Assessing and Identifying Viral Trends on Twitter

Deniz Karakaya
deniz.karakaya@studenti.unipd.it
Md Rubayet Afsan
mdrubayet.afsan@studenti.unipd.it
Omer Faruk Caki
omerfaruk.caki@studenti.unipd.it
Lanlan Gao
lanlan.gao@studenti.unipd.it

Abstract

In today's media age, content spreads rapidly on social platforms, significantly impacting public opinion, especially when it contains misleading information. To mitigate the negative effects of viral tweets on Twitter, early detection is crucial. Traditional methods for identifying viral tweets often fail in accuracy and categorization. Our report introduces a new metric for identifying viral tweets by evaluating existing metrics and analyzing data from Twitter's "Viral Tweets" topic. We identified a threshold based on the ratio of retweets to followings to minimize false positives. Additionally, we developed a transformer-based model that enhances viral content detection, providing researchers, social media managers, and fact-checkers with tools to combat misinformation. This model demonstrated high accuracy and recall in predicting viral tweets.

Keywords: Viral Trends; Twitter; Social Media Analysis; Influence Measurement; Content Spread; Retweet Patterns; Fact-checking; Virality Detection.

1. Introduction

Social media platforms, like Twitter, have the power to quickly shape public opinion and generate widespread discussion through viral posts, regardless of the author's popularity. This rapid spread can both expand influence and spread false information. Understanding viral communication patterns is crucial for early detection and effective fact-checking, which becomes less effective once misinformation is widely disseminated. Due to the vast amount of information, fact-checkers must prioritize potentially explosive claims. Previous research on predicting viral tweets often relied on human labeling or retweet counts, which can be inaccurate. Twitter's "Viral Tweets" topic, introduced in 2021, offers a more reliable dataset. Our project addresses the challenge of spotting fake news before it goes viral by analyzing tweet-sharing behaviors and using powerful AI to predict which tweets will explode. This approach helps catch false information early, maintaining honest information on social media and reflecting real public opinion.

2. Related Works

Understanding and predicting viral tweets has been a significant focus of recent research. Various studies have attempted to define and forecast viral tweets using different methodologies and criteria.

Jenders et al. highlighted the inaccuracies of using arbitrary retweet count thresholds to predict viral tweets. They experimented with different thresholds but found them subjective and limited.

Zadeh et al. used multivariate Hawkes processes to predict tweet popularity, considering retweets, replies, and likes. Their regression-based approach provided a continuous measure of engagement.

Samuel et al. analyzed over a million tweets, identifying key factors like content and timing that influence tweet success.

Hoang et al. studied socio-political tweets from Singapore's 2011 election, introducing metrics to measure tweet spread and identifying viral messages and authors.

Hasan et al. examined how viral tweets affect user behavior and visibility, analyzing tweet activity and follower changes among scientists.

Rameez et al. introduced ViralBERT, a BERT-based model focusing on user-specific factors to predict tweet virality, improving prediction accuracy.

Our study builds on these works by using reliable data from Twitter and advanced machine-learning models to enhance viral tweet detection, aiming to improve information management and fact-checking on social media.

3. Data

To study viral trends on Twitter, we curated popular tweets. We used tweets from 2022 because there was no specific API endpoint available. We gathered the most recent tweets from each topic, over 430,000 tweets in total.

Our analysis revealed that most viral tweets didn't come from famous accounts or require millions of retweets, in-

dicating that anyone can achieve virality. This extensive dataset helps us understand what drives virality, enabling us to develop better tools for predicting and tracking viral trends on Twitter.

4. Measuring Virality

In this chapter, we discuss in detail several traditional methods for assessing viral tweets and propose a more accurate and scientific assessment method by analyzing in depth the advantages and disadvantages of these methods. Our goal is to enhance the understanding of the communication characteristics of viral tweets through this improved assessment method to provide more reliable data support and a theoretical basis for related research.

4.1. Traditional Measures of Virality

4.1.1 ReTweets >Thresholds

Simply counting retweets can be misleading. A celebrity's tweet may get many retweets within their following without spreading widely. This method misses tweets gaining traction among regular users without high retweet numbers.

4.1.2 ReTweets / Median ReTweets

This method considers a user's typical performance. If a user usually gets 20 retweets but one tweet gets 200, it indicates significant virality. This "retweet ratio" gives a personalized view by dividing retweets by the user's median count.

4.1.3 ReTweets Percentile

A tweet is viral if its retweet count is above a certain percentile of the user's tweets. This requires extensive historical data and assumes each user has a consistent number of viral tweets, which may not be accurate.

4.1.4 ReTweets / Followers

This method adjusts retweets based on follower count, providing a balanced measure of virality relative to the user's audience size without needing tweet history.

4.2. Identifying the Optimal Measure of Virality

Our report assessed viral tweets by leveraging real Twitter data and employing various methods. We aimed to accurately identify genuinely viral tweets by analyzing detailed graphs and metrics.

We found that simply counting retweets can overlook significant tweets and overemphasize ordinary ones. The most effective measures were those that accounted for both retweets and a user's follower base, such as "Retweets / Followers."

4.3. Optimal Threshold

The best value for the hard threshold ReTweets >Threshold was about 3088 retweets which favors big accounts. This value gives us a more well-rounded picture of what's truly viral. We can catch viral tweets more accurately, no matter who posted them. This helps us manage information flow better and supports fact-checkers in their fight against misinformation.

5. Detection

5.1. Motivation Problem

Our approach aims to predict viral tweets by focusing solely on tweet content, ignoring engagement metrics like "likes." The goal is to enhance fact-checking efficiency by identifying tweets likely to spread quickly and reach a large audience. This method assumes real-time monitoring of specific users known for sharing misleading content. Fact-checkers can then verify and assess the urgency of these tweets. Given the vast number of tweets, it's impractical to fact-check all, so prioritizing those with viral potential is crucial. Our problem statement is: "Given tweets from a set of users, which are likely to go viral?" We created a dataset including both viral and non-viral tweets from the same authors on the same day, focusing on English tweets to ensure consistency in our analysis.

5.1.1 Data Preparation

We compiled a dataset of about 1000 viral tweets and about 432,000 tweets in total. We randomly sampled from the non-viral tweets to create a balanced dataset suitable for training our models. This resulted in a training set of about 1,200 tweets, evenly divided between viral and non-viral tweets, and a test set of almost 300 tweets, also balanced.

5.1.2 Prioritizing Facts

Our strategy focuses on analyzing tweet content and using a balanced dataset to accurately predict which tweets are likely to become viral. This approach lays a strong foundation for developing predictive models that help fact-checkers prioritize critical tasks, ensuring potentially viral and misleading information is quickly identified and addressed. We plan to deploy advanced machine learning models, particularly transformer-based ones, to examine tweet content features. These models will be trained and evaluated using the curated dataset to achieve high accuracy in predicting viral tweets. The results will improve the efficiency and proactivity of fact-checking processes, supporting the integrity of information on social media platforms.

5.2. Feature Engineering

To predict viral tweets with accuracy, we prioritize the textual content of the tweets. Using transformer-based language models, which are highly effective in natural language processing, we analyze this text. However, focusing only on the text might not cover all aspects that drive a tweet's virality. Therefore, we add other features that these models might miss, enhancing our overall predictive performance.

5.2.1 Textual Content Features

At the center of our feature engineering strategy are transformer-based language models. These models are specifically designed to process extensive text data and are highly effective at discerning context, meaning, and the intricate relationships between words. For our purposes, we use pre-trained models that have been fine-tuned with datasets specifically curated to understand the subtleties of tweet content. This method ensures we accurately capture the critical linguistic and contextual factors that contribute to a tweet's potential to go viral.

5.2.2 Additional Features

To further refine our model, we incorporate several supplementary features that have been observed to slightly boost performance. These features include:

Media Presence: A boolean indicator that shows whether a tweet includes media elements like images, videos, or GIFs. Tweets featuring media typically engage more users, which increases their chances of going viral.

Hashtags: A boolean indicator of whether the tweet includes hashtags. Hashtags can significantly amplify a tweet's reach by linking it to broader conversations and trending topics.

Mentions: A boolean indicator of whether the tweet mentions other users. Mentions can increase engagement through interactions and retweets from those mentioned or their followers.

Sentiment Analysis: We perform sentiment analysis using the distilbert-base-uncased-fine tuned-sst-2-english model. This model assesses whether a tweet expresses positive or negative sentiment, providing a confidence score for each sentiment. Sentiment can play a crucial role in virality, as tweets evoking strong emotions are more likely to be shared.

Verified Account: A boolean indicator specifies whether the tweet comes from a verified account. Verified accounts typically have a larger audience and greater influence, which can enhance the likelihood of their tweets going viral.

5.3. Experimental Result

To assess the effectiveness of our approach in predicting viral tweets, we employed a variety of transformer-based language models available through HuggingFace. The models selected for our experiments include BERT-Base, RoBERTa, TinyBERT, and BERTweet. Each of these models has unique strengths, and their performance varies based on the specific characteristics of the text data they analyze.

5.3.1 Selection of Models

BERT-Base: Transformer-based model developed by Google for natural language understanding, consisting of 12 layers, 768 hidden units, and 12 attention heads and is pre-trained on a large corpus of text.

RoBERTa: Enhanced version of BERT that improves performance by training on more data and for longer periods, using dynamic masking and removing the next sentence prediction task.

TinyBERT: Compressed, lighter version of BERT, designed to retain performance while significantly reducing model size and computationalrequirements through knowledge distillation.

BERTweet: Pre-trained language model specifically designed for processing and understanding Twitter data, built on the architecture of RoBERTa and optimized for social media text.

5.3.2 Experimental Setup

We opted for case-sensitive versions of these models. Our observation indicated that the usage of uppercase letters in tweets often signifies emphasis on certain words or phrases, which can be critical in understanding the tweet's intent and potential virality.

5.3.3 Model Configurations

We conducted experiments using two configurations for each model:

Text-Only Models: These models rely solely on the textual content of the tweets to make predictions.

Text + Additional Features Models: These models incorporate the supplementary features (media presence, hashtags, mentions, sentiment, and verified account status) alongside the text content to enhance prediction accuracy.

5.3.4 Evaluation Metrics

To rigorously evaluate the performance of each model, we used the following metrics:

Accuracy: The overall correctness of the model's predictions.

Precision: Percentage of correctly identified viral tweets out of all tweets identified as viral.

Recall: The percentage of true viral tweets, correctly recognized out of all the actual viral tweets.

F1 Score: The harmonic mean of precision and recall, providing a balanced measure of the model's performanc

5.3.5 Results and Analysis

The results of our experiments are summarized in Table 1. Here are the key observations:

BERT-Base: Showed worst performance among all models across all metrics.

RoBERTa: Exhibited better performance than BERT-Base, due to its optimized training.

TinyBERT: Provided the best recall and second-best F1 score behind BERTweet.

BERTweet: Outperformed all other models, achieving the highest F1 score. Because of its specialized training in Twitter data, makes it adept at understanding the unique linguistic patterns found in tweets.

When we integrated the additional features into the models, we noticed an overall improvement in performance. For instance, the F1 score for BERTweet increased by 0.119, highlighting the value of incorporating features beyond text content.

Model	Precision	Recall	F1 Score	F1 Score (with additional features)
BERT-Base	0.6646	0.6943	0.6791	0.7134
RoBERTa	0.7000	0.7134	0.7060	0.7514
TinyBERT	0.6484	0.9045	0.7553	0.7575
BERTweet	0.7277	0.8344	0.7774	0.7893

Figure 1. Performance Comparison

The experimental results underscore the importance of using specialized models like BERTweet for analyzing tweets. Additionally, incorporating extra features related to tweet content and user profile information significantly enhances model performance. These findings validate our approach and provide a robust framework for predicting viral tweets, thereby aiding fact-checkers and social media managers in effectively managing the dissemination of information on Twitter.

6. Future Directions

Automated Content Generation: Building on the success of our predictive models, future research could explore the automatic generation of tweets likely to go viral. This could help content creators and fact-checkers craft more engaging and effective messages.

Media Content Analysis: Given that viral tweets often include media elements, further investigation into how images, videos, and GIFs contribute to virality could provide deeper insights. This includes analyzing the type and quality of media that enhances engagement.

Cross-Platform Virality: Extending the current research to other social media platforms could validate the universality of our metrics and models. By applying our methodologies to platforms like Facebook, we can understand how virality manifests in different environments.

Real-Time Detection: Developing real-time detection systems based on our models could greatly benefit social media monitoring and fact-checking efforts. Such systems could provide timely alerts about potentially viral content, enabling quicker responses to misinformation.

User Behavior Post-Virality: Investigating how users' behavior changes after their content goes viral could offer insights into the long-term effects of virality on social media engagement and influence.

By continuing to refine and expand upon our findings, we can contribute to a more informed and responsible use of social media, ultimately fostering a healthier online environment where information is accurately disseminated and effectively managed.

7. Conclusion

This study enhances our understanding of social media virality by evaluating existing metrics, proposing a new one, and utilizing advanced machine learning models to predict viral tweets. By analyzing data from Twitter, we developed a framework for identifying potentially viral tweets. Traditional virality metrics often rely on arbitrary thresholds that can be manipulated and may not accurately reflect true reach. Our research introduces a new metric based on the ratio of retweets to followers, offering a more accurate measure of impact. Combined with transformer-based language models, this metric demonstrated high precision and recall in identifying viral content.

Our predictive models consider tweet text, media presence, hashtags, mentions, sentiment, and account verification status, with the BERTweet model proving the most effective. These methodologies and metrics, adaptable to platforms without view counts, are relevant across various contexts. The study also suggests future research, such as developing automated content generation systems to help experts counteract misinformation and exploring the impact of media elements on tweet virality.

References

[1] et al. Ahmed. User-specific sentiment and clusters over time on twitter during the pandemic. *PLOS ONE*, 2021.

[2] Sinan Aral, Soroush Vosoughi, and Deb Roy. False news on twitter. *MIT News*, 2018.

[3] Skunkan Boon-Itt. Topic modeling on covid-19 tweets. *PLOS ONE*, 2020.

[4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[5] Kiran Garimella and Robert West. Hot streaks on social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 13, pages 170–180, 2019.

[6] PLOS ONE Study Group. Academic information on twitter: A user survey. *PLOS ONE*, 2022.

[7] Weibo Study Group. Who creates trends in online social media: The crowd or opinion leaders? *Journal of Computer-Mediated Communication*, 2020.

[8] Zhijiang Guo, Michael Schlichtkrull, and Andreas Vlachos. A survey on automated fact-checking. *Transactions of the Association for Computational Linguistics*, 10:178–206, 2022.

[9] Rakibul Hasan, Cristobal Cheyre, Yong-Yeol Ahn, Roberto Hoyle, and Apu Kapadia. The impact of viral posts on visibility and behavior of professionals: A longitudinal study of scientists on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 323–334, 2022.

[10] Tuan-Anh Hoang, Ee-Peng Lim, Palakorn Achananuparp, Jing Jiang, and Feida Zhu. On modeling virality of twitter content. In *International Conference on Asian Digital Libraries*, pages 212–221. Springer, 2011.

[11] et al. Maldonado-Sifuentes. Measuring and detecting virality on social media: The case of twitter's viral tweets topic. *arXiv preprint arXiv:2303.06120*, 2021.

[12] Felix Naumann, Maximilian Jenders, and Gjergji Kasneci. Analyzing and predicting viral tweets. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 7, pages 465–468, 2013.

[13] et al. Rajput. Sentiment analysis of tweets during covid-19 pandemic. *PLOS ONE*, 2020.

[14] Rikaz Rameez and Hossein A. Rahmani. Viralbert: A user-focused bert-based approach to virality prediction, 2022.

[15] Amir Zadeh and Ramesh Sharda. How can our tweets go viral? point-process modelling of brand content. *Information Management*, 59(2):103594, 2022.

A. Appendix

A.1. Ethical Considerations

In this research, we adhered strictly to ethical guidelines to ensure the responsible use of data and user privacy protection. Our study was designed with a strong commitment to ethical principles, focusing on transparency, integrity, and respect for the individuals whose data was analyzed.

A.1.1 Data Usage

We utilized data exclusively from public profiles that Twitter had amplified under the "Viral Tweets" topic. This means that all the tweets included in our analysis were already made publicly accessible and highlighted by Twitter due to their significant reach and engagement. By restricting our data sources to these public and amplified tweets, we ensured that we were not infringing on the privacy of Twitter users.

A.1.2 Data Disclosure

In presenting our findings, we disclosed only the tweet IDs from the data we collected. This approach allowed us to share the necessary information for verification and reproducibility of our study without compromising the privacy of the individuals who authored the tweets. Tweet IDs serve as unique identifiers that can be used to locate the original tweets, but they do not reveal any additional personal information about the users beyond what is publicly available on Twitter.

A.1.3 Anonymity and Privacy

We took care to anonymize the dataset as much as possible, focusing on aggregate data and trends rather than individual user details. Any analysis or presentation of data was done in a manner that ensures individual users cannot be easily identified or targeted. This practice is crucial in maintaining the integrity and confidentiality of user data.

A.1.4 Ethical Research Practices

Throughout the research process, we were guided by ethical research practices, including obtaining data legally and transparently, ensuring the accuracy and honesty of our findings, and respecting the rights of the individuals whose data we analyzed. We also considered the broader implications of our research, aiming to contribute positively to the field of social media analysis and fact-checking without causing harm or misuse of the data.

A.1.5 Compliance with Regulations

By adhering to these standards, we aim to uphold the highest levels of research integrity and public trust. Our commitment to ethical considerations was paramount in every stage of this research. By using publicly available data, anonymizing our dataset, and disclosing only the necessary tweet IDs, we ensured that our study respects the privacy and rights of Twitter users while providing valuable insights into the dynamics of viral trends on social media.

Final Work Plan and Time Spent Report

Tasks and Time Spent

Task	Description	Estimated Time	Actual Time
Literature Review	Compiled a comprehensive review of at least 4-5 relevant papers on viral tweet identification methods.	15 hours	20 hours
Data Collection	Collected real data from Twitter's 'Viral Tweets' section, aiming for a dataset of at least 50,000 tweets.	10 hours	15 hours
Implementing Transformer-Based Model and Custom Model Development	Implemented BERT as a baseline model. Experimented with other transformer-based models and custom models to achieve a target F1 score of 0.75.	25 hours	30 hours
Preparing Report	Prepared the introduction and methodology sections of the final report.	20 hours	25 hours
Code Documentation	Documented the codebase for clarity and maintainability.	10 hours	12 hours
Presentation Preparation	Crafted the introduction and methodology slides for the final presentation.	8 hours	10 hours

Mandatory Objectives

- Completed the literature review.
- Acquired the dataset from Twitter.
- Developed a basic version of the algorithm.
- Prepared the results and other sections of the report.

Secondary Objectives

- Enhanced the algorithm to improve the F1 score beyond 0.75.

Member-Specific Subtasks

Deniz Karakaya:

- Focused on data collection and initial algorithm coding (25 hours).
- Assisted with code documentation (8 hours).
- Helped with presentation preparation (5 hours).
- Assisted with literature review (5 hours).

Omer F. Caki:

- Assisted with algorithm coding (20 hours).
- Helped with code documentation (6 hours).
- Prepared the introduction and methodology sections of the final report (10 hours).
- Helped with data collection (5 hours).

Lanlan Gao:

- Led the literature review (20 hours).
- Prepared the report (10 hours).

- Crafted the introduction and methodology slides for the final presentation (8 hours).
- Assisted with algorithm coding (5 hours).

Md Rubayet Afsan:

- Worked on literature review summaries (10 hours).
- Assisted with the report (8 hours).
- Assisted with presentation preparation (7 hours).
- Implemented BERT as a baseline model and experimented with other transformer-based models and custom models (15 hours).
- Helped with data collection (5 hours).

Member Time Distribution

- Deniz Karakaya: 43 hours
- Omer Caki: 41 hours
- Lanlan Gao: 43 hours
- Md Rubayet Afsan: 45 hours