

# A Misoginia no YouTube Brasileiro: Um Estudo de Caso sobre o Conteúdo Produzido pela Comunidade Red Pill

Victor Martins<sup>1</sup>, Sophia Eduarda V. Serafim<sup>1</sup>, Livia Caroline R. Pereira<sup>1</sup>,  
Amanda F. Alves<sup>1</sup>, Carlos H. G. Ferreira<sup>2</sup>, Jussara M. Almeida<sup>1</sup>  
victor.martins@dcc.ufmg.br, sophiaevieira@dcc.ufmg.br, liviacarolinerp@dcc.ufmg.br,  
amandafernandes07@ufmg.br, chgferreira@ufop.edu.br, jussara@dcc.ufmg.br

<sup>1</sup>Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, Brasil

<sup>2</sup>Universidade Federal de Ouro Preto (UFOP), João Monlevade, Brasil

## ABSTRACT

YouTube, as one of the most influential video platforms in Brazil, has become a central space for the circulation of ideas and, consequently, a fertile ground for the spread of hate speech, particularly misogyny, within the Manosphere ecosystem. The growing visibility and influence of content creators highlight the urgency of understanding how their messages are constructed and received. While previous studies have examined online misogyny in other platforms and contexts, little is known about its large-scale manifestations in the Brazilian YouTube ecosystem, especially regarding the practices of particular communities. This study addresses this gap by examining the presence and activity of the Red Pill community on YouTube Brazil, aiming to characterize how misogynistic discourse is constructed, interpreted, and propagated. We analyzed 18,034 videos and over 2.2 million comments from 28 channels published between January 2024 and June 2025, applying state-of-the-art natural language processing techniques to investigate engagement patterns and discursive dynamics. Our findings reveal that this community reaches a wide audience through diverse content strategies shaped by creators themselves. While titles and descriptions predominantly convey neutral sentiment, possibly to maximize reach and avoid automated moderation, user comments display a pronounced prevalence of negative sentiment and elevated toxicity. These results underscore the role of content creators in exploring YouTube to amplify extremist narratives and contribute to the understanding of online misogyny in Brazil by shedding light on its mechanisms of construction, reception, and dissemination.

## KEYWORDS

Red Pill, YouTube, Hate Speech, NLP, Misogyny

## 1 INTRODUÇÃO

A popularização das redes sociais revolucionou as formas de socialização e de acesso à informação, consolidando esses ambientes como espaços centrais para a circulação de ideias que influenciam dinâmicas sociais contemporâneas [40]. Porém, esse cenário de ampla conectividade também favoreceu a disseminação de discursos de ódio, que encontraram na esfera digital um terreno fértil para fortalecimento, radicalização e expansão [30, 45]. É nesse contexto que a presente pesquisa se aprofunda, especificamente em uma

das manifestações de discurso de ódio, a misoginia, articulada no ecossistema conhecido como Manosfera (ou Machosfera).

A Manosfera é uma rede composta por diversas comunidades virtualmente interconectadas que promovem ideologias misóginas, funcionando como espaço de reforço ideológico e acolhimento para indivíduos simpatizantes [26, 34]. Diferentes subculturas coexistem neste ecossistema, como os Celibatários Involuntários (Incels), *Men Going Their Own Way (MGTOW)* e Red Pill, com diferenças sutis nas narrativas adotadas<sup>1</sup>. Dentre estas, a comunidade Red Pill defende que os homens são vítimas de ódio e injustiças por parte das mulheres, e incentiva seus seguidores a adotarem atitudes de “machos alfa”, a se desenvolverem pessoalmente para se tornarem “homens destacados” ou “de alto valor”. As ideias sexistas promovidas, muitas vezes, justificam o abuso e a violência contra as mulheres<sup>2</sup>. Enquanto alguns trabalhos colocam a filosofia Red Pill como central à Manosfera [21], outros apontam que esta comunidade age como portões de entrada para ideologias conspiratórias, reforçando narrativas anti-gênero, anticiência e antidemocráticas<sup>3</sup>. De fato, relatórios recentes [11, 43] apontam que canais brasileiros com conteúdo misógino — muitos alinhados à ideologia Red Pill — já alcançam até 229 mil inscritos e mais de 105 mil vídeos no YouTube, com média de 612 mil visualizações cada, evidenciando seu protagonismo discursivo e alcance para além de nichos extremistas.

A Manosfera vem expandindo, conquistando novos adeptos em todo o mundo e motivando uma série de estudos focados em misoginia no geral ou em subculturas específicas e suas manifestações em plataformas digitais como o Reddit [17, 24, 26], o Telegram [12] e o YouTube [36]. A análise no contexto brasileiro, por sua vez, é especialmente importante considerando a posição de destaque do Brasil entre os países com maiores índices de violência contra as mulheres<sup>4</sup>. Além disso, a presença da misoginia no ambiente digital brasileiro tem-se intensificado de forma significativa, refletida tanto pelo crescimento expressivo nas denúncias<sup>5</sup> quanto pela ampla circulação de conteúdos misóginos em plataformas sociais [43]. Apesar deste crescimento, poucos são os estudos que buscam caracterizar a misoginia digital no contexto brasileiro [11–13, 22, 42, 43].

Neste contexto, este trabalho visa caracterizar a Manosfera brasileira, com foco na atuação de uma de suas comunidades mais

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2025). Rio de Janeiro, Brazil. Porto Alegre: Brazilian Computer Society, 2025.

© 2025 SBC – Brazilian Computing Society.

ISSN 2966-2753

<sup>1</sup><https://catarinass.info/red-pill-incele-sigma-conheca-subculturas-da-machosfera/>

<sup>2</sup><https://jornal.usp.br/atualidades/discursos-misoginos-nas-redes-sociais-geram-retrocesso-nas-conquistas-feministas/>

<sup>3</sup><https://jornal.usp.br/artigos/como-comunidades-redpill-e-anti-woke-capturam-jovens-para-redes-de-odio/>

<sup>4</sup><https://worldpopulationreview.com/country-rankings/femicide-rates-by-country>

<sup>5</sup><https://www1.folha.uol.com.br/cotidiano/2024/01/denuncias-de-misoginia-na-internet-cresceram-quase-30-vezes-em-cinco-anos-no-brasil.shtml>

centrais, a Red Pill, no YouTube. A escolha da plataforma se justifica por sua recorrente identificação como um ambiente propício à expansão de movimentos conspiratórios e de disseminação de discursos de ódio [6, 19, 31, 35], incluindo conteúdos misóginos [22, 36, 42, 43]. Para isso, coletamos dados públicos de 18.034 vídeos publicados por 28 canais entre janeiro de 2024 e junho de 2025, além de mais de 2,2 milhões de comentários. Para cada vídeo e comentário, foram extraídos a data de publicação, os números de curtidas, visualizações, a duração e o conteúdo textual, permitindo análises de engajamento, recepção e estrutura discursiva.

Nossas análises abrangem três eixos principais: (1) a evolução temporal dos canais, suas estratégias de atração de audiência e os níveis de engajamento gerados; (2) as características textuais dos metadados dos vídeos (especialmente títulos e descrições); e (3) as características textuais dos comentários recebidos. As análises de conteúdo textual (itens 2 e 3) utilizam técnicas estado-da-arte em Processamento de Linguagem Natural (PLN), incluindo: análise de sentimentos com XLM-T [1], análise psicolinguística com LIWC [48] e modelagem de tópicos com BERTopic [10]. Além disso, para identificar a presença de conteúdo misógeno, utilizamos a ferramenta Detoxify combinada à anotação manual de três usuárias da plataforma, com o objetivo de capturar nuances do discurso ofensivo que ferramentas automatizadas isoladamente podem não detectar. Em síntese, oferecemos uma análise em larga escala da atuação da comunidade Red Pill no YouTube Brasil, contribuindo para o entendimento sobre como seu discurso é construído, recebido e propagado, revelando os padrões de engajamento e as dinâmicas de discurso que sustentam esse ecossistema. Em comparação aos recentes relatórios sobre misoginia no YouTube brasileiro [22, 42, 43], nosso estudo apresenta escopo muito mais amplo, ao incorporar análises inéditas que não se limitam aos vídeos, mas abrangem também todo o ecossistema que os circunda, em especial os comentários publicados por usuários, fundamentais para compreender a recepção, ressonância e amplificação do discurso.

A seguir, a Seção 2 discute trabalhos relacionados, enquanto a metodologia de coleta de dados adotada é descrita na Seção 3. As análises realizadas são discutidas na Seção 4, enquanto conclusões e trabalhos futuros são apresentados na Seção 5.

## 2 TRABALHOS RELACIONADOS

A literatura sobre discurso de ódio online é bastante ampla. Alguns trabalhos focam na análise de padrões textuais [44], outros nas relações com eventos violentos [33, 39] ou em como o discurso de ódio se manifesta dependendo do alvo [14]. Outros trabalhos exploram as características dos autores destes discursos [15, 37, 41] e propõem soluções para detecção automática [9].

Quanto à misoginia online, Ging [21] identificou as principais categorias e características da Manosfera, buscando teorizar as formas de masculinidade que caracterizam esse espaço discursivo. Outros trabalhos focaram em plataformas específicas. Por exemplo, no Reddit, trabalhos identificaram um aumento significativo de violência e hostilidade contra mulheres [17], mapearam a intensificação da linguagem misógina e tóxica ao longo do tempo [26] e associaram a participação nessas comunidades a um aumento da radicalização [24]. No YouTube, Papadamou *et al.* [36] caracterizaram a comunidade Incel e sua ideologia “blackpill”, baseada em visões fatalistas

e tradicionais sobre gênero, que, apesar de compartilhar posicionamentos antifeministas, se distingue da Red Pill, que enfatiza o autodesenvolvimento masculino.

Especificamente no contexto brasileiro, alguns trabalhos recentes investigaram a presença de misoginia em plataformas sociais (e.g., [12, 13, 29, 42, 51]). Por exemplo, em [12], os autores apresentam um relatório sobre grupos brasileiros do Telegram voltados à propagação de narrativas misóginas. Já em [42], foram analisadas as estratégias de produção e disseminação de conteúdo em diversas plataformas com foco em um caso específico de grande repercussão midiática (as acusações de assédio contra o humorista Marcius Melhem). Oriundos de áreas diversas, como Ciências Sociais e Comunicação Social, estes trabalhos analisam especialmente as narrativas e expressões comuns da comunidade, frequentemente a partir da análise de nuvens de palavras e da modelagem de tópicos. Entre os trabalhos mais próximos ao nosso, podemos citar dois relatórios técnicos recentes sobre misoginia no YouTube brasileiro [22, 43]. Em [22], os autores investigaram 89 vídeos da comunidade Red Pill, revelando que seu conteúdo reforça narrativas conservadoras e violentas contra mulheres, além de mostrar como o sistema de recomendação do YouTube contribui para a consolidação desse grupo. Já em [43], o foco de estudo é a monetização de criadores misóginos no YouTube e as comunidades da Manosfera brasileira.

À luz dos trabalhos discutidos, nossa pesquisa se diferencia por focar na comunidade Red Pill no YouTube brasileiro. Como destacado em [22], esta é uma das comunidades mais ativas e influentes da Manosfera. Outros trabalhos abordam misoginia digital de forma generalizada [26, 43] ou focam em outras plataformas [12, 17, 24]. Em comparação com estudos anteriores sobre o YouTube brasileiro [22, 43], nosso trabalho apresenta uma análise mais ampla, cobrindo um volume de dados muito mais expressivo (em comparação a [22]) e um conjunto de análises mais ricas que incluem propriedades não somente dos vídeos mas também dos comentários associados a eles.

## 3 METODOLOGIA DE COLETA DE DADOS

Utilizamos a YouTube Data API V3<sup>6</sup> para a coleta de dados públicos referentes a vídeos, canais, comentários e metadados da plataforma. A coleta de dados relacionados à temática Red Pill se mostrou bastante desafiadora devido à dificuldade de identificação de um vocabulário descritivo e discriminativo do conteúdo relacionado. A seguir, descrevemos as principais etapas realizadas.

Inicialmente, foi feito um estudo sobre os principais conceitos, terminologias, mecanismos de disseminação e influenciadores da comunidade Red Pill [13, 22, 43]. A partir deste estudo, foi criada uma lista ampla com 66 expressões comuns na comunidade. Exemplos incluem “O que é Red Pill?” e “O que é Valor Sexual de Mercado?”<sup>7</sup> Dessa forma, cada expressão foi submetida à API de busca do YouTube, com parâmetros restringindo o mês de publicação do vídeo (entre 01/2024 e 05/2025), o idioma (português) e região (Brasil). Para cada expressão e mês de publicação, foram coletados os 50 vídeos mais relevantes retornados pela plataforma. Em uma avaliação preliminar dos dados coletados, observou-se um volume

<sup>6</sup><https://developers.google.com/youtube/v3/getting-started>

<sup>7</sup>A lista completa é omitida por restrições de espaço mas pode ser encontrada em [https://drive.google.com/drive/folders/1KfwrqIN\\_B0bLfe2NylA5NDom3Vjav9ot?usp=sharing](https://drive.google.com/drive/folders/1KfwrqIN_B0bLfe2NylA5NDom3Vjav9ot?usp=sharing). Ademais, os dados públicos coletados e anonimizados estarão disponíveis mediante solicitação, devido à natureza altamente sensível de seus conteúdos.

**Tabela 1: Termos utilizados para filtragem de dados.**

Beta	Betinha	Red Pill	Valor Sexual de Mercado
Chad	Alpha	Black Pill	Celibato Involuntário
Coomer	Miqueinha	Femimimi	Masculinidade Forte
80/20	Mulh3res	Macho Alfa	Homem de Valor
Polo Feminino	Mulher 304	Mask Pill	Monkey Branching
Polo Masculino	Bostileira	Manosphere	Homem Beta
Antiotário	Manosfera	Machosfera	Alfa de Verdade
	Coomecrata	Coomecracia	Hipergamia

relevante de vídeos não relacionados à temática, possivelmente devido a múltiplos significados de algumas consultas. Por exemplo, a expressão “The Wall”, que na comunidade Red Pill, representa o declínio do “valor sexual de mercado” feminino ao atingir 30 anos, também pode se referir ao álbum da banda Pink Floyd. Sendo assim, foi feita uma filtragem dos vídeos, mantendo apenas os com títulos ou descrições contendo termos mais específicos, listados na Tabela 1, resultando em 4.158 vídeos associados a 2.444 canais distintos.

Uma nova análise exploratória dos dados filtrados levou a três observações: 1) muitos canais tinham apenas um vídeo na coleta. Ao analisar manualmente estes canais, verificou-se que enquanto alguns de fato publicaram na temática Red Pill apenas esporadicamente, outros tinham muitos vídeos relacionados que não apareciam nos dados coletados, devido ao uso de termos mais genéricos como *mulher*, *homem maduro* e *feminismo*; 2) a despeito dos filtros, vários vídeos tinham conteúdo não apoiador à comunidade e 3) vários vídeos tinham conteúdo em outras línguas. Estas observações demonstraram a necessidade de uma curadoria ainda mais cuidadosa a fim de se obter uma coleção de dados confiável.

Para tanto, restringimos a análise a um subconjunto de canais com evidências mais claras de vínculo com a comunidade Red Pill: todos com ao menos cinco vídeos coletados, além de seis outros canais frequentemente mencionados em mídias e sites sobre a Manosfera. Todos eles foram manualmente analisados para conferir que de fato eram dedicados à divulgação da temática Red Pill no Brasil, totalizando 28 para o corpus final. Com a lista de canais definida, foi feita nova busca utilizando a API do YouTube para coletar *todos os vídeos públicos* de cada canal, publicados entre 1º de janeiro de 2024 e 30 de junho de 2025. No total, foram coletados 18.034 vídeos, com 3.037.221 comentários compartilhados por 613.146 usuários.

Para cada vídeo, foram coletados: identificador do vídeo, título, descrição, tags, identificador do canal de origem, data de publicação, duração total, números de curtidas, comentários e visualizações totais, lista de comentários e contas associadas (comentadores). Para preservar a privacidade dos usuários, todas as menções a emails e contas do YouTube (identificados pela presença do @), incluindo canais monitorados e autores de comentários, foram anonimizados.

A fim de analisar as propriedades do conteúdo de cada vídeo, utilizou-se a concatenação do título e da descrição como representação de seu conteúdo, removendo apenas tags e os caracteres ‘\t’ e ‘\n’, que poderiam afetar o desempenho das técnicas de análise de textos usadas. O mesmo foi feito para os comentários, mantendo apenas aqueles com 30 ou mais caracteres. Assim, a coleção de dados analisada consiste em 18.034 vídeos de 28 canais, com 2.238.836 comentários compartilhados por 486.010 usuários distintos.

## 4 ECOSSISTEMA RED PILL NO YOUTUBE BRASILEIRO

Esta seção apresenta os resultados da análise dos 28 canais coletados. Ela oferece inicialmente uma visão geral dos dados focando nas estratégias de produção de conteúdo pelos canais e na recepção pela comunidade (Seção 4.1). Em seguida, ela apresenta uma caracterização multifacetada dos conteúdos textuais associados aos vídeos (Seção 4.2) e aos comentários (Seção 4.3).

### 4.1 Produção e Consumo de Conteúdo

Esta seção apresenta uma visão geral dos dados, focando tanto na produção de conteúdo pelos canais monitorados quanto em como este conteúdo foi recebido pela comunidade. Ambas discussões são baseadas nos resultados mostrados na Tabela 2, que apresenta para cada canal: o número total de vídeos, o número de meses, entre janeiro de 2024 e junho de 2025, de atividade (i.e., meses com alguma postagem), a média e o desvio padrão (dp) do número de vídeos por mês de atividade, os números totais, médias e desvios padrões de comentários, visualizações e curtidas e, por fim, o total de inscritos.

A análise dos 28 canais selecionados revela um ecossistema bastante heterogêneo quanto às estratégias de produção de conteúdo adotadas, refletidas na frequência de postagem, meses de atividade e duração média dos vídeos. Por exemplo, a média de vídeos mensais varia de apenas 2 (canal 27) a 139 (canal 11), com desvios-padrão que indicam desde alta regularidade até grande variação. A duração média dos vídeos também difere drasticamente: vai de apenas 33,8 segundos (canal 28) a mais de 84 minutos (canal 10). Alguns canais, como o 28, estiveram ativos por poucos meses, enquanto outros tiveram postagens em todos os meses da coleta (e.g., Canal 1).

Para complementar os resultados da Tabela 2, a Figura 1 mostra a evolução do número de vídeos postados, tanto no total quanto para três canais selecionados, entre janeiro de 2024 e meados de junho de 2025. O pico em março de 2025 coincide com a repercussão da série “Adolescência” (Netflix), que aborda a radicalização masculina por grupos misóginos. Já a queda atípica no fim de junho pode ser devido à saturação do tema e ao feriado de Corpus Christi (19 de junho). A Figura 1(b) ilustra, de forma mais detalhada, as diferentes estratégias de publicação ao longo do tempo: O canal 15 exibe uma baixa variação de produção, mantendo uma faixa estável, o que é evidenciado pelo seu baixo desvio padrão na média de vídeos mensais (4,1, na Tabela 2). Já o canal 16 demonstra uma frequência mais variável e um pico significativo em março de 2025. O canal 23, por sua vez, representa um padrão mais esporádico, com longos períodos de inatividade seguidos por breves surtos de conteúdo, alinhado à uma frequência relativamente baixa de somente 19,9 vídeos/mês. A ausência de padrão único desses 3 canais reforça a diversidade da comunidade.

No que diz respeito à recepção dos conteúdos pela comunidade, os dados indicam um público numeroso e altamente exposto ao material analisado. Essa recepção é avaliada em duas dimensões complementares: alcance, representado pelo número de inscritos, e engajamento, medido por visualizações, curtidas e comentários. A distinção entre essas dimensões é importante: o número de inscritos indica um vínculo mais duradouro com o canal, sugerindo interesse contínuo no conteúdo. Já curtidas, visualizações e comentários refletem interações pontuais, que podem ocorrer mesmo sem um comprometimento mais firme por parte do espectador.

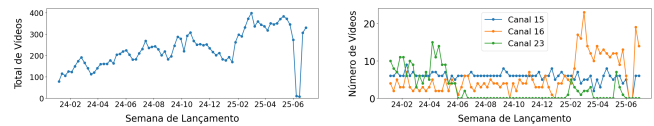
Tabela 2: Visão Geral dos Dados.

Canal	# total de vídeos	Meses de atividade	Média # vídeos/mês (dp)	Duração média em seg (dp)	# total de comentários	Média #coment/vídeo (dp)	# total de visualizações	Média vis/vídeo (dp)	# total de curtidas	Média curtidas/vídeo (dp)	# de inscritos
1	2134	18	118,6 (48,3)	94,7 (107,1)	332305	155,7 (332,5)	102759571	48153,5 (165719,5)	8833125	4139,2 (14375,0)	325000
2	1698	15	113,2 (42,1)	75,6 (42,5)	387503	228,2 (526,9)	141435162	83295,1 (301671,7)	11832309	6968,4 (23493,0)	211000
3	1498	18	83,2 (22,9)	116,1 (213,1)	25418	17,0 (87,2)	26280517	17543,7 (104589,6)	1098259	733,2 (3632,5)	77200
4	1277	13	98,2 (62,6)	166,7 (113,1)	16600	13,0 (37,4)	5802479	4543,8 (95742,2)	232493	182,1 (3220,9)	17000
5	1147	18	63,7 (13,1)	52,3 (28,4)	143482	125,1 (419,5)	78657146	68576,4 (435370,4)	5002954	4361,8 (20002,6)	423000
6	1065	18	59,2 (18,3)	96,3 (280,1)	26679	25,1 (54,1)	11530336	10826,6 (33455,8)	793787	745,3 (1352,3)	51100
7	992	17	58,4 (27,5)	108,3 (252,6)	59881	60,4 (254,6)	126248207	127266,3 (732266,8)	101651	102,5 (2254,8)	101000
8	960	9	106,7 (56,2)	72,8 (38,3)	45674	47,6 (199,5)	27007213	28132,5 (188179,9)	1821735	1897,6 (13885,8)	46700
9	808	11	73,5 (45,1)	128,7 (103,4)	3826	4,7 (32,7)	448015	554,5 (2491,7)	22555	27,9 (111,6)	1960
10	709	18	39,4 (23,0)	5067,2 (4430,6)	132537	186,9 (457,7)	33125276	46721,8 (217032,8)	2137261	3014,5 (15724,0)	320000
11	695	5	139,0 (61,4)	68,4 (29,4)	66729	96,0 (271,0)	65901706	94822,6 (539565,5)	3112495	4478,4 (13554,0)	92700
12	600	15	40,0 (23,6)	135,7 (134,7)	1201	2,0 (11,2)	1576104	2626,8 (7331,1)	90478	150,8 (402,7)	6790
13	540	18	30,0 (4,5)	315,0 (259,8)	127717	236,5 (230,1)	13533089	25061,3 (66584,0)	1160032	2148,2 (2257,1)	59000
14	500	18	27,8 (10,5)	262,4 (311,2)	11986	24,0 (113,3)	14963322	29926,6 (132856,4)	923997	1848,0 (6471,8)	125000
15	453	18	25,2 (4,1)	681,4 (518,1)	210594	464,9 (433,2)	19805561	43720,9 (49987,2)	2103546	4643,6 (4396,6)	250000
16	448	18	24,9 (16,7)	49,5 (33,2)	297920	665,0 (981,4)	114572778	255742,8 (827169,1)	7141192	15940,2 (46120,5)	435000
17	410	18	22,8 (5,8)	832,5 (410,3)	210494	513,4 (219,9)	6188898	15094,9 (7050,9)	1199838	2926,4 (991,3)	188000
18	332	4	83,0 (90,4)	54,8 (13,1)	881	2,7 (6,8)	413389	1245,1 (1418,1)	22010	66,3 (60,4)	1620
19	301	18	16,7 (7,9)	873,6 (424,8)	36739	122,1 (221,0)	5383422	17885,1 (27198,9)	666183	2213,2 (2836,3)	890000
20	285	4	71,2 (11,1)	118,9 (116,4)	222	0,8 (2,6)	135477	475,4 (1566,0)	9332	32,7 (117,2)	669
21	241	17	14,2 (9,2)	824,8 (244,2)	25918	107,5 (115,4)	3971367	16478,7 (16050,0)	441387	1831,5 (1432,4)	104000
22	233	7	33,3 (15,2)	1106,8 (254,3)	2798	12,0 (28,6)	847678	3638,1 (9828,3)	65056	279,2 (671,6)	17300
23	199	10	19,9 (16,7)	46,6 (37,3)	19275	96,9 (227,5)	8992426	45188,1 (123076,2)	554086	2784,4 (7277,3)	76400
24	185	11	16,8 (6,8)	1481,9 (3029,4)	23621	127,7 (163,1)	1901023	10275,8 (32393,0)	118107	638,4 (1116,0)	10900
25	160	18	8,9 (2,6)	839,3 (415,2)	22153	138,5 (129,5)	973025	6081,4 (6425,2)	96842	605,3 (492,9)	12700
26	110	4	27,5 (10,2)	456,8 (374,7)	5827	53,0 (117,4)	934919	8499,3 (23543,6)	42652	387,7 (1093,2)	8380
27	32	16	2,0 (1,2)	623,8 (518,3)	757	23,7 (49,6)	36792	1149,8 (2602,2)	2872	89,8 (173,4)	2230
28	22	2	11,0 (14,1)	33,8 (16,0)	99	4,5 (7,8)	77135	3506,1 (3831,9)	3319	150,9 (215,7)	138

A Tabela 2 destaca a magnitude da recepção do conteúdo: mais de 813 milhões de visualizações, 49 milhões de curtidas e 2 milhões de comentários no total, além de um número expressivo de inscritos em vários canais. Observa-se, mais uma vez, grande variação nessas métricas (nos valores totais, médias e desvios-padrão). Por exemplo, alguns canais, como o 20, geraram baixíssimo engajamento: em média, menos de 1 comentário, 475 visualizações e 33 curtidas por vídeo, além de um alcance bastante limitado (669 inscritos). Em contraste, canais como o 19 alcançaram números expressivos em todas as dimensões, com 890 mil inscritos e altos índices de visualizações, curtidas e comentários. No entanto, a relação entre alcance e engajamento não é linear. Embora o canal 19 tenha o maior número de inscritos, não está entre os canais com maior engajamento médio em nenhuma das três métricas, indicando que um grande público não garante, necessariamente, interações frequentes com o conteúdo.

Um outro exemplo é o canal 2, o primeiro em curtidas, comentários e visualizações totais, embora esteja apenas em 7º lugar em número de inscritos. Ao elaborar vídeos curtos (média de menos de 1 minuto e meio e baixa variabilidade) em alta frequência (113 vídeos/mês), este canal tem uma estratégia baseada em volume e consistência, em um perfil caracterizado por vídeos sempre no mesmo estilo: reações (visuais, não verbalizadas) a situações do dia a dia envolvendo relações de gênero. Em contrapartida, o canal 10 adota formato predominantemente de Podcast, com média de duração superior a 84 minutos (com alta variabilidade devido a presença também de vídeos de 15–20 minutos). Além disso, é o 5º colocado em inscritos e permanece consistente entre os sete primeiros colocados em métricas totais e médias de engajamento.

Esses dados evidenciam uma forte assimetria entre os canais, em termos de produção e de recepção. A análise dos coeficientes de correlação de Spearman ( $\rho$ ) entre as métricas associadas à produção de conteúdo (frequência, duração) e as métricas de engajamento e alcance não revelou associações fortes, o que sugere que fatores ligados ao conteúdo em si exercem papel mais determinante na



(a) Total de vídeos (b) Vídeos para 3 canais  
 Figura 1: Evolução temporal do número de vídeos coletados.

atração da audiência. Ainda assim, destaca-se uma correlação moderada entre a frequência de postagem e o total de visualizações ( $\rho = 0,445$ ), sugerindo uma associação entre publicar com maior regularidade e um aumento da visibilidade do canal.

## 4.2 Título e Descrição dos Vídeos

A análise dos vídeos considerou uma representação de conteúdo baseada na concatenação do título e da descrição, por serem estes os principais elementos usados pelo canal para comunicar temas e atrair o público. O tamanho médio destas representações foi de 491 caracteres, chegando ao máximo de 4.780. Nesta seção, utilizamos o termo conteúdo do vídeo em referência a esta representação.

A análise dos vídeos buscou capturar aspectos linguísticos e simbólicos que compõem o discurso Red Pill, incluindo: (i) uso de emojis; (ii) sentimento; (iii) atributos psicolinguísticos; (iv) modelagem de tópicos e (v) toxicidade. Os resultados são descritos a seguir.

**4.2.1 Uso de Emojis.** Os top 10 emojis mais populares entre os vídeos, extraídos por meio da biblioteca emoji<sup>8</sup>, são: 📺 (22,7%), 📺 (21,3%), 📺 (17,2%), 📺 (15,9%), 📺 (15,5%), 📺 (15,4%), 📺 (14,5%), 📺 (14,4%), 📺 (14,4%) e 📺 (14,3%). A interpretação dos seus significados baseou-se em guias sobre simbologia digital contemporânea [28].

Os emojis 📺, 📺 e 📺 são comumente usados como chamadas para ação ou marcação de destaque. Outros emojis da lista se destacam por carregar significados simbólicos fortemente associados à

<sup>8</sup><https://pypi.org/project/emoji>

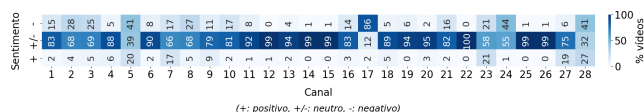


Figura 2: Sentimentos dos Vídeos.

ideologia Red Pill. Exemplos são: (i) 🤖, que é frequentemente interpretado como símbolo de força estoica e emocional, características valorizadas na retórica de masculinidade racional e autocontrolada; (ii) ⚠️, que simboliza um alerta para comportamentos problemáticos, e no discurso Red Pill marca atitudes femininas consideradas ameaças à ordem de gênero; (iii) 🚫, que aparece em contextos envolvendo censura, tabu ou crítica sexual, evocando o controle de fronteiras morais; (iv) 🏛️, que sugere autoridade simbólica e a hierarquia, como emblema de poder masculino; e (v) 🧘, que representa meditação, podendo indicar o autoconhecimento masculino defendido pela ideologia Red Pill. Outros emojis ligados à temática Red Pill ainda aparecem em proporções relevantes: 🚀 (5,8%), associado a ascensão e sucesso rápidos; 🍷 (3,6%) e 🩸 (3,3%), ambos remetendo ao conceito de “tomar a pílula vermelha”, central da ideologia Red Pill.

**4.2.2 Análise de Sentimentos.** A análise de sentimentos foi feita com o modelo XLM-T [1], uma especialização do modelo multilíngue XLM-RoBERTa (Cross Lingual Language Model - Robustly Optimized BERT-Pretraining Approach) [7]. Este modelo foi pre-treinado em um conjunto com milhões de tweets em múltiplas línguas, sendo inglês e português as mais frequentes. Estudos demonstram sua eficácia também em outras plataformas, como Reddit [38], evidenciando sua versatilidade em vários domínios. O modelo produz probabilidades para três sentimentos - positivo, negativo e neutro - e um texto é associado ao sentimento com maior probabilidade.

A Figura 2 mostra a distribuição de vídeos por sentimento, para cada canal, revelando a predominância de tom neutro na maioria deles, especialmente nos canais 12, 14, 15, 22, 25 e 26, para os quais praticamente todos os vídeos seguem este padrão. No total, mais de 77% dos vídeos apresentam títulos e descrições com sentimento neutro. Essa neutralidade *pode* indicar uma estratégia discursiva deliberada: ao usar títulos e descrições neutros, e muitas vezes enigmáticos, os canais promovem conteúdos controversos sob uma aparência de objetividade, o que pode ampliar o alcance e evitar moderação automatizada. Exemplos de vídeos classificados como neutros incluem: “Pedi a Deus uma princesa e ele me deu logo a Branca de Neve”, “A verdade sobre mulheres na academia” e “Pegar mulher casada é mais fácil | Qual a sua opinião?”. Apesar desse tom, a visualização destes vídeos revela discursos agressivos contra mulheres. Uma linha promissora de investigação futura é analisar se o tom neutro se mantém também nas transcrições dos vídeos.

Entre as exceções ao padrão geral, destacam-se os canais 5, 17 e 28, com prevalência de sentimentos negativos. O canal 17, por exemplo, tem mais de 86% dos vídeos com esse tom. Com 188 mil inscritos, apresenta-se muito ativo durante o período de coleta, com uma média de 22 vídeos mensais (14 minutos/vídeo) e uma das maiores médias de comentários por vídeo, indicando alto engajamento da comunidade. O canal 5, por sua vez, com mais de 420 mil inscritos, também é muito ativo, publicando cerca de 64 vídeos curtos por mês, e se destaca em curtidas e visualizações por vídeo.

Exemplos de (trechos de) títulos/descrições de vídeos destes canais com sentimento negativo incluem: “A ARROGÂNCIA DAS BALZACAS ARREPENDIDAS!”, “O EMPODERAMENTO DAS MULHERES JÁ ESTÁ PREJUDICANDO-AS!” e “O feminismo e sua histeria com relação a abusos contra mulheres levou a que praticamente todas as mulheres entrassem em um modo bomba-relógio.”

Por fim, os canais 7 e 23, a despeito de uma maioria neutra, apresentam um equilíbrio entre vídeos positivos e negativos. Com uma base mediana de inscritos (70-100 mil) e frequência moderada de postagens, eles apresentam desempenho apenas mediano em visualizações. Vídeos destes canais têm conteúdos como: “DEUSA RARA Da Beleza Absoluta Mandou a BRABA Em Podcast” (positivo, com probabilidade de 0,7) e “agora elas perderam a noção de vez” (negativo, com probabilidade de 0,91).

**4.2.3 Atributos Psicolinguísticos.** O LIWC (Linguistic Inquiry and Word Count) [20] é uma ferramenta para identificar termos ligados a atributos linguísticos e psicológicos [2]. Nesta análise, usamos a versão LIWC-22 em português<sup>9</sup> [5, 18] para calcular a frequência de palavras associadas aos atributos do modelo. No geral, os resultados mostraram maior frequência dos atributos *social*, *bio*, *humans* e *insight*, reforçando a ênfase da comunidade em temas como estilo de vida, relacionamentos românticos, autoestima e cuidados corporais. Para comparar canais, normalizamos os valores por z-score (-1 a 1) em cada atributo, medindo o desvio em relação à média geral.

A Figura 3 mostra os resultados para os principais atributos. Os canais 7 e 23 apresentam valores mais altos nos atributos *insight*, *achieve*, *work*, *friend* e *humans*, indicando um discurso voltado a conselhos pessoais, busca de sucesso e relações interpessoais — características típicas do perfil “guru” na comunidade Red Pill. Ambos têm mais de 75 mil inscritos e produzem vídeos orientando homens sobre como agir “corretamente”, em temas como carreira, autoestima e relacionamentos. Os baixos valores em *family* apontam menor foco em relações familiares e maior ênfase em uma masculinidade centrada na conquista e na performance individual. Entretanto, sua concentração nos canais 1, 2, 6, 8 e 18 — responsáveis por 34,32% do corpus — indica que, apesar da baixa frequência global, a temática familiar é estrategicamente relevante nos segmentos de maior produção e engajamento do conteúdo *Red Pill*. O Canal 6 tem um padrão similar mas com maior destaque para atributos *sexual*, *relig* e *bio* (e menos para *work* e *achieve*), sugerindo um foco maior em relacionamentos românticos e estilo de vida. Entre os canais com mais inscritos (19, 16, 5 e 1), observam-se padrões mais homogêneos, enquanto os menores (28, 20, 18 e 9) mostram maior heterogeneidade, sugerindo que canais consolidados já possuem um estilo definido, enquanto os menores ainda estão construindo sua identidade, experimentando formatos e temas.

**4.2.4 Análise de Tópicos.** Para identificar as principais vertentes discursivas nos vídeos, aplicou-se modelagem de tópicos, utilizando o modelo BERTopic [23], reconhecido pelo seu desempenho estado da arte nesta tarefa [23, 27, 47, 50]. Este método usa representação vetorial (*embeddings*) para preservar a essência semântica das sentenças e agrupamento das representações (após redução de

<sup>9</sup><http://143.107.183.175:21380/portlex/index.php/pt/projetos/liwc>



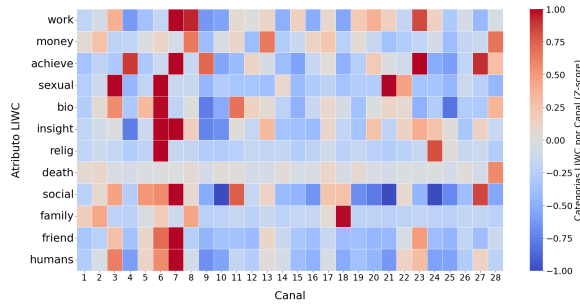


Figura 3: Atributos Psicolinguísticos Presentes nos Vídeos.

dimensionalidade) com base em suas similaridades. Para a parametrização, seguimos a documentação do BERTopic<sup>10</sup> e uma exploração leve para equilibrar número de tópicos e tamanho do corpus. A configuração final empregou o modelo de embeddings neuralmind/bert-base-portuguese-cased [46], UMAP com  $n\_neighbors = 8$ ,  $n\_components = 10$ ,  $min\_dist = 0,0$  e métrica cosseno, HDBSCAN com  $min\_cluster\_size = 150$ ,  $min\_samples = 150$  e métrica euclidiana, além de vetorização c-TF-IDF filtrando termos com frequência abaixo de 5% ou acima de 95% no corpus. Definimos  $min\_topic\_size = 150$  e  $top\_n\_words = 30$ , e ativamos  $diversity = 0,8$  para aplicar Maximal Marginal Relevance (MMR), promovendo diversidade entre palavras-chave e reduzindo redundâncias.

A Tabela 3 sintetiza os 14 tópicos mais representativos, que concentram cerca de 80% de todos os vídeos, destacando-se como centrais para caracterizar as principais narrativas do corpus. O tópico 0, embora o mais frequente, não expressa um conteúdo ideológico específico, mas sim elementos de engajamento (e.g., cumprimentos e interações típicas do YouTube [4]), indicando a busca dos criadores por manter conexão contínua com o público. Já o tópico 1 evidencia o papel central da família e dos relacionamentos na ideologia *Red Pill*, consistente com o observado na Seção 4.2.3, em que o atributo *family* apresentou participação relevante nos canais de maior produção e engajamento. Esse tópico é complementado pelo tópico 11, que confronta pautas progressistas como o feminismo. A interconexão entre valorização da família e oposição ao feminismo é recorrente na filosofia *Red Pill* [52]. O desenvolvimento pessoal masculino constitui outro pilar discursivo. Os tópicos 2, 3, 4, 7, 8, 9 e 12 tratam da busca por conhecimento e crescimento individual, reforçando a ideia, amplamente promovida por influenciadores *Red Pill* [22, 43], de construção de uma identidade masculina resiliente.

Uma nuance importante observada é a estratégia de conquista e a ideia de que os homens devem ser “anti-otários”. Tópicos como os 5 e 10, que utilizam explicitamente os termos “redpill” e “redpilled”, abordam essas concepções, promovendo a adoção de uma postura de “macho alfa” alinhada a técnicas de sedução. Por fim, os tópicos 6, 9 e 13 revelam a forte presença de conteúdos comerciais e de marketing, incluindo venda de cursos, produtos especializados e materiais escritos, além de estratégias de propaganda cruzada para outras plataformas, evidenciando o viés empreendedor do movimento. Em suma, os tópicos revelam um ecossistema discursivo multifacetado, com pautas ideológicas centradas em família e gênero, estratégias de autoaperfeiçoamento e conquista, e monetização.

<sup>10</sup> [https://maartengr.github.io/BERTopic/getting\\_started/parameter%20tuning/parametertuning.html](https://maartengr.github.io/BERTopic/getting_started/parameter%20tuning/parametertuning.html)

Tabela 3: Principais Tópicos Presentes nos Vídeos.

ID	# vídeos	Palavras-chave
0	3362	vamos, perder, nenhum, vídeo, bem, vindo
1	2318	marido, esposa, filho, pai, mãe, casa
2	1844	red, flags, lições, código, grupo, combomanual
3	928	reflexão, responsabilidade, namoro, problema, vale
4	615	homemdealtovalor, desenvolvimentomascullino, alpha, método
5	588	redpilled, conquistarmulheres, machoalpha, macho, namorosafado
6	557	masters, mercado, descontohttps, kiwify, hotmart, instagram
7	498	atualidade, assuntos, dinâmicas, contribuição, informado, analiso
8	498	homemmoderno, homemdevalor, antitrouxa, motivação, truques
9	446	conselho, livros, agenda, consultoria, conversar, hábitos
10	423	antiotario, redpill, redpillbrasil, opnião, esposa, marido
11	408	contra, feminista, feminismo, sistema, filhos, crianças
12	402	masculinidade, respeito, lei, relacionamentos, confiança, comunicação
13	385	antiotário, livro, vendido, país, desenvolvimento, pessoal

**4.2.5 Toxicidade.** Para analisar a toxicidade dos conteúdos dos vídeos, utilizou-se o Detoxify [25], modelo baseado na BERT multilíngue que atribui um *score* entre 0 e 1 para diferentes dimensões de negatividade textual (quanto maior o valor, mais tóxico). Um limiar  $\theta$  aplicado a cada dimensão define o ponto a partir do qual o conteúdo é considerado tóxico. Este estudo foca na dimensão toxicidade, aplicada em duas tarefas: detecção de misoginia e detecção de toxicidade em geral. Essas tarefas foram motivadas pela identificação, em experimentos preliminares, de conteúdos tóxicos não necessariamente misóginos, como discursos homofóbicos, ataques a homens e críticas políticas.

Embora existam recomendações na literatura para definição do limiar  $\theta$  [3], o contexto *Red Pill* e o foco em misoginia exigem calibração específica. Para isso, foi conduzido um experimento de ajuste de  $\theta$  contextualizado a cada tarefa. Todas as representações de vídeos foram avaliadas com o Detoxify e separadas em 10 grupos estratificados de acordo com os *scores* recebidos (e.g., 0-0,1; 0,1-0,2; etc). De cada grupo, 22 vídeos foram amostrados aleatoriamente e foram posteriormente rotulados manualmente por três avaliadores independentes (todas mulheres, como em [43]) quanto à presença de 1) misoginia e 2) toxicidade. A concordância entre avaliadoras, medida pelo coeficiente Fleiss Kappa  $\kappa$  [16], foi de 0,52 para misoginia e 0,55 para toxicidade, indicando concordância moderada, mas aceitável, considerando a ambiguidade de muitos textos.

Cada vídeo da amostra foi então rotulado com o resultado mais frequente entre as avaliadoras (misógino/tóxico ou não). Estes rótulos foram tratados como *ground truth* para tarefas de classificação utilizando o Detoxify com valores de  $\theta$  diferentes (entre 0,1 e 0,9, com passos de 0,1). Escolheu-se então, para cada tarefa, o valor de  $\theta$  que resultou em maior Macro-F1<sup>11</sup> da classificação:  $\theta_{misoginia} = 0,6$  (Macro-F1 = 0,59) e  $\theta_{toxicidade} = 0,3$  (Macro-F1 = 0,69). Os valores de Macro-F1 moderados, especialmente para misoginia, evidenciam a dificuldade da tarefa, em concordância com os valores de  $\kappa$ . De fato, nós testamos outros modelos de toxicidade – ToxiGuardrail PT [8] e um classificador de toxicidade pronto pré-treinado com os dados do OLID-BR [49] – em um projeto experimental similar, mas os resultados foram piores, justificando a escolha do Detoxify.

A Figura 4 apresenta o percentual de vídeos classificados como tóxicos e misóginos por canal. No geral, os percentuais são baixos (menor que 4% vídeos misóginos e 10% tóxicos para a maioria dos canais), reforçando o argumento de textos mais neutros e menos

<sup>11</sup> Macro-F1 é a média aritmética do F1-score de cada classe, onde o F1-score é a média harmônica entre precisão e revocação, calculados separadamente para cada classe

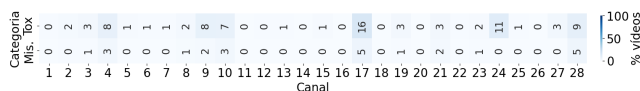


Figura 4: Percentuais de Vídeos Misóginos e Tóxicos.

explícitos. Também observam-se níveis de toxicidade sempre superiores aos de misoginia, sugerindo uma retórica que pode ser agressiva, provocativa ou ofensiva de maneira mais ampla.

Especificamente, observa-se uma grande variação entre os canais, com diferentes intensidades de discursos potencialmente prejudiciais no ecossistema analisado. Por exemplo, os canais 17 e 24 lideram em toxicidade, com 15,9% e 11,4% dos vídeos tóxicos, respectivamente. Ambos produzem vídeos longos e recebem alto volume de comentários, favorecendo discussões mais elaboradas e controversas, assim como o canal 10 (7,5%). Por outro lado, os canais 4 e 9, também em destaque quanto à toxicidade, apresentam vídeos mais curtos e com baixo engajamento, o que pode refletir um discurso muito inflamado e pouco recebido mesmo pela própria comunidade.

Quanto à misoginia, o canal 17 também apresenta o maior percentual (5,1%), seguido pelos canais 28, 4 e 10. Nenhum vídeo do canal 24 foi classificado como misógino, apesar do alto índice de toxicidade. Isto ocorre devido a um foco maior em depreciação masculina, como no vídeo “POR ISSO os HOMENS estão frouxos MASCULINIDADE TÓXICA e FEMINISMO”.

A análise manual mostrou que o modelo identifica bem mensagens tóxicas, mesmo sem linguagem explicitamente agressiva. Textos como “Mãe não dá mole para filhas preguiçosas e abusadas” e “Da uma olhada nesse vídeo. Ela joga a aliança onde? No lixo.” (Canal 2) foram corretamente classificados como tóxicos, apesar da ausência de xingamentos diretos. Essa sensibilidade é crucial para captar formas sutis de toxicidade (e.g., julgamentos e depreciações veladas) frequentes na retórica Red Pill. Exemplos explícitos de misoginia, que reforçam estereótipos negativos, como “O cara desistiu de sair comigo porque sou gorda” (Canal 1) e “DANIEL ALVES INOCENTADO E OTÁVIO MESQUITA ACUSADO: AMBOS VÍTIMAS DA INSANIDADE DAS MODERNETES” (Canal 17), também foram corretamente classificados como misóginos e tóxicos.

Porém, ressalta-se que, a despeito do ajuste fino de  $\theta$  para cada tarefa, os modelos também apresentaram falhas relevantes. Por exemplo, o modelo teve dificuldade em detectar misoginia específica, como no trecho tóxico acima com menções a “filhas preguiçosas e abusadas”. Apesar do discurso depreciativo contra mulheres, o vídeo não foi classificado como misógino, sugerindo que o modelo está mais calibrado para toxicidade geral do que para nuances linguísticas e culturais da misoginia contextualizada. Outros exemplos são textos com frases curtas, gírias ou termos internos da comunidade Red Pill, assim como ironias e críticas implícitas como em “A DEUSA DA BELEZA DIZ QUE É MUITO BELA POR ISSO OS HOMENS NÃO CHEGAM!! CORRAM PARA AS COLINAS!” (Canal 9), “SOL QUER ENCONTRAR ALGUÉM PARA SEGUIR NESSA CAMINHADA! MIQUEINHA SE QUISER SUA HORA CHEGOU!” (canal 4), “MIQUEINHA CORNO CONFORMADO!” (Canal 4) e “E você namoraria uma mulher rodada?” (Canal 11), que foram erroneamente classificados como não tóxicos/misóginos. Estes resultados reforçam a dificuldade de ambas tarefas de detecção, motivando esforços futuros para o desenvolvimento de modelos especializados.



Figura 5: Sentimentos dos Comentários.

### 4.3 Comentários

Nesta seção, a nossa atenção se torna para os mais de 2,2 milhões de comentários associados aos vídeos, com o objetivo de comparar a linguagem dos produtores de conteúdo com a de seus públicos. Após filtragem descrita na Seção 3, os comentários variam de 30 a 9.703 caracteres, com média de 143. Nossas análises, discutidas a seguir, abordam as mesmas dimensões e técnicas exploradas na caracterização dos vídeos. Omitimos, porém, os resultados sobre atributos psicolinguísticos por restrição de espaço e também pela similaridade com as observações feitas para vídeos (Seção 4.2.3).

**4.3.1 Emojis.** A diversidade de emojis usados nos comentários é bem maior do que nos vídeos, com uma distribuição de popularidade com cauda muito longa. Para ilustrar, os top 10 emojis mais frequentes nos comentários são: 😊 (8,5%), 😊 (2,4%), ❤️ (1,7%), 😊 (1,2%), 😊 (0,86%), 😊 (0,81%), 🤔 (0,8%), 🤔 (0,61%), 🤔 (0,51%) e 🤔 (0,42%). Diferentemente dos vídeos, em que emojis mais populares são chamadas para ação por parte dos criadores de conteúdo ou remetem à temática Red Pill, nos comentários, a audiência destes conteúdos tendem a empregar símbolos expressando emoções, estados de espírito e aprovação. Ressalta-se que 🤔 e 🤔 são comumente usados em conjunto para representar uma figura masculina de alta capacidade intelectual<sup>12</sup>, justificando as frequências similares.

**4.3.2 Análise de Sentimentos.** Diferentemente dos vídeos, que, em vasta maioria, apresentam um sentimento neutro, os comentários são muito mais concentrados no sentimento negativo (53% no total). A Figura 5 mostra que esta concentração predomina em 26 dos 28 canais. Ressalta-se que não foi observada uma Correlação de Spearman [32] muito forte entre os percentuais de sentimento negativo entre os vídeos e entre os comentários de um mesmo canal ( $\rho = 0,39$ ), o que evidencia que o sentimento dos comentários pode divergir do vídeo devido à diversidade de interpretações, motivações e estratégias discursivas dos usuários. Alguns exemplos interessantes são os canais 4 e 12 que, a despeito dos percentuais de vídeos negativos muito baixos, têm alta frequência de negatividade em seus comentários (60% e 49,5%, respectivamente). Outro exemplo é o Canal 17, destaque por apresentar mais de 86% dos vídeos negativos, mas que tem um percentual de comentários negativos inferior a 60%.

**4.3.3 Análise de Tópicos.** Mais uma vez, utilizamos o BERTopic, como na Seção 4.2.4, mas com parâmetros ajustados ao maior volume e diversidade do corpus. Usamos o modelo de *embeddings* BERTimbau [46], UMAP com *n-neighbors* 10, *n-components*=10, *min-dist*=0,0 e métrica cosseno, HDBSCAN com *min\_cluster\_size*= 500, *min\_samples*= 500 e métrica euclidiana, e vetorização c-TF-IDF filtrando termos com frequência abaixo de 5% ou acima de 95%. Definimos *min\_topic\_size*= 500 e *top\_n\_words*= 30, com *diversity*= 0,7.

<sup>12</sup><https://www.techtudo.com.br/noticias/2023/01/o-que-significa-entenda-significado-do-meme-com-emojis.ghtml>

**Tabela 4: Principais Tópicos Presentes nos Comentários.**

ID	# de comentários	Palavras-chave
0	250132	mãe, casei, esposo, graças, são
1	128234	soldado, acompanhado, promovido, exigem, oferecer, perceber
2	71580	cristo, salvador, graças, amém, abençoe, parabéns
3	58787	capitalismo, socialismo, população, países, debate, ciência
4	40425	cristão, católica, bíblia, pecado, provérbios, adão, eva
5	28440	interesseiras, brasileiras, solteiras, bonitas, reclamam, prestam
6	24253	mamães, educar, postura, parabéns, maravilhoso, felicidades
7	22892	nojenta, ridícula, vergonhoso, lamentável, horror, inacreditável
8	22683	maquiagem, unhas, sobrancelhas, cílios, altura
9	21304	encenação, pegadinha, meme, zoeira, teatro, montagem
10	19798	daria, surra, mandava, deixava, pegava, dava
11	19088	proteja, abençoado, continue, felicidades, ilumine, abençoe
12	18631	safada, chifres, gaia, mané, comendo, bota
13	18381	escolas, professores, educar, autoridade, adolescentes, adultos
17	17308	novinhas, velhas, auge, maturidade, faixa
40	10697	escravoceta, mangina, feminista, beta, gados, castigo
47	9429	pensões, alimentícia, socioafetiva, paternidade, biológico

Os resultados obtidos revelam uma diversidade muito maior de tópicos, comparativamente à dos vídeos, com uma distribuição de popularidade com cauda longa. Sendo assim, restringimos a nossa discussão a um subconjunto destes tópicos, mostrados na Tabela 4. A tabela apresenta os 14 tópicos mais populares, além de outros três que, embora menos frequentes, evidenciam pontos característicos da filosofia Red Pill [22, 43]. No conjunto, esses 17 tópicos representam 35% de todos os comentários analisados, oferecendo uma visão e detalhada das pautas e da linguagem adotadas pelo grupo.

Assim como para os vídeos, a estrutura familiar é um tema central entre comentários, destacada pelos tópicos 0 e 6, que exaltam a família tradicional, a maternidade e o casamento. Esse resultado sugere que tanto criadores de conteúdo quanto sua audiência valorizam bastante a temática familiar. O julgamento do comportamento feminino também é recorrente. Alguns tópicos (e.g., 5) abordam a estética, discutindo características físicas e procedimentos realizados por mulheres; outros (e.g., 17) enfatizam a idade feminina, revelando, além da misoginia, um viés etarista. Por outro lado, o tom de conflito e crítica permeia temas como traição e ataques a homens alinhados a pautas feministas (e.g., tópicos 12, 40). Questões financeiras, como pensões, também aparecem como relevantes no debate (e.g., tópico 47). A religião, especialmente a cristã, surge em diversos tópicos (e.g., 2, 4, 11), refletindo uma interseção com o discurso da comunidade. Há ainda debates políticos e sociais, bem como conteúdos de humor e entretenimento, reforçando que o grupo também aborda pautas que moldam sua visão de mundo (e.g., tópico 3, 9) [22], além de certos tópicos (e.g., 1, 13) que revelam um discurso focado na educação, disciplina e valorização de instituições de autoridade. Por fim, alguns tópicos evidenciam uma faceta mais violenta, com uso de vocabulário ofensivo para atacar indivíduos ou comportamentos desaprovados, além de incitação à violência (e.g., tópicos 7, 10).

Em suma, como para vídeos, os temas discutidos nos comentários, revelam um discurso variado, abrangendo desde a valorização da família tradicional até debates sobre gênero, religião, política e humor, com manifestações de misoginia, etarismo e violência verbal.

**4.3.4 Toxicidade.** O Detoxify foi usado mais uma vez para classificar comentários em tóxicos (ou não) e misóginos (ou não). Para tanto, adotou-se a mesma metodologia para determinação do melhor limiar  $\theta$  para cada tarefa, baseada na criação de um *ground truth* rotulado por avaliadoras mulheres. A rotulação foi feita em uma amostra de 250 comentários. Para a tarefa de misoginia, obteve-se

**Figura 6: Percentuais de Comentários Misóginos e Tóxicos.**

um Fleiss Kappa  $\kappa=0,68$  (concordância substancial), com o melhor resultado para  $\theta_{misoginia} = 0,6$  (Macro-F1 = 0,52). Para toxicidade, obteve-se  $\kappa = 0,53$  (moderada) e  $\theta_{toxicidade} = 0,4$  (Macro-F1 = 0,6).

A Figura 6 mostra os percentuais de comentários tóxicos e misóginos por canal. Mais uma vez, observam-se grandes diferenças entre canais. Além disto, a comparação com os resultados para vídeos (Figura 4) revela percentuais bem mais altos para os comentários, chegando a 23% de comentários tóxicos e quase 18% de misóginos (ambos percentuais para o Canal 18). Isto sugere, mais uma vez, que os criadores de conteúdo tendem a ser mais contidos nos títulos e descrições de seus vídeos. De fato, os coeficientes  $\rho$  de Spearman [32] indicam baixa correlação tanto entre o percentual de vídeos tóxicos e o de comentários tóxicos por canal ( $\rho = 0,24$ ), quanto entre os percentuais de vídeos e comentários misóginos ( $\rho = 0,19$ ).

Assim como observado para vídeos, os modelos conseguiram classificar corretamente casos de misoginia e toxicidade explícitas como “Vamos falar sério. Quem quer uma vagina mais usada que banheiro público? E ainda tem de investir nisso? Nem pensar”. Porém, cometeram erros em textos irônicos, com discursos depreciativos implícitos ou gírias, como “me nego a dar minha atenção e meus recursos a uma mulher rodada e ainda 30+”, classificado como tóxico e não misógino. Mais uma vez, estes resultados sugerem a necessidade de uma investigação mais aprofundada a fim de desenvolver soluções mais robustas, especializadas para o contexto em análise.

## 5 CONCLUSÕES E TRABALHOS FUTUROS

Esta pesquisa buscou aprofundar a compreensão do movimento Red Pill no YouTube Brasil, mapeando como seu discurso é elaborado, interpretado e disseminado na plataforma. Os resultados revelam um cenário de ampla exposição a esse tipo de conteúdo, com mais de 813 milhões de visualizações e 49 milhões de curtidas, e grande diversidade nas estratégias de publicação e alcance, com uso de símbolos próprios da comunidade, identificados entre os emojis mais utilizados. Evidenciou-se também as distinções entre o material textual – título e descrição – dos criadores e a reação de sua audiência – comentários, sendo mais comuns mensagens explicitamente negativas e tóxicas no segundo grupo. Apesar disso, os tópicos analisados variam amplamente em ambos os âmbitos de análise. Esses achados ajudam pesquisadores, plataformas e formuladores de políticas a entender a escala, as estratégias e os mecanismos de engajamento que sustentam a disseminação de discursos misóginos.

Trabalhos futuros incluem extensões das análises para as transcrições dos vídeos, para um conjunto mais amplo de canais e para outras comunidades misóginas, bem como o projeto de modelos de detecção de misoginia adaptados para o contexto brasileiro.

## 6 AGRADECIMENTOS

Este trabalho contou com apoio do CNPq, FAPEMIG, CAPES e Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional e Tratamento e Disseminação de Informação (INCT-TILD-IAR).



## REFERÊNCIAS

- [1] Francesco Barbieri, Luis Espinosa Anke, and Jose Camacho-Collados. 2022. XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. European Language Resources Association, 258–266. <https://aclanthology.org/2022.lrec-1.27/>
- [2] Ryan L Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W Pennebaker. 2022. The development and psychometric properties of LIWC-22. *Austin, TX: University of Texas at Austin 10* (2022), 1–47.
- [3] Caroline Brun and Vassilina Nikoulina. 2024. FrenchToxicityPrompts: a Large Benchmark for Evaluating and Mitigating Toxicity in French Texts. In *Proceedings of the Fourth Workshop on Threat, Aggression & Cyberbullying @ LREC-COLING-2024*. ELRA and ICCL, 105–114. <https://aclanthology.org/2024.trac-1.12/>
- [4] Jean Burgess and Joshua Green. 2009. *YouTube: Online Video and Participatory Culture*. Polity Press, 140 pages.
- [5] Flavio Carvalho, Rafael Rodrigues, Gabriel Santos, Pedro Cruz, Lilian Ferrari, and Gustavo Guedes. 2019. Avaliação da versão em português do LIWC Lexicon 2015 com análise de sentimentos em redes sociais. In *Anais do VIII Brazilian Workshop on Social Network Analysis and Mining*. SBC, 24–34.
- [6] Matteo Cinelli, Andraž Pelicon, Igor Mozetič, Walter Quattrociocchi, Petra K. Novak, and Fabiana Zollo. 2021. Dynamics of online hate and misinformation. *Scientific Reports* 11 (2021), 22197.
- [7] Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised Cross-lingual Representation Learning at Scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 8440–8451. <https://aclanthology.org/2020.acl-main.747/>
- [8] Nicholas Kluge Corrêa. 2024. *Dynamic normativity: Necessary and sufficient conditions for value alignment*. Tese de Pós Graduação. PUCRS, Porto Alegre, Brasil.
- [9] Thomas Davidson, Dana Warmley, Michael Macy, and Ingmar Weber. 2017. Automated Hate Speech Detection and the Problem of Offensive Language. *Proceedings of the International AAAI Conference on Web and Social Media* 11 (2017).
- [10] Muriel de Groot, Mohammad Aliannejadi, and Marcel R. Haas. 2022. Experiments on Generalizability of BERTopic on Multi-Domain Short Text. *ArXiv abs/2212.08459* (2022).
- [11] Christian C. de Lima. 2023. *Movidos pelo ódio: a machosfera e o desejo de supremacia masculina*. Trabalho de Conclusão de Curso. Universidade de Brasília.
- [12] Ergon C. de M. Silva. 2024. Anti-woke agenda, gender issues, revisionism and hate speech communities on Brazilian Telegram: from harmful reactionary speech to the crime of glorifying Nazism and Hitler. *CoRR* (2024).
- [13] Bruna C. S. Lima e Silva. 2023. *Masculinismo: misoginia e redes de ódio no contexto da radicalização política do Brasil*. Tese de Doutorado. Pontifícia Universidade Católica de Minas Gerais.
- [14] Mai ElSherief, Vivek Kulkarni, Dana Nguyen, William Yang Wang, and Elizabeth Belding. 2018. Hate Lingo: A Target-Based Linguistic Analysis of Hate Speech in Social Media. *Proceedings of the International AAAI Conference on Web and Social Media* 12, 1 (2018).
- [15] Mai ElSherief, Shirin Nilizadeh, Dana Nguyen, Giovanni Vigna, and Elizabeth Belding. 2018. Peer to Peer Hate: Hate Speech Instigators and Their Targets. *Proceedings of the International AAAI Conference on Web and Social Media* 12, 1 (2018).
- [16] Rosa Falotico and Piero Quatto. 2015. Fleiss' kappa statistic without paradoxes. *Quality & Quantity* 49, 2 (2015), 463–470.
- [17] Tracie Farrell, Miriam Fernandez, Jakub Novotny, and Harith Alani. 2019. Exploring Misogyny across the Manosphere in Reddit. In *Proceedings of the 10th ACM Conference on Web Science*. Association for Computing Machinery, 87–96.
- [18] Pedro P. Balage Filho, Thiago Alexandre Salgueiro Pardo, and Sandra M. Aluisio. 2013. An Evaluation of the Brazilian Portuguese LIWC Dictionary for Sentiment Analysis. In *Proceedings of the 9th Brazilian Symposium on Information and Human Language Technology*.
- [19] Tommaso Fornaciari, Omar G. Morales, Auxiliadora Alcántara, Simona Frenda, and Massimo Piccardi. 2020. Linguistic and content features to identify fake news on YouTube. In *Proceedings of the 7th Italian Conference on Computational Linguistics (CLiC-it 2020)*, Vol. 2699.
- [20] Martha E Francis and James W Pennebaker. 1992. Putting stress into words: The impact of writing on physiological, absentee, and self-reported emotional well-being measures. *American Journal of Health Promotion* 6, 4 (1992), 280–287.
- [21] Debbie Ging. 2019. Alphas, Betas, and Incels: Theorizing the Masculinities of the Manosphere. *Men and Masculinities* 22 (2019), 638–657.
- [22] Amaro Grassi and Marco A. Ruediger. 2023. *Movimento Red Pill no Youtube: eixos argumentativos e recomendações algorítmicas*. Technical Report. FGV ECMI (Escola de Comunicação, Mídia e Informação).
- [23] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794* (2022).
- [24] Hussam Habib, Padmini Srinivasan, and Rishab Nithyanand. 2022. Making a Radical Misogynist: How Online Social Engagement with the Manosphere Influences Traits of Radicalization. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 450 (2022), 28 pages.
- [25] Laura Hanu and Unitary team. 2020. Detoxify. Github. <https://github.com/unitaryai/detoxify>.
- [26] Manoel Horta Ribeiro, Jeremy Blackburn, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, Summer Long, Stephanie Greenberg, and Savvas Zannettou. 2021. The Evolution of the Manosphere Across the Web. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 15. 196–207.
- [27] Larissa Malagoli, Giovana Piorino, Carlos HG Ferreira, and Ana Paula Couto da Silva. 2024. Twitter and the 2022 Brazilian elections portrait: A network and content-driven analysis. In *Brazilian Symposium on Multimedia and the Web (WebMedia)*. SBC, 283–291.
- [28] Manor High School. 2025. Guide to Emojis. Disponível em: <https://www.manorhigh.leics.sch.uk/wp-content/uploads/sites/4/2025/04/Guide-to-emojis.pdf>. Acesso em: 13 de julho de 2025.
- [29] Alice E. Marwick and Robyn Caplan. 2018. Drinking male tears: language, the manosphere, and networked harassment. *Feminist Media Studies* 18 (2018), 543–559.
- [30] Binny Mathew, Ritam Dutt, Pawan Goyal, and Animesh Mukherjee. 2019. Spread of Hate Speech in Online Social Media. In *Proceedings of the 10th ACM Conference on Web Science* (Boston, Massachusetts, USA). Association for Computing Machinery, New York, NY, USA, 173–182.
- [31] Daniel T. Miller. 2021. Characterizing QAnon: Analysis of YouTube comments presents new conclusions about a popular conservative conspiracy. *First Monday* 26 (2021).
- [32] Leann Myers and Maria J Sirois. 2004. Spearman correlation coefficients, differences between. *Encyclopedia of statistical sciences* 12 (2004).
- [33] Alexandra Olteanu, Carlos Castillo, Jeremy Boy, and Kush Varshney. 2018. The Effect of Extremist Violence on Hateful Speech Online. *Proceedings of the International AAAI Conference on Web and Social Media* 12 (2018).
- [34] ONU Mulheres. 2025. O que é a “machosfera” e por que devemos nos preocupar? <https://www.onumulheres.org.br/noticias/o-que-e-a-machosfera-e-por-que-devemos-nos-preocupar/>. Acesso em: 20 de julho de 2025.
- [35] John C. Paoillo. 2018. The Flat Earth phenomenon on YouTube. *First Monday* 23 (2018).
- [36] Kostantinos Papadamou, Savvas Zannettou, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Michael Sirivianos. 2021. 'How over is it?' Understanding the Incel Community on YouTube. *Proc. ACM Hum. Comput. Interact.* 5, CSCW2, Article 412 (2021).
- [37] Samuel Pinto, José Campolina, João Sena, Gabriel Félix, Lucas Ferreira, and Julio Reis. 2024. Caracterização e Predição de Usuários Tóxicos no Twitter/X durante as Eleições Brasileiras de 2022. In *Proc. XIII Brazilian Workshop on Social Network Analysis and Mining*.
- [38] Giovana Piorino, Vitor Moreira, Luiz Lima, Adriana Pagano, and Ana Silva. 2024. Análise de sentimentos de conteúdo compartilhado em comunidades brasileiras do Reddit: Avaliação de um conjunto de dados rotulados por humanos. In *Proceedings of the 30th Brazilian Symposium on Multimedia and the Web*. SBC, 54–62.
- [39] Kunal Relia, Zhengyi Li, S. H. Cook, and Rumi Chunara. 2019. Race, Ethnicity and National Origin-Based Discrimination in Social Media and Hate Crimes across 100 U.S. Cities. *Proceedings of the International AAAI Conference on Web and Social Media* 13 (2019), 417–427.
- [40] Cláudia da Cunha Ribas and Paula Ziviani. 2008. Mediação, circulação e uso da informação no contexto das redes sociais. *ETD - Educação Temática Digital* 9 (2008), 1–19.
- [41] Manoel Ribeiro, Pedro Calais, Yuri Santos, Virgílio Almeida, and Wagner M. Meira Jr. 2018. Characterizing and detecting hateful users on Twitter. *Proceedings of the International AAAI Conference on Web and Social Media* 12, 1 (2018).
- [42] Débora Salles, Rose Marie Santini, Carlos Eduardo Barros, Bruno Martins, Marina Loureiro, Felipe Grael, Fernando Ferreira, Danielle Mello, Vitor do Carmo, Bianca Melo, Thamires Magalhães, Daniel Loliola, João Gabriel Haddad, Mathheus Gomes, Renata Seade, Daphne Silva, and Lucas Murakami. 2023. Orquestração Multiplataforma da Misoginia: O Caso Marcus Melhem.
- [43] Rose M. Santini, Débora Salles, Luciane L. Belin, Adriano Belisário, Bruno Mattos, Stéphanie G. Medeiros, Danielle Mello, Felipe Grael, Renata Seade, Amanda Borges, Lucas Murakami, Rafael Cardoso, Erick Dau, Felipe Loureiro, Bernardo Yoneshigue, Vitor D. Carmo, and Felipe Maia. 2024. "Aprenda a evitar 'esse tipo' de mulher": estratégias discursivas e monetização da misoginia no Youtube. Technical Report. Laboratório de Estudos de Internet e Redes Sociais, Universidade Federal do Rio de Janeiro (UFRJ).
- [44] Vinicius Santos, Felipe Henriques, and Gustavo Guedes. 2022. O discurso de ódio homofóbico no Twitter a partir da análise de dados. In *Proc. XI Brazilian Workshop on Social Network Analysis and Mining*.
- [45] Tamara Shepherd, Alison Harvey, Tim Jordan, Sam Sraay, and Kate Miltner. 2015. Histories of Hating. *Social Media + Society* 1 (2015), 1–10.

- [46] Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. 2020. BERTimbau: pretrained BERT models for Brazilian Portuguese. In *Brazilian conference on intelligent systems*. Springer, 403–417.
- [47] Richardy R Tanure, Aline M Dias, Lucas A Camelo, Jussara Almeida, Helen CSC Lima, and Carlos HG Ferreira. 2025. Caracterização do debate online sobre cigarro eletrônico no Brasil: Uma análise de tópicos de discussão no YouTube. In *Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*. SBC, 54–64.
- [48] Yla Tausczik and James Pennebaker. 2010. The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods. *Journal of Language and Social Psychology* 29 (2010), 24–54.
- [49] Douglas Trajano, Rafael H Bordini, and Renata Vieira. 2024. OLID-BR: offensive language identification dataset for Brazilian Portuguese. *Language Resources and Evaluation* 58, 4 (2024), 1263–1289.
- [50] Otávio R Venâncio, Gabriel HS Gonçalves, Carlos HG Ferreira, and Ana Paula C da Silva. 2024. Evidências de disseminação de conteúdo no Telegram durante o ataque aos órgãos públicos brasileiros em 2023. In *Brazilian Symposium on Multimedia and the Web (WebMedia)*. SBC, 385–389.
- [51] Gracila Vilaça and Carlos d'Andréa. 2021. Da manosphere à machosfera: Práticas (sub)culturais masculinistas em plataformas anonimizadas. *Revista Eco-Pós* 24, 2 (2021), 410–440. [https://revistaecopos.eco.ufrj.br/eco\\_pos/article/view/27703](https://revistaecopos.eco.ufrj.br/eco_pos/article/view/27703)
- [52] Gracila Vilaça and Carlos d'Andréa. 2021. Da manosphere à machosfera: Práticas (sub)culturais masculinistas em plataformas anonimizadas. *Revista Eco-Pós* 24, 2 (2021), 410–440.