Artificial Intelligence for Virality Analysis on Short-Form Video Platforms: A Taxonomy, Challenges, and Research Agenda

Sthefany do Socorro Souza da Silva[1], **Tiago de Melo**[1], **Carlos M. S. Figueiredo**[1], **Israel Gondres Torné**[1]

[1] PPGEEL - Programa de Pós-graduação em Engenharia Elétrica, Universidade do Estado do Amazonas, Manaus, 69050-020, Brasil

{sdssds.mee26,tmelo,cfigueiredo,itorne}@uea.edu.br

Abstract. *The growing popularity of short-video platforms such as TikTok, Instagram Reels, and YouTube Shorts has fostered the development of Artificial Intelligence techniques for content virality analysis. However, the diversity of approaches, metrics, and virality definitions still hinders the consolidation of knowledge in this field. This paper presents a structured exploratory review on virality analysis in short-video and related social media platforms. The reviewed studies are organized into a taxonomy comprising textual, visual, and multimodal approaches, enabling the identification of trends, limitations, and research gaps. Based on this analysis, challenges related to the definition of virality, class imbalance, and multimodal modeling are discussed, and future research directions are proposed.*

Resumo. *A popularização de plataformas de vídeos curtos, como TikTok, Instagram Reels e YouTube Shorts, impulsionou o uso de técnicas de Inteligência Artificial para analisar a viralização de conteúdos digitais. Porém, a variedade de abordagens, métricas e definições de viralização ainda dificulta a consolidação do conhecimento na área. Este trabalho apresenta uma revisão exploratória estruturada sobre análise de viralização em plataformas de vídeos curtos e redes sociais correlatas. Os estudos são organizados em uma taxonomia com abordagens textuais, visuais e multimodais, o que permite identificar tendências, limitações e lacunas de pesquisa. A partir dessa análise, discutem-se desafios na definição de viralização, no desbalanceamento de classes e na modelagem multimodal, além de se proporem direções futuras para o avanço da área.*

1. Introdução

O TikTok tornou-se uma das plataformas de mídia social mais influentes da atualidade, impulsionando o crescimento do consumo de vídeos curtos em escala global [Saheb et al. 2024]. Seu modelo baseado em recomendações algorítmicas permite que conteúdos alcancem grandes audiências em curtos intervalos de tempo, mesmo quando produzidos por criadores com poucos seguidores. Como consequência, a plataforma passou a exercer uma influência significativa em áreas como entretenimento, marketing digital, notícias, comportamento e formação de tendências [Cheng and Li 2024].

Nesse contexto, a viralização tornou-se um dos fenômenos mais relevantes das redes sociais contemporâneas. Vídeos que alcançam elevados níveis de visualização,

engajamento e compartilhamento podem gerar impactos econômicos e sociais expressivos. Entretanto, compreender quais fatores levam determinados conteúdos a se tornarem virais permanece um desafio complexo, envolvendo características do conteúdo, comportamento dos usuários, mecanismos de recomendação e dinâmicas específicas de cada plataforma [Jenders et al. 2013, Xu and Qian 2023].

Nos últimos anos, diferentes abordagens baseadas em Inteligência Artificial (IA) têm sido propostas para enfrentar esse problema. Técnicas de aprendizado de máquina, aprendizado profundo, Processamento de Linguagem Natural (PLN) [Devlin et al. 2019] e modelos multimodais [Vasanthi and Viswanatham 2024] são usadas para analisar informações textuais, visuais e contextuais associadas ao desempenho de conteúdos [Saheb et al. 2024]. Apesar dos avanços, persistem desafios importantes: a ausência de uma definição padronizada de viralização, a heterogeneidade dos conjuntos de dados e a dificuldade de integrar diferentes modalidades em uma única estrutura preditiva.

Diante desse cenário, este trabalho apresenta uma revisão exploratória estruturada sobre viralização em plataformas de vídeos curtos e redes sociais correlatas, abordando conceitos, técnicas de IA, fontes de dados, métricas e desafios de pesquisa. Sua principal contribuição é uma taxonomia de abordagens textuais, visuais e multimodais, sintetizando métodos, fontes de dados, limitações e desafios. Durante o levantamento bibliográfico, não foram identificadas revisões voltadas à análise de viralização em vídeos curtos com foco em técnicas de IA. Assim, o artigo oferece uma síntese temática e uma estrutura analítica para orientar novos estudos sobre análise e predição de viralização em vídeos curtos.

2. Metodologia da Revisão

A revisão segue uma abordagem exploratória estruturada, visando mapear e organizar contribuições recentes sobre análise de viralização e engajamento em vídeos curtos e redes sociais. Priorizaram-se estudos sobre viralização, popularidade, engajamento ou difusão de conteúdos com técnicas de IA, aprendizado de máquina, aprendizado profundo, PLN, visão computacional e modelos multimodais, aderentes à análise ou predição de desempenho em plataformas como TikTok, Instagram Reels, YouTube e correlatas, com preferência por publicações recentes e com descrição suficiente de dados, métodos e métricas. Excluíram-se trabalhos sem relação direta com análise computacional de engajamento, estudos puramente conceituais e publicações sem informações metodológicas mínimas.

A seleção dos trabalhos foi realizada por meio de buscas exploratórias em mecanismos de busca acadêmica, como Google Acadêmico, baseando-se nos termos-chave: *engagement prediction tiktok, short videos engagement prediction, virality tiktok, social media metrics, virality social media predicting, virality social media definition, content analysis social media, visual content analysis, popularity prediction*, entre outros. Em seguida, foram examinadas as referências dos estudos inicialmente identificados, utilizando combinações de termos, em português e inglês, relacionados a viralização, engajamento, vídeos curtos, redes sociais, aprendizado de máquina, aprendizado profundo e modelagem multimodal.

Cada estudo foi caracterizado por dimensões comuns: modalidade predominante

dos dados (textual, visual ou multimodal), tarefa (classificação, regressão ou predição de popularidade e engajamento), técnicas e arquiteturas, fontes de dados e plataformas, métricas de avaliação e limitações. Essa caracterização possibilitou comparar trabalhos geralmente avaliados de forma isolada, fundamentar a síntese comparativa e a taxonomia da Seção 4 e orientar a identificação de desafios e lacunas recorrentes. Por se tratar de uma revisão exploratória estruturada, e não de uma revisão sistemática exaustiva, o objetivo não foi quantificar toda a produção, mas organizar os principais eixos metodológicos da literatura e evidenciar lacunas que motivam novas investigações. Assim, os resultados devem ser entendidos como uma síntese estruturada das principais abordagens metodológicas, e não como um levantamento exaustivo de toda a literatura sobre o tema.

3. Fundamentos da Viralização e Métricas de Engajamento

A viralização tem sido estudada por diferentes áreas, sobretudo nas redes sociais. A expansão dessas plataformas e de seus algoritmos de recomendação intensificou a disseminação de conteúdos em larga escala, gerando altos níveis de visualizações e engajamento em pouco tempo [Denisova 2023]. O fenômeno relaciona-se aos processos de difusão e propagação de conteúdo, dependendo não só de aspectos visuais ou quantitativos, mas também de fatores sociais ligados à interação e ao comportamento coletivo dos usuários [Rathje and Van Bavel 2025]. Nessa perspectiva, importa menos o conteúdo em si e mais os efeitos de sua circulação, que fortalecem conexões sociais e ampliam a participação em ambientes digitais [Blommaert and Varis 2015]. Embora viralização, popularidade e engajamento sejam frequentemente usados de forma relacionada na literatura, este trabalho os trata como conceitos próximos, porém não equivalentes: popularidade refere-se ao alcance ou volume de atenção; engajamento, às interações dos usuários; e viralização, ao processo de disseminação acelerada de um conteúdo.

Um desafio recorrente é a falta de consenso sobre o que determina a viralização e como mensurá-la, já que cada plataforma tem seus próprios parâmetros [Roring 2024], sendo TikTok, YouTube e Twitter/X os contextos mais recorrentes na literatura. Essa dificuldade é reforçada por estudos que mostram que eventos virais nem sempre geram crescimento sustentado de engajamento, pois muitas vezes produzem apenas picos temporários de atenção seguidos de uma queda rápida, indicando que alcance elevado e impacto duradouro não são equivalentes [Sangiorgio et al. 2025].

As métricas de viralização variam por plataforma e contexto, sem consenso único; porém, visualizações, curtidas, comentários, compartilhamentos e seguidores são parâmetros recorrentes para avaliar o desempenho de conteúdos e campanhas [Trunfio and Rossi 2021]. Para Lemel (2021) [Lemel 2021], visualizações são o nível mais básico de engajamento e indicam o alcance inicial, diferentemente das impressões; curtidas e salvamentos expressam aprovação; comentários revelam interação mais profunda; e compartilhamentos indicam potencial de propagação. Além de indicadores de desempenho, essas métricas são centrais na economia da atenção, pois funcionam como sinais de relevância que influenciam os mecanismos de recomendação e o comportamento dos criadores [Mears and Beauvais 2025]. Assim, a viralização é geralmente operacionalizada por indicadores simples ou compostos, conforme o objetivo do estudo e a disponibilidade de dados.

4. Estado da Arte

A popularização das plataformas de vídeos curtos tem impulsionado abordagens em IA para compreender fatores de engajamento e viralização de conteúdos digitais. Pesquisadores exploram diferentes fontes de informação (textuais, visuais e contextuais) e métodos que vão de modelos estatísticos tradicionais a arquiteturas avançadas de aprendizado profundo [Ling et al. 2022]. Esta seção organiza o estado da arte por modalidade de dados (abordagens textuais, visuais e multimodais) e conclui com uma síntese comparativa que fundamenta a taxonomia proposta. Embora o foco sejam vídeos curtos, também são considerados estudos sobre viralização em redes correlatas, como Twitter/X, YouTube e plataformas baseadas em imagens, quando ajudam a explicar fatores de engajamento. A Figura 1 resume a organização conceitual adotada, relacionando fontes de dados, modelos de IA, formas de análise e desafios recorrentes.

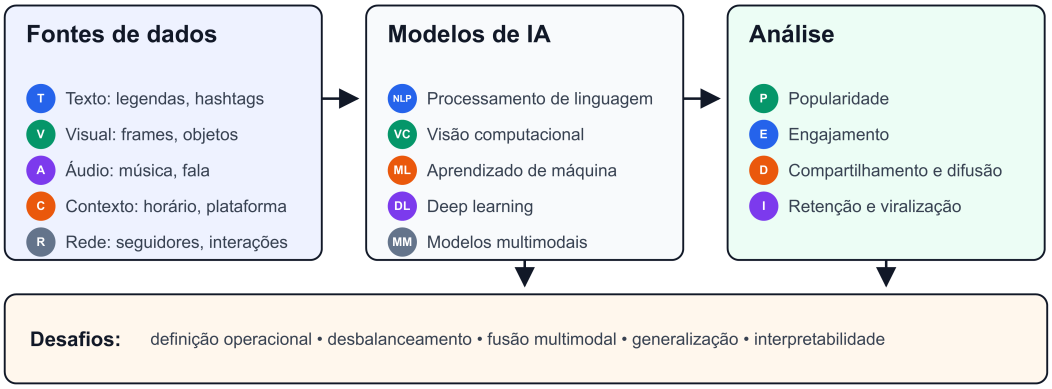


Figura 1. Framework conceitual para análise de viralização em vídeos curtos.

4.1. Abordagens Baseadas em Conteúdo Textual

As abordagens baseadas em conteúdo textual fundamentam-se nas características intrínsecas da mensagem, tais como estilo, semântica e carga emocional, que podem contribuir para o engajamento e a propagação em redes sociais e em componentes textuais de plataformas audiovisuais. Diferentemente das análises focadas na estrutura da rede, esta categoria explora dados linguísticos, hashtags e metadados para capturar a relevância de uma postagem [Elmas et al. 2023], com aplicações na antecipação da popularidade em marketing, saúde pública e disseminação de notícias.

A área evoluiu de modelos estatísticos tradicionais, de alta interpretabilidade, ao correlacionar termos específicos com a viralização [Mo et al. 2021], para arquiteturas de aprendizado profundo. Redes recorrentes com memória de longo prazo (LSTM) tratam o texto como sequências temporais e superam algoritmos clássicos como Naive Bayes e Voted Perceptron, com F1-score superior a 0.84 em notícias [Maldonado-Sifuentes et al. 2021b]. Mais recentemente, modelos baseados em *Transformers*, sobretudo variantes do BERT, tornaram-se predominantes: o ViralBERT combina a representação semântica do texto com análise de sentimentos e recursos numéricos (como contagem de seguidores e menções), lidando com a linguagem informal típica dessas plataformas [Rameez et al. 2022]; e modelos como o RuPERTa (RoBERTa em espanhol) mostram que o conteúdo textual, isoladamente, pode ser um preditor mais robusto que modelos centrados na rede, atingindo F1-scores de até 0.873 [Maldonado-Sifuentes et al. 2021a].

Apesar dos avanços, persistem limitações metodológicas críticas, sendo a principal delas o desequilíbrio extremo de classes e a inflação de zeros, dado que a vasta maioria das postagens nunca atinge níveis significativos de compartilhamento [Rameez et al. 2022]. Observa-se que a generalização desse tipo de modelo é frequentemente restrita a domínios específicos (notícias, saúde) ou a idiomas particulares, enfrentando desafios de escalabilidade devido ao alto custo computacional dos *Transformers* em comparação com métodos estatísticos mais rápidos, porém menos precisos [Maldonado-Sifuentes et al. 2021b, Maldonado-Sifuentes et al. 2021a].

4.2. Abordagens Baseadas em Conteúdo Visual

As abordagens baseadas em conteúdo visual consideram atributos como composição, enquadramento, presença de personagens, expressões faciais, cores dominantes, saliência visual, complexidade da cena e apelo estético, que influenciam a retenção de atenção do usuário e, consequentemente, podem afetar métricas de engajamento, como curtidas, visualizações, compartilhamentos ou republicações [Jo et al. 2025].

A literatura revela uma transição de descritores visuais interpretáveis para modelos computacionais mais flexíveis. Em estudos sobre a viralização de memes, a análise visual emprega atributos semanticamente compreensíveis (composição, escala da imagem, tipo de sujeito, presença de personagens, expressões emocionais e quantidade de texto embutido): imagens com sujeitos identificáveis, enquadramentos próximos e personagens expressando emoções tendem a se disseminar mais, enquanto imagens sem foco definido ou com excesso de texto são menos compartilhadas [Ling et al. 2021].

Outra vertente explora a extração automática de características visuais por visão computacional e aprendizado de máquina, transformando rótulos de objetos, tópicos visuais e cores representativas em variáveis estruturadas. [Jeong et al. 2024] demonstram que informações visuais extraídas com a Google Cloud Vision API melhoram a predição de popularidade de postagens baseadas em imagem, sobretudo combinadas a variáveis não visuais, e que modelos não lineares (Random Forest, XGBoost) superam os lineares, indicando que a influência do visual depende de interações complexas entre atributos da imagem, do usuário e do contexto. Complementarmente, [Göring and Raake 2023] investigam o *image appeal* sobre um novo conjunto de dados multiplataforma, considerando atributos como nitidez, contraste, saturação, saliência, profundidade e sentimento visual; os resultados mostram que o apelo visual é modelável computacionalmente, mas sua relação com curtidas e visualizações é limitada, sendo a estética apenas uma das dimensões da popularidade.

Em conjunto, as abordagens visuais avançaram em duas direções complementares: atributos interpretáveis derivados de teorias de composição e percepção, e recursos extraídos automaticamente por APIs e redes profundas, mais escaláveis. Ainda assim, são insuficientes para explicar isoladamente a viralização, exigindo a incorporação de movimento, áudio, texto sobreposto e interações entre modalidades.

4.3. Abordagens Multimodais

As abordagens multimodais representam uma das principais tendências na predição de viralização e engajamento, especialmente em ambientes dominados por conteúdos audiovisuais. Diferentemente dos métodos anteriores, que exploram uma única fonte de

informação, os modelos multimodais integram simultaneamente conteúdo visual, áudio, texto, metadados e sinais contextuais. A motivação decorre da natureza complexa da popularidade nas redes sociais: em vídeos curtos, a retenção da atenção depende não só da qualidade visual, mas também de trilha sonora, legendas, títulos, descrições e contexto de publicação, tornando a modelagem multimodal particularmente adequada para capturar os elementos associados ao engajamento [Roring 2024].

Trabalhos anteriores já indicavam que a previsão de viralidade não depende de uma única plataforma: [Vallet et al. 2015] investigaram conjuntamente métricas de popularidade no YouTube e sinais de propagação no Twitter, mostrando que características de engajamento têm forte capacidade preditiva para identificar conteúdos simultaneamente populares e virais, e que integrar informações de diferentes sistemas amplia essa capacidade.

Os estudos recentes evidenciam uma evolução das abordagens multimodais, que partem da simples combinação de atributos para arquiteturas capazes de modelar interações complexas entre características diversas. Em predição de popularidade de vídeos online [Sun et al. 2025, He and Li 2024], atributos do conteúdo audiovisual são combinados com metadados, estatísticas históricas e informações contextuais. A mesma tendência aparece no YouTube, onde categoria, duração, definição da mídia, curtidas, comentários e outras métricas de interação alimentam modelos de aprendizado de máquina para estimar a popularidade futura [Nisa et al. 2021]. Os resultados mostram que incorporar múltiplas fontes de informação tende a superar conjuntos restritos de atributos, evidenciando o valor da complementaridade entre modalidades.

O avanço recente também está associado a arquiteturas de aprendizado profundo e grandes modelos multimodais. O desafio VQualA 2025[1] para predição de engajamento em vídeos curtos demonstra que os modelos mais competitivos utilizam simultaneamente quadros do vídeo, trilha sonora, títulos, descrições e informações semânticas extraídas automaticamente. Os melhores sistemas combinam representações visuais e auditivas de grandes modelos multimodais, como VideoLLaMA2 e Qwen2.5-VL, com mecanismos de fusão entre modalidades: a integração de vídeo, áudio e texto produz ganhos consistentes, enquanto modelos exclusivamente textuais ou visuais apresentam desempenho inferior [Li et al. 2025]. Os experimentos também revelam que informações não visuais (títulos, descrições e classificações de áudio) contribuem significativamente para a previsão, reforçando a natureza multimodal da viralização. Nessa linha, [Li et al. 2024] propõem a base SnapUGC, com cerca de 90 mil vídeos curtos do Snapchat, e métricas específicas de engajamento, demonstrando que a fusão de características visuais, sonoras e textuais por atenção multimodal supera abordagens exclusivamente visuais.

Outra direção integra informações comportamentais e contextuais aos sinais multimodais do conteúdo. Estudos de detecção de microviralidade em TikTok mostram que combinar metadados do criador, características temporais, atributos de rede, informações textuais e elementos contextuais (como músicas em tendência e duração) produz modelos eficazes na identificação precoce de conteúdos com potencial de disseminação [Anggraeni et al. 2026]. Nesses cenários, métodos de ensemble como Random Forest e XGBoost têm desempenho elevado ao explorar interações não lineares

[1] https://openaccess.thecvf.com/ICCV2025_workshops/VQualA

entre grupos de atributos, reforçando a relevância da integração de múltiplas fontes de informação.

Estudos também investigam atributos independentes do conteúdo audiovisual: ao analisar o comportamento da plataforma e as interações dos usuários em diferentes regiões geográficas, mostra-se que indicadores como curtidas e comentários ajudam a prever a entrada de vídeos em listas de tendências, complementando as informações extraídas do conteúdo multimodal [Halim et al. 2022].

Outras pesquisas combinam conteúdo e comportamento, integrando métricas de engajamento e indicadores de influência social aos atributos do conteúdo, e apontam que a propagação depende simultaneamente da mensagem, da estrutura da rede, dos mecanismos de recomendação e do perfil dos criadores [Xu et al. 2025].

4.4. Síntese Comparativa e Taxonomia Proposta

A Tabela 1 apresenta uma síntese comparativa dos principais trabalhos revisados, destacando plataformas, métodos empregados e limitações reportadas pelos autores.

Tabela 1. Comparativo de estudos sobre análise e predição de viralização.

Plataforma	Método Principal	Limitação	**Referências**
Snapchat	Large Multimodal Models	Sem validação multiplataforma	**[Sun et al. 2025]**
Vine	CNN Multimodal Fusion	Contexto e plataforma específicos	**[He and Li 2024]**
Multiplataforma	Retrieval-Augmented GNN	Alto custo computacional	**[Xu et al. 2025]**
Snapchat	Deep Learning Multimodal	Dataset proprietário	**[Li et al. 2025]**
Twitter/X	BERT + Análise de Sentimentos	Dependência de atributos do usuário	**[Rameez et al. 2022]**
TikTok	Ensemble Learning	Dataset pequeno	**[Ling et al. 2022]**
TikTok	Random Forest	Dados desbalanceados	**[Anggraeni et al. 2026]**
Twitter/X	Transformer	Apenas uma plataforma	**[Elmas et al. 2023]**
YouTube	XGBoost	Dependência de metadados	**[Nisa et al. 2021]**
Multiplataforma	Machine Learning Classification	Dataset pequeno	**[Ling et al. 2021]**
Twitter/X	RNN-LSTM	Apenas dados textuais	**[Maldonado-Sifuentes et al. 2021b]**
Twitter/X	RoBERTa	Apenas dados textuais	**[Maldonado-Sifuentes et al. 2021a]**
Twitter/X	Generalized Linear Model	Dependência de texto	**[Mo et al. 2021]**
Instagram	XGBoost	Dependência de API externa	**[Jeong et al. 2024]**
Multiplataforma	CNN	Baixa correlação com popularidade	**[Göring and Raake 2023]**
Twitter/X	Regression Analysis	Restrito ao setor de moda	**[Jo et al. 2025]**
Multiplataforma	Hybrid Framework	Dataset temporal limitado	**[Vallet et al. 2015]**
Snapchat	Multimodal Deep Learning	Dataset proprietário	**[Li et al. 2024]**
YouTube	ANN + SVM + KNN	Apenas fatores não visuais	**[Halim et al. 2022]**

A literatura apresenta grande diversidade de plataformas e técnicas de IA: estudos recentes exploram arquiteturas multimodais, Transformers e modelos de linguagem de larga escala, enquanto trabalhos anteriores concentram-se em métodos estatísticos e algoritmos tradicionais. Independentemente da abordagem, limitações de disponibilidade de dados, dependência de plataformas específicas, desbalanceamento e dificuldades de generalização permanecem recorrentes, sugerindo que a modalidade dos dados é um dos principais fatores que diferenciam as abordagens. Com base nesses padrões, propõe-se uma taxonomia que organiza a literatura segundo a principal modalidade de informação utilizada, agrupando estudos metodologicamente semelhantes e facilitando a identificação de tendências e lacunas.

Como resultado, foram identificadas três categorias principais: abordagens **textuais** (legendas, hashtags e comentários, via PLN), **visuais** (atributos de imagens e vídeos

relacionados à atenção do usuário) e **multimodais** (integração de texto, imagem, metadados e sinais de interação social em representações mais abrangentes).

A Tabela 2 sintetiza as principais características de cada categoria, dados utilizados, métodos predominantes e limitações mais frequentes, permitindo comparar abordagens geralmente avaliadas de forma isolada.

Tabela 2. Taxonomia de IA para análise de viralização.

Abordagem	Dados utilizados	Principais métodos	Limitações
Textual	Tweets, legendas, hashtags, comentários e termos textuais	Regressão Logística, Hurdle Model, LSTM e modelos baseados em Transformers, como BERT, ViralBERT e RoBERTa	Não incorpora adequadamente informações visuais, sonoras ou contextuais, limitando a compreensão dos fatores associados à viralização.
Visual	Imagens, objetos detectados, cores dominantes, composição visual e atributos estéticos	Google Vision API, extração de características visuais, XGBoost, modelos estatísticos e redes neurais profundas	Depende de características visuais isoladas e apresenta dificuldade para capturar aspectos semânticos e contextuais do conteúdo.
Multimodal	Metadados, características visuais, informações textuais, métricas de engajamento e contexto social	Fusão de atributos, XGBoost, CNNs multimodais, aprendizado profundo e integração de múltiplas fontes de informação	Apresenta maior complexidade computacional, menor interpretabilidade e necessidade de modalidades de dados sincronizadas.

A taxonomia evidencia uma evolução gradual da literatura, que vai dos estudos textuais com modelos estatísticos tradicionais, passando pela incorporação de informações visuais com o avanço da visão computacional, até a migração recente para modelos multimodais que combinam texto, imagem, vídeo, áudio e sinais contextuais. Ela também explicita limitações recorrentes. As abordagens unimodais capturam apenas parte da natureza multifacetada da viralização, enquanto os modelos multimodais, apesar do maior potencial preditivo, impõem desafios de complexidade computacional, interpretabilidade e a necessidade de conjuntos de dados mais abrangentes e adequadamente anotados.

Embora construída a partir da modalidade dos dados, a literatura também pode ser analisada sob perspectivas complementares — a evolução temporal das técnicas, a plataforma investigada ou os fatores considerados para explicar a viralização. A Tabela 3 ilustra essas dimensões analíticas.

Assim, além de organizar o conhecimento existente, a taxonomia fornece uma base para a identificação das principais lacunas da literatura, discutidas a seguir.

5. Desafios de Pesquisa e Direções Futuras

Apesar dos avanços relevantes na aplicação de técnicas de IA para compreender a viralização nas redes sociais, a taxonomia proposta na Seção 4 evidencia desafios ainda

Tabela 3. Perspectivas complementares de classificação da literatura.

Perspectiva	Categorias possíveis
Modalidade dos dados	Textual, Visual, Multimodal
Evolução temporal	Métodos estatísticos, aprendizado profundo, Transformers, modelos multimodais
Plataforma	Twitter, TikTok, YouTube, Multiplataforma
Fatores analisados	Conteúdo, Usuário, Contexto sociocultural, Plataforma

persistentes, que podem ser organizados em cinco eixos. O primeiro é a **definição operacional de viralização**, já que diferentes trabalhos adotam visualizações, curtidas, compartilhamentos ou engajamento composto, sem um critério único e comparável. O segundo é o **desequilíbrio e a inflação de zeros**, pois a maioria dos conteúdos nunca viraliza, o que enviesa o treinamento e a avaliação dos modelos preditivos. O terceiro é a **multimodalidade e a fusão de sinais**, isto é, a integração de texto, vídeo, áudio, metadados e contexto em representações conjuntas. O quarto é a **generalização entre plataformas**, dado que TikTok, Instagram Reels, YouTube Shorts e X/Twitter possuem dinâmicas distintas que limitam a transferência de modelos. O quinto é a **interpretabilidade e a explicabilidade**, pois modelos multimodais preveem o desempenho, mas nem sempre explicam quais fatores impulsionam a viralização. Considerando o espaço disponível, são aprofundados a seguir três desafios centrais — a definição de viralização, o desbalanceamento dos conjuntos de dados e a modelagem multimodal —, por representarem as lacunas mais recorrentes e diretamente associadas ao desenvolvimento de métodos de IA para a análise de viralização.

5.1. Definição Operacional de Viralização

A ausência de uma definição consensual de viralização é uma das principais lacunas da literatura: diferentes estudos caracterizam conteúdos virais por visualizações, curtidas, compartilhamentos ou combinações dessas variáveis, com relevância que varia entre plataformas, levando a critérios de rotulagem incompatíveis que dificultam comparações diretas e *benchmarks* padronizados. Como a variável-alvo condiciona todas as etapas da modelagem, uma direção promissora é desenvolver *frameworks* formais que incorporem múltiplas métricas de engajamento, aspectos temporais e padrões de propagação, aumentando a reprodutibilidade e a comparabilidade.

5.2. Desequilíbrio de Classes

Os conjuntos de dados de viralização são fortemente desbalanceados, pois apenas uma pequena parcela dos conteúdos alcança disseminação elevada. Esse predomínio de exemplos não virais produz modelos enviesados para a classe majoritária, com acurácia alta mas baixa capacidade de identificar conteúdos virais, dificultando ainda a comparação entre estudos com diferentes estratégias de amostragem e avaliação. Pesquisas futuras podem explorar amostragem avançada, aprendizado sensível a custos, métricas adequadas a eventos raros e *datasets* mais representativos.

5.3. Multimodalidade

Embora a literatura tenha evoluído de abordagens unimodais para multimodais, a modelagem conjunta ainda é um dos maiores desafios. Plataformas como TikTok e Instagram

Reels combinam vídeo, áudio, texto, metadados e contexto, e a viralização emerge da interação entre esses fatores, de modo que modelos restritos a uma modalidade capturam apenas uma representação parcial. Destacam-se, como oportunidades, arquiteturas baseadas em atenção e grandes modelos multimodais capazes de aprender representações conjuntas de vídeo, áudio e texto, cujo avanço depende de dados bem anotados e de estratégias que conciliem desempenho, interpretabilidade e custo computacional.

6. Conclusão

Este trabalho apresentou uma revisão exploratória estruturada sobre análise de viralização em plataformas de vídeos curtos, investigando os principais conceitos, abordagens e desafios relacionados ao uso de Inteligência Artificial nesse contexto. A revisão permitiu identificar uma evolução gradual das estratégias utilizadas pela literatura, desde abordagens baseadas em modalidades isoladas até modelos multimodais capazes de integrar informações textuais, visuais e contextuais para compreender fatores associados ao engajamento e à disseminação de conteúdos digitais.

Como contribuição central, o trabalho sistematiza a literatura em uma taxonomia de abordagens textuais, visuais e multimodais, sintetizando dados, métodos e limitações. A análise também evidenciou desafios persistentes ligados à definição operacional de viralização, ao desbalanceamento dos conjuntos de dados e à modelagem multimodal, que continuam limitando a comparabilidade e a generalização dos resultados.

Com base nessas lacunas, discutiram-se direções futuras voltadas a *frameworks* mais consistentes para caracterizar a viralização, à construção de conjuntos de dados multimodais representativos e ao avanço de arquiteturas capazes de integrar modalidades de forma eficiente e interpretável. Assim, a taxonomia proposta oferece uma estrutura para comparar abordagens e orientar novos estudos empíricos, sobretudo voltados a bases multimodais, à definição operacional de viralização e à avaliação comparável entre plataformas.

Agradecimentos

Os autores agradecem o apoio da Universidade do Estado do Amazonas (UEA), por meio da Gratificação de Produtividade Acadêmica (GPA), conforme a Portaria nº 1177/2025-GR/UEA. Este trabalho contou com apoio do Instituto Nacional de Ciência e Tecnologia em Inteligência Artificial Responsável para Linguística Computacional e Tratamento e Disseminação de Informação (INCT-TILD-IAR), financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), processo nº 408490/2024-1.

Declaração sobre o Uso de Inteligência Artificial

Durante a preparação deste trabalho, os autores utilizaram ferramentas de Inteligência Artificial como apoio à revisão textual e à formatação. Todo o conteúdo científico foi concebido, revisado e validado pelos autores.

Referências

Anggraeni, R., Purwadi, and Subarkah, P. (2026). Detection of micro-viral content on TikTok through social listening and machine learning. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, 11(3):775–783.

Blommaert, J. and Varis, P. (2015). Conviviality and collectives on social media: Virality, memes, and new social structures. *Multilingual Margins: A journal of multilingualism from the periphery*, 2(1):31–31.

Cheng, Z. and Li, Y. (2024). Like, comment, and share on TikTok: Exploring the effect of sentiment and second-person view on the user engagement with TikTok news videos. *Social Science Computer Review*, 42:201–223.

Denisova, A. (2023). Viral journalism. strategy, tactics and limitations of the fast spread of content on social media: Case study of the united kingdom quality publications. *Journalism*, 24:1919–1937.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, pages 4171–4186. Association for Computational Linguistics.

Elmas, T., Stephane, S., and Houssiaux, C. (2023). Measuring and detecting virality on social media: The case of Twitter's viral tweets topic. In *Companion Proceedings of the ACM Web Conference 2023*, pages 314–317. ACM.

Göring, S. and Raake, A. (2023). Image appeal revisited: Analysis, new dataset, and prediction models. *IEEE Access*, 11:69563–69585.

Halim, Z., Hussain, S., and Ali, R. H. (2022). Identifying content unaware features influencing popularity of videos on YouTube: A study based on seven regions. *Expert Systems with Applications*, 206:117836.

He, Z. and Li, D. (2024). The short video popularity prediction using internet of things and deep learning. *IEEE Access*, 12:47508–47517.

Jenders, M., Kasneci, G., and Naumann, F. (2013). Analyzing and predicting viral tweets. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 657–664. ACM.

Jeong, D., Son, H., Choi, Y., and Kim, K. (2024). Enhancing social media post popularity prediction with visual content. *Journal of the Korean Statistical Society*, 53:844–882.

Jo, W., Kim, H., and Choi, J. (2025). A picture's worth a thousand shares: An empirical analysis of logo sizes in social media posts and their impact on content virality. *Marketing Letters*, 36:243–257.

Lemel, R. (2021). Determining which metrics matter in social media marketing. *Journal of Business & Retail Management Research*, 15.

Li, D., Li, W., Lu, B., Li, H., Ma, S., Krishnan, G., and Wang, J. (2024). Delving deep into engagement prediction of short videos. In *European Conference on Computer Vision*, pages 289–306. Springer.

Li, D., Ma, S., Hua, H., Li, W., Wang, J., Zhou, C. W., Guan, F., Li, X., Yu, Z., Lu, Y., et al. (2025). VQualA 2025 challenge on engagement prediction for short videos: Methods and results. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3391–3401.

Ling, C., AbuHilal, I., Blackburn, J., De Cristofaro, E., Zannettou, S., and Stringhini, G. (2021). Dissecting the meme magic: Understanding indicators of virality in image memes. *Proceedings of the ACM on human-computer interaction*, 5(CSCW1):1–24.

Ling, C., Blackburn, J., De Cristofaro, E., and Stringhini, G. (2022). Slapping cats, bopping heads, and oreo shakes: Understanding indicators of virality in tiktok short videos. In *Proceedings of the 14th ACM Web Science Conference 2022*, pages 164–173.

Maldonado-Sifuentes, C. E., Angel, J., Sidorov, G., Kolesnikova, O., and Gelbukh, A. (2021a). *Virality Prediction for News Tweets Using RoBERTa*, pages 81–95. Springer International Publishing.

Maldonado-Sifuentes, C. E., Sidorov, G., and Kolesnikova, O. (2021b). Improved twitter virality prediction using text and rnn-lstm. *International Journal of Combinatorial Optimization Problems and Informatics*, 12(3):50–62.

Mears, A. and Beauvais, T. (2025). Learning to like the likes and the hate: The labor of internet fame in the new attention economy. *Social Problems*, page spaf028.

Mo, C., Yin, J., Fung, I. C.-H., and Tse, Z. T. H. (2021). Aggregating Twitter text through generalized linear regression models for tweet popularity prediction and automatic topic classification. *European Journal of Investigation in Health, Psychology and Education*, 11:1537–1554.

Nisa, M. U., Mahmood, D., Ahmed, G., Khan, S., Mohammed, M. A., and Damaševičius, R. (2021). Optimizing prediction of YouTube video popularity using XGBoost. *Electronics*, 10:2962.

Rameez, R., Rahmani, H. A., and Yilmaz, E. (2022). ViralBERT: A user focused BERT-based approach to virality prediction. In *Adjunct Proceedings of the 30th ACM Conference on User Modeling, Adaptation and Personalization*, pages 85–89. ACM.

Rathje, S. and Van Bavel, J. J. (2025). The psychology of virality. *Trends in Cognitive Sciences*.

Roring, R. S. (2024). Decoding social media virality: Measuring impact, influence, and engagement in the digital age. *SSRN Electronic Journal*.

Saheb, T., Sidaoui, M., and Schmarzo, B. (2024). Convergence of artificial intelligence with social media: A bibliometric & qualitative analysis. *Telematics and Informatics Reports*, 14:100146.

Sangiorgio, E., Di Marco, N., Etta, G., Cinelli, M., Cerqueti, R., and Quattrociocchi, W. (2025). Evaluating the effect of viral posts on social media engagement. *Scientific Reports*, 15(1):639.

Sun, W., Cao, L., Cao, Y., Zhang, W., Wen, W., Zhang, K., Chen, Z., Lu, F., Min, X., and Zhai, G. (2025). Engagement prediction of short videos with large multimodal models. In *2025 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 3443–3452. IEEE.

Trunfio, M. and Rossi, S. (2021). Conceptualising and measuring social media engagement: A systematic literature review. *Italian Journal of Marketing*, 2021:267–292.

Vallet, D., Berkovsky, S., Ardon, S., Mahanti, A., and Kafaar, M. A. (2015). Characterizing and predicting viral-and-popular video content. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 1591–1600.

Vasanthi, P. and Viswanatham, V. M. (2024). Sentiment analysis of multi social media using machine and deep learning models: a review. *Multimedia Tools and Applications*, 83:90033–90051.

Xu, X., Zhang, Y., Zhou, F., and Song, J. (2025). Improving multimodal social media popularity prediction via selective retrieval knowledge augmentation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39:932–940.

Xu, Z. and Qian, M. (2023). Predicting popularity of viral content in social media through a temporal-spatial cascade convolutional learning framework. *Mathematics*, 11:3059.