

Mapeando Pontes da Misoginia: Comunidades Temáticas e Circulação Tóxica no Reddit

Mateus Brito¹, Thiago Braga¹, Naan Vasconcelos¹, Elisa Tuler¹,

Diego Roberto Colombo Dias², Leonardo Rocha¹

¹Universidade Federal de São João del-Rei, ²Universidade Federal do Espírito Santo

mateusdeoliveirabrito@aluno.ufsj.edu.br; thiagobragaamorim@aluno.ufsj.edu.br; naan.vasconcelos@aluno.ufsj.edu.br; etuler@ufsj.edu.br; diego.dias@ufes.br; lcrocha@ufsj.edu.br

ABSTRACT

Misogynistic online communities are rarely isolated spaces. Rather, they are embedded in broader thematic ecosystems through which toxic narratives and gendered hostility may circulate beyond explicitly extremist forums. In this paper, we investigate how manosphere-related discourse spreads on Reddit and how users aligned with pro- and anti-mansphere communities navigate adjacent thematic spaces. Our analysis combines semantic topic modeling, multidimensional toxicity assessment, and co-authorship network expansion across more than 500,000 posts from more than 1,700 subreddits. Starting from 1,000 seed subreddits identified through manosphere vocabulary, we characterize the thematic environments in which this lexicon appears and measure toxicity across these contexts. We then isolate 83 highly immersed communities and trace the peripheral diasporas of ideologically opposed pro- and anti-mansphere groups. The results show that manosphere vocabulary spreads across a broad polythematic ecosystem in which discussions about gender, sexuality, and affective relationships concentrate the highest toxicity levels. We also find distinct peripheral trajectories: pro-mansphere users circulate through 167 exclusive subreddits associated with self-improvement and individual performance, whereas anti-mansphere users reach 697 exclusive subreddits more strongly connected to social support, identity, and collective engagement. We identify 26 shared subreddits operating as intersection zones where ideologically opposed users coexist and where toxicity remains high. These findings frame misogyny on Reddit as a distributed and relational phenomenon structured by thematic bridges, user mobility, and conflict zones.

KEYWORDS

misogyny, manosphere, toxic discourse, online communities

1 INTRODUÇÃO

As plataformas de mídia social reconfiguraram a circulação de informações e a formação de comunidades online, mas também ampliaram a visibilidade e a persistência de discursos tóxicos e polarizados. Em ambientes mediados por recomendações algorítmicas e por interações recorrentes entre

usuários com interesses convergentes, formam-se ecossistemas discursivos densos, nos quais identidades coletivas, jargões especializados e narrativas antagonistas tendem a se reforçar mutuamente [8, 28]. Entre essas subculturas, a *manosfera* ocupa posição central por articular uma rede antifeminista e misógina que compartilha vocabulário, repertórios interpretativos e formas recorrentes de hostilidade contra mulheres e pautas igualitárias. A literatura já mostrou que esse ecossistema, organizado em torno de grupos como *incels*, *MGTOWs* e *Red Pills*, combina ressentimento de gênero, vitimização masculina e rejeição ao feminismo, associando-se ainda a processos de radicalização, intensificação do assédio e, em casos extremos, a formas de violência offline [9, 13, 14, 33, 36, 38].

Apesar desses avanços, a maior parte dos estudos ainda privilegia a análise da manosphere como um conjunto de comunidades relativamente autocontidas, descrevendo seu léxico, sua taxonomia interna, seus temas recorrentes e seus movimentos de crescimento e migração [5, 15, 20, 25, 33]. Permanece menos explorada, contudo, a maneira como esse vocabulário e seus usuários extrapolam os limites do núcleo explicitamente misógino e se conectam a outros ambientes temáticos da plataforma. Essa questão é particularmente relevante no Reddit, cuja arquitetura baseada em subreddits permite observar com maior precisão como as discussões se organizam em torno de comunidades temáticas, como os usuários circulam entre elas e onde se formam zonas de proximidade, de transbordamento ou de confronto ideológico. Reconhecida pela literatura como espaço privilegiado para o estudo de polarização, subculturas digitais marginais e processos de radicalização [4, 7, 23, 30, 31, 39], a plataforma oferece um contexto especialmente adequado para investigar a misoginia como fenômeno relacional, isto é, como um conjunto de sentidos, práticas e estilos discursivos que se dispersam por diferentes regiões da rede e não apenas por fóruns explicitamente alinhados à manosphere.

Este trabalho investiga a misoginia no Reddit para além de suas comunidades explicitamente centrais, buscando compreender tanto os contextos temáticos em que o vocabulário da manosphere circula quanto as trajetórias periféricas de usuários vinculados a subreddits ideologicamente pró- e anti-mansphere. Para isso, partimos de um conjunto semente de comunidades, identificadas por meio do vocabulário da manosphere, e combinamos modelagem temática, análise multidimensional de toxicidade e expansão de rede por coautoria. Buscamos responder às seguintes perguntas de pesquisa:

In: Proceedings of the Brazilian Symposium on Multimedia and the Web (WebMedia'2026). Lavras/MG, Brazil. Porto Alegre: Brazilian Computer Society, 2026.

© 2026 SBC – Brazilian Computing Society.
ISSN 2966-2753

RQ1: *Em que contextos temáticos o vocabulário da manosfera se concentra no Reddit, e como a toxicidade discursiva se distribui nesses ambientes?*

RQ2: *Para quais comunidades e eixos temáticos circulam os usuários vinculados a subreddits pró- e anti-manosfera, e que diferenças estruturais há entre essas trajetórias?*

Nossos resultados mostram, em primeiro lugar, que o vocabulário da manosfera não permanece restrito a nichos explicitamente misóginos, mas se dispersa por um ecossistema politemático amplo, observado em 1.000 subreddits-semente e mais de 238 mil publicações válidas, no qual as discussões sobre gênero, sexualidade e relações afetivas concentram os maiores níveis de toxicidade. Mostram também que essa hostilidade não se comporta como um desvio localizado em poucos fóruns, mas como uma propriedade estrutural e relativamente descentralizada da rede. Em segundo lugar, evidenciamos que usuários pró- e anti-manosfera apresentam diásporas periféricas distintas: enquanto o primeiro grupo circula por 167 fóruns exclusivos mais ligados ao autoaperfeiçoamento, ao desempenho e à independência individual, o segundo alcança 697 fóruns exclusivos marcados por suporte social, identidade e engajamento coletivo. Além disso, identificamos uma zona de interesse composta por 26 subreddits compartilhados, nos quais usuários de espectros opostos coexistem e a toxicidade atinge níveis elevados, caracterizando arenas de confronto ideológico.

De forma resumida, o artigo apresenta três contribuições principais. Primeiro, propõe uma leitura da misoginia online como fenômeno de rede interconectada, e não apenas como atributo de comunidades extremistas isoladas. Segundo, integra análise semântica, toxicidade multidimensional e topologia de coautoria em um único pipeline, permitindo observar conjuntamente temas, estilos de hostilidade e padrões de circulação. Terceiro, evidencia que a dispersão da misoginia no Reddit ocorre tanto por transbordamento de vocabulário quanto por mobilidade de usuários entre comunidades, contribuindo para compreender como ideologias misóginas se mantêm, se adaptam e entram em contato com outros ecossistemas discursivos da plataforma.

2 TRABALHOS RELACIONADOS

A literatura relacionada a este trabalho pode ser organizada em três frentes: (i) estudos sobre polarização, bolhas e discursos tóxicos em plataformas digitais; (ii) pesquisas sobre a manosfera, seu vocabulário e suas dinâmicas discursivas; e (iii) abordagens baseadas em redes para compreender a circulação, a migração e a radicalização em ecossistemas online.

Na primeira frente, diferentes estudos mostram que a arquitetura das plataformas digitais favorece a formação de bolhas discursivas e a intensificação da polarização. Sistemas de recomendação e padrões recorrentes de interação tendem a reduzir a exposição a posições divergentes e a reforçar comunidades ideologicamente homogêneas [8, 28]. Mesmo quando há contato com posições opostas, esse encontro pode acentuar, e não reduzir, a polarização [1]. Em plataformas como o Reddit, a moderação descentralizada e a lógica de visibilidade baseada em *upvotes* favorecem a consolidação de comunidades tóxicas e identidades extremistas [7, 23].

Nesse cenário, trabalhos recentes apontam que a compreensão desses ecossistemas exige abordagens que articulem texto, contexto e estrutura relacional [4, 27].

Na segunda frente, a literatura sobre a *manosfera* já produziu contribuições importantes para descrever esse ecossistema como uma rede de comunidades antifeministas organizadas em torno de grupos como *incels*, *MGTOWs* e *Red Pills* [9]. Estudos linguísticos mostram que esse universo mobiliza um vocabulário altamente codificado, usado para normalizar hostilidade contra mulheres, reforçar hierarquias de gênero e construir narrativas de vitimização masculina [13, 14, 22, 38]. Em paralelo, análises temáticas indicam que a misoginia não aparece apenas em sua forma explícita, mas frequentemente se mistura a discursos de autoajuda, masculinidade, sexualidade e relacionamentos [5, 25]. Trabalhos comparativos também mostram que o Reddit abriga comunidades semanticamente próximas, mas ideologicamente opostas, como fóruns pró e anti-feministas que disputam temas semelhantes com níveis distintos de toxicidade e enquadramento discursivo [15].

Na terceira frente, estudos baseados em redes têm sido particularmente relevantes para compreender a expansão desses ecossistemas. Pesquisas mostram que a manosfera não é estática, mas sim marcada por processos de crescimento, migração e radicalização progressiva [33]. Trabalhos recentes também evidenciam que interações cruzadas entre comunidades marginais e espaços mais amplos da plataforma podem funcionar como vetores de difusão ideológica [20, 36]. Essas abordagens são fundamentais porque deslocam a análise do conteúdo isolado para a topologia das conexões entre usuários e comunidades, permitindo observar como subculturas extremistas se articulam com ambientes mais amplos e heterogêneos.

Nosso trabalho se aproxima dessas três frentes ao investigar misoginia, toxicidade, polarização e circulação em uma plataforma fortemente estruturada por comunidades temáticas. Aproxima-se, em particular, dos estudos sobre a manosfera ao reconhecer a centralidade de seu vocabulário e de seus enquadramentos discursivos, e dialoga com a literatura sobre redes ao considerar que a expansão ideológica depende não apenas do conteúdo, mas também das trajetórias dos usuários e das conexões entre fóruns. Ao mesmo tempo, diferencia-se dessas abordagens em um ponto central: em vez de tratar a misoginia apenas como atributo de comunidades explicitamente extremistas, investigamos sua circulação como fenômeno relacional e distribuído. Para isso, combinamos modelagem temática, toxicidade multidimensional e expansão por coautoria para mapear, simultaneamente, os contextos temáticos em que o vocabulário da manosfera se dispersa, os padrões de hostilidade associados a esses contextos e as diásporas periféricas de usuários pró- e anti-manosfera. Com essa estratégia, o artigo avança sobre a literatura ao mostrar não apenas onde a misoginia se concentra, mas também por quais pontes ela circula, em que temas se ancora e onde se estabelecem zonas de confronto entre ecossistemas ideologicamente opostos.

3 METODOLOGIA

A metodologia deste estudo foi concebida para investigar a misoginia no Reddit para além das fronteiras de comunidades

explicitamente alinhadas à manosphere. Para isso, estruturamos o pipeline em três frentes complementares: (i) identificação de um ecossistema semente a partir do vocabulário da manosphere; (ii) construção de um pipeline analítico para extração, limpeza, modelagem semântica e mensuração multidimensional de toxicidade; e (iii) aplicação desse pipeline em um funil topológico de dois estágios, composto pelo isolamento de um núcleo focal ideologicamente polarizado e pela expansão periférica via coautoridade. Em conjunto, essas etapas permitem observar não apenas onde o léxico da manosphere se concentra, mas também como ele se dispersa tematicamente, como a toxicidade se distribui e para quais espaços circulam usuários vinculados a comunidades pró- e anti-mansphere.

3.1 Identificação do Ecossistema Semente

Para definir as comunidades potencialmente associadas ao ecossistema da manosphere sem depender de uma seleção manual arbitrária, partimos do glossário oficial sobre a *manosphere* publicado pela ONU Mulheres [40]. Essa lista reúne termos amplamente empregados por essas subculturas, como “*red pill*”, “*incel*”, “*Chad*” e “*looksmaxxing*”. Com base nesse vocabulário, utilizamos a API PullPush [32] para realizar uma busca histórica no Reddit. Para cada palavra-chave, extraímos as 1.000 menções mais recentes, juntamente com metadados como autoria, subreddit e data de publicação. Em seguida, agregamos os resultados por subreddit e contabilizamos quantos termos distintos do glossário apareciam em cada comunidade. Em vez de considerar a mera ocorrência de um termo como evidência de alinhamento, utilizamos essa contagem como medida de densidade vocabular. A partir dela, selecionamos os 1.000 subreddits com maior variedade de termos do glossário, definindo esse conjunto como o ecossistema semente do estudo. Essa estratégia tem duas vantagens metodológicas. Primeiro, permite capturar comunidades nas quais o vocabulário da manosphere aparece de forma recorrente e diversificada, e não apenas em menções ocasionais. Segundo, evita limitar a análise a subreddits explicitamente rotulados como misóginos, o que seria incompatível com nosso objetivo de estudar a misoginia como fenômeno relacional e distribuído.

3.2 Pipeline Analítico de Conteúdo

3.2.1 Extração e Pré-processamento Textual. Uma vez definido o ecossistema semente, extraímos de cada subreddit as 300 publicações com maior *score* por meio da API PullPush [32]. No Reddit, o *score* corresponde à diferença entre votos positivos e negativos, funcionando como um indicador da relevância e da aceitação de uma postagem na comunidade. Essa escolha metodológica buscou priorizar os conteúdos mais centrais à dinâmica discursiva local. Ao final dessa etapa, obtivemos 294.986 publicações. De cada postagem, coletamos os seguintes atributos: identificador (*id*), título, conteúdo textual, autor, subreddit de origem, *score*, data de publicação, número de comentários, URL, indicador de postagem textual (*is_self*), marcador de conteúdo adulto (*over_18*) e categoria de remoção (*removed_by_category*). Como uma parcela relevante das postagens no Reddit é composta por imagens ou vídeos com pouco ou nenhum texto no corpo, concatenamos

os campos de título e conteúdo textual em um único atributo textual, utilizado como base para as análises semânticas. Em seguida, realizamos a identificação automática de idioma com o modelo *XLNet-RoBERTa* [3], mantendo apenas as publicações classificadas como escritas em inglês. A uniformização linguística foi adotada para aumentar a consistência das representações semânticas geradas nas etapas posteriores. Após isso, aplicamos um pipeline de pré-processamento com: (i) lematização; (ii) conversão para caixa baixa e remoção de acentuação; (iii) normalização de repetições de caracteres; (iv) remoção de URLs, pontuação e numerais; (v) exclusão de palavras com até três caracteres; e (vi) remoção de *stopwords* [29, 41]. Por fim, mantivemos apenas as postagens não nulas e com mais de um token no campo textual resultante. Esse filtro foi necessário para garantir um conteúdo semântico mínimo para a modelagem. Ao final, o corpus válido para análise totalizou 238.856 publicações.

3.2.2 Modelagem de Tópicos e Agrupamento Semântico. Para identificar os temas predominantes no ecossistema semente, empregamos a modelagem de tópicos com BERTopic [10], modelo consolidado na literatura por sua capacidade de produzir agrupamentos semanticamente coesos e interpretáveis [2, 17, 19]. Como etapa preparatória, removemos textos duplicados, evitando que repetições em massa fossem interpretadas como densidade temática artificial [26]. Em seguida, os textos únicos foram convertidos em embeddings com o modelo *all-MiniLM-L6-v2*, escolhido por sua eficiência e adequação ao inglês. A configuração do BERTopic foi definida por busca em grade (*grid search*), variando os parâmetros *umap_neighbors* entre 30, 50 e 100, mantendo *umap_components* em 5 e *umap_min_dist* em 0.0, além de testar valores de *k* entre 5 e 25 *clusters* ($k \in \{5, 7, 10, 12, 15, 18, 20, 22, 25\}$). O desempenho foi avaliado com base em três critérios complementares: coerência de tópicos [24], coeficiente de silhueta [35] e média normalizada agregada entre ambas as métricas. Como nenhuma medida isolada é suficiente para capturar simultaneamente qualidade semântica e separação estrutural dos agrupamentos [12], também realizamos inspeção qualitativa das cinco melhores parametrizações em cada critério. A configuração final selecionada foi composta por *umap_neighbors* = 50, *umap_components* = 5 e *k* = 22 *clusters*, alcançando coerência de 0,592, silhueta de 0,0045 e média agregada de 0,6231. Os tópicos foram propagados de volta às publicações do corpus completo, incluindo as repetições.

Para uma interpretação mais concisa, agregamos tópicos semanticamente próximos em macrotópicos. Essa etapa foi realizada com o modelo GPT-5.5 Thinking, que recebeu, para cada tópico, suas 10 palavras-chave mais representativas, três documentos centrais segundo o BERTopic e sete postagens adicionais selecionadas aleatoriamente. Com base nesse material, o modelo foi instruído a gerar um resumo descritivo de cada tópico e sugerir agrupamentos em macrotemas. O *prompt* utilizado é apresentado na Figura 1. Aplicado ao conjunto semente, esse procedimento consolidou os tópicos originais em 9 macrotópicos substantivos, além de um grupo residual de tópicos não substantivos descartado das análises subsequentes.

Prompt

```
# CONTEXT
You are a researcher and data analyst specializing in social networks. I
am analyzing a corpus of [N_1] posts published in [N_2] subreddits on the
Reddit platform, acquired from [descrição].
I applied topic modeling (BERTopic with multilingual embeddings + UMAP
+ K-Means) and obtained [N_3] topics. Each topic below is represented by
its 10 most characteristic keywords (via c-TF-IDF, in lemmatized form), 3
representative posts, and 7 randomly selected post examples.
Note: To optimize processing, the posts have been truncated to a maximum
of 200 characters. Do not attempt to complete the sentences; evaluate only
the visible context.
# TASK Based on the keywords and post examples for each topic, do TWO
things:
1) SUMMARY — for EACH topic, provide a short label (3 to 6 words)
and a one-line description explaining what the topic is about. 2) MACRO-
TOPICS — propose a grouping of the topics into coherent macro-topics
(macro-themes). Suggest an appropriate number of macro-topics (generally
between 8 and 10) and justify them. For each macro-topic, list the IDs
of the topics that compose it and explain, in 1 to 2 lines, the semantic
criterion that unites them.
- IMPORTANT: write your ENTIRE answer in Brazilian Portuguese (pt-BR),
even if these instructions are in English.
# ANSWER FORMAT (content in Portuguese)
## Part 1 — Topic Labels Topic {id} | {label} | {1-line description} ##
Part 2 — Proposed Macrotopics MACRO {macrotopic name} Topics: {list
of IDs} Criteria: {1-2 lines}
```

Figura 1: Estrutura do Prompt submetido à LLM para sumarização de tópicos e geração de macrotópicos.

3.2.3 Classificação Multidimensional de Toxicidade. Para avaliar a agressividade textual do ecossistema, adotamos uma abordagem multidimensional de toxicidade, decompondo a agressividade textual em vetores distintos, como *Severe Toxicity*, *Identity Attack*, *Insult*, *Profanity* e *Threat*, permitindo capturar nuances que uma nota única ocultaria. A escolha do classificador foi baseada em experimentos comparativos prévios entre duas ferramentas amplamente utilizadas na literatura, Detoxify [11] e Perspective API [16], conduzidos sobre uma amostra de 60 mil publicações. Embora os modelos tenham apresentado uma forte correlação estatística na dimensão de toxicidade geral (Coeficiente de Spearman de 0,7950, $p < 0,001$), as auditorias qualitativas revelaram diferenças relevantes. O Detoxify mostrou maior suscetibilidade a falsos positivos em textos emocionalmente intensos, porém não hostis, e menor sensibilidade a formas implícitas de hostilidade misógina. A Perspective API, em contraste, demonstrou maior robustez contextual, especialmente em zonas cinzentas de linguagem orgânica dos fóruns [6, 27]. Com base nessa comparação, adotamos a Perspective API como ferramenta principal para mensuração de toxicidade. Para cada publicação, extraímos tanto a pontuação de toxicidade geral quanto as subdimensões supracitadas, que são fornecidas no intervalo [0, 1], constituindo uma base comparável para caracterizar diferentes formas de hostilidade textual no ecossistema analisado. Contudo, tais modelos podem herdar vieses estruturais de treinamento, atribuindo pontuações artificialmente infladas a termos de identidade e discursos de minorias sociais [21, 37].

3.3 Mapeamento Topológico em Estágios

Após a caracterização temática e toxicológica do ecossistema semente, aplicamos um funil topológico em dois estágios para investigar a circulação de usuários entre comunidades ideologicamente polarizadas e seus espaços periféricos.

3.3.1 Isolamento do Núcleo Focal e Alinhamento Ideológico.

O primeiro estágio consistiu em isolar o subconjunto de subreddits com maior imersão na subcultura da manosphere. Para isso, calculamos, para cada comunidade, a proporção combinada de postagens pertencentes aos dois macrotópicos que apresentaram as maiores médias de toxicidade: *Gênero, sexualidade e padrões de atratividade* e *Relações afetivas, familiares e vida cotidiana*. Essa medida foi escolhida por captar a densidade relativa de temas fortemente associados ao núcleo discursivo da manosphere. A partir dessa proporção, calculamos o z -score de cada comunidade segundo a expressão $z = \frac{x - \mu}{\sigma}$, em que x representa a proporção observada, μ a média do ecossistema e σ o desvio padrão. Adotamos o limiar $z \geq 2$ — pelo menos dois desvios-padrão acima da média global — para selecionar as comunidades mais imersas nessas pautas centrais, já que este filtro restringe os fóruns situados na cauda superior da distribuição. Como mostrado na Figura 2, esse procedimento resultou em 83 subreddits que compõem o núcleo focal do ecossistema.

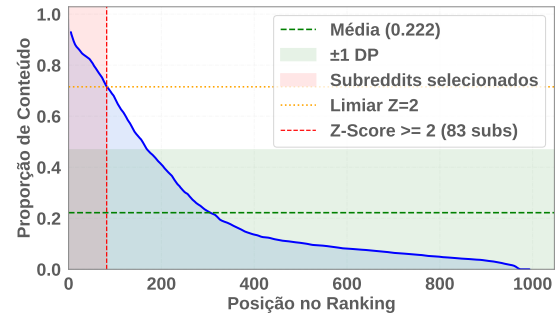


Figura 2: Distribuição dos subreddits em relação à proporção de conteúdo temático, com destaque para o limiar estatístico de z -score ≥ 2 aplicado para a seleção das comunidades núcleo.

Em seguida, classificamos essas comunidades em três grupos ideológicos: pró-manosphere, contra-manosphere e neutro/a-fim. Para essa tarefa, utilizamos o modelo GPT-5.5 Thinking, que recebeu, para cada subreddit, as 10 postagens com maior *score*, interpretadas como as narrativas de maior aceitação local. O modelo foi instruído a agrupar os fóruns segundo seu alinhamento ideológico em relação à manosphere, com base no nome da comunidade e no conteúdo semântico das postagens. O *prompt* empregado nessa etapa é apresentado na Figura 3.

3.3.2 Expansão Periférica e Mapeamento da Diáspora. No segundo estágio, expandimos separadamente as redes pró- e contra-manosphere para mapear os ambientes periféricos alcançados pelos usuários mais ativos de cada núcleo. Para isso, adotamos uma estratégia baseada em coautoria, que mapeia relações estruturais entre comunidades a partir do compartilhamento de usuários ativos. Em cada subreddit do núcleo focal, selecionamos os 10 autores mais ativos na amostra de 300 publicações, utilizando como critérios de ordenação o número de postagens e, em caso de empate, a soma dos *scores*. Em seguida, coletamos as 10 publicações de maior relevância no histórico de cada um desses usuários,

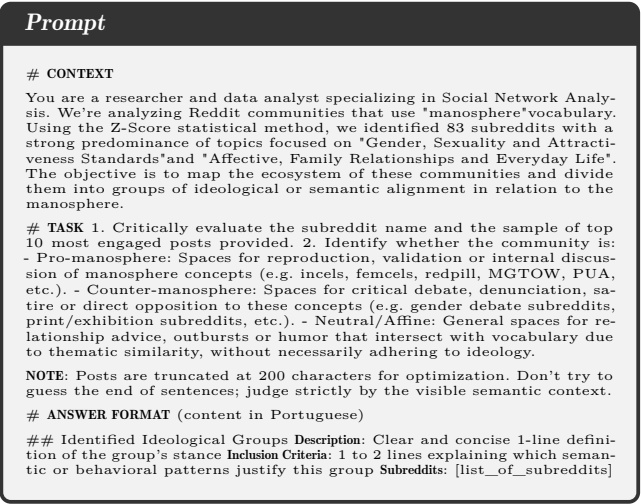


Figura 3: Estrutura do *prompt* submetido à LLM para a categorização ideológica das comunidades do núcleo focal.

registrando os subreddits de origem. Publicações associadas ao identificador *[deleted]* foram excluídas, pois esse rótulo agrega múltiplas contas apagadas sob um mesmo identificador, comprometendo a interpretação topológica. A partir da relação bipartida entre usuários e subreddits, construímos uma projeção unipartida, na qual um subreddit se conecta a outro quando compartilham usuários em comum. Esse procedimento permitiu revelar as diásporas periféricas associadas a cada polo ideológico. As comunidades periféricas identificadas por esse processo foram então submetidas ao mesmo pipeline descrito anteriormente: extração, limpeza textual, modelagem semântica e classificação de toxicidade.

4 RESULTADOS

Esta seção apresenta os resultados obtidos, organizados em torno das duas perguntas de pesquisa que orientam o estudo. A primeira subseção responde à **RQ1**, caracterizando os contextos temáticos em que o vocabulário da manosfera se concentra e examinando como a toxicidade se distribui nesses ambientes. A segunda subseção responde à **RQ2**, investigando as trajetórias periféricas de usuários vinculados a comunidades pró- e anti-manosfera e contrastando os perfis temáticos e toxicológicos desses ecossistemas. Ao estruturar a análise dessa maneira, abordamos este ecossistema de forma multidimensional, distinguindo explicitamente a presença lexical, a toxicidade e a misoginia como dimensões interconectadas, porém não equivalentes.

4.1 RQ1: Contextos temáticos e toxicidade

Para responder à **RQ1**, analisamos os 1.000 subreddits identificados com base no vocabulário da manosfera e modelamos o conteúdo de 238.856 publicações válidas. Os tópicos resultantes do BERTopic foram agregados em 9 macrotópicos substantivos, apresentados na Tabela 1. Em conjunto, eles mostram que o léxico da manosfera não se restringe a espaços explicitamente misóginos, mas aparece em uma rede

com variedade temática ampla, que abrange desde relações afetivas, gênero e sexualidade até política, identidade racial, trabalho, entretenimento, jogos e estilos de vida.

Tabela 1: Nome, critério de agrupamento e volume de publicações dos macrotópicos identificados no ecossistema semente.

Nome (Volume)	Critério de Agrupamento
1. Entretenimento, cultura e produção criativa (21,4%)	Reúne consumo e produção de bens culturais e midiáticos, incluindo cinema, séries, música, vídeos, livros, fotografias, pinturas e arte digital.
2. Relações afetivas, familiares e vida cotidiana (19,8%)	Reúne relatos pessoais e pedidos de aconselhamento relacionados a namoro, casamento, família, parentalidade, convivência doméstica e conflitos interpessoais.
3. Jogos e esportes competitivos (12,6%)	Integra atividades estruturadas por competição, desempenho e estratégia, abrangendo videogames, jogos de cartas, futebol, basquete, ligas e partidas esportivas.
4. Identidade, raça e relações internacionais (11,4%)	Agrupar discussões sobre raça, etnia, nacionalidade, pertencimento cultural, países, imigração, estereótipos e conflitos geopolíticos.
5. Gênero, sexualidade e padrões de atratividade (10,7%)	Agrupar discussões sobre relações de gênero, feminismo, identidades trans, masculinidade, aparência feminina e hierarquias de beleza ou desejabilidade.
6. Política, religião e disputas ideológicas (10,3%)	Contém debates explicitamente politizados sobre eleições, conservadorismo, aborto, religião, vacinação e políticas públicas, frequentemente marcados por polarização moral ou partidária.
7. Trabalho, economia e condições materiais (5,7%)	Abrange carreira, emprego, empresas, investimentos, custo de vida, preços, aluguel, patrimônio e desigualdades econômicas.
8. Corpo, saúde e apresentação pessoal (4,1%)	Reúne práticas de transformação e cuidado corporal, como emagrecimento, exercícios, saúde, cabelo, barba, calvície, tatuagens e alterações da aparência.
9. Alimentação, ética e estilos de vida (3,9%)	Concentra debates sobre veganismo, consumo de produtos animais, ética alimentar, culinária e escolhas individuais associadas a estilos de vida.

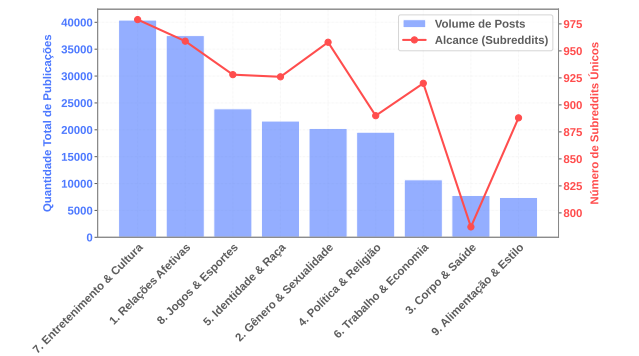


Figura 4: Relação entre o volume absoluto de publicações e a dispersão de subreddits por macrotópico.

Como mostra a Figura 4, os temas variam tanto em volume de postagens quanto em alcance entre os subreddits. “Entretenimento, cultura e produção criativa” e “Relações afetivas, familiares e vida cotidiana” aparecem como eixos de alta volumetria, enquanto “Jogos e esportes competitivos”, “Identidade, raça e relações internacionais”, “Gênero, sexualidade e padrões de atratividade” e “Política, religião e disputas ideológicas” se distribuem em patamares intermediários. Já “Trabalho, economia e condições materiais”, “Corpo, saúde e apresentação pessoal” e “Alimentação, ética e estilos de vida” apresentam menor produção absoluta. O aspecto mais expressivo, contudo, está na combinação entre volume e capilaridade.

Os macrotópicos mais centrais, somados às discussões de gênero, aparecem em quase todos os fóruns analisados. Mesmo os eixos com menor volume absoluto mantêm elevada penetração na rede. Isso indica que os subreddits associados ao vocabulário da manosphere não se organizam como bolhas monotemáticas; ao contrário, tendem a combinar múltiplos assuntos em sua composição discursiva. Em outras palavras, trata-se de um ecossistema politêmico, no qual o mesmo subreddit interliga múltiplas narrativas em um único ambiente.

Entretanto, essa ampla dispersão não implica ausência de concentração. A Figura 5 utiliza a Curva de Lorenz para demonstrar a concentração das publicações por macrotópico, revelando que a maior parte do conteúdo se concentra em um subconjunto relativamente pequeno de subreddits. Em vários casos, mais de 60% das postagens de um tema (em alguns macrotópicos, até 80%) estão concentradas nos 20% de fóruns mais ativos naquele eixo. Esse resultado sugere uma estrutura em duas camadas: embora os subreddits sejam politêmicos, cada comunidade tende a apresentar uma assinatura temática dominante, organizando sua produção em torno de poucos eixos centrais.

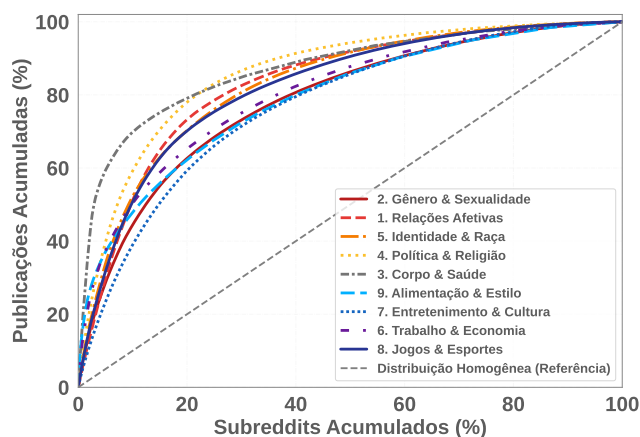


Figura 5: Curva de Lorenz [18] para a concentração de publicações em subreddits por macrotópico.

A análise de toxicidade aprofunda esse diagnóstico. A Figura 6 apresenta a média das diferentes dimensões da Perspective API por macrotópico. Observamos que “Gênero, sexualidade e padrões de atratividade” e “Relações afetivas, familiares e vida cotidiana” lideram os índices de toxicidade geral, configurando os dois eixos mais sensíveis do ecossistema. Em seguida, aparecem tópicos ligados a identidade e política, indicando que debates socialmente marcados por conflito simbólico tendem a mobilizar níveis mais altos de agressividade do que temas cotidianos ou de entretenimento, como jogos e cultura pop. A decomposição multidimensional da toxicidade mostra ainda que diferentes temas ativam formas distintas de hostilidade. Nos macrotópicos de maior agressividade geral, predominam insultos e profanidades, sugerindo uma retórica de ataque direto e hostilidade verbal explícita. Em “Identidade, raça e relações internacionais”, por outro lado, destaca-se uma elevação mais acentuada em

Identity Attack, indicando que o teor agressivo nesse caso assume forma mais diretamente dirigida a grupos sociais e identidades coletivas. Assim, a toxicidade não apenas varia em intensidade, mas também em perfil discursivo, de acordo com o alvo e o enquadramento temático do debate.

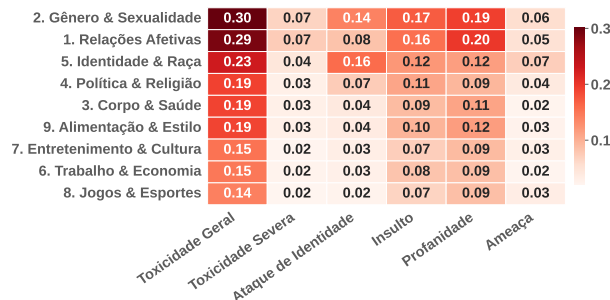


Figura 6: Mapa de calor para a média de cada perspectiva de toxicidade em relação com o macrotópico.

Por fim, a Figura 7 mostra que a concentração de toxicidade acumulada apresenta um comportamento estrutural distinto da concentração de postagens. Observa-se que as curvas de Lorenz para a toxicidade situam-se mais próximas da linha de distribuição homogênea do que as curvas de volume de publicações. Isso indica que a agressividade discursiva é menos localizada do que a atividade temática e se distribui de maneira relativamente mais descentralizada ao longo da rede. Em outras palavras, a toxicidade não parece ser um atributo exclusivo de poucos fóruns extremos, mas uma propriedade estrutural do ecossistema em que esses jargões circulam.

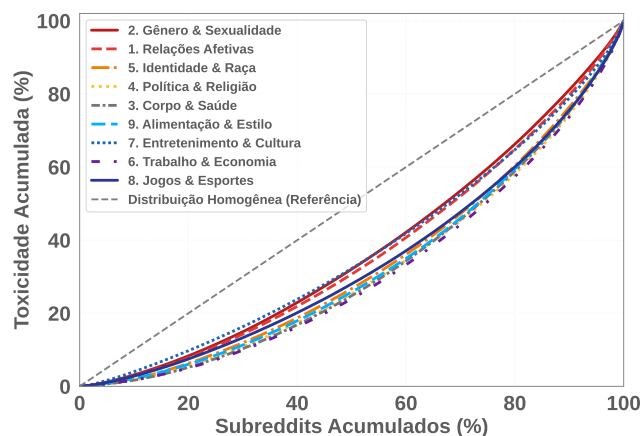


Figura 7: Curva de Lorenz [18] da concentração de toxicidade por subreddits e macrotópico.

Em conjunto, os resultados permitem responder à **RQ1** em duas dimensões complementares. Do ponto de vista temático, mostramos que o vocabulário da manosphere se dispersa por um ecossistema amplo e politêmico, embora sua atividade discursiva se concentre mais fortemente em torno de relações afetivas, gênero, sexualidade, política e identidade. Do ponto de vista toxicológico, mostramos que a hostilidade não está confinada a poucos nichos, mas se distribui de forma

relativamente descentralizada, assumindo diferentes perfis conforme o tema. Esses achados reforçam a ideia de que a misoginia no Reddit não opera apenas em comunidades explicitamente extremistas, mas também circula por diversos ambientes da plataforma, com níveis variados, porém persistentes, de agressividade verbal.

4.2 RQ2: Ecossistemas periféricos

Para responder à RQ2, partimos dos 83 subreddits identificados como núcleo focal do ecossistema e os classificamos em grupos pró-, contra- e neutro/afim em relação à manosfera. A classificação resultou em sete comunidades pró-manosfera (*askmrp*, *blackpillscience*, *marriedredpill*, *redpillwomen*, *seduction*, *semenretention*, *theredpill*) e onze comunidades contra-manosfera (*askfeminists*, *askgaybros*, *askgaybrosover30*, *asktransgender*, *exredpill*, *incelexit*, *letgirlshavefun*, *menslib*, *mtf*, *mypartneristrans*, *straighttransgirls*). A partir dos usuários mais ativos desses grupos, reconstruímos as diásporas periféricas por coautoria. A Figura 8 sintetiza esse movimento. Os sete subreddits semente pró-manosfera originaram uma diáspora de 167 fóruns exclusivos, alcançados por 126 autores. Já os onze subreddits semente anti-manosfera produziram uma diáspora muito mais ampla, com 697 fóruns exclusivos alcançados por 481 autores. Além disso, identificamos uma *Zona de Interseção* composta por 26 subreddits compartilhados, nos quais 33 autores oriundos da semente pró e 31 oriundos da semente contra publicaram simultaneamente. Essa interseção é um dos resultados mais relevantes do artigo, pois evidencia que ecossistemas ideologicamente opostos não permanecem completamente separados: há espaços em que usuários de ambas as vertentes coexistem, transformando certos fóruns em possíveis arenas de contato e confronto.

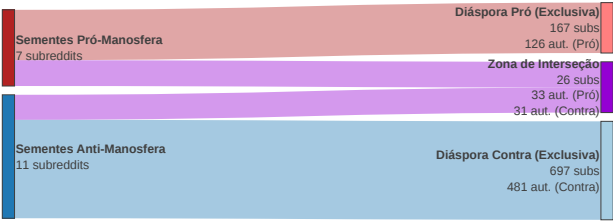


Figura 8: Diagrama de Sankey [34] de autores únicos e distribuição de comunidades entre as sementes e as diásporas temáticas.

Para caracterizar semanticamente essas diásporas, aplicamos o mesmo pipeline de modelagem temática aos subreddits periféricos de cada rede. As Tabelas 2 e 3 mostram que os dois polos não apenas circulam por espaços distintos, mas se orientam por lógicas temáticas distintas.

A diáspora pró-manosfera apresenta forte concentração em eixos individualistas, especialmente “Desenvolvimento Masculino e Autoaperfeiçoamento”, “Relações de Gênero e Sexualidade” e “Trabalho, Carreira e Independência Financeira”. Esses temas sugerem uma circulação periférica centrada em desempenho, disciplina, aprimoramento individual e interpretação competitiva das relações sociais. Já a diáspora

Tabela 2: Taxonomia e critérios de agrupamento dos macrotópicos da rede pró-manosfera.

Nome (Volume)	Critério de Agrupamento
Desenvolvimento Masculino e Autoaperfeiçoamento (23,66%)	Tópicos centrados na melhoria individual, incluindo disciplina, saúde física, controle de hábitos, sobriedade e construção de uma vida considerada mais saudável e autônoma.
Relações de Gênero e Sexualidade (10,57%)	Discussões sobre comportamento masculino e feminino, atração, namoro, relacionamentos e interpretações das dinâmicas entre os sexos, um dos temas centrais da manosfera.
Política, Cultura e Conflitos Ideológicos (11,20%)	Debates sobre política, liberdade de expressão, moderação online, identidade política e críticas a instituições ou comunidades percebidas como adversárias.
Trabalho, Carreira e Independência Financeira (6,38%)	Conteúdos relacionados à ascensão profissional, empreendedorismo, tecnologia, qualificação técnica e obtenção de autonomia econômica.
Entretenimento e Cultura Digital (38,26%)	Participação em fandoms, consumo de mídia e cultura gamer, representando espaços de socialização e entretenimento recorrentes entre os usuários.
Conhecimento, Educação e Produção Intelectual (4,37%)	Discussões sobre leitura, escrita, aprendizado de idiomas, literatura e aquisição de conhecimento.
Hobbies e Consumo de Nicho (5,56%)	Interesses especializados voltados para café, perfumaria e apreciação de produtos associados a estilos de vida e identidade pessoal.

Tabela 3: Taxonomia e critérios de agrupamento dos macrotópicos da rede anti-manosfera.

Nome (Volume)	Critério de Agrupamento
Comunidade, Humor e Interações Cotidianas (16,09%)	Conteúdos amplamente sociais, informais e característicos da cultura Reddit, incluindo humor, memes e relatos cotidianos sem forte especialização temática.
Relacionamentos, Família e Vida Pessoal (15,56%)	Tópicos centrados em vínculos interpessoais, relacionamentos amorosos, dinâmicas familiares, parentalidade, autoestima e experiências pessoais.
Gênero, Sexualidade e Feminismo (5,01%)	Discussões sobre identidades de gênero, direitos LGBTQ+, feminismo, masculinidades e conflitos ideológicos relacionados ao gênero — um eixo central em comunidades anti-manosfera.
Política, Ideologia e Questões Sociais (12,97%)	Debates políticos, nacionalismo, religião, cidadania, identidade coletiva e disputas ideológicas frequentemente associadas a temas de justiça social e direitos civis.
Saúde, Corpo e Bem-Estar (5,89%)	Questões relacionadas à saúde física e pública, incluindo exercícios, emagrecimento, hábitos saudáveis e impactos da pandemia.
Educação, Trabalho e Economia (7,57%)	Temas ligados a carreira, estudos, mercado de trabalho, empreendedorismo digital, criação de conteúdo e mobilidade socioeconômica.
Cultura Pop e Entretenimento (15,47%)	Consumo e discussão de produtos culturais, incluindo séries, filmes, quadrinhos, videogames e música.
Expressão Criativa e Estética (7,91%)	Conteúdos focados em aparência, moda, arte visual, tatuagens, fotografia e formas de expressão estética e identitária.
Estilo de Vida, Alimentação e Ambiente (8,91%)	Temas relacionados à natureza, sustentabilidade, animais, culinária, alimentação e hábitos de vida.
Mobilidade e Transporte (4,63%)	Discussões específicas sobre direção, transporte, trânsito e obtenção de habilitação, formando um tema suficientemente distinto dos demais.

anti-manosfera apresenta composição mais voltada à vida relacional e ao engajamento coletivo. Destacam-se macrotópicos como “Comunidade, Humor e Interações Cotidianas”, “Relacionamentos, Família e Vida Pessoal”, “Gênero, Sexualidade e Feminismo” e “Política, Ideologia e Questões Sociais”. Em vez de priorizar repertórios centrados na performance individual, esse polo se projeta mais fortemente em espaços

de suporte social, discussão identitária, justiça social e experiências cotidianas. Assim, a oposição entre os grupos não se manifesta apenas nas comunidades centrais, mas nas direções periféricas dos usuários. Essa diferença também se manifesta no modo como a política é tematizada: na rede pró-manosfera, os debates associam-se a liberdade de expressão, moderação e antagonismo a adversários. Já na anti-manosfera, a política articula-se a nacionalismo, religião, direitos civis e justiça social. Esse contraste revela que os polos divergem na forma de conceber o debate público: um adota postura reativa de confronto digital, enquanto o outro organiza-se em torno de estruturas institucionais e reivindicações coletivas.

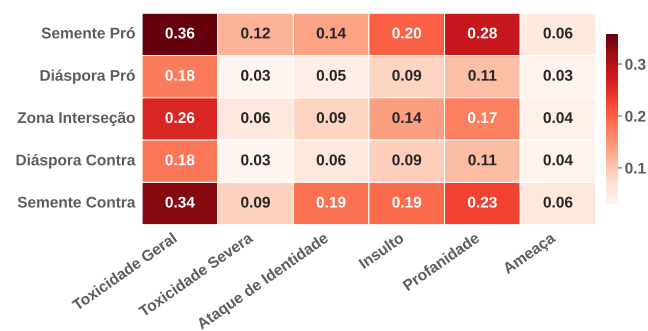


Figura 9: Mapa de calor para a média de cada perspectiva de toxicidade em relação com o conjunto de subreddits.

A análise de toxicidade por região, apresentada na Figura 9, reforça esse diagnóstico estrutural. As duas sementes (pró e contra) apresentam os níveis mais elevados de agressividade verbal em todas as dimensões avaliadas, o que indica que os núcleos opostos funcionam como regiões de maior intensificação retórica. O polo pró-manosfera registra a maior média de *Toxicity* geral e *Profanity*, enquanto o polo anti-manosfera apresenta o maior pico de *Identity Attack*, superando inclusive o grupo oposto nessa dimensão. A *Zona de Interseção* emerge como a terceira região mais hostil da rede, o que reforça a interpretação de que os fóruns compartilhados funcionam, em parte, como arenas de confronto ideológico. Esse padrão é reforçado ao observar os subreddits mais tóxicos dessa região, liderados por *trueunpopularopinion* (0,415), *allpilldebate* (0,411) e *purplepilldebate* (0,388). Tanto o nome dessas comunidades quanto suas pontuações sugerem ambientes explicitamente organizados em torno do embate entre perspectivas opostas. Em contraste, as diásporas exclusivas apresentam níveis de toxicidade mais baixos. Isso sugere que o teor hostil das sementes não se reproduz integralmente quando os usuários transitam para espaços periféricos dedicados a outros temas. Em vez de uma exportação direta e uniforme da agressividade, observamos um padrão em que a hostilidade tende a ser mitigada quando os usuários se deslocam para fóruns mais amplos, menos explicitamente ideológicos ou voltados a outras agendas temáticas.

Esses resultados respondem à **RQ2** ao mostrar que os usuários de comunidades pró- e anti-manosfera seguem trajetórias periféricas estruturalmente distintas. Enquanto o primeiro grupo circula mais por espaços ligados ao autoaperfeiçoamento, ao desempenho e à independência individual,

o segundo se projeta com maior intensidade em ambientes de suporte social, de identidade e de engajamento coletivo. Identificamos uma zona de interseção onde esses ecossistemas entram em contato e a toxicidade se mantém elevada, caracterizando um espaço de confronto. Esses achados reforçam que a circulação da misoginia no Reddit não depende apenas de onde o discurso se concentra, mas também das pontes estabelecidas pelos próprios usuários ao transitar entre comunidades ideologicamente próximas, opostas ou intermediárias.

5 CONCLUSÃO

Este trabalho investigou a misoginia no Reddit para além de comunidades explicitamente alinhadas à manosfera, tratando-a como um fenômeno distribuído, estruturado por pontes temáticas, mobilidade dos usuários e zonas de confronto. Para isso, combinamos a identificação vocabular de comunidades semente, a modelagem temática, a mensuração multidimensional de toxicidade e a expansão topológica por coautoria.

Os resultados mostraram que o vocabulário da manosfera se dispersa por um ecossistema politêmico amplo, no qual discussões sobre gênero, sexualidade e relações afetivas concentram os maiores níveis de toxicidade. Mostraram também que essa hostilidade não se restringe a poucos fóruns extremos, mas se distribui de forma relativamente descentralizada pela rede. Além disso, evidenciamos que usuários pró- e anti-manosfera seguem trajetórias periféricas distintas: enquanto o primeiro grupo circula mais por espaços ligados a autoaperfeiçoamento, o segundo se projeta com maior intensidade em fóruns de suporte social. Por fim, identificamos zonas de interseção em que usuários de espectros opostos coexistem e nas quais a toxicidade permanece elevada, caracterizando arenas de confronto ideológico.

Em conjunto, esses achados reforçam a ideia de que a misoginia online não deve ser compreendida apenas como atributo de comunidades extremistas isoladas, mas como um padrão discursivo que circula por diferentes regiões da plataforma. Como trabalhos futuros, destacam-se três direções principais: incorporar uma perspectiva longitudinal, aprofundar a análise das zonas de interseção e aplicar o mesmo método a outros ecossistemas de extremismo online, de modo a compreender em que medida a lógica de dispersão temática e mobilidade de usuários observada aqui se repete em outras subculturas digitais radicalizadas.

6 LIMITAÇÕES

Destacamos algumas limitações metodológicas neste estudo. Primeiramente, o filtro por autores mais ativos e postagens de alto *score* prioriza conteúdos centrais, podendo omitir fluxos periféricos. Em segundo lugar, a classificação ideológica via LLM carece de anotação humana cega e independente, impedindo a aferição empírica de eventuais vieses introduzidos pelo modelo. Por fim, as pontuações da Perspective API configuram indicadores probabilísticos de características linguísticas e não devem ser interpretadas como juízos normativos definitivos acerca do conteúdo ou da intenção comunicativa.

7 CUIDADOS ÉTICOS

Este estudo utilizou exclusivamente dados publicamente disponíveis no Reddit, coletados por meio da API PullPush e analisados em nível agregado, com foco em padrões temáticos, estruturais e toxicológicos de comunidades e de ecossistemas discursivos. Não houve interação com usuários, intervenção em comunidades, tentativa de reidentificação individual ou exposição de perfis específicos nos resultados. Considerando esse enquadramento, a pesquisa foi conduzida conforme os princípios de finalidade, necessidade e minimização previstos na LGPD (Lei n.º 13.709/2018) e em consonância com as diretrizes da Resolução CNS n.º 510/2016 para estudos baseados em informações de acesso público.

Os dados foram tratados apenas na medida necessária aos objetivos da pesquisa. Identificadores de usuários foram empregados exclusivamente para a construção das redes de coautoria e para análises agregadas de circulação entre comunidades, sendo anonimizados nas bases analíticas e omitidos de qualquer apresentação de resultados. O acesso aos dados permaneceu restrito à equipe de pesquisa, e a eventual disponibilização ocorrerá apenas em versão anonimizada, sem informações que permitam a reidentificação de indivíduos. Embora os dados analisados sejam públicos, reconhecemos que discussões em plataformas sociais podem envolver expectativas contextuais de privacidade e riscos de exposição indevida. Por isso, evitamos reproduzir postagens identificáveis, nomear usuários ou destacar casos individuais. As análises são apresentadas em nível coletivo, priorizando padrões agregados e reduzindo os riscos de estigmatização, exposição ou amplificação de discursos hostis.

ACKNOWLEDGMENTS

Esta pesquisa foi financiada por: CNPq, Capes, Fapemig, Fapesp, AWS, NVIDIA, CHIA-Saúde e INCT-TILD-IAR.

Ferramentas de Inteligência Artificial Generativa foram utilizadas como suporte à revisão gramatical de todo o conteúdo e, conforme descrito na Seção 3.2.2, na etapa de sumarização e agrupamento semântico dos tópicos. Além disso, como detalhado na Seção 3.3.1, modelos de linguagem foram empregados para classificar os subreddits a partir de seu alinhamento ideológico. Os autores realizaram uma auditoria humana e uma revisão completa de todo o conteúdo gerado pelo modelo, assumindo responsabilidade pela integridade das informações apresentadas.

REFERÊNCIAS

- [1] Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115, 37 (2018), 9216–9221. doi:10.1073/pnas.1804840115
- [2] A Bhuvaneswari, M Kumudha, et al. 2024. Topic Modeling Based Clustering of Disaster Tweets Using BERTopic. In *2024 MIT Art, Design and Technology School of Computing International Conference (MITADTSOCiCon)*. IEEE, 1–6.
- [3] 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*. 8440–8451.
- [4] Gustavo F Cunha and Ana Paula Couto da Silva. 2024. Caracterizando polarização em redes sociais: Um estudo de caso das discussões no reddit sobre as eleições brasileiras de 2018 e 2022. In *Brazilian Symposium on Multimedia and the Web (WebMedia)*. SBC, 365–369.
- [5] 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*. 87–96.
- [6] Paula Fortuna, Juan Soler, and Leo Wanner. 2020. Toxic, hateful, offensive or abusive? What are we detecting? An expert survey. In *Proceedings of the 12th Language Resources and Evaluation Conference*. European Language Resources Association (ELRA), Marseille, France, 6799–6808.
- [7] Tiana Gaudette, Ryan Scrivens, Garth Davies, and Richard Frank. 2021. Upvoting extremism: Collective identity formation and the extreme right on Reddit. *New Media & Society* 23, 12 (2021), 3491–3508.
- [8] Yingqiang Ge, Shuya Zhao, Honglu Zhou, Changhua Pei, Fei Sun, Wenwu Ou, and Yongfeng Zhang. 2020. Understanding echo chambers in e-commerce recommender systems. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*. 2261–2270.
- [9] Debbie Ging. 2019. Alphas, betas, and incels: Theorizing the masculinities of the manosphere. *Men and masculinities* 22, 4 (2019), 638–657.
- [10] Maarten Grootendorst. 2022. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv:2203.05794* (2022).
- [11] Laura Hanu and Unitary team. 2020. Detoxify. Github. <https://github.com/unitaryai/detoxify>.
- [12] Alexander Hoyle, Pranav Goel, Andrew Hian-Cheong, Denis Peskov, Jordan Boyd-Graber, and Philip Resnik. 2021. Is automated topic model evaluation broken? the incoherence of coherence. *Advances in neural information processing systems* 34 (2021), 2018–2033.
- [13] Callum Jones, Verity Trott, and Scott Wright. 2020. Sluts and soyboys: MGTOW and the production of misogynistic online harassment. *New media & society* 22, 10 (2020), 1903–1921.
- [14] Alexander Krendel. 2020. The men and women, guys and girls of the ‘manosphere’: A corpus-assisted discourse approach. *Discourse & Society* 31, 6 (2020), 607–630.
- [15] Jack LaViolette and Bernie Hogan. 2019. Using platform signals for distinguishing discourses: The case of men’s rights and men’s liberation on Reddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 323–334.
- [16] Alyssa Lees, Vinh Q. Tran, Yi Tay, Jeffrey Sorensen, Jai Gupta, Donald Metzler, and Lucy Vasserman. 2022. A New Generation of Perspective API: Efficient Multilingual Character-level Transformers. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (Washington DC, USA) (KDD ’22)*. Association for Computing Machinery, New York, NY, USA, 3197–3207. doi:10.1145/3534678.3539147
- [17] Lucas Licurgo, Ghabriel Souza, Pedro de Moraes, Fabio Malini, Giovanni Comarella, and Athus Cavalini. 2025. Discourse Dynamics on X and Bluesky Amid Brazil’s 2024 Environmental and Political-Digital Crises. In *Brazilian Symposium on Multimedia and the Web (WebMedia)*. SBC, 540–544.
- [18] Max O Lorenz. 1905. Methods of measuring the concentration of wealth. *Publications of the American statistical association* 9, 70 (1905), 209–219.
- [19] Ana Cláudia Machado, Gabriel Prenassi, Elisa Tuler, and Leonardo Rocha. 2025. Análise Aprimorada das Percepções dos Usuários por Meio de Abordagens de Processamento de Linguagem Natural. In *Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia)*. SBC, 33–36.
- [20] Robin Mamié, Manoel Horta Ribeiro, and Robert West. 2021. Are anti-feminist communities gateways to the far right? Evidence from Reddit and YouTube. In *Proceedings of the 13th ACM Web Science Conference 2021*. 139–147.
- [21] Marta Marchiori Manerba, Riccardo Guidotti, Lucia Passaro, and Salvatore Ruggieri. 2022. Bias discovery within human raters: A case study of the jigsaw dataset. In *Proceedings of the 1st Workshop on Perspectivist Approaches to NLP@ LREC2022*. 26–31.
- [22] Victor Martins, Sophia Eduarda V Serafim, Livia Caroline R Pereira, Amanda F Alves, Carlos HG Ferreira, and Jussara M

- Almeida. 2025. A Misoginia no YouTube Brasileiro: Um Estudo de Caso sobre o Conteúdo Produzido pela Comunidade Red Pill. In *Brazilian Symposium on Multimedia and the Web (WebMedia)*. SBC, 28–37.
- [23] Adrienne Massanari. 2017. # Gamergate and The Fapping: How Reddit's algorithm, governance, and culture support toxic technocultures. *New media & society* 19, 3 (2017), 329–346.
- [24] David Mimno, Hanna Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. Optimizing semantic coherence in topic models. In *Proc. EMNLP '11*.
- [25] Joseph B Mountford. 2018. Topic modeling the red pill. *Social Sciences* 7, 3 (2018), 42.
- [26] Yida Mu, Mali Jin, Xingyi Song, and Nikolaos Aletras. 2024. Enhancing data quality through simple de-duplication: Navigating responsible computational social science research. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. 12477–12492.
- [27] Francisco Assis Ricarte Neto, Rafael Torres Anchieta, Raimundo Santos Moura, Pedro de A dos Santos Neto, and Andre Macedo Santana. 2025. Investigating Contextual Word Embeddings in Semi-Supervised Learning for Toxic Comment Detection. In *Brazilian Symposium on Multimedia and the Web (WebMedia)*. SBC, 321–330.
- [28] Tien T Nguyen, Pik-Mai Hui, F Maxwell Harper, Loren Terveen, and Joseph A Konstan. 2014. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd international conference on World wide web*. 677–686.
- [29] Antônio Pereira, Felipe Viegas, Marcos André Gonçalves, and Leonardo Rocha. 2023. Evaluating the limits of the current evaluation metrics for topic modeling. In *Proceedings of the 29th Brazilian Symposium on Multimedia and the Web*. 119–127.
- [30] Giovana Piorino, Luiz Henrique Quevedo Lima, Adriana Silvina Pagano, and Ana Paula Couto da Silva. 2025. Toxicidade e Gatilhos: Um Estudo de Caso em Comunidades do Reddit no Brasil. In *Brazilian Symposium on Multimedia and the Web (WebMedia)*. SBC, 575–579.
- [31] Giovana Piorino, Vitor Moreira, Luiz Henrique Quevedo Lima, Adriana Silvina Pagano, and Ana Paula Couto da Silva. 2024. Análise de sentimentos de conteúdo compartilhado em comunidades brasileiras do Reddit: Avaliação de um conjunto de dados rotulados por humanos. In *Brazilian Symposium on Multimedia and the Web (WebMedia)*. SBC, 54–62.
- [32] PullPush. 2025. PullPush Reddit API Documentation. <https://pullpush.io> Accessed: 2026-06-02.
- [33] 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.
- [34] Patrick Riehmann, Manfred Hanfler, and Bernd Froehlich. 2005. Interactive sankey diagrams. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005*. IEEE, 233–240.
- [35] Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20 (1987).
- [36] Giuseppe Russo, Manoel Horta Ribeiro, and Robert West. 2024. Stranger danger! cross-community interactions with fringe users increase the growth of fringe communities on reddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 18. 1342–1353.
- [37] Joni Salminen, Sercan Sengün, Juan Corporan, Soon-gyo Jung, and Bernard J Jansen. 2020. Topic-driven toxicity: Exploring the relationship between online toxicity and news topics. *PloS one* 15, 2 (2020), e0228723.
- [38] Dagmara Solska and Jan Sawicki. 2024. Decoding gender bias through a textual exploration of Reddit/r/MensRights community. *Beyond Philology An International Journal of Linguistics, Literary Studies and English Language Teaching* 21/1 (2024), 167–202.
- [39] Fernanda Luiza Tobias and Ana Paula Couto da Silva. 2025. Análise Linguística de Comentários de Humanos e de Modelos de Linguagem a Postagens em Comunidades Brasileiras do Reddit. In *Brazilian Symposium on Multimedia and the Web (WebMedia)*. SBC, 521–525.
- [40] UN Women. 2024. Glossary: The manosphere. <https://www.unwomen.org/en/articles/glossary/glossary-the-manosphere> Accessed: 09 de junho de 2026.
- [41] Alper Kursat Uysal and Serkan Gunal. 2014. The impact of preprocessing on text classification. *Information processing & management* 50, 1 (2014), 104–112.