LGBTfobia em Aplicativos de Mensageria Instantânea: Uma Análise por Redes de Co-compartilhamento

Fernanda Ferreira do Nascimento[1], **José Maria Monteiro**[1], **Javam Machado**[1]

[1]Universidade Federal do Ceará (UFC)
Av. Humberto Monte, s/n, Pici - CEP 60440-593 – Fortaleza – CE – Brasil

fernanda.nascimento@alu.ufc.br

{jose.monteiro, javam.machado}@lsbd.ufc.br

Resumo. *A LGBTfobia no ambiente digital tem sido investigada predominantemente em plataformas abertas, onde interações explícitas facilitam a modelagem de redes. Em aplicativos de mensageria instantânea, contudo, a circulação de conteúdos ocorre de forma mais opaca, exigindo estratégias específicas de análise. Este trabalho propõe um método para modelar a dinâmica do compartilhamento de mensagens LGBTfóbicas no WhatsApp e no Telegram por meio de redes de co-compartilhamento. Utilizamos o dicionário PrejudicePT-br para selecionar as mensagens de interesse e construímos uma rede usuário–usuário com arestas ponderadas indicando o co-compartilhamento de mensagens idênticas. A estrutura resultante foi avaliada por meio de medidas globais, componentes conexas e detecção de comunidades, complementadas por uma análise qualitativa que distingue gatilhos lexicais ambíguos de termos pejorativos explícitos. Os resultados revelam um padrão global fragmentado, com um núcleo coeso sustentado por um conjunto reduzido de mensagens. O estudo oferece um pipeline replicável e evidências empíricas sobre a estrutura da circulação de conteúdo LGBTfóbico em aplicativos de mensageria instantânea.*

Abstract. *LGBTphobia in the digital environment has been investigated predominantly on open platforms, where explicit interactions facilitate network modeling. In instant messaging applications, however, content circulation occurs in a more opaque manner, requiring specific analytical strategies. This work proposes a method to model the dynamics of LGBTphobic message sharing on WhatsApp and Telegram through co-sharing networks. We employ the PrejudicePT-br dictionary to select messages of interest and construct a user-to-user network with weighted edges indicating the co-sharing of identical messages. The resulting structure is evaluated through global measures, connected component analysis, and community detection, complemented by a qualitative analysis that distinguishes ambiguous lexical triggers from explicit pejorative terms. The results reveal a globally fragmented pattern, with a cohesive core sustained by a reduced set of messages circulating across platforms. The study provides a replicable pipeline and empirical evidence on the structural organization of LGBTphobic content circulation in instant messaging applications.*

1. Introdução

A expansão das plataformas de mensageria instantânea transformou profundamente a forma como informações e discursos circulam na sociedade brasileira. O WhatsApp mantém presença quase ubíqua entre usuários de *smartphones*, e o Telegram consolidou-se como alternativa relevante para grupos políticos e de mobilização [Kansaon et al. 2025, Machado et al. 2019]. Essas plataformas tornaram-se ecossistemas comunicacionais onde conteúdos de natureza política, religiosa e ideológica se propagam em larga escala, muitas vezes de forma coordenada e com baixa transparência.

O termo LGBTfobia é um construto científico amplo que designa o conjunto de atitudes, crenças, comportamentos e estruturas sociais de caráter hostil, discriminatório ou violento direcionados a pessoas lésbicas, gays, bissexuais, transgêneras e demais identidades que divergem das normas heterossexuais e cisgêneras hegemônicas [Warner 1991]. Trata-se de um fenômeno multidimensional que opera simultaneamente nos níveis individual, interpessoal, institucional e cultural [Herek 2004]. Infelizmente, a LGBTfobia também ocorre nos meios digitais e se propaga por meio de aplicativos de mensageria instantânea [Alorainy et al. 2022]. Relatórios institucionais recentes documentam um aumento expressivo nas denúncias de ataques motivados por orientação sexual e identidade de gênero em ambientes digitais no Brasil [SaferNet Brasil 2025].

A LGBTfobia no ambiente digital tem sido investigada majoritariamente em plataformas abertas, em especial o Twitter/X, nas quais as interações explícitas (p.ex., retuítes, comentários, menções, curtidas) facilitam a modelagem de redes de difusão [Bozhidarova et al. 2023, Unlu et al. 2025, Fan et al. 2024]. Em aplicações de mensageria instantânea, contudo, a circulação de conteúdos ocorre de forma mais opaca e frequentemente se manifesta por meio de replicação textual (copiar/colar ou encaminhamento), o que exige estratégias específicas de identificação e análise.

Este trabalho propõe uma estratégia para modelar e analisar a dinâmica do compartilhamento de mensagens relacionadas à LGBTfobia em aplicativos de mensagens instantâneas (WhatsApp e Telegram) a partir de redes de co-compartilhamento. Inicialmente, utilizamos os termos do dicionário PrejudicePT-br relacionados à LGBTFobia para selecionar as mensagens de interesse a partir de um corpus de 2.977.323 mensagens coletadas de grupos públicos do WhatsApp e Telegram [Sousa et al. 2025b]. Em seguida, construímos uma rede do tipo usuário–usuário em que as arestas ponderadas indicam o co-compartilhamento de mensagens idênticas. Avaliamos a estrutura resultante por meio de medidas globais, de componentes conexas, da decomposição núcleo-periferia e da detecção de comunidades. Adicionalmente, conduzimos uma análise qualitativa das mensagens que sustentam o núcleo da rede, distinguindo gatilhos lexicais ambíguos de termos pejorativos explícitos.

Os resultados evidenciaram um padrão fragmentado no nível global (1.272 componentes em um grafo de 1.640 nós e 1.439 arestas), com um núcleo estrutural coeso e modular ($Q = 0{,}82$), sustentado por um conjunto reduzido de 175 mensagens únicas. A componente gigante concentra 1.268 das 1.439 arestas totais, e sua decomposição núcleo-periferia revela um clique completo de 22 usuários no *shell* máximo ($k = 21$). A análise qualitativa das 20 mensagens mais compartilhadas evidencia que 70% do volume de codifusão é produzido por narrativas político-eleitorais, de desinformação eleitoral e religiosas com gatilhos lexicais ambíguos, e não por discurso abertamente LGBTfóbico.

As principais contribuições deste estudo são: (i) um *pipeline* replicável para modelar a difusão de conteúdo em aplicativos de mensageria instantânea a partir de redes de co-compartilhamento; (ii) evidências empíricas sobre a organização estrutural da LGBTfobia em grupos públicos do WhatsApp e Telegram; e (iii) uma análise léxico-estrutural que revelou a instrumentalização da LGBTfobia como elemento retórico de mobilização política.

O restante deste artigo está organizado da seguinte forma. A Seção 2 discute os principais trabalhos relacionados. A Seção 3 descreve os conjuntos de dados utilizados. A Seção 4 detalha a metodologia proposta. A Seção 5 apresenta e discute os resultados obtidos. Por fim, a Seção 6 apresenta as conclusões e perspectivas de trabalhos futuro.

2. Trabalhos Relacionados

A detecção automática de linguagem ofensiva e de discurso de ódio tem sido extensivamente investigada em plataformas abertas como Twitter/X, Instagram e YouTube, nas quais a disponibilidade de APIs facilita a coleta [Vargas et al. 2022, Leite et al. 2020, Vargas et al. 2021]. Em português, destacam-se o corpus HateBR [Vargas et al. 2022], com 7.000 comentários anotados do Instagram em nove categorias de ódio, e o ToLD-Br [Leite et al. 2020], com 21.000 tweets rotulados em sete tipos de toxicidade. Ambos alcançaram melhores resultados com modelos baseados em BERT (*F1* até 0,88).

Para WhatsApp e Telegram, a escassez de bases rotuladas é uma limitação reconhecida. [Sousa et al. 2025b] apresentaram os primeiros *datasets* públicos de mensagens preconceituosas em português nessas plataformas: PrejudiceWhatsApp.Br e PrejudiceTelegram.Br, com 3.000 mensagens cada. Adicionalmente, os autores propuseram o dicionário PrejudicePT-br, contendo 842 termos em nove categorias. Os experimentos realizados obtiveram um *F1* de 0,868 com Gradient Boosting na tarefa de classificação de mensagens com ou sem preconceito. O presente trabalho explora os *datasets* PrejudiceWhatsApp.Br e PrejudiceTelegram.Br, além do dicionário PrejudicePT-br, com a finalidade de caracterizar a estrutura e a dinâmica do compartilhamento de mensagens LGBTFóbicas.

A análise da propagação de conteúdos em redes sociais por meio de grafos de co-compartilhamento tem raízes na literatura de comportamento coordenado inautêntico (*Coordinated Inauthentic Behavior* – CIB). [Pacheco et al. 2021] propuseram um arcabouço não supervisionado para detectar coordenação a partir de traços comportamentais como sincronização temporal e similaridade de conteúdo construindo redes ponderadas de usuários cujas arestas codificam a intensidade da atividade sincronizada.

Em contexto mais amplo, [Kansaon et al. 2025] demonstraram empiricamente que campanhas coordenadas no WhatsApp brasileiro operam predominantemente dentro de janelas de *burstiness* de até 60 segundos, validando a premissa de que a co-difusão síncrona é um sinal robusto de coordenação nessa plataforma. [Cinelli et al. 2022] investigaram o papel de contas coordenadas na cascata informacional do Twitter, mostrando que contas coordenadas tendem a ocupar posições mais altas nas cascatas e a infectar mais usuários não coordenados. O resultado reforça a relevância de identificar o núcleo estrutural das redes de co-difusão, objetivo central do presente trabalho.

A investigação da LGBTfobia digital concentra-se majoritariamente no Twitter/X. [Bozhidarova et al. 2023] construíram grafos de conhecimento combinando análise de sentimento e modelagem de tópicos (BERTopic) sobre *tweets*, correlacionando picos de toxicidade online com o aumento de crimes de ódio LGBTQ+ registrados pelo FBI. [Unlu et al. 2025] mapearam comunidades de ódio contra populações LGBTQ+ em redes de *retweet* finlandesas, identificando câmaras de eco com segregação política severa por meio do algoritmo de Leiden e análise de correspondência. [Alorainy et al. 2022] analisaram a robustez de redes de ódio e demonstraram que a remoção de nós com alto grau de saída (*super-spreaders*) é mais eficaz para fragmentar a rede do que a remoção dos produtores originais de conteúdo.

No contexto brasileiro, [Lima-Lopes 2025] estudaram padrões léxicos do conservadorismo religioso em reações a uma exposição de arte *queer* no Instagram, revelando como termos como "criança", "família" e "pedofilia" co-ocorrem para construir narrativas de desumanização LGBTQIA+. Esse achado é consistente com o que observamos nas mensagens mais compartilhadas na componente gigante da nossa rede.

A análise dos trabalhos relacionados confirma a forte concentração dos estudos na plataforma Twitter/X, a predominância de redes de *retweet* e a integração ainda limitada entre a análise estrutural e a análise semântica, com escassez particular de estudos em aplicativos de mensageria instantânea e, em particular, na realidade brasileira. Este trabalho aborda diretamente essas lacunas.

3. Conjuntos de Dados

Com o objetivo de analisar a circulação de mensagens relacionadas à LGBTfobia em grupos públicos de WhatsApp e Telegram, utilizamos quatro conjuntos de dados coletados pela plataforma **BATMAN** [1](Big dATa platforM for misinformAtion moNitoring) [de Sá et al. 2023]. Esses conjuntos reúnem todas as mensagens que circularam nos grupos observados. Na Tabela 1, apresentamos as estatísticas básicas calculadas para cada *dataset*, incluindo o número de grupos, de usuários, de mensagens e o período de coleta.

Tabela 1. Estatísticas Básicas dos Conjuntos de Dados Utilizados.

Dataset	Grupos	Usuários	Mensagens	Período
dataset_zap_1	306	10.243	163.748	ago–out/2022
dataset_zap_2	229	9.102	191.377	nov/2022–fev/2023
telegram_dataset_01	177	7.721	472.430	set–out/2022
telegram_dataset_02	154	9.479	640.588	nov/2022–fev/2023
Total	**516**	**29.563**	**1.468.143**	**ago/2022–fev/2023**

A coleta foi realizada por meio do monitoramento automatizado de *grupos e canais públicos* (acessíveis por link de convite). Foram selecionados grupos nos quais circularam conteúdos associados à extrema-direita brasileira, mais precisamente material relacionado a um dos seguintes temas: nacionalismo, militarismo, anticomunismo e ataques ao sistema judiciário brasileiro. Esse recorte temático é relevante para a interpretação dos resultados, pois o corpus não constitui uma amostra aleatória do discurso sobre LGBTfobia no Brasil, mas sim um recorte de sua circulação em um ecossistema político-ideológico específico.

[1]Disponível em: `https://www.faroldigital.info/`

Os aplicativos de mensageria não fornecem, de forma consistente, um identificador de encaminhamento de mensagens (*forward*) entre grupos distintos. Por esta razão, operacionalizamos o co-compartilhamento por meio de *templates* textuais. Assim, mensagens com textos idênticos são mapeadas para o mesmo identificador de conteúdo, o que permite detectar replicação literal (copiar/colar e/ou encaminhamento) entre os usuários e os grupos observados. As limitações dessa estratégia decorrem da observabilidade: (i) captura apenas co-compartilhamentos dentro do universo de grupos públicos monitorados, não incluindo grupos privados; e (ii) em canais configurados com postagem restrita (p.ex., apenas administradores), pode haver muitos leitores e poucos publicadores, o que afeta a distribuição de atividade e a conectividade observada. Cabe destacar ainda que o total de mensagens após a fase de pré-processamento (1.468.143) é inferior à soma bruta dos quatro arquivos originais (2.977.323), uma vez que mensagens vazias, duplicatas e *spans* foram descartadas (ver Seção 4.1).

Com o objetivo de proteger a privacidade dos usuários e atender às exigências da Lei Geral de Proteção de Dados Pessoais (LGPD), bem como das políticas de privacidade das plataformas, os dados coletados foram anonimizados mediante a aplicação de funções *hash* sobre atributos identificadores, como números de telefone e identificadores de grupos. Logicamente, a anonimização das informações sensíveis impõe limitações à interpretação semântica dos resultados. Em particular, torna-se inviável vincular narrativas específicas a atores identificáveis. Apesar dessa restrição, a estrutura da rede e as métricas de co-difusão preservam informações suficientes para as análises conduzidas.

Para a identificação de mensagens relacionadas à LGBTfobia, utilizamos o dicionário **PrejudicePT-br** [Sousa et al. 2025a], desenvolvido especificamente para a detecção de linguagem preconceituosa em plataformas de mensageria instantânea em português. O dicionário é composto por 842 termos organizados em nove categorias de preconceito, dentre as quais a categoria **LGBTFobia**.

4. Metodologia

A Figura 1 ilustra o *pipeline* utilizado neste trabalho, o qual é composto por quatro etapas sequenciais: (i) pré-processamento dos textos, (ii) seleção de mensagens relacionadas à LGBTFobia, (iii) construção da rede de co-compartilhamento e (iv) análise estrutural. As subseções a seguir detalham cada uma dessas etapas. Todos os códigos e os resultados obtidos estão disponíveis em nosso repositório *online*[2].

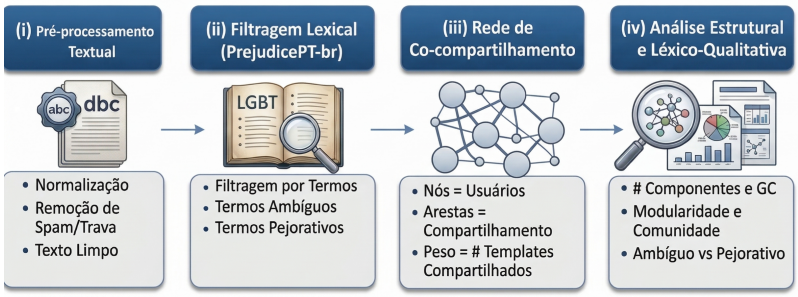

Figura 1. Pipeline Utilizado para Análise de Mensagens LGBTfóbicas.

[2]https://github.com/jmmfilho/complexnetworks-whatsapp-2022elections/tree/main/SBBD2026/LGBTfobia

4.1. Pré-processamento de Textos

Na etapa de pré-processamento, o texto de cada mensagem capturada (1.468.143 mensagens, ver Seção 3) passa por um conjunto de transformações: conversão para minúsculas; substituição de URLs pelo seu domínio (*e.g.*, `folha.uol.com.br`); remoção de pontuação e caracteres especiais; colapso de múltiplos espaços; tokenização com `word_tokenize` da biblioteca NLTK [Bird et al. 2009]; remoção de *stopwords*; conversão de emojis em *tokens* via `emoji.demojize`[3]; além da remoção de mensagens de *spam* ou de *trava-zap*. O resultado é uma cadeia de caracteres padronizada.

4.2. Seleção de Mensagens Relacionadas à LGBTFobia

Para identificar mensagens potencialmente relacionadas à LGBTfobia, utilizamos a seguinte estratégia: caso uma determinada mensagem contenha pelo menos um dos termos presentes na categoria **LGBTFobia** do dicionário **PrejudicePT-br** [Sousa et al. 2025b], esta deveria ser selecionada. É importante destacar que a presença de um termo do dicionário não implica, necessariamente, que a mensagem seja LGBTfóbica, pois vários termos têm múltiplos sentidos. Essa questão é analisada em profundidade na Seção 5.6.

Após este processo de filtragem, obtivemos 8.468 mensagens (selecionadas a partir do conjunto inicial de 1.468.143 mensagens (Tabela 1). Em seguida, removemos as mensagens que continham identificadores de usuário inválidos (valores `nan`, `None` ou vazio) ou que não puderam ser corretamente anonimizados, resultando em 4.376 mensagens de 1.640 usuários e 239 grupos distintos. Dentre as 4.376 mensagens foram identificadas 2.269 mensagens distintas. A distribuição por plataforma é a seguinte: 2.070 mensagens do WhatsApp (47,3%) e 2.174 do Telegram (49,7%), com 132 mensagens onde não foi possível determinar a plataforma de origem (3,0%).

4.3. Construção da Rede de Co-compartilhamento

A partir dos pares (usuário, mensagem), construímos uma rede bipartida implícita $\mathcal{B} = (U, M, E_B)$, onde U é o conjunto de usuários, M é o conjunto de templates de mensagens (identificadores de conteúdo) distintos e $(u, m) \in E_B$ se o usuário u postou pelo menos uma vez o template m. Posteriormente, a partir de $\mathcal{B}$, geramos a projeção usuário–usuário: um grafo não-direcionado ponderado $G = (V, E, W)$, onde $V = U$. Uma aresta $(u, v) \in E$ existe se u e v co-compartilharam ao menos um template, e seu peso w_{uv} é o número de templates distintos compartilhados por ambos (u e v):

$$w_{uv} = |\{\, m \in M \;:\; (u, m) \in E_B \wedge (v, m) \in E_B \}|$$

A rede resultante contém 1.640 nós e 1.439 arestas, com densidade $\delta = 2|E|/(|V|(|V|-1)) = 0{,}00107$. Nós sem arestas (usuários que postaram apenas conteúdo único, sem co-compartilhamento) permanecem isolados na rede, preservando a contagem total de usuários.

[3]https://carpedm20.github.io/emoji/docs/index.html

4.4. Análise Estrutural

Inicialmente, aplicamos o algoritmo k-core [Batagelj and Zaveršnik 2003] para realizar a decomposição núcleo-periferia da rede G, o que permite responder três questões fundamentais: (i) *Quão coeso é o núcleo?* (medido pela densidade do subgrafo induzido por $k_{\max}$); (ii) *Quão concentrado é o núcleo?* (medido pela razão entre o número de nós em $k_{\max}$ e o total de nós da rede); e (iii) *Se o núcleo é homogêneo relação às plataformas?* (verificando a quantidade de usuários de cada plataforma presentes no núcleo). Em seguida, obtemos a componente gigante (GC), ou seja, a maior componente conexa de G, e calculamos as seguintes métricas: Grau, número de vizinhos de cada nó; Grau Ponderado, soma dos pesos das arestas incidentes em um determinado nó; além da Centralidade de Intermediação [Freeman 1977], para cada nó v, mede-se a fração de caminhos mínimos entre todos os pares de nós que passam por v em relação ao número de pares possíveis. Posteriormente, aplicamos o algoritmo de Louvain [Blondel et al. 2008] em G para identificar comunidades, ou seja, agrupamentos naturais de nós em que as conexões dentro de cada agrupamento sejam densas e as entre agrupamentos distintos sejam esparsas. Trata-se, portanto, de um método de particionamento de grafos orientado pela maximização de uma função de qualidade denominada modularidade.

Para investigar diferenças estruturais entre WhatsApp e Telegram, construímos redes de co-compartilhamento independentes para cada plataforma, aplicando o mesmo procedimento da Seção 4.3. Em seguida, calculamos as mesmas métricas para cada rede. Adicionalmente, quantificamos as mensagens que forem compartilhadas exclusivamente em cada plataforma, bem como aqueles que foram compartilhadas em ambas as plataformas (cross-plataforma). Com a finalidade de analisar a distribuição das mensagens ao longo do tempo, agregamos o volume de mensagens, padronizamos esses valores utilizando z-score e correlacionamos as medidas computadas com os principais eventos do calendário político-eleitoral de 2022 e 2023.

Já para caracterizar o conteúdo compartilhado pelos usuários do núcleo da rede, isolamos as mensagens dos usuários da componente gigante (GC) e as comparamos com as mensagens compartilhadas pelos usuários da periferia da rede. Para cada subconjunto $S \in \{\text{GC}, \text{out}\}$ e cada termo t do dicionário, calculamos a taxa de ocorrência por 1.000 mensagens. O fator de enriquecimento do t no núcleo é dado por:

$$e(t) = \frac{\text{taxa}_{\text{GC}}(t)}{\text{taxa}_{\text{out}}(t)}$$

com $e(t) > 1$ indicando sobre-representação na GC. Os termos são classificados em dois grupos: (i) ambíguos: cujo sentido é frequentemente não-LGBTfóbico fora de contexto (*e.g., gay*, *ativo**, *passivo**, *todes*); e (ii) pejorativos explícitos os demais termos da categoria. Para as 20 mensagens com maior frequência na GC, recuperamos o seu texto e as classificamos manualmente em quatro categorias narrativas: (POL) político-governamental, (REL) religioso-moral, (LGBT) LGBTfóbico explícito e (DES) desinformação eleitoral.

5. Resultados e Discussão

5.1. Estrutura Global da Rede

A rede de co-compartilhamento G possui 1.640 nós e 1.439 arestas, com densidade $\delta = 0{,}00107$, o que é um valor típico de redes de difusão em mensageria instantânea, nas quais a organização em grupos isolados impede a formação de uma rede densa. A Tabela 2 resume as principais métricas calculadas para a rede de co-compartilhamento. Note que a rede se divide em 1.272 componentes, das quais a componente gigante (GC) concentra 256 nós (15,6%) e 1.268 arestas (88,1% do total). Essa assimetria (uma minoria de usuários responsável por quase toda a conectividade) é um primeiro indicativo de estrutura núcleo–periferia, confirmada nas subseções seguintes. A modularidade $Q = 0{,}822$ indica uma estrutura muito bem definida: o discurso LGBTfóbico não circula de forma homogênea, mas está segmentado em câmaras de eco relativamente independentes.

A análise de centralidade de intermediação sobre a GC revela uma oligarquia de *bridges*: quatro nós concentram *betweenness* de 0,459, 0,437, 0,418 e 0,416, com queda abrupta para 0,195 no quinto nó. Esses quatro usuários são estruturalmente indispensáveis à coesão da GC. A remoção desses usuários fragmentaria a difusão de conteúdos na rede, tornando-os alvos prioritários para intervenções estruturais.

Tabela 2. Principais Métricas da Rede de Co-compartilhamento.

Métrica	Valor
Nós (usuários)	1.640
Arestas	1.439
Densidade	0,00107
Componentes conexas	1.272
Componente gigante (GC)	256 nós · 1.268 arestas
GC / total	15,6% nós · 88,1% arestas
Modularidade Q (total)	0,822
Modularidade Q (GC)	0,782
k máximo	21
Top betweenness (GC)	0,459 · 0,437 · 0,418 · 0,416

5.2. Decomposição Núcleo–Periferia

A Figura 2 ilustra a distribuição dos *k-shells* resultantes do algoritmo de decomposição núcleo-periferia *k-core*. Esse algoritmo produziu *shells* com k variando de 0 a 21. A distribuição é fortemente assimétrica: 1.204 nós (73,4%) possuem $k = 0$, sem estrutura de co-difusão relevante. Os *shells* intermediários ($1 \leq k \leq 16$) distribuem-se gradualmente, e apenas os *shells* superiores ($k = 20$: 20 nós; $k = 21$: 22 nós) concentram os usuários de maior *embeddedness*. O resultado mais expressivo é o do núcleo estrito ($k = 21$, 22 usuários, 1,3% do total): a densidade do subgrafo induzido por esses 22 nós é 1,0 formam um **clique completo**, onde cada usuário co-compartilhou conteúdo com todos os outros 21. Todos pertencem a grupos do WhatsApp. A correlação entre número de core e grau ponderado na GC é $r = 0{,}923$, confirmando que posição estrutural e volume de co-difusão são praticamente equivalentes nesse subgrupo. Essa estrutura responde às três perguntas levantadas na metodologia: o núcleo é máximo em coesão (densidade = 1,0), altamente concentrado (1,3% dos nós) e homogêneo em plataforma (100% WhatsApp). O critério $k \geq k_{\max} - 1$ (núcleo expandido: 42 nós, 2,6%) oferece um limiar operacional para sistemas de monitoramento de comportamento coordenado.

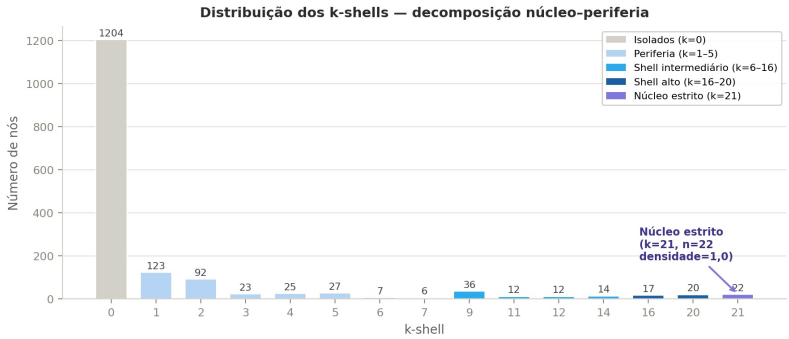

Figura 2. Distribuição dos K-shells da Decomposição Núcleo-Periferia.

5.3. Análise de Comunidades

O algoritmo de Louvain identificou comunidades com modularidade $Q = 0{,}822$ no grafo completo e $Q = 0{,}782$ na GC. Esses valores são considerados muito altos, indicando câmaras de eco bem delimitadas [Newman and Girvan 2004]. A Figura 3 exibe a estrutura de comunidades da GC. Já a Tabela 3 detalha as maiores comunidades da GC.

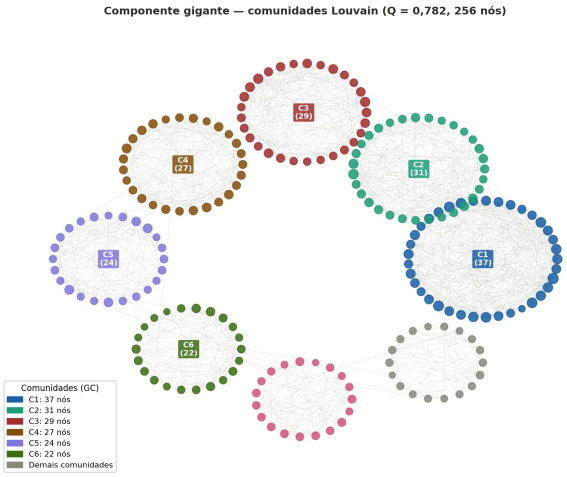

Figura 3. Componente Gigante com Comunidades Identificadas pelo Algoritmo de Louvain ($Q = 0{,}782$, 256 nós). Cada cor representa uma comunidade e o tamanho do nó é proporcional ao grau.

A GC é composta por 15 comunidades detectadas, das quais as 8 maiores somam 207 nós (80,9% da GC). As comunidades são relativamente homogêneas em tamanho: a maior (C1, 37 nós) representa 14,5% da GC e a menor das principais (C8, 18 nós) representa 7,0%, sem dominância de uma única comunidade. Esse padrão de distribuição equilibrada, combinado com alta modularidade, indica que a co-difusão de conteúdo LGBTfóbico está distribuída por múltiplos subgrupos coesos, sem que um único *cluster* concentre a maior parte da atividade. A elevada modularidade tem implicações diretas para estratégias de moderação: intervenções em uma comunidade têm impacto limitado nas demais. Os quatro *bridges* identificados pela centralidade de intermediação são, portanto, os principais vetores de propagação cross-comunidade e os pontos de maior alavancagem para reduzir a difusão de conteúdo.

Tabela 3. Maiores Comunidades na Componente Gigante (Louvain, $Q = 0{,}782$).

Comunidade	Nós	% da GC
C1 (id=36)	37	14,5%
C2 (id=52)	31	12,1%
C3 (id=20)	29	11,3%
C4 (id=9)	27	10,5%
C5 (id=33)	24	9,4%
C6 (id=22)	22	8,6%
C7 (id=10)	19	7,4%
C8 (id=0)	18	7,0%
Demais (7)	49	19,1%
Total	**256**	**100%**

5.4. Comparação entre WhatsApp e Telegram

A Figura 4 e a Tabela 4 comparam as redes construídas independentemente para cada plataforma investigada. O WhatsApp apresenta uma rede maior e mais densa em termos absolutos (1.095 arestas vs. 173 no Telegram), com GC proporcionalmente maior (9,3% vs. 6,2%). O Telegram, contudo, exibe um *bridge* de centralidade consideravelmente mais elevado (betweenness $= 0{,}668$ vs. $0{,}441$), indicando que, nessa plataforma, um único ator domina o fluxo entre comunidades de forma muito mais intensa. Consequentemente, as estratégias de moderação adequadas diferem: no WhatsApp, a estrutura distribuída exige endereçamento de múltiplos *bridges*; no Telegram, um único ponto de intervenção pode ser suficiente.

Quanto à sobreposição de conteúdo, das 2.183 mensagens identificadas, apenas 32 (1,5%) circularam simultaneamente nas duas plataformas. No entanto, entre as 20 mensagens mais compartilhadas (virais) da GC, 12 (60%) são cross-plataforma. As duas plataformas operam como ecossistemas majoritariamente independentes, mas o conteúdo que consegue atravessar essa barreira é desproporcionalmente o mais replicado. Uma lista das top-100 mensagens mais codifundidas capturaria a maior parte do conteúdo coordenado entre o WhatsApp e o Telegram, podendo ser atualizada periodicamente e compartilhada entre equipes de moderação.

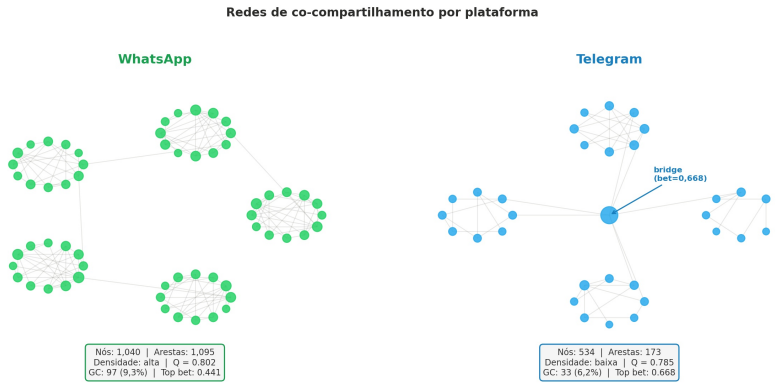

Figura 4. Redes de Co-compartilhamento por Plataforma.

Tabela 4. Comparação Estrutural das Redes por Plataforma.

Métrica	WhatsApp	Telegram	Unificada
Nós	1.040	534	1.640
Arestas	1.095	173	1.439
Densidade	0,0020	0,0012	0,0011
GC (nós / %)	97 / 9,3%	33 / 6,2%	256 / 15,6%
Modularidade Q	0,802	0,785	0,822
Top betweenness	0,441	0,668	0,459

5.5. Dinâmica Temporal

A Figura 5 apresenta o número de mensagens compartilhadas semanalmente. As 4.376 mensagens selecionadas cobrem 27 semanas (19/08/2022 a 14/02/2023), com volume médio de 162,1 mensagens por semana (mediana: 144,0). Três semanas apresentam volume estatisticamente discrepante ($z > 1{,}5$), todas em outubro de 2022 (Tabela 5). O pico ocorre na semana anterior ao primeiro turno (424 mensagens, 190 usuários, $z = 2{,}44$). O volume permanece elevado nas duas semanas em torno do segundo turno, indicando que a LGBTfobia circulou como elemento de mobilização ao longo de todo o mês de outubro. Após o resultado eleitoral, o volume cai em novembro, recupera-se em dezembro e volta a subir em janeiro de 2023 (período dos acampamentos nos quartéis).

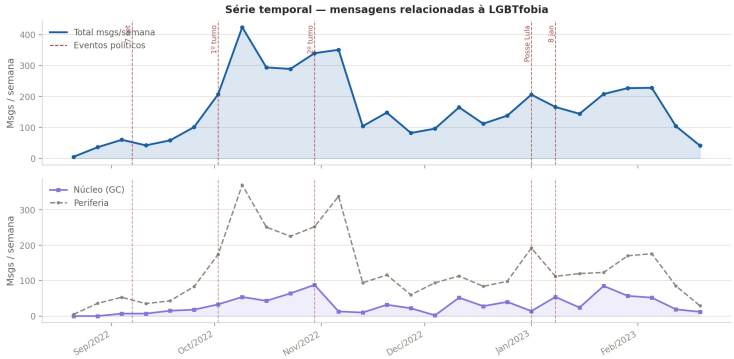

Figura 5. Número de Mensagens Compartilhadas Semanalmente: Volume Total (Painel Superior) e Núcleo vs. Periferia (Painel Inferior). Linhas tracejadas indicam eventos político-eleitorais.

Tabela 5. Semanas com Volume Atípico ($z > 1{,}5$) e Eventos Políticos.

Semana	Mensagens	z	Evento
03/10/2022	424	2,44	1º turno (02/10)
24/10/2022	340	1,66	2º turno (30/10)
31/10/2022	351	1,76	2º turno (30/10)

A decomposição núcleo–periferia revela papéis assimétricos: na semana do primeiro turno, a periferia produziu 370 mensagens contra 54 do núcleo (amplificadora de ocasião em picos eleitorais). O núcleo atinge seu volume mais alto na semana de 16/01/2023 (85 mensagens), articulando-se em momentos de crise institucional. A correlação entre os volumes semanais dos dois subconjuntos é moderada ($r = 0{,}449$): respondem ao mesmo contexto político, mas em momentos distintos.

5.6. Análise de Conteúdo

Na GC, 52,0% das mensagens contêm ao menos um termo ambíguo (*e.g., gay, ativo, passivo*) e 48,6% contêm ao menos um termo pejorativo explícito. Na periferia, a proporção se inverte: 44,8% de ambíguos e 55,7% de pejorativos. O núcleo difunde proporcionalmente mais linguagem codificada; a periferia usa mais vocabulário abertamente hostil. Os termos mais frequentes no núcleo são *bichas* ($e = 4{,}87$), *bichona* ($e = 2{,}43$) e *gays* ($e = 1{,}87$). Os menos recorrentes são *todes* ($e = 0{,}14$), *fresco* ($e = 0{,}25$) e *viado* ($e = 0{,}64$). O núcleo opera por estigmatização coletiva, a periferia por insultos individualizados. A Figura 6 apresenta as 20 mensagens mais frequentes na GC, classificadas em quatro categorias narrativas. A análise revela o achado central deste trabalho: 70% das mensagens virais na GC têm gatilho lexical ambíguo, ou seja, não são conteúdos abertamente LGBTfóbico. As mensagens explicitamente LGBTfóbicas (29% das ocorrências) têm gatilhos inequívocos. Assim, o discurso abertamente LGBTfóbico, embora presente, é menos replicado do que o discurso conservador que o envolve. Esse padrão caracteriza o que denominamos LGBTfobia instrumental: o discurso de ódio circula embutido em narrativas político-conservadoras mais amplas, não como tema autônomo. Neste cenário, sistemas de detecção automática de LGBTFobia baseados apenas em termos pejorativos podem não ser suficientes, sendo necessário a utilização de abordagens contextuais.

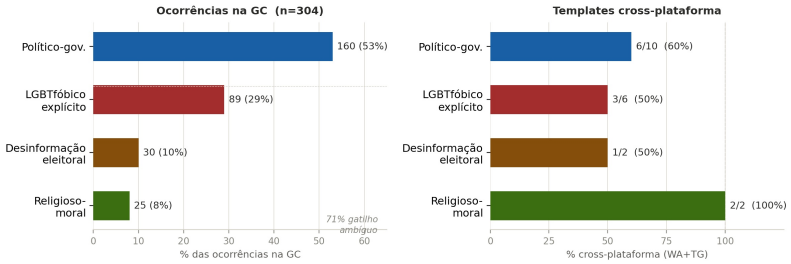

Figura 6. Distribuição das Top-20 Mensagens Virais da GC por Categoria Narrativa: Ocorrências Totais (Esquerda) e Proporção cross-plataforma (Direita).

6. Conclusão

Este trabalho apresentou uma estratégia para modelar e analisar a dinâmica do compartilhamento de mensagens relacionadas à LGBTfobia em aplicativos de mensagens instantâneas a partir de redes de co-compartilhamento. Os resultados evidenciaram um padrão fragmentado no nível global (1.272 componentes em um grafo de 1.640 nós e 1.439 arestas), com um núcleo estrutural coeso e modular ($Q = 0{,}82$), sustentado por um conjunto reduzido de 175 mensagens únicas. A componente gigante concentra 1.268 das 1.439 arestas totais, e sua decomposição núcleo-periferia revela um clique completo de 22 usuários no *shell* máximo ($k = 21$). A análise qualitativa das 20 mensagens mais compartilhadas evidencia que 70% do volume de co-difusão é produzido por narrativas político-eleitorais, de desinformação eleitoral e religiosas com gatilhos lexicais ambíguos, e não por discurso abertamente LGBTfóbico. A análise temporal mostrou que os picos de co-difusão estão diretamente acoplados ao calendário eleitoral.

Declaração sobre uso de IA Generativa

Ferramentas de IA generativa foram utilizadas para auxílio na estruturação do texto e revisão linguística. A responsabilidade pelo conteúdo técnico, experimentos e verificação das referências bibliográficas permanece integralmente com os autores.

Referências

Alorainy, W. et al. (2022). Disrupting networks of hate: Characterising hateful networks and removing critical nodes. In *Social Network Analysis and Mining*.

Batagelj, V. and Zaveršnik, M. (2003). An $o(m)$ algorithm for cores decomposition of networks. *arXiv preprint cs/0310049*.

Bird, S., Klein, E., and Loper, E. (2009). Natural language processing with python. O'Reilly Media.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008.

Bozhidarova, M. et al. (2023). Hate speech and hate crimes: A data-driven study of evolving discourse around marginalized groups. In *Proc. IEEE International Conference on Big Data*.

Cinelli, M., Cresci, S., Quattrociocchi, W., Tesconi, M., and Zola, P. (2022). Coordinated inauthentic behavior and information spreading on Twitter. *Decision Support Systems*, 160:113819.

de Sá, I. C., Galic, L., Franco, W., Gadelha, T., Monteiro, J. M., and Machado, J. C. (2023). BATMAN: A big data platform for misinformation monitoring. In Filipe, J., Smialek, M., Brodsky, A., and Hammoudi, S., editors, *Proceedings of the 25th International Conference on Enterprise Information Systems, ICEIS 2023, Volume 1, Prague, Czech Republic, April 24-26, 2023*, pages 237–246. SCITEPRESS.

Fan, L. et al. (2024). Characterizing online toxicity during the 2022 Mpox outbreak: A computational analysis of topical and network dynamics. *arXiv preprint*.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41.

Herek, G. M. (2004). Beyond "homophobia": Thinking about sexual prejudice and stigma in the twenty-first century. *Sexuality Research & Social Policy*, 1(2):6–24.

Kansaon, D. et al. (2025). From fake news to real protests: WhatsApp's role in brazilian political coordination. In *Proc. Int. AAAI Conf. on Web and Social Media (ICWSM)*.

Leite, J. A., Silva, D., Bontcheva, K., and Scarton, C. (2020). Toxic language detection in social media for Brazilian Portuguese: New dataset and multilingual analysis. In *Proc. 1st Conf. of the Asia-Pacific Chapter of the ACL and 10th IJCNLP*, pages 914–924.

Lima-Lopes, R. E. d. (2025). Lexical patterns of religious conservatism: A study of social media reactions to an art exhibition in Brazil. *Digital Studies / Le champ numérique*, 15(1).

Machado, C., Kira, B., Narayanan, V., Kollanyi, B., and Howard, P. N. (2019). A study of misinformation in WhatsApp groups with a focus on the Brazilian presidential elections. In *Companion Proc. The Web Conference (WWW)*, pages 1–4.

Newman, M. E. J. and Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113.

Pacheco, D., Hui, P.-M., Torres-Lugo, C., Truong, B. T., Flammini, A., and Menczer, F. (2021). Uncovering coordinated networks on social media: Methods and case studies. In *Proc. Int. AAAI Conf. on Web and Social Media (ICWSM)*, volume 15, pages 455–466.

SaferNet Brasil (2025). Indicadores da central nacional de denúncias de crimes cibernéticos – 2024. Technical report, SaferNet Brasil.

Sousa, M., do Nascimento, F. F., Martins, G., Monteiro, J. M., and Machado, J. C. (2025a). An approach for the automatic detection of prejudice in instant messaging applications. In *Proceedings of the 14th International Conference on Data Science, Technology and Applications (DATA 2025) - Volume 1: DATA*, pages 108–119, Bilbao, Spain. INSTICC, SciTePress.

Sousa, M., Nascimento, F., Martins, G., Monteiro, J. M., and Machado, J. (2025b). An approach for the automatic detection of prejudice in instant messaging applications. In *Proc. DATA 2025*.

Unlu, A. et al. (2025). Mapping the terrain of hate: Identifying and analyzing online communities and political parties engaged in hate speech. In *International Journal of Data Science and Analytics*.

Vargas, F., Carvalho, I., Rodrigues de Goes, F., Pardo, T., and Benevenuto, F. (2022). HateBR: A large expert annotated corpus of Brazilian Instagram comments for offensive language and hate speech detection. In *Proc. 13th Language Resources and Evaluation Conference (LREC)*, pages 7174–7183.

Vargas, F., Rodrigues de Goes, F., Carvalho, I., Benevenuto, F., and Pardo, T. (2021). Contextual-lexicon approach for abusive language detection. In *Proc. RANLP 2021*, pages 1438–1447.

Warner, M. (1991). Introduction: Fear of a queer planet. *Social Text*, (29):3–17.