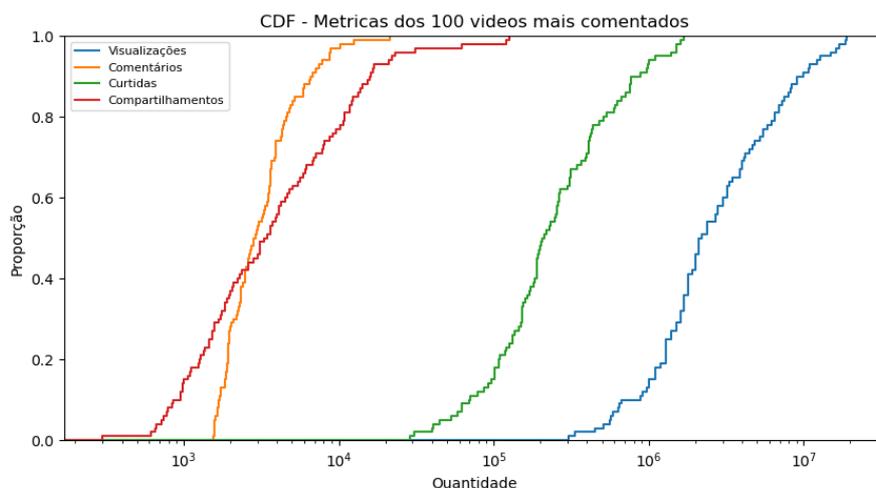


Figura 12 – CDF Interações nos top 100 mais comentados.

Ao observarmos a distribuição dos vídeos do top 100 em relação às hashtags de coleta, conforme apresentado na Tabela 1, podemos corroborar a análise anterior, constatando que os vídeos coletados da #autista são os mais frequentes no top 100.

Tabela 1 – Tabela de Hashtags para Top 100

Hashtag de Coleta	Quantidade de Vídeos
autista	62
autismoinfantil	22
autismobrasil	11
autismobr	3
espectroautista	1
souautista	1

4.3.1 Linguistic Inquiry and Word Count (LIWC)

O Linguistic Inquiry and Word Count (LIWC) é uma ferramenta de análise de texto desenvolvida para categorizar textos com base na linguagem utilizada nos mesmos. Criado por [Pennebaker, Booth e Francis 2007], o LIWC tem como objetivo capturar estados emocionais e características psicológicas do autor do texto com base nas palavras utilizadas na sua construção.

Para este trabalho, utilizou-se o dicionário LIWC de 2007, o qual foi traduzido para o português, conforme descrito em [LIWC - Linguistic Inquiry and Word Count].

Os resultados do LIWC podem ser divididos em duas frentes: uma contagem léxica das palavras no texto e outra focada nos processos linguísticos representados no texto. O nosso foco de análise estará na segunda parte.

Para viabilizar a execução da ferramenta em tempo hábil, mas ainda em uma quantidade significativa de comentários, foi necessário realizar mais um recorte na base. Os 500 comentários mais curtidos de cada um dos vídeos no top 100 mais comentados foram filtrados, totalizando 50 mil comentários. Essa seleção foi utilizada como base para a aplicação do LIWC, após um tratamento básico que incluiu a remoção de *stopwords* e emojis.

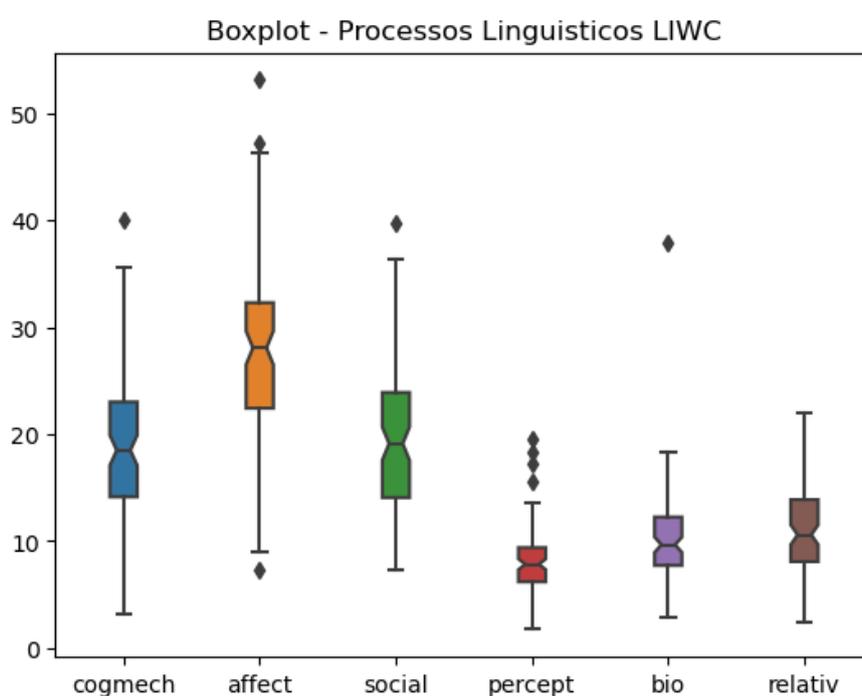


Figura 13 – Processos Linguísticos LIWC

Analisando todos os processos linguísticos apontados pela ferramenta, conforme ilustrado na Figura 13, torna-se evidente a predominância do processo afetivo, com a maior variação e mediana. Esses processos estão relacionados às emoções expressas no texto. Em seguida, o processo social se destaca, relacionado a palavras que expressam interações sociais e relacionamentos interpessoais. Por fim, outro processo de relevância é o Cognitivo-Mecânico (cogmech), que examina como o

autor realiza suas análises mentais, destacando palavras relacionadas a pensamento lógico, análise crítica e processos mentais.

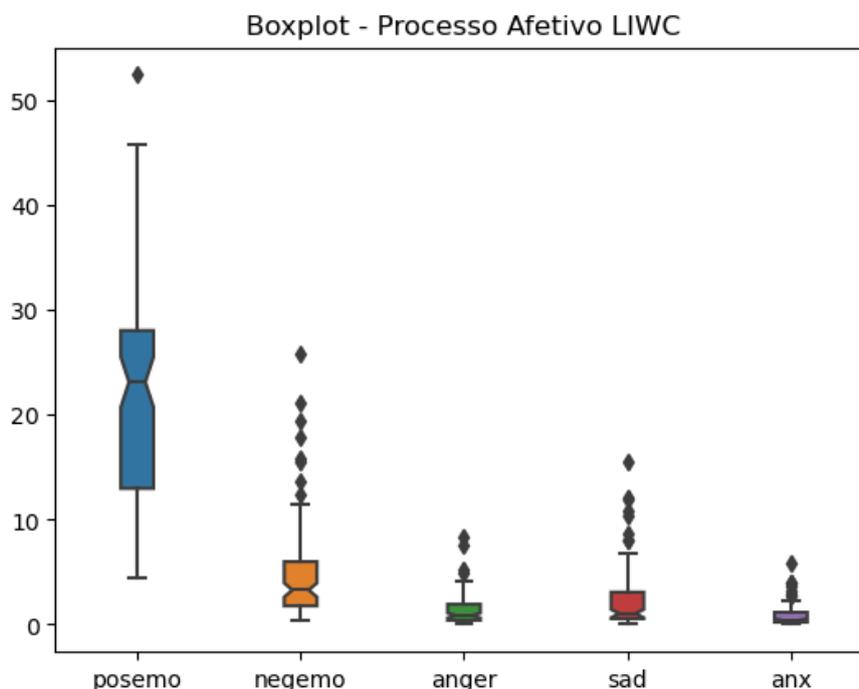


Figura 14 – Processo Afetivo LIWC.

Na Figura 14, aprofundamos nos componentes do processo afetivo, e é possível perceber como as palavras relacionadas a emoções positivas (*posemo*) são as mais predominantes. Em contrapartida, as palavras que expressam emoções negativas (*negemo*) são consideravelmente inferiores. As outras emoções, como raiva (*anger*), tristeza (*sad*) e ansiedade (*anx*), são muito próximas de zero.

4.3.2 Análise de Sentimentos

Para a execução da análise de sentimentos, utilizou-se a biblioteca VADER [Hutto e Gilbert 2014], cuja adaptação para a língua portuguesa foi realizada por [Almeida 2018]. Essa biblioteca classifica o texto em três variáveis representando a emoção: positiva, neutra e negativa, de acordo com a intensidade. Além disso, traz a variável *compound*, que representa o valor normalizado da emoção, variando entre -1 (extremamente negativo) e 1 (extremamente positivo).

A base de dados utilizada compreendeu os 500 comentários mais curtidos dos 100 vídeos mais comentados.

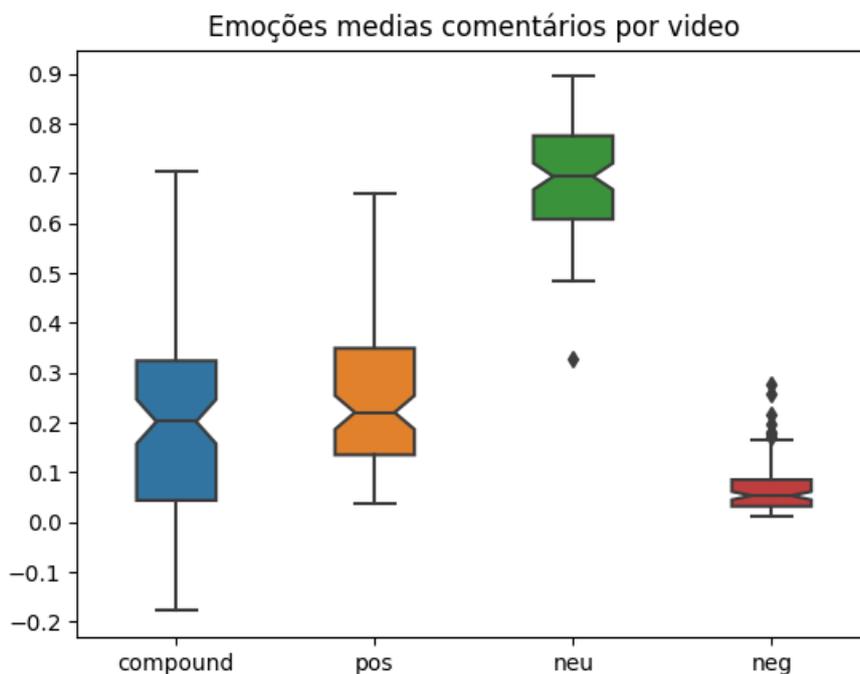


Figura 15 – Análise de Sentimento - Média por video

Na Figura 15, apresenta-se a distribuição média dos comentários para cada um dos 100 vídeos. A partir disso, podemos confirmar as impressões anteriores, observando a predominância de comentários neutros e positivos. Os comentários negativos têm uma presença menor, com uma variação significativamente menor.

A variável *compound* pode ser utilizada para descrever o sentimento predominante no texto. Valores acima de 0.05 configuram um sentimento positivo, entre 0.05 e -0.05 indicam um sentimento neutro, e abaixo de -0.05 representam um sentimento negativo.

Portanto, podemos extrair que, em média, a emoção expressa nos principais comentários dos vídeos mais comentados é positiva.

4.3.3 BERTopic

BERTopic é uma biblioteca desenvolvida por [Grootendorst 2022] para a clusterização de texto em tópicos utilizando embeddings BERT (Bidirectional Encoder Representations from Transformers) pré-treinados. A abordagem empregada é não supervisionada, visando a clusterização eficiente do texto. A base de dados utilizada compreendeu os top 500 comentários dos vídeos mais comentados. Na fase de preparação, além das etapas básicas de remoção de stopwords, pontuação e sujeiras do texto, realizou-se a lematização dos comentários.

Em seguida, foram calculados os embeddings, e o BERTopic foi executado com uma limitação de 10 tópicos, focando a análise. Em uma execução sem restrições, foram gerados 1076 tópicos distintos. A Figura 16 apresenta a distribuição entre os 10 tópicos.

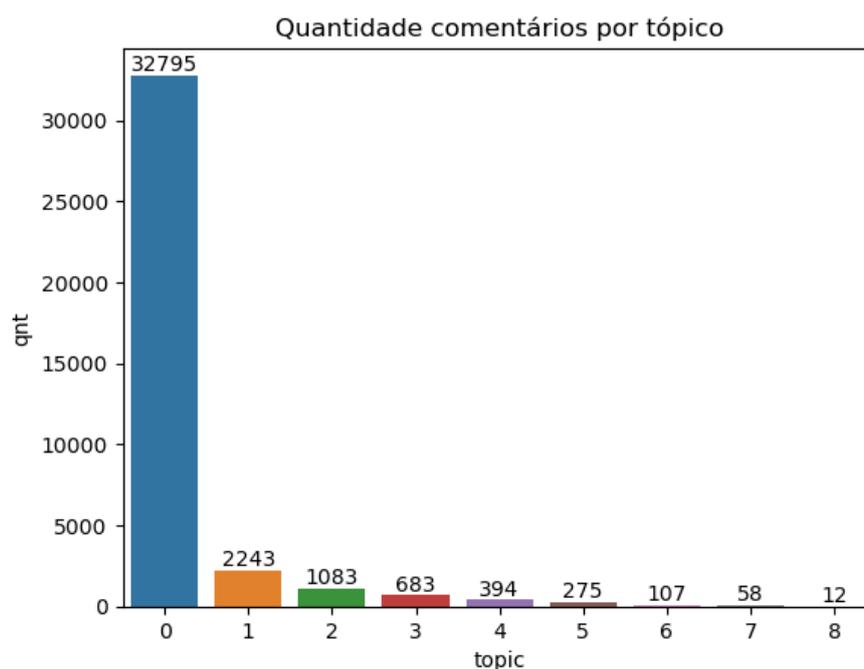


Figura 16 – Quantidade de Comentários por Tópico

Ao analisar o gráfico, torna-se evidente que a concentração principal dos comentários reside no tópico 0, contendo mais de 100 vezes o número de comentários do que o segundo tópico mais representado.

Observando a Figura 17, que ilustra a dinâmica no espaço bidimensional para os tópicos, destaca-se a proximidade dos tópicos 1, 3 e 6 com o tópico 0 (pintado de vermelho) no terceiro quadrante. Isso sugere que esses tópicos estão mais relacionados e apresentam proximidade semântica. Por outro lado, os outros tópicos distribuem-se de maneira mais esparsa; os tópicos menores 7, 2 e 4 reúnem-se um pouco mais próximos ao centro do gráfico, mas cada um em um quadrante distinto. Por fim, os tópicos 8 e 5 se distanciam mais, com o tópico 5 posicionado mais à direita do gráfico no primeiro quadrante.

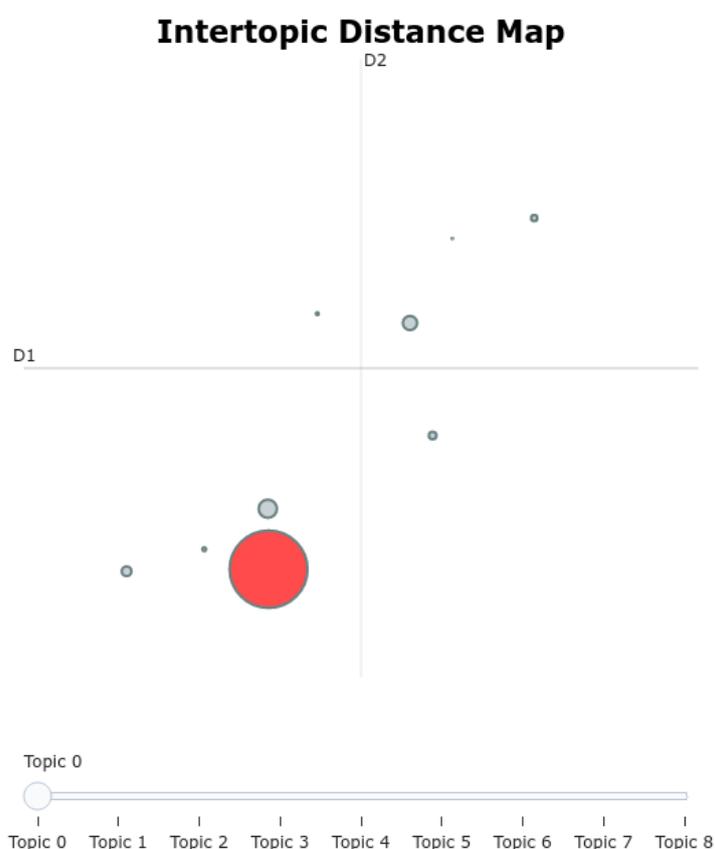


Figura 17 – Disposição espacial dos tópicos

É relevante compreender as palavras-chave de cada um dos tópicos. A Figura 18 fornece uma representação visual dessas palavras-chave. A partir dessa análise, podemos concluir que o tópico 0 engloba predominantemente comentários de cunho religioso ou elogios. O tópico 1 concentra-se em comentários diretamente

relacionados ao autismo, abordando relações e ideias sobre o tema. O tópico 2 parece estar centrado em comentários que discutem o vídeo como conteúdo. Por sua vez, o tópico 3 aborda hábitos e restrições alimentares associadas a pessoas autistas, um tema recorrente no meio. Finalmente, os tópicos 4 e 5 parecem ser mais específicos, relacionados a um ou dois vídeos sobre temas particulares envolvendo pessoas autistas. O tópico 4 aborda assuntos relacionados a denúncias ou reportagens, enquanto o tópico 5 reúne comentários sobre viagens de avião e reserva de assentos.

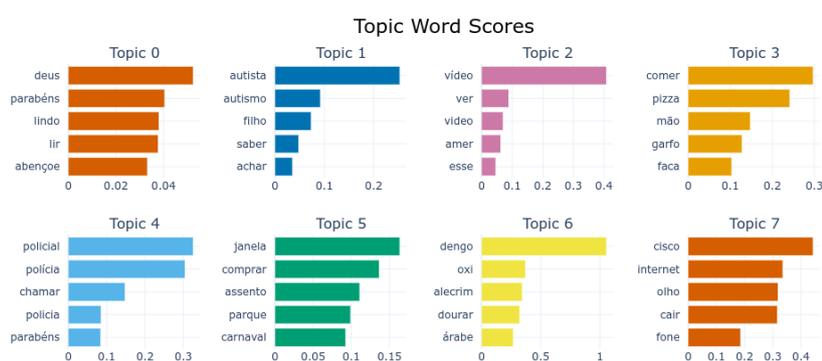


Figura 18 – Palavras mais frequentes por tópico

5 Considerações Finais e Trabalho Futuro

Ao finalizar esta análise abrangente da base de dados de conteúdos em vídeos relacionados ao Transtorno do Espectro Autista (TEA), emergem conclusões significativas. A alta concentração de emoções positivas e neutras reflete uma profunda empatia dentro da comunidade analisada. Observa-se, preliminarmente, uma diversidade notável nos tópicos de discussão, indicando que as pessoas autistas não se limitam a falar apenas sobre o autismo, evidenciando a importância de dar voz a essa comunidade. Destaca-se a presença de contas influentes que compartilham experiências relacionadas ao transtorno, aproveitando o espaço online para proporcionar visibilidade a um público historicamente sub-representado.

Com o avanço inicial no entendimento do conteúdo veiculado, os próximos passos envolverão uma colaboração estreita com especialistas em saúde para classificar esses conteúdos. Essa abordagem permitirá aprofundar a análise em um conjunto menor de vídeos, compreendendo padrões em vídeos considerados desinformação e realizando uma análise de tópicos mais refinada. Além disso, será possível explorar mais profundamente os comentários para compreender como os usuários percebem esses conteúdos, bem como realizar comparações entre diferentes classes de conteúdos. Essa abordagem contribuirá de maneira significativa para o combate à desinformação e promoverá um ambiente mais informado e inclusivo.

Os resultados preliminares desta análise fornecem uma base sólida para futuras investigações, consolidando a importância de uma compreensão aprofundada do conteúdo online relacionado ao TEA e seu impacto na comunidade.

Referências

- [Almeida 2018]ALMEIDA, R. J. A. *LeIA - Léxico para Inferência Adaptada*. [S.l.]: GitHub, 2018. <https://github.com/rafjaa/LeIA>.
- [Alper et al. 2023]ALPER, M. et al. Tiktok as algorithmically mediated biographical illumination: Autism, self-discovery, and platformed diagnosis on# autistok. *New Media & Society*, SAGE Publications Sage UK: London, England, p. 14614448231193091, 2023.
- [Grootendorst 2022]GROOTENDORST, M. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*, 2022.
- [Hutto e Gilbert 2014]HUTTO, C.; GILBERT, E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: *Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. Ann Arbor, MI: [s.n.], 2014.
- [Jensen e Borgen 2023]JENSEN, I.; BORGEM, V. H. *# Autism: A Cross-Sectional Study of the Quality of Diagnostic Information on TikTok*. Dissertação (B.S. thesis) — uis, 2023.
- [LIWC - Linguistic Inquiry and Word Count]LIWC - Linguistic Inquiry and Word Count. Acesso em: 4 de dezembro de 2023. Disponível em: <<http://143.107.183.175:21380/portlex/index.php/pt/projetos/liwc>>.
- [Pennebaker, Booth e Francis 2007]PENNEBAKER, J. W.; BOOTH, R. J.; FRANCIS, M. E. Liwc2007: Linguistic inquiry and word count. *Austin, Texas: liwc. net*, 2007.
- [Pineda et al. 2023]PINEDA, K. et al. Using bert to understand tiktok users' adhd discussion. In: *2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science (IRI)*. [S.l.: s.n.], 2023. p. 213–214.
- [Singhal, Kumar e Sharma 2021]SINGHAL, K.; KUMAR, N.; SHARMA, V. Analysis of online media content using data science techniques. In: *2021 3rd International*

Conference on Advances in Computing, Communication Control and Networking (ICAC3N). [S.l.: s.n.], 2021. p. 55–59.

[Yeung, Ng e Abi-Jaoude 2022]YEUNG, A.; NG, E.; ABI-JAOUDE, E. Tiktok and attention-deficit/hyperactivity disorder: A cross-sectional study of social media content quality. *The Canadian Journal of Psychiatry*, v. 67, n. 12, p. 899–906, 2022. PMID: 35196157. Disponível em: <<https://doi.org/10.1177/07067437221082854>>.