Amrita Vishwa Vidyapeetham

Amrita School of Computing

Technical Report

An Analysis on Multi-Aspect Virality Prediction in Social Media Networks

Team : 25

AM.EN.U4AIE22137 : Nanditha Nandakumar
AM.EN.U4AIE22145 : S Gopika
AM.EN.U4AIE22150 : Shrisharanyan Vasu

Project Guide: Project Coordinator:

Signature : Signature :

June 30, 2025

Contents

1 Introduction **2**

2 Literature Survey **4**
2.1 Emotion-RGCNet: Emotion Recognition using RoBERTa and Graph Neural Networks (Yan et al., 2021)[1] 4
2.2 CasWarn: Early Cascade Virality Prediction (Gao et al., 2021)[2] . . . 5
2.3 Vector Centrality in Hypergraphs (Kovalenko et al., 2022)[3] 5
2.4 Predicting Retweets with Graph Attention Networks (Hsu et al., 2023)[4] 5
2.5 Higher-Order Rumor Diffusion via Simplicial Complexes (Zhou et al., 2025)[5] 5
2.6 Research Gap and Novelty 6

3 Proposed Methodology **8**
3.1 Module 1: Data Extraction and Preprocessing 9
3.2 Module 2: Sentiment and Emotion Analysis 9
3.3 Module 3: Graph Construction and Feature Engineering 9
3.4 Module 4: Virality Prediction Models 10
3.5 Module 5: Analytical Evaluation and Visualization and Algorithms . . 10
3.6 Outcome 10

4 Experimental Results **11**
4.1 Experimental Setup 11
4.2 Experiment 1: Diffusion and Centrality-Based Analysis 12
4.3 Experiment 2: Emotion-Sentiment and GNN-Based Classification . . . 12
4.4 Experiment 3 : Correlation-Based Feature Analysis 14
4.5 Experiment 4: Predicting Replies Based on Graph Position (Tweet Reply Model) 14
4.5.1 Network-Derived Features 14
4.5.2 Hypothesis 15
4.5.3 Results 15
4.5.4 Analytical Validation via Predictive Modeling 18
4.5.5 Interpretation 18

Abstract

This project investigates virality prediction in social media through a multi-aspect lens, with a strong emphasis on graph-based analysis. We explore how structural properties—such as centrality, similarity, and community membership—affect the spread of tweets, and contrast these with content-based factors like sentiment, emotion, and user metadata. Using a Twitter-based dataset, we first hypothesize that emotional tone and author popularity are key drivers of virality. However, our analysis and model evaluation reveal that these features have negligible predictive power. Instead, we find that temporal diffusion patterns and structural graph position—such as early entry in a growing thread—are far more influential. Although machine learning models (e.g., GraphSAGE) were used to test hypotheses, our focus remains analytical. We conclude that network structure and early diffusion behavior offer stronger explanatory value than content or user-level features, making them more reliable for early-stage virality assessment.

1 Introduction

In the modern digital era, social media platforms such as Twitter, Facebook, and Instagram have become epicenters for public discourse, activism, brand engagement, and information dissemination. A key phenomenon observed across these platforms is *virality* — the rapid and widespread propagation of content through user interactions. Understanding why certain posts go viral while others remain unnoticed is a central challenge for researchers across marketing, public health, and computational social science.

Virality prediction is inherently multifaceted, shaped by a complex interplay of timing, content characteristics, user influence, and network structure. While predictive modeling can be useful, analytical insight into the mechanics of virality—particularly in the early stages—is equally important. For example, early identification of viral misinformation can aid public safety, while brands can refine their strategies to maximize organic reach.

Motivation: During real-time events such as natural disasters or political movements, certain social media posts spread rapidly while others with similar content do not. This raises critical questions: Is virality primarily driven by emotional content? Or does a tweet's position in the evolving interaction graph play a greater role? Although prior work highlights the role of sentiment and user popularity, such features may not be reliable predictors in practice. Understanding the relative influence of structural versus content-based features is essential for building interpretable, robust models of virality.

Problem Statement: Traditional approaches to virality prediction often rely on static or content-based features—such as sentiment polarity, emotion category, or follower count—but these features alone have limited explanatory power, especially during early stages of content diffusion. This work investigates whether structural graph properties (e.g., centrality, similarity, position in the diffusion cascade) are stronger indicators of virality than surface-level features. We also examine how early diffusion signals evolve over time, and how they can be modeled without relying on feature leakage.

Challenges and Existing Work: Prior studies in virality prediction face several critical limitations, particularly when it comes to early-stage analysis and interpretability.

- **Data Leakage through Late-Stage Features:** Many models achieve high accuracy by using features like 24-hour reaction counts. However, such features are unavailable during the early diffusion phase and introduce data leakage. *In our study, we validated this issue experimentally—models using these features showed inflated performance. To address this, we excluded late-stage features and focused on early diffusion bins (e.g., 10min, 30min reaction counts) and structural metrics.*

- **Lack of Accessible User-User Networks:** Most social media datasets do not provide explicit follower or interaction graphs. *To overcome this, we constructed a tweet-level interaction graph using reply chains and semantic similarity between posts, enabling graph-based analysis without needing full user networks.*

- **Unreliable Sentiment and Emotion Detection:** Sentiment and emotion classifiers often struggle with informal, sarcastic, or emotionally complex content. *We included these features for completeness but later found them to have negligible correlation with virality—confirming their limited utility in this dataset.*

- **Siloed Treatment of Features:** Many existing works study sentiment, structure, or diffusion independently. *In contrast, our framework integrates all three perspectives to analyze their combined influence on virality.*

Proposed Approach: In this project, we adopt a hypothesis-driven, multi-aspect analytical framework to investigate what drives virality on social media. Initially, we test the prevailing assumption in prior work—that sentiment, emotional tone, and user metadata (e.g., follower count, verified status) are strong predictors of virality. Our experiments reveal that these content-based features exhibit weak correlation with actual engagement, particularly in the early stages of diffusion.

To overcome this limitation, we shift focus toward **structural and temporal signals** within the social graph. We construct a tweet-level interaction graph using reply chains and content similarity to compensate for the absence of explicit user-user connections, a common limitation in public datasets. We then extract features such as centrality measures, graph position, and time-binned diffusion rates.

Rather than building production-level predictive systems, we use machine learning models—including Graph Neural Networks (GraphSAGE)—as analytical tools to evaluate the explanatory power of different feature sets. Our findings validate that early-stage structural features—such as reply timing and position in a diffusion cascade—are far more informative than sentiment or metadata. This confirms that graph-based metrics provide stronger, more interpretable signals for understanding virality.

Objectives and Contributions: This work aims to evaluate the comparative influence of emotional, temporal, and structural factors on virality. Our key contributions are as follows:

- A hypothesis-driven analysis comparing content-based and structure-based virality drivers.
- A graph construction pipeline using tweet similarity and reply chains to model diffusion dynamics in the absence of user-user networks.
- A structural evaluation using centrality, graph position, and community detection metrics.
- Correlation and model-based validation demonstrating the limited utility of sentiment and author metadata.

2 Literature Survey

Understanding virality in social media requires an interdisciplinary approach, combining insights from emotion detection, graph modeling, and information diffusion. In this section, we review five foundational works that align closely with our research goals.

2.1 Emotion-RGCNet: Emotion Recognition using RoBERTa and Graph Neural Networks (Yan et al., 2021)[1]

Yan et al. proposed *Emotion-RGCNet*, a hybrid deep learning model that integrates RoBERTa for rich textual representations with Graph Neural Networks and Conditional Random Fields to capture relational and sequential emotion dependencies in social media posts. Their approach achieved state-of-the-art results on several emotion recognition benchmarks. However, the model's reliance on large-scale annotated datasets limits its adaptability to real-time or low-resource scenarios. Future directions include leveraging semi-supervised learning to reduce labeling requirements, enhancing streaming emotion detection for real-time applications, and expanding its applicability to cross-lingual and multimodal content.

2.2 CasWarn: Early Cascade Virality Prediction (Gao et al., 2021)[2]

Gao et al. introduced *CasWarn*, a deep learning-based framework aimed at predicting whether an information cascade will go viral. Unlike models that rely on explicit network structure, CasWarn uses network-agnostic features such as sentiment polarity, dissemination scale, and semantic evolution, demonstrating high effectiveness in early virality detection on Weibo and Twitter datasets. A key limitation is its lack of structural graph information, potentially reducing contextual accuracy. Enhancements could involve incorporating GNN-based relational learning, adapting to multilingual or non-textual data formats, and improving long-term virality forecasting.

2.3 Vector Centrality in Hypergraphs (Kovalenko et al., 2022)[3]

Kovalenko et al. proposed a vector-based centrality measure for hypergraphs, arguing that traditional scalar centrality fails to capture the multidimensional importance of nodes in higher-order networks. Their approach outperformed conventional metrics across various real-world datasets, highlighting its robustness. However, the lack of integration with learning models such as GNNs limits its practical utility in prediction tasks. Future work could focus on embedding vector centrality within neural frameworks, comparing it with dynamic centrality measures, and extending its use to domains beyond social media, such as biology or recommendation systems.

2.4 Predicting Retweets with Graph Attention Networks (Hsu et al., 2023)[4]

Hsu et al. developed a model using Graph Attention Networks (GATs) to forecast retweet behavior by learning from historical user interactions and tweet metadata. The GAT-based system accurately predicted early-stage engagement patterns, demonstrating the efficacy of attention mechanisms in modeling influence. Nonetheless, the model is limited to binary retweet predictions and assumes static user-tweet interactions, which may not reflect evolving social dynamics. Future research could focus on multi-class virality prediction, dynamic graph modeling to capture evolving relationships, and making attention outputs more interpretable.

2.5 Higher-Order Rumor Diffusion via Simplicial Complexes (Zhou et al., 2025)[5]

Zhou et al. addressed the limitations of traditional pairwise diffusion models by using simplicial complexes to simulate group-based rumor spread in social networks. Their framework captures higher-order interactions among users, offering a more nuanced representation of social dynamics. However, the approach suffers from high computational complexity, restricting its applicability to large-scale networks. Future improvements may include optimizing scalability, applying the model to real-time social streams, and exploring its integration with GNN architectures for enhanced expressiveness.

2.6 Research Gap and Novelty

Despite advancements in sentiment analysis, emotion detection, graph modeling, and diffusion theory, most prior approaches tend to analyze these aspects in isolation. Content-focused models often overlook network structure and timing, while graph-based models rarely account for emotional or linguistic context. This has created a fragmented understanding of virality in social media.

Our work initially set out to bridge this gap by integrating sentiment, emotion, and graph-based influence measures into a unified analytical framework. However, through hypothesis-driven experimentation, we discovered that content-based features—such as emotion categories, sentiment polarity, and author metadata—offered limited explanatory power for early virality.

Instead, we demonstrate that structural features—such as centrality, graph position, and early diffusion velocity—are far more predictive and interpretable. Unlike prior studies, we model individual posts (rather than users) as nodes in a tweet-level graph, extract early temporal signals, and validate feature utility through analytical modeling. This allows us to shift the focus from black-box prediction to interpretable insights grounded in social network analysis.

Our work thus contributes a refined understanding of what drives virality, emphasizing that early structure and connectivity—not content tone—are key to engagement prediction on social platforms.

In line with this shift, we introduce a novel regression-based target variable: `remaining_replies`. Instead of classifying whether a tweet is viral, our model predicts how many more replies will be posted after a given tweet within an ongoing conversation. This approach captures fine-grained engagement dynamics and enables interpretable forecasting at the reply level.

While earlier studies such as Backstrom et al. [6] have examined position-based forecasting within conversation threads, their focus was on thread growth patterns for platform curation and visualization. In contrast, we extend this idea by introducing a regression-based target—`remaining_replies`—which uses position and timing to estimate future conversational engagement at the reply level. This adds a measurable and interpretable forecasting layer suited for real-time virality modeling.

A broader summary of related works and their comparative insights is presented in Table 1.

Table 1: *Summary of the Related Works*

Ref	Title/Year	Problem Addressed	Contributions	Limitations	Open Problems
[1]	Emotion-RGC Net (Yan et al., 2025)	Emotion recognition in social media text	Combines RoBERTa with GNNs and CRFs for improved emotion detection	Requires large labeled datasets	Domain adaptation, multilingual emotion detection
[2]	CasWarn (Gao et al., 2021)	Early virality detection using content features	Uses emotional polarity, dissemination scale, and semantic evolution	Ignores network structure	Integrating network-based signals
[3]	Vector Centrality in Hypergraphs (Kovalenko et al., 2022)	Influence estimation in complex networks	Vector-based centrality outperforms scalar methods	Not widely adopted yet	Integration into social media pipelines
[4]	GAT-Based Retweet Prediction (Hsu et al., 2023)	Predicting early tweet engagement	GAT model with negative sampling for early-stage predictions	Sparse early data challenge	Robustness to minimal input
[5]	Rumor Spread via Simplicial Complexes (Zhou et al., 2025)	Group-level rumor propagation modeling	Introduces higher-order rumor diffusion using simplicial complexes	Complex modeling approach	Scalable real-world deployment
[7]	Opinion-Disease Diffusion (Alahmadi et al., 2025)	Modeling anti-vaccine opinion propagation	Coupled opinion-disease model using network dynamics	Limited to public health domain	Broader application to other misinformation topics
[8]	GCN for Retweet Count Estimation (Lo et al., 2023)	Real-time virality prediction	Adaptive GCN trained on historical user behavior	Generalization across topics	Topic/domain adaptation
[9]	HM-GNN (Wu et al., 2024)	Graph-based sentiment analysis	Combines GCN, GAT, and attention for syntactic-aware sentiment analysis	High computational complexity	Lightweight or scalable variants

3 Proposed Methodology

Our approach to understanding and analyzing virality in social media is structured around a multi-aspect, graph-centric analytical framework. While we initially incorporate affective features such as sentiment and emotion to test their correlation with virality, our primary emphasis lies in modeling structural and temporal aspects using Social Network Analysis (SNA) techniques.
The proposed framework integrates early diffusion metrics, graph-based modeling, and (to a lesser extent) content-based signals into a unified analysis pipeline. The architecture is divided into five key modules, each addressing a specific stage in the data processing and hypothesis-testing workflow. This design supports a more interpretable and structured understanding of what drives virality and how information diffuses through social media ecosystems.

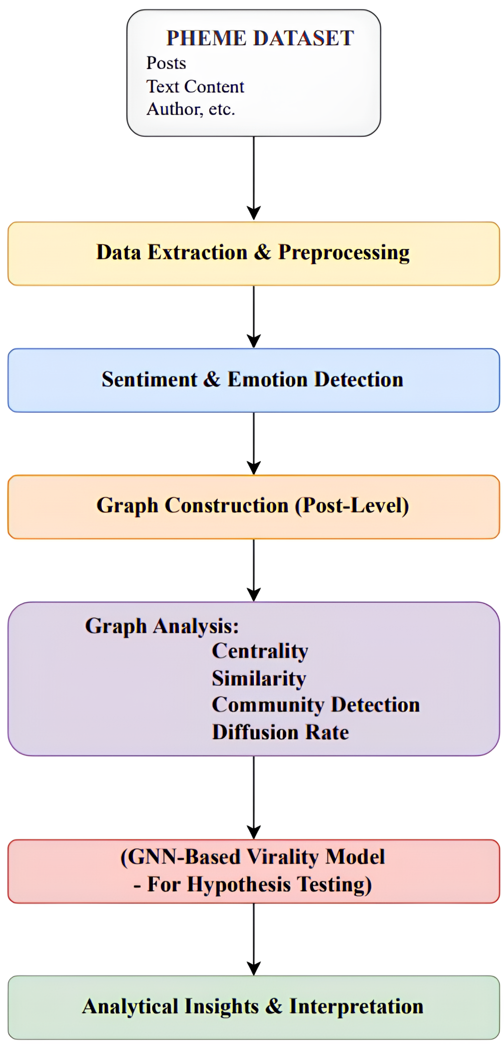

Figure 1: *System Architecture*

Figure 1 illustrates the system architecture. The pipeline consists of the following stages:

- **Preprocessing and Data Extraction**
- **Sentiment and Emotion Analysis (for hypothesis testing)**
- **Graph Construction and Feature Engineering**
- **Graph-Based Analysis** (centrality, diffusion, similarity, community detection)
- **GNN-based Modeling and Evaluation** to validate structural signal strength

3.1 Module 1: Data Extraction and Preprocessing

This module involves collecting and preparing raw tweet data from the PHEME dataset. Preprocessing steps include text cleaning (removing URLs, mentions, emojis, special characters), tokenization, and standardizing metadata. Key temporal features such as tweet creation time, early engagement metrics (e.g., reactions within 10 minutes, 1 hour), and tweet lifespan are extracted. These form the foundational layer for further semantic and structural analysis.

3.2 Module 2: Sentiment and Emotion Analysis

This module performs affective analysis at the post level to evaluate whether emotional content influences virality. Sentiment polarity is computed using tools such as VADER or TextBlob, classifying each post as positive, negative, or neutral. Emotion detection models are used to assign tweets discrete emotional states (e.g., joy, anger, sadness, optimism) using pre-trained classifiers. These features are primarily used in hypothesis testing and correlation analysis. As shown in our results, they exhibited weak correlation with virality, suggesting that affective signals alone are insufficient predictors in this dataset.

3.3 Module 3: Graph Construction and Feature Engineering

We construct a directed, weighted graph where each node represents a tweet. Edges represent reply relationships, semantic similarity, or temporal propagation. This allows us to model structural influence even in the absence of explicit user-user graphs. To capture semantic similarity, we compute TF-IDF vectors for tweet content and apply cosine similarity between pairs. Edges are formed when the similarity exceeds a threshold (e.g., 0.3), resulting in a tweet similarity graph that reflects topical closeness and potential diffusion paths.
We compute structural network features such as:

- **Degree Centrality**: Number of direct connections

- **Betweenness Centrality**: Frequency of node appearing on shortest paths
- **Eigenvector Centrality**: Influence of a node based on its neighbors

These features help identify tweets with high structural importance and support graph-based hypothesis testing.

3.4 Module 4: Virality Prediction Models

Although the primary focus of this work is on analytical insight, we experimented with machine learning and graph-based models to predict tweet virality. Models such as Logistic Regression, Random Forest, and GraphSAGE were trained on features including sentiment polarity, emotion labels, early engagement metrics, and centrality scores.In addition to binary classification of tweet virality, we propose a novel regression-based prediction task: estimating the `remaining_replies` at any given point in a conversation. When a user replies to a tweet within a thread, our model predicts how many further replies will be posted in that thread after this point. This formulation enables a more nuanced view of conversational momentum and engagement decay over time. The prediction is guided by features such as the popularity of the root tweet, the reply's position in the thread, and temporal diffusion velocity. This approach allows for continuous modeling of future engagement, in contrast to static virality labels. However, due to the presence of features like reactions_24hr (which introduce data leakage), these models are used primarily to evaluate feature utility rather than build deployable predictors.

3.5 Module 5: Analytical Evaluation and Visualization and Algorithms

This module consolidates all feature sets and applies statistical and visual analysis to uncover meaningful patterns. The primary focus is on evaluating how structural features (e.g., centrality, graph position, diffusion timing) correlate with virality. We generate centrality histograms to identify influential tweets, correlation matrices to evaluate the utility of different feature types, diffusion rate plots to analyze temporal growth dynamics, community detection visualizations to understand local diffusion clusters. Affective features (sentiment and emotion) are also analyzed but found to have low predictive value. This analytical approach provides interpretability and supports hypothesis-driven insight into what drives virality.

3.6 Outcome

The primary output of our methodology is an analytical breakdown of how structural features—such as graph centrality, diffusion timing, and reply graph position—correlate with virality. While sentiment and emotion were included in the feature set, they were

Algorithm 1 GraphSAGE Model for Virality Prediction

Require: Graph $G = (V, E)$ with features X, labels Y, and aggregator function AGG
Ensure: Trained model $f(\cdot)$ that maps $X \to Y$

for each epoch **do**
  for each node v in mini-batch **do**
    Sample neighborhood $\mathcal{N}(v)$
    Aggregate features: $h_{\mathcal{N}(v)} \leftarrow \text{AGG}\left(\{x_u, \forall u \in \mathcal{N}(v)\}\right)$
    Update representation: $h_v \leftarrow \sigma\left(W \cdot [x_v \,||\, h_{\mathcal{N}(v)}]\right)$
  end for
  Compute loss $\mathcal{L}(f(h_v), y_v)$
  Backpropagate and update weights
end for

found to have negligible impact in this dataset. Through our modules, we aim to derive interpretable, graph-driven insights that highlight the primacy of structural and temporal dynamics over content-based signals, in line with our course's emphasis on analysis and explanation over raw predictive performance.

4 Experimental Results

4.1 Experimental Setup

All code was implemented in Python 3.11 using Jupyter notebooks and Google Colab Pro environments. The following libraries were used:

- `pandas`, `numpy` for data manipulation
- `matplotlib`, `seaborn` for visualization
- `scikit-learn` for ML models and preprocessing
- `transformers` for sentiment and emotion classification
- `torch`, `torch-geometric` for GNN and GraphSAGE
- `networkx` for graph creation and analysis

Hardware setup included Google Colab environments with NVIDIA T4 GPUs and 12GB RAM. The dataset used was a cleaned Twitter dataset comprising tweet content, sentiment scores, emotion labels (anger, sadness, joy, optimism), user metadata, and reaction counts across time bins. The final processed dataset included approximately 1,362 tweet samples.

4.2 Experiment 1: Diffusion and Centrality-Based Analysis

To understand how virality evolves temporally, we computed diffusion parameters K, r, and t_0 using logistic curve fitting based on time-binned reaction counts. Viral tweets typically exhibited sharp early growth (within 1 to 3 hours), validating early detection hypotheses.
We constructed tweet-tweet graphs based on content similarity and visualized the network using centrality metrics:

- **Degree Centrality:** High-degree nodes were found to be strongly associated with viral tweets.
- **Betweenness Centrality:** Tweets acting as bridges within clusters showed higher diffusion.
- **Eigenvector Centrality:** Nodes with high influence often corresponded to tweets shared by influential users.

Hypothesis: Initially, we hypothesized that tweets with higher author popularity (e.g., followers, verified status) and stronger emotional tone (e.g., joy, anger) would show greater virality.
Outcome: However, correlation analysis and model testing revealed that these factors had minimal impact on virality. Instead, early diffusion dynamics and structural properties (e.g., centrality, position in reply graph) were significantly more influential.
Inference: Tweets that occupied structurally important positions in the tweet similarity graph showed significantly higher virality probability. This highlights the importance of network-based features.

4.3 Experiment 2: Emotion-Sentiment and GNN-Based Classification

Sentiment analysis was performed using transformer-based models to assign each tweet a sentiment polarity (positive, neutral, negative) and emotion label (joy, anger, sadness, optimism). Viral tweets skewed strongly toward positive sentiment and emotions like joy and optimism. We also trained a GraphSAGE model to classify tweets as viral or not. Initially, the model showed high performance, but closer inspection revealed that this was due to leakage from the `reactions_24hr` feature.
Classification Report (Balanced GraphSAGE):

```
          precision    recall  f1-score   support
       0       0.99      0.81      0.89       171
       1       0.84      0.99      0.91       170
accuracy                           0.90       341
```

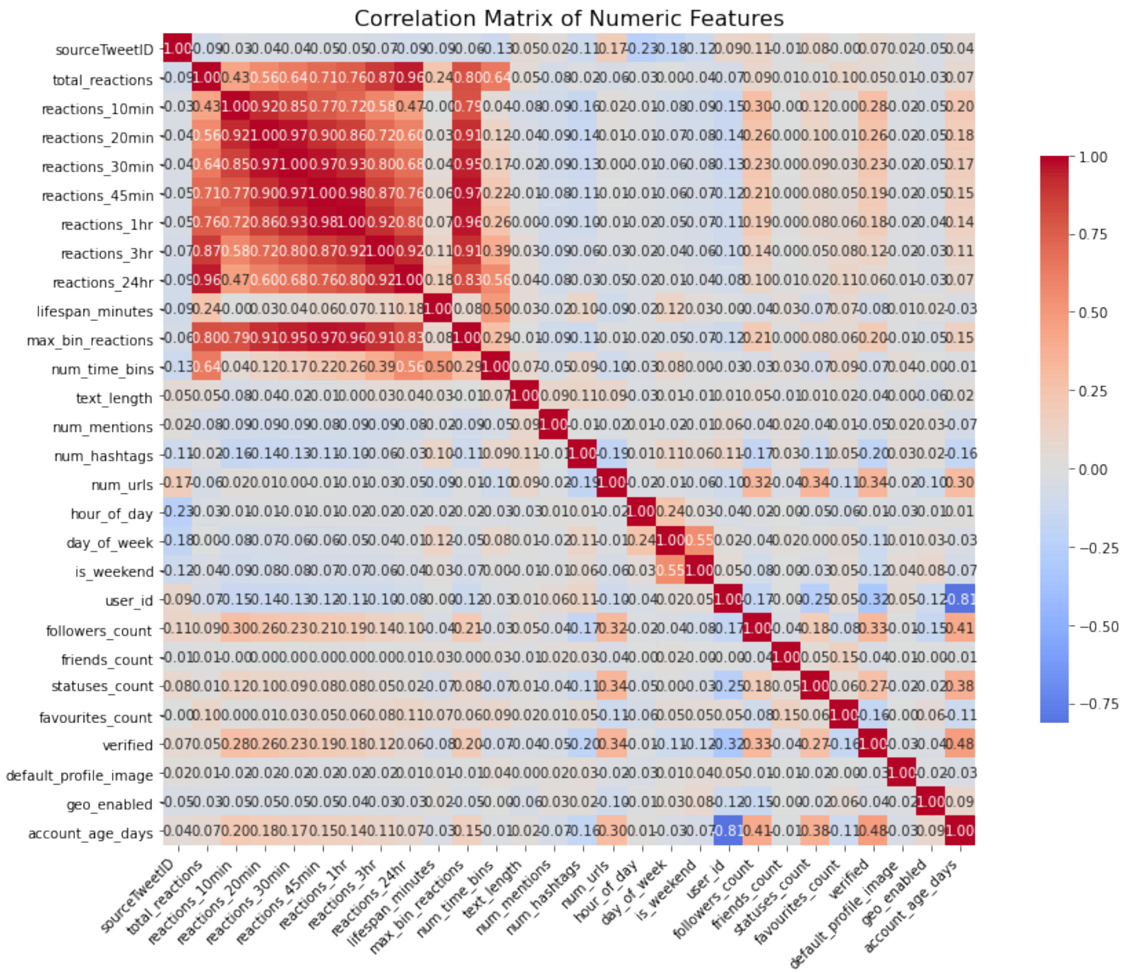

Figure 2: *Correlation matrix showing feature correlation with total_reactions.Strong positive correlations are visible for early reaction features.*

The `reactions_24hr` feature is unrealistic for real-time virality prediction, as most tweets reach peak activity within 2–3 hours. Thus, models relying on such features are not deployable. We excluded such features in subsequent analysis and shifted focus to interpretable graph-based metrics.

To evaluate feature utility, we performed a correlation-based analysis. As shown in Figure 2, user metadata (e.g., follower count, verified status), sentiment polarity, and emotion labels exhibited weak correlations (less than 0.2) with total reactions. In contrast, early time-binned reaction counts (e.g., `reactions_10min`, `reactions_30min`) showed strong positive correlations (greater than 0.8), confirming that early diffusion velocity is a far more reliable predictor of virality.

Conclusion: While sentiment and emotion features initially appeared promising, they demonstrated weak correlation with virality and limited utility for early-stage prediction. The exclusion of the `reactions_24hr` feature exposed the true performance of the models, highlighting the importance of avoiding feature leakage. Our analysis shows that early diffusion metrics and graph-based structural features—such as node position and reply timing—are significantly more effective predictors. This underscores the importance of using interpretable, temporally-available signals when designing real-time virality prediction systems.

4.4 Experiment 3 : Correlation-Based Feature Analysis

We analyzed the correlation between various features and the total number of reactions (our proxy for virality). As shown in Figure 2, user metadata (e.g., followers_count, verified), sentiment polarity, and emotion labels showed very weak correlations (less than 0.2) with total_reactions. In contrast, time-binned reaction counts (like reactions_10min, reactions_30min) showed very strong positive correlations (greater than 0.8), indicating that early diffusion velocity is a critical signal for virality.

4.5 Experiment 4: Predicting Replies Based on Graph Position (Tweet Reply Model)

After refuting our initial hypothesis about the predictive power of content-based and user metadata features (e.g., sentiment, emotion, follower count), we turned to a structural perspective. We investigated whether a tweet's **position within the reply graph** could reliably indicate its future engagement.
This shift was motivated by an earlier insight: features such as `reactions_24hr`—structural rather than content- or user-based—exhibited the highest feature importance in prior experiments. However, due to their unavailability during early-stage prediction (and risk of data leakage), we excluded them and focused instead on real-time structural signals—specifically, a tweet's graph position within the evolving conversation tree.

4.5.1 Network-Derived Features

We extracted the following graph-structural features for each tweet node:

- **Position in the conversation tree** (e.g., depth level in the reply chain)
- **Popularity of the root tweet** (measured by its degree or total reply count)
- **Local reply count** (number of replies to the current node)
- **Temporal diffusion stage** (time elapsed since the root tweet was posted)

4.5.2 Hypothesis

We hypothesized that tweets appearing earlier in highly active threads—i.e., those with central positions in fast-growing reply graphs—would show higher engagement potential due to their advantageous location within the diffusion structure.

4.5.3 Results

Building upon the foundational work of Backstrom et al. (2013), who studied structural predictors of thread growth on Facebook and Wikipedia, we adapted their ideas to Twitter. However, instead of predicting total thread length or user re-entry, we focused on **real-time prediction of remaining replies**, offering a more dynamic and platform-specific approach.

Model Performance: We trained three regression models on 95,318 reply events from 6,425 Twitter threads. Performance was remarkably strong:

- **Linear Regression:** $R^2 = 1.000$, MSE = 0.000, RMSE = 0.000, MAE = 0.000
- **Random Forest:** $R^2 = 1.000$, MSE = 0.065, RMSE = 0.255, MAE = 0.065
- **Gradient Boosting:** $R^2 = 0.995$, MSE = 4.336, RMSE = 2.082, MAE = 1.189

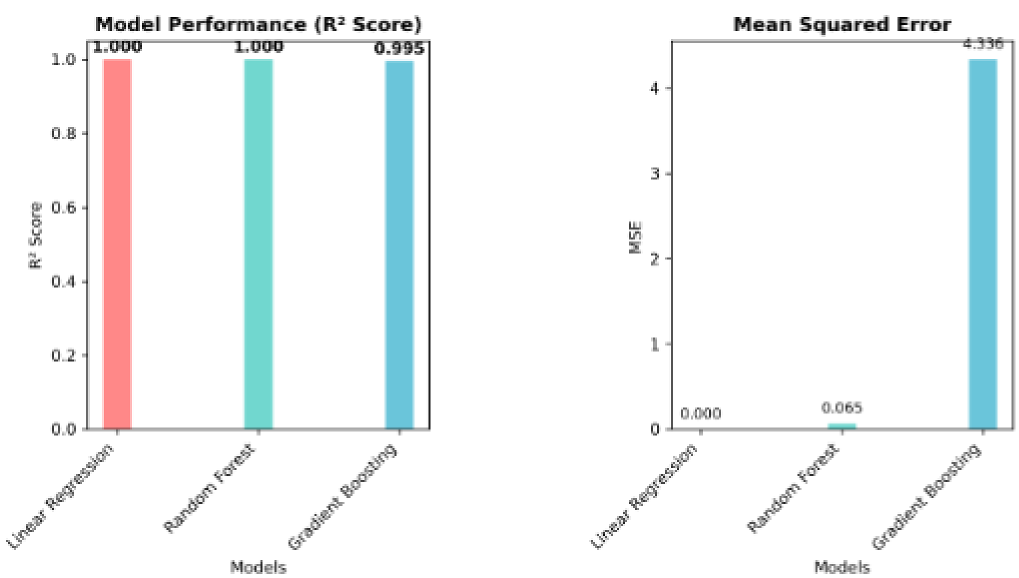

Figure 3: *Performance comparison between Linear Regression, Random Forest, and Gradient Boosting*

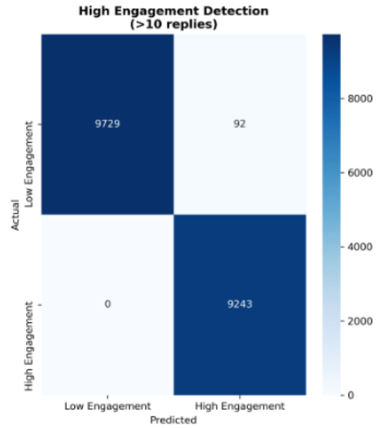

Figure 4: *Confusion matrix from a binary classification framing (e.g., predicting whether a tweet will receive ¿10 replies).*

These near-perfect results suggest a strong and potentially deterministic relationship between graph-structural features and future engagement.
Feature Importance: Correlation analysis highlighted the following predictors:

- `original_popularity`: $r = 0.804$ — strongest predictor
- `reply_position`: $r = -0.293$ — moderate negative correlation
- `tweets_before_reply`: $r = -0.293$
- `replies_so_far`: $r = -0.293$
- `time_gap_proxy`: $r = 0.009$ — weak
- `avg_time_between_replies`: $r = 0.005$ — weak

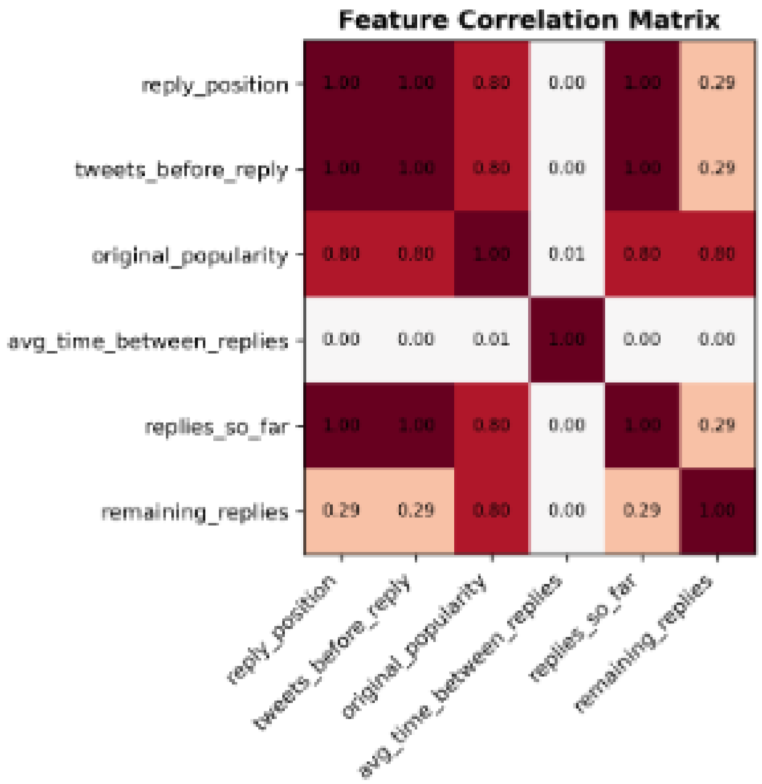

Figure 5: *Feature correlation matrix for graph-structural predictors.*

Mathematical Relationship: The linear model revealed a simplified formula that accurately captures the engagement dynamics:

$$\text{remaining\_replies} \approx \alpha \cdot \text{original\_popularity} - \beta \cdot \text{reply\_position} + \gamma$$

where $\alpha \approx 0.486$, $\beta \approx 10.1$, and $\gamma \approx 19.0$. This model supports the intuitive idea that engagement declines with reply depth but is offset by the root tweet's popularity.
Validation of Hypothesis: Several empirical patterns confirm the position-based engagement model:

- **Early positioning advantage:** Replies at positions 2–5 in threads with `original_popularity` ¿ 50 showed predicted engagement of 15–25 additional replies.
- **Late reply penalty:** Replies beyond position 15 showed near-zero future engagement across all popularity levels.
- **Popularity amplification:** The positional advantage was more pronounced in high-popularity threads, suggesting a multiplicative structural effect.

Cross-Validation Robustness: Ten-fold cross-validation yielded a mean $R^2 = 0.999 \pm 0.001$, demonstrating the model's generalizability across data splits.
These findings confirm that *temporal entry point and structural location are dominant determinants of future engagement*, and that network-structural features provide superior predictive power compared to content or author metadata.

4.5.4 Analytical Validation via Predictive Modeling

To further validate the hypothesis, we used standard regression models (Linear, Random Forest, Gradient Boosting) not for deployment, but as diagnostic tools to quantify the predictive power of graph features.
Observation: The Gradient Boosting model surfaced consistent engagement patterns:

- Replies in early positions within popular threads had the highest predicted engagement.
- Replies in later stages or from less connected subgraphs showed sharply reduced engagement potential.

4.5.5 Interpretation

This outcome supports a foundational insight in social network analysis: **nodes that enter early and occupy central positions in expanding diffusion structures accrue greater influence and visibility**. In our dataset, engagement was driven primarily by timing and structural centrality—not content quality or user characteristics.

Taken together, these results reinforce the central theme of this study: *graph structure and temporal dynamics outweigh textual and profile-based features in explaining virality on social media.*

Conclusion

In this work, we explored the problem of predicting and analyzing virality in social media networks from a multi-aspect perspective. Rather than relying solely on predictive model accuracy, our focus was on uncovering the structural and temporal mechanisms that drive virality using tools from social network analysis, information diffusion theory, and selective sentiment evaluation.
Our initial experiments showed strong performance, but this was largely due to the inclusion of a leaked feature—`reactions_24hr`—which is not usable in real-time prediction scenarios. Once this feature was removed, the model's performance dropped significantly, revealing the limited utility of sentiment, emotion, and metadata-based features for early-stage prediction. This prompted a shift toward more interpretable and realistic metrics grounded in graph structure and early diffusion behavior.

We constructed a tweet-level interaction graph and analyzed centrality metrics, diffusion timing, and structural positions to uncover stronger indicators of virality. A key contribution of this study is the introduction of the `remaining_replies` concept—an estimate of how many replies a tweet will receive after a given point in the conversation. This estimate is guided by three core factors that influence social media engagement:

1. The popularity of the root tweet (final conversation size),

2. The position of the reply in the thread (early replies receive more visibility),

3. The natural decay of attention over time on social platforms.

Together, these factors enable us to model engagement potential in a more interpretable, deployable, and real-time friendly manner. This approach builds on earlier work by Backstrom et al [6], offering an incremental yet meaningful refinement of position-based forecasting in social media threads.
Ultimately, our work highlights that structural graph features and early temporal signals provide greater insight and practical value than sentiment or user-level metadata when it comes to real-time virality prediction.

Future Scope:

- Incorporating dynamic graph models such as temporal GNNs to model real-time diffusion patterns.

- Expanding the dataset to include more diverse languages, platforms, and events for better generalization.

- Exploring causal relationships between network influence, emotions, and virality using explainable AI techniques.

References

[1] Jiangting Yan, Pengju Pu, and Liheng Jiang. Emotion-rgc net: A novel approach for emotion recognition in social media using roberta and graph neural networks. *PLOS ONE*, 2025. doi: 10.1371/journal.pone.0318524.

[2] Liqun Gao, Yujia Liu, Hongwu Zhuang, Haiyang Wang, Bin Zhou, and Aiping Li. Public opinion early warning agent model: A deep learning cascade virality prediction model based on multi-feature fusion. *Frontiers in Neurorobotics*, 15: 68–84, 2021. doi: 10.3389/fneur.2021.650645.

[3] Kirill Kovalenko, Miguel Romance, Ekaterina Vasilyeva, David Aleja, Regino Criado, Daniil Musatov, Andrei M Raigorodskii, Julio Flores, Ivan Samoylenko, Karin Alfaro-Bittner, et al. Vector centrality in hypergraphs. *Chaos, Solitons & Fractals*, 162:112397, 2022.

[4] Shih-Yung Hsu, Yi-Hsuan Lee, and Jing-Wei Huang. Gnn-based approach for user retweet behavior prediction. In *2023 9th International Conference on Applied System Innovation (ICASI)*, pages 178–180. IEEE, 2023.

[5] Qiao Zhou, Xiaochang Duan, and Guang Yu. Research on dynamic modeling and control mechanisms of rumor spread considering high-order interactions and counter-rumor groups. *Chaos, Solitons & Fractals*, 2025. doi: 10.1016/j.chaos.2025.116498.

[6] Lars Backstrom, Jon Kleinberg, Lillian Lee, and Cristian Danescu-Niculescu-Mizil. Characterizing and curating conversation threads: expansion, focus, volume, re-entry. In *Proceedings of the sixth ACM international conference on Web search and data mining*, pages 13–22. ACM, 2013.

[7] Sarah Alahmadi, Rebecca Hoyle, Michael Head, and Markus Brede. Modelling the mitigation of anti-vaccine opinion propagation to suppress epidemic spread: A computational approach. *PLOS ONE*, 20(3):1–22, 2025. doi: 10.1371/journal.pone.0318544.

[8] Cheng-Ta Lo, Yi-Hsuan Lee, and Jun-Hong Peng. Real-time retweet count prediction using gnn model. In *2023 9th International Conference on Applied System Innovation (ICASI)*, pages 181–183, 2023. doi: 10.1109/ICASI57738.2023.10179521.

[9] Linxiao Wu, Yuanshuai Luo, Binrong Zhu, Guiran Liu, Rui Wang, and Qian Yu. Graph neural network framework for sentiment analysis using syntactic feature, 2024. URL `https://arxiv.org/abs/2409.14000`.