PREDICTING RUMOUR SPREAD USING TEXTUAL AND SOCIAL CONTEXT IN PROPAGATION GRAPH WITH GNN

A Major Project Report Submitted
In Partial Fulfilment of the Requirements for the Degree of

BACHELOR OF TECHNOLOGY
in
INFORMATION TECHNOLOGY

by

HARDIK SACHAN (2007340130025)
SATYAM DUBEY (2007340130052)
AYUSH PANDEY (2007340130019)
MANSI GAUTAM (2007340130033)

Under the Supervision of

Dr. Siddhartha Arjaria
(Assistant Professor)
Rajkiya Engineering College, Banda

to the

DEPARTMENT OF INFORMATION TECHNOLOGY

DR. A. P. J. ABDUL KALAM TECHNICAL UNIVERSITY, LUCKNOW
(Formerly Uttar Pradesh Technical University, Lucknow)

June 2024

DECLARATION

We hereby declare that the work presented in this report entitled "PREDICTING RUMOUR SPREAD USING TEXTUAL AND SOCIAL CONTEXT IN PROPAGATION GRAPH WITH GNN", was carried out by us. We have not submitted the matter embodied in this report for the award of any other degree or diploma of any other University or Institute.

We have given due credit to the original authors/sources for all the words, ideas, diagrams, graphics, computer programs, experiments, and results, that are not our original contributions. We have used quotation marks to identify verbatim sentences and given credit to the original authors/sources.

We affirm that no portion of our work is plagiarized, and the experiments and results reported in the report are not manipulated. In the event of a complaint of plagiarism and the manipulation of the experiments and results, we shall be fully responsible and answerable.

Name: **Candidate's Signature**

Hardik Sachan (2007340130025)

Satyam Dubey (2007340130052)

Ayush Pandey (2007340130019)

Mansi Gautam (2007340130033)

CERTIFICATE

Certified that **Hardik Sachan** (2007340130025), **Satyam Dubey** (2007340130052), **Ayush Pandey** (2007340130019), and **Mansi Gautam** (2007340130033) has carried out the research work presented in this major project report entitled "PREDICTING RUMOUR SPREAD USING TEXTUAL AND SOCIAL CONTEXT IN PROPAGATION GRAPH WITH GNN" for the award of Bachelor of Technology from Dr. APJ Abdul Kalam Technical University, Lucknow under our supervision. The major project report embodies results of original work, and studies are carried out by the students themselves and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Signature:

Dr. Siddhartha Arjaria

(Assistant Professor)

(Rajkiya Engineering College, Banda)

Date:

ABSTRACT

In today's age of digital advancement, it becomes very easy for people to share their opinion on internet. Social media platform like twitter, Facebook etc. connects people to share the latest news in the world. Rumour refers to unverified information or speculation that is spread widely among users but has not been confirmed by credible sources. With this freedom of sharing tweet and opinion, it becomes very easy to spread a rumour. Rumour can create chaos in public domain that can trigger fear, damage person's image and can also affect the others sectors like economic etc. To control the damage, a strong rumour detection system is needed so that it can detect rumour with better accuracy. This project aims to develop a rumour detection system that uses previous history of user's, propagating pattern of tweet, semantic relation of text content. It starts with a hybrid data collection approach using API interfaces and third-party crawlers to extract Twitter rumour information. Feature extraction combines content and context analysis, considering textual content and social context features of users involved in the tweet. Textual embeddings are generated through advanced models like Word2Vec and BERT, facilitating an understanding of semantic relationships within tweet text. A unique contribution involves constructing a propagation graph with the tweet as the root node and users as leaf nodes, followed by the application of Graph Neural Networks (GNNs) for effective feature extraction. The GNN process includes message passing, node updating, and iterative refinement, capturing complex relationships in the network. An attention-based mechanism then combines textual embeddings with GNN-derived user involvement embeddings to create a precise and less redundant tweet representation. Finally, a neural classifier is trained on integrated embeddings to predict the likelihood of a tweet being a rumour. This comprehensive methodology harnesses the strengths of graph-based and textual analysis for nuanced and accurate rumour detection on social media, particularly Twitter.

Keywords: Rumour detection, Graph neural network, Social Media Network, Twitter, BERT

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present report of the B. Tech Project undertaken during B.Tech. Final Year. We owe special debt of gratitude to our project guide **Dr. Siddhartha Arjaria**, Assistant Professor, Rajkiya Engineering College, Banda and for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavors have seen light of the day.

We also do not like to miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our lab assistant for their contribution in the completion of the project.

Signature:

Hardik Sachan (2007340130025)

Satyam Dubey (2007340130052)

Ayush Pandey (2007340130019)

Mansi Gautam (2007340130033)

TABLE OF CONTENTS

Content	Page No.
Declaration	ii
Certificate	iii
Abstract	iv
Acknowledgement	v
Table of Content	vi-vii
List of Tables	viii
List of Figures	ix

CHAPTER 1: INTRODUCTION……………………………………. 1-7

1.1 BACKGROUND 1
1.2. PROBLEM STATEMENT 2
1.3 RUMOUR 2
1.4 IMPACT OF RUMOUR 4
1.5 ROLE OF X IN RUMOUR PROPAGATION 5
1.6. NEED OF RUMOUR DETECTION SYSTEM 6
1.7. STUDIES 7
1.8. OBJECTIVE 8

CHAPTER 2: LITERATURE REVIEW…………………………....9-13

CHAPTER 3: METHODOLOGY 14-27

3.1. TWEEP TENDENCY ENCODING 15
3.1.1. WORD2VEC 16
3.1.1.1 CBOW 17
3.1.1.2 Skip Gram 19
3.1.2. BERT 20
3.2. NETWORK DYNAMIC EXTRACTION 23
3.3. EXTRACTED INFORMATION INTEGRATION 23
3.3.1. GRAPHICAL NEURAL NETWORK 23
3.3.1.1 Basic Architecture Of GNNs 24
3.3.1.2. GNNs Message Passing Mechanism 24

CHAPTER 4: RESULT…………………………………………… 28-32

CHAPTER 5: CONCLUSION AND FUTURE WORK……………33

REFERENCES **34**

APPENDIX **35-46**

LIST OF TABLES

TABLE NO	TABLE NAME	PAGE NO.
Table no. 4.1	Dataset and Graph Statistics	28
Table no. 4.2	Performance Evaluation	30
Table no. 4.3	Rumour Detection Performance on Two Dataset	31

LIST OF FIGURES

FIG. NO	FIGURE NAME	PAGE NO.
Fig 1.1.	Propagation of Rumour on X	3
Fig 3.1.	Tweep Tendency Rumour Detection Model Framework	15
Fig 3.2.	CBOW	18
Fig 3.3.	Skip Gram	20
Fig 3.4.	Bert Model Architecture	22

CHAPTER 1

INTRODUCTION

1.1. BACKGROUND

In today's digital age, where social media platforms like X and Facebook have become integral parts of our daily lives, the dependency on these platforms to disseminate information has skyrocketed. However, this increased reliance also brings with it the risk of the rapid spread of rumours, capable of causing significant disruptions in individuals' lives. The pervasive nature of social media makes it a fertile ground for the propagation of misinformation, necessitating a critical examination of the impact rumours can have on society.

The potential for rumours to bring about massive changes in the lives of individuals cannot be understated. False information, once unleashed on social media, can quickly gain traction, leading to widespread panic, confusion, and damage to reputations. Recognizing the urgency of the situation, the need for a robust rumour detection system has become increasingly apparent. As rumours can have far-reaching consequences, from affecting personal relationships to influencing political climates, the development of effective detection mechanisms has become a pressing concern.

Mitigating the effects of rumours is not only essential for safeguarding individual integrity but also for maintaining the stability of communities and societies at large. The destructive potential of false information necessitates proactive measures to identify and counteract rumours before they can cause irreparable harm. As a result, the focus on rumour detection has intensified in recent years, with researchers and technologists exploring various techniques to address this burgeoning issue.

The proposed methodology integrates multiple stages to form a comprehensive framework for rumour detection. Firstly, a hybrid data collection strategy combining API interfaces and third-party crawler programs is employed to extract X rumour information from the X public page. Feature extraction encompasses both content and context analysis, incorporating textual content and social context features of users involved in the tweet, such as followers, interests, and verified accounts.

Textual embeddings are generated using advanced models like Word2Vec and Bidirectional Encoder Representations from Transformers (BERT), enabling the system to understand the semantic relationships within tweet text. A unique contribution lies in the construction of a propagation graph, where the tweet serves as the root node and involved users as leaf nodes. Graph Neural Networks (GNNs) are then applied to extract features capturing the influence and engagement of users. This involves message passing, node updating, and iterative refinement processes, allowing complex relationships within the network to be captured effectively.

Furthermore, an attention-based mechanism combines textual embeddings with GNN-derived user involvement embeddings to obtain a precise and less redundant tweet representation. Finally, a neural classifier is trained on the integrated embeddings to predict the probability of a tweet being a rumour. This comprehensive methodology offers an advanced and adaptive solution, leveraging the strengths of both graph-based and textual analysis for more accurate and nuanced rumour detection on social media platforms, particularly X.

1.2. PROBLEM STATEMENT

To make a prediction model using GNN that utilizes historical post data a propagation graph to enhance accuracy in identifying rumours.

1.3. RUMOUR

A rumour refers to unverified or speculative information that is circulated among users but has not been confirmed by credible sources. Rumours on social media can encompass a wide range of content, including false news, gossip, or unconfirmed reports about events, individuals, or issues. These rumours often spread rapidly through the interconnected network of social media platforms, facilitated by features such as retweets, shares, and likes.

Rumours can originate from various sources, including misunderstandings, misinterpretations, intentional misinformation, or the distortion of facts. The fast-paced and viral nature of social media makes it challenging to control the spread of rumours, and they can have significant consequences by influencing public opinion, shaping perceptions, and even impacting real-world events.

Rumour, a term that resonates deeply within human society, refers to unverified information that spreads informally from person to person. Often, rumours fill the gaps left by a lack of concrete information, particularly during times of uncertainty or crisis. They can cover a broad spectrum of topics, from personal anecdotes to global events, and can be either benign or malicious in nature. Unlike verified news, rumours thrive on ambiguity and speculation, making them a powerful but unpredictable force in shaping public perception and opinion.

The origins of a rumour can be traced to various sources, including miscommunication, deliberate misinformation, or even simple curiosity. Human nature plays a significant role in the propagation of rumours. People tend to share information that evokes strong emotions or aligns with their pre-existing beliefs, contributing to the rapid spread of rumours. This phenomenon is often exacerbated by social media platforms, where information can go viral in a matter of minutes, reaching a global audience. The anonymity provided by the internet also emboldens individuals to create and share rumours without immediate repercussions, further fueling their spread.

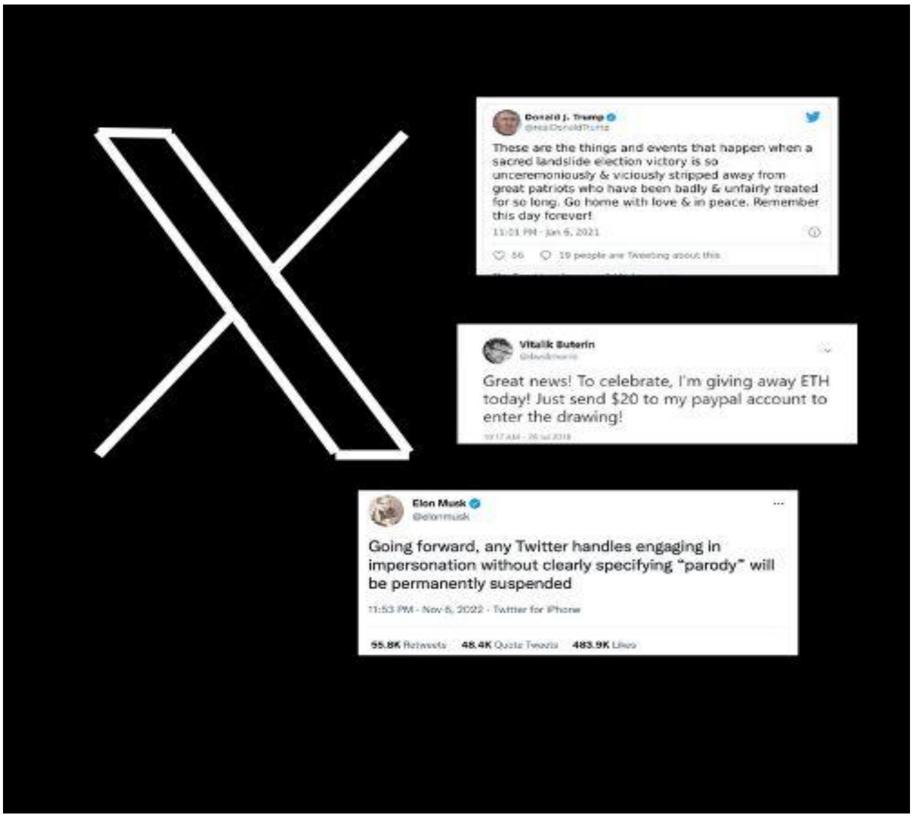

Fig. 1.1 Propagation of Rumour on X

The impact of rumours can be profound and multifaceted. On a personal level, rumours can damage reputations, erode trust, and cause emotional distress. In a broader context, they can influence public opinion, sway political outcomes, and even incite social unrest. Historical examples abound, such as the widespread panic caused by rumours during wartime or the economic turmoil resulting from speculative financial information. In contemporary times, the rapid dissemination of rumours related to public health, such as during the COVID-19 pandemic, has highlighted the potential for misinformation to hinder effective response efforts and endanger lives.

Despite their often negative connotations, rumours can also serve useful purposes. They can act as an early warning system, bringing attention to issues that might otherwise remain unnoticed. In some cases, the circulation of a rumour can prompt authorities to address the underlying concerns, thereby mitigating potential problems. Furthermore, studying the patterns and content of rumours can provide valuable insights into societal fears, values, and communication dynamics. This understanding can inform strategies to manage misinformation and enhance public information campaigns.

In conclusion, rumours are an intrinsic part of human communication, driven by our innate curiosity and the need to make sense of the world around us. While they can cause significant harm, they also offer opportunities for learning and improvement. Addressing the challenges posed by rumours requires a nuanced approach that balances debunking false information with fostering open, transparent communication. By understanding the mechanisms behind rumour propagation, society can better navigate the complexities of information in the digital age.

1.4. IMPACT OF RUMOUR

Rumours can have a profound and diverse impact on individuals and communities. One significant effect is the psychological toll they take on people. Rumours often generate anxiety, fear, and uncertainty, as individuals grapple with the ambiguity surrounding the information. This emotional distress can lead to a breakdown in trust and cohesion within communities, as people struggle to navigate the blurred lines between fact and fiction.

Moreover, rumours can have tangible consequences on social dynamics and relationships. The spread of false or misleading information can result in the isolation and stigmatization of individuals or groups, as unfounded rumours shape public opinion. This can have lasting effects on the reputations and well-being of those targeted. In extreme cases, false rumours can even incite hostility and discrimination, exacerbating existing tensions within society.

Economically, the impact of rumours can be substantial. In today's interconnected world, where information travels rapidly, rumours can affect financial markets, causing fluctuations in stock prices and impacting businesses. Investors may make decisions based on inaccurate information, leading to financial losses. The uncertainty created by rumours can also disrupt supply chains and business operations, further amplifying the economic repercussions.

Lastly, rumours can influence political landscapes and decision-making processes. In the realm of politics, false information can sway public opinion, manipulate elections, and undermine the legitimacy of governments. The weaponization of rumours for political gain can have far reaching consequences, eroding the foundations of democracy and creating a climate of mistrust. In essence, the diverse effects of rumours extend beyond individual emotions, permeating various aspects of society and highlighting the need for vigilant information management and media literacy.

1.5 ROLE OF X IN RUMOUR PROPAGATION

Twitter, now rebranded as X, plays a significant role in the dissemination of information across the globe, and this includes the spread of rumours. As a microblogging platform with millions of active users, X allows for rapid sharing and amplification of content. The platform's design, which emphasizes short, real-time updates, makes it an ideal environment for rumours to flourish. This dynamic nature of X can lead to both positive and negative outcomes, particularly in the context of rumour propagation.

The spread of rumours on X is facilitated by several platform-specific mechanisms. Retweets, replies, and likes are fundamental features that allow users to amplify messages rapidly. When a rumour is tweeted by an influential user or a verified account, it can quickly gain traction as followers engage with the content. Hashtags and trending topics further enhance visibility, enabling rumours to reach a broader audience beyond the original followers of the account. The algorithmic nature of X's feed, which promotes engaging content, can inadvertently prioritize sensational or misleading information, accelerating the spread of rumours.

User behaviour on X significantly impacts the propagation of rumours. Users often share information without verifying its accuracy, driven by the urgency to be the first to comment or share breaking news. This behaviour is exacerbated by the echo chamber effect, where individuals are exposed primarily to information that aligns with their beliefs. Influential users, including celebrities, journalists, and politicians, can significantly amplify rumours due to their large followings. When these users share or comment on a rumour, it lends credibility to the information, regardless of its veracity, and encourages further sharing among their followers.

The spread of rumours on X can have profound implications. In the context of public health, misinformation can lead to harmful behaviours, as seen during the COVID-19 pandemic. False rumours about treatments or preventive measures spread widely, causing confusion and undermining public health efforts. In political contexts, rumours can influence elections and shape public opinion, potentially destabilizing democratic processes. Additionally, rumours can damage reputations, incite violence, and create widespread panic. The rapid dissemination of unverified information underscores the need for effective measures to mitigate the impact of rumours on X.

Addressing the spread of rumours on X requires a multifaceted approach. Platform policies must prioritize the identification and removal of misleading information. This includes employing advanced algorithms to detect and flag potential rumours, as well as partnerships with fact-checking organizations to verify content. User education is equally important; X should promote digital literacy initiatives that encourage users to critically evaluate information before sharing. Transparency in how information is curated and presented can also help users make informed decisions. Ultimately, a combination of technological, policy-driven, and educational strategies is necessary to curb the propagation of rumours on X.

In conclusion, X's role in rumour propagation is shaped by its platform features, user behaviours, and the influence of high-profile accounts. The rapid and widespread dissemination of unverified information on X poses significant challenges, with serious implications for public health, politics, and society at large. Effective mitigation strategies must involve a concerted effort by the platform, its users, and external organizations to foster a more informed and responsible information-sharing environment. By addressing these challenges, X can better balance its role as a conduit for real-time information with the need to prevent the spread of harmful rumours.

1.6 NEED OF RUMOUR DETECTION SYSTEM

Rumour detection plays a crucial role in today's interconnected and digital society, where information spreads rapidly through various online platforms. The advent of social media has accelerated the dissemination of information, but it has also given rise to the proliferation of rumours. The need for rumour detection arises from the potential harm that false or misleading information can cause to individuals, communities, and even entire societies. Rumours can incite panic, fear, and misinformation, leading to real-world consequences such as social unrest, economic instability, or even threats to public safety.

Moreover, in the era of information overload, distinguishing between accurate and inaccurate information has become increasingly challenging. Rumour detection helps in maintaining the integrity of information by providing tools and methodologies to verify the authenticity of news and updates. This is particularly important in fields such as journalism, where the spread of false information can erode public trust in media outlets. By identifying and mitigating the impact of rumours, rumour detection contributes to the overall credibility of information sources.

Additionally, rumour detection is vital for safeguarding democratic processes. False information, especially during elections, can influence public opinion and undermine the democratic principles of informed decision-making. Political rumour-mongering can manipulate perceptions, sway votes, and compromise the fairness of electoral processes. Effective rumour detection mechanisms empower individuals to make well-informed choices, preserving the integrity of democratic systems.

In the realm of cybersecurity, rumour detection is also essential. False alarms and unfounded speculations can lead to unnecessary panic and resource allocation. By accurately identifying and debunking rumours related to cyber threats, organizations can focus their efforts on real security concerns, thus enhancing their resilience to potential cyberattacks. In summary, the need for rumour detection arises from its pivotal role in preserving societal well-being, maintaining the integrity of information, safeguarding democratic processes, and enhancing cybersecurity measures in our increasingly interconnected world.

1.7. STUDIES

Sociological and psychological studies on journalism have explored the correlation between users' tendencies and their behaviours in consuming online tweets. Consumers tend to believe their perceptions of reality are the only accurate views. Those who disagree are often considered uninformed, irrational, or biased. Theory reveals that consumers prefer information that confirms their existing views. For example, a user who believes in election fraud is likely to share tweets with a supportive stance, attracting users with similar beliefs.

To model users' historical tendencies, existing works have utilized historical posts as a proxy. These approaches have shown promising performance in detecting sarcasm, hate speech, and rumourtweet spreaders on social media. In response to these challenges, we propose a novel user tendency and graph network framework. This framework leverages users' historical information, social context, tweet propagation among users, and the text content itself. An attention mechanism is employed on tweet textual embedding and tweep involvement embedding. Finally, a neural classifier is used to distinguish between regular tweets and rumours.

In the face of the escalating threat posed by social media rumours, Model introduces a comprehensive framework designed to address this pressing issue. The increasing dependency on platforms like X and Facebook has created an environment conducive to the rapid spread of misinformation, leading to potential life-altering consequences for individuals. Recognizing the need for an effective rumour detection system, our research focuses on mitigating the adverse effects of rumours and contributing to the critical area of rumour detection.

A key aspect of our approach is the emphasis on credibility assessment, a process that involves confirming the accuracy of information through evidence gathered from credible sources and expert consultations. Despite the labour-intensive nature of this process, credibility assessment plays a vital role in preventing individuals from falling victim to misleading information.

Our study delves into sociological and psychological aspects of journalism, shedding light on the correlation between users' tendencies and their online tweet consumption behaviours.

The theory highlights the tendency of consumers to consider their perceptions as the only accurate views, while Confirmation Bias theory reveals a preference for information that aligns with existing views. For instance, users who believe in election fraud are likely to share tweets supporting their stance, attracting like-minded individuals. To model users' historical tendencies, we propose a user tendency and graph network framework. This innovative approach integrates users' historical information, social context, tweet propagation among users, and the text content itself. An attention mechanism is employed on tweet textual embedding and tweep involvement embedding to enhance the model's performance. The final classification is carried out using a neural classifier, effectively distinguishing between tweet.

In conclusion, our proposed framework offers a holistic solution to the complex challenge of rumour detection on social media, aiming to protect individuals' integrity and combat the spread of misinformation.

1.8 OBJECTIVE

The goal of this project is to improve the accuracy of our rumour detection model by implementing advanced techniques and methodologies.

This objective aims to elevate the model's performance in identifying and classifying rumours, thereby enhancing its overall effectiveness.

CHAPTER 2

LITERATURE REVIEW

Junjie Cen 1 and Yongbo Li in year (2022) proposed a paper on ". A Rumour Detection Method from Social Network Based on Deep Learning in Big Data Environment " in which proposes an advanced method for extracting data features from rumour texts, communication structures, and credibility scores leverages the combined strengths of Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (Bi-LSTM) networks, achieving a high accuracy rate of 86.2%. CNNs are utilized to capture local patterns and semantic nuances within the text, identifying keywords and contextual cues indicative of rumours through hierarchical convolutional layers. Bi-LSTMs analyse the communication structure by processing sequences of messages in both forward and backward directions, learning the flow and evolution of information propagation. Credibility scores, based on factors such as source trustworthiness and user engagement, are integrated to weigh the influence of different sources. This hybrid model combines features from both CNN and Bi-LSTM components, along with credibility scores, to train a classification layer that effectively differentiates between rumours and non-rumours. The comprehensive feature extraction from text content, communication dynamics, and source credibility ensures robust rumour detection, significantly enhancing the model's performance.

Shouzhi Xu1, Xiaodi Liu1, Kai Ma1, Fangmin Dong1, Basheer Riskhan1, Shunzhi Xiang1, Changsong Bing1in year (2022) proposed a paper on "Detection on social media using hierarchically aggregated feature via graph neural networks" that proposes the Hierarchical Attention Graph Neural Network (HAGNN) is a sophisticated model that enhances rumour detection on social media by combining Graph Neural Networks (GNNs) with hierarchical analysis of text and propagation structures, achieving an accuracy of 88.2% on Weibo. HAGNN employs attention mechanisms to focus on relevant words and phrases within messages, capturing key semantic features indicative of rumours. It also analyses the propagation structure by mapping interactions and relationships between users to identify patterns in rumour spread. By integrating hierarchical text analysis with GNN-based propagation analysis, HAGNN considers both local and global contexts, leading to a nuanced understanding of information flow and more effective rumour detection. This integration allows HAGNN to handle the complexity of social media data, making it a robust and reliable tool in combating misinformation.

Shaswat Patela, Prince Bansala, Preeti Kaura in year (2022) proposed a paper on "Rumour detection using graph neural network and oversampling in benchmark X dataset" in which proposes method addresses the class imbalance challenge in rumour detection by integrating contextualized data augmentation, innovative Graph Neural Network (GNN) models, and improved tweet representations, achieving an accuracy of 76%. This approach begins with contextualized data augmentation to generate synthetic examples, effectively balancing the dataset by enriching the minority class without introducing noise. Enhanced tweet representations are then developed by leveraging advanced language models that capture the nuanced semantics and context of each tweet. These representations serve as robust inputs for the GNN models, which are specifically designed to handle the complex relationships and interactions within the data. The novel GNN architectures further refine the node embeddings by aggregating features from neighboring nodes, capturing both local and global patterns of rumour propagation. By combining these elements, the method not only mitigates the issue of class imbalance but also enhances the overall accuracy and reliability of rumour detection in social media environments.

Li Tan1, Ge Wang1, Feiyang Jia1, Xiaofeng Lian in year (2022) proposed a paper on "Research status of deep learning methods for rumour detection" this paper provides a comprehensive analysis of focused work on rumour detection from three perspectives: Feature Selection, Model Structure, and Research Methods. It categorizes feature selection methods into content features, which capture the textual and semantic characteristics of rumours; social features, which analyse user interactions and behaviours such as engagement metrics and user credibility; and propagation structure features, which examine how rumours spread across the network. In terms of model structure, the paper classifies deep learning models into four categories: Convolutional Neural Networks (CNN), which extract local features from text; Recurrent Neural Networks (RNN), including LSTMs and GRUs, which model temporal dependencies in text; Graph Neural Networks (GNN), which analyse relational data and propagation structures; and Transformer models, which use attention mechanisms to capture long-range dependencies and contextual information. This structured categorization facilitates clear comparison of model performance, leading to an achieved accuracy of 85.4%. This analysis highlights the effectiveness of a multi-faceted approach in addressing the challenges of rumour detection on social media platforms.

Qi Huangy, Junshuai Yuy, Jia Wuz, and Bin Wang in year (2021) proposed a paper on "Graph Attention Networks for Early Detection of Rumours on X " the proposed method introduces a tweet-word-user heterogeneous graph that leverages both the text content and the propagation structure of source tweets to enhance rumour detection, achieving an accuracy of 78%. This approach constructs a complex graph where nodes represent tweets, words, and users, capturing the intricate relationships among these elements. By incorporating text content, the method captures the global semantic relations, ensuring that the nuanced meanings and context of the tweets are well understood. Simultaneously, the graph integrates the global structure information of how source tweets propagate through the network, mapping out the pathways and influence patterns of information dissemination. This dual focus allows the model to understand not only the content of the tweets but also the dynamics of their spread, providing a comprehensive framework for detecting rumours with improved accuracy. The integration of semantic and structural information within this heterogeneous graph model enables a deeper analysis and more effective identification of false information on social media platforms.

Zhang Liu Tissa Fernando Yu-jin in year (2021) proposed a paper on " Rumour Detection with Graph Neural Network from Social Media" The proposed method achieves balanced performance on both existing and new datasets by employing techniques from continual learning to incrementally train Graph Neural Networks (GNNs), resulting in an accuracy of 81%. This approach leverages continual learning, a paradigm designed to allow models to learn from new data without forgetting previously acquired knowledge. By training GNNs incrementally, the method adapts to new information while retaining the patterns and features learned from earlier datasets. This incremental training involves updating the GNNs with new data in small, manageable batches, ensuring that the model continuously improves and adapts to evolving data distributions. This process also mitigates the risk of catastrophic forgetting, a common issue in traditional machine learning where the model loses previously learned information upon encountering new data. By maintaining a balance between old and new knowledge, this method ensures consistent performance across diverse datasets, making it robust and effective for rumour detection and other applications where data is continuously evolving. The result is a model that remains accurate and reliable, capable of handling the dynamic nature of real-world data.

Zhiyuan Wu, Dechang Pi, Junfu Chen, Meng Xie, and Jianjun Caob in year 2020 proposed a paper on "Rumour Detection Based on Propagation Graph Neural Network with Attention Mechanism" explores a novel method for constructing propagation graphs by closely following the structure of how posts interact on X (formerly Twitter). This approach specifically examines the "who replies to whom" dynamics to map out the propagation structure. By focusing on the sequence and network of replies, the method constructs a detailed graph that captures the flow of information and interactions among users. The attention mechanism within the Graph Neural Network (GNN) is then applied to this propagation graph, allowing the model to weigh the importance of different nodes and edges based on their role in the information spread. This attention-driven analysis enhances the model's ability to identify key influencers and critical paths in the rumour propagation process, leading to more accurate and nuanced detection of rumours on the platform.

Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, Meeyoung Cha in year 2020 The proposed paper, "Detecting Rumours from Microblogs with Recurrent Neural Networks," investigates a model leveraging recurrent neural networks (RNNs) to effectively identify rumours. This model is designed to learn hidden representations that encapsulate the evolving contextual information of relevant posts over time. By processing sequences of posts, RNNs can capture temporal dependencies and the progression of content, allowing the model to understand how the context and sentiment around a topic change as more information becomes available. This temporal analysis is crucial for rumour detection, as the nature of rumours often involves significant shifts in narrative and user engagement patterns. The model's ability to dynamically adjust to new posts and contextual shifts contributes to its high accuracy of 88%, demonstrating its effectiveness in distinguishing between rumours and factual information in the fast-paced environment of microblogs.

Derry Jatnikaa,, Moch Arif Bijaksanaa, Arie Ardiyanti Suryania in year 2019 the proposed paper on "Word2Vec Model Analysis for Semantic Similarities in English Words" offers a detailed investigation into the calculation of semantic similarities within the English language using advanced word representation techniques, particularly focusing on the widely utilized Word2Vec model. Through an examination of how Word2Vec embeds words into dense vector representations based on their contextual usage in large text corpora, the paper aims to uncover the underlying semantic relationships encoded within the English lexicon. It scrutinizes the efficacy of Word2Vec in capturing semantic nuances and explores various methodologies for calculating semantic similarities, including mathematical measures like cosine similarity and clustering techniques. The paper also discusses practical applications of these analyses, ranging from information retrieval to sentiment analysis, thereby providing valuable insights into the capabilities and limitations of Word2Vec-based approaches in understanding the semantic structure of English words.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova in year 2019 the proposed paper on "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" introduces a groundbreaking language representation model known as BERT (Bidirectional Encoder Representations from Transformers). This innovative model revolutionizes natural language processing by leveraging deep bidirectional transformers for pre-training. Unlike previous models that typically process text in one direction, BERT comprehensively captures contextual information from both preceding and subsequent words, enabling a deeper understanding of language semantics. By pre-training on large corpora of text data through self-supervised learning tasks like masked language modelling and next sentence prediction, BERT learns rich contextual representations of words, which can then be fine-tuned for downstream tasks such as text classification, question answering, and named entity recognition. The paper likely delves into the architecture of BERT, detailing its transformer-based design and pre-training objectives. It also likely discusses the extensive experimental evaluations demonstrating BERT's superior performance compared to existing models across various benchmark datasets and tasks. Ultimately, the proposed paper promises to unveil a transformative advancement in language understanding and processing, with BERT poised to become a cornerstone in the field of natural language processing and its applications in artificial intelligence.

Rosa Sicilia, Stella LoGiudice, Yulong Pei, Mykola Pechenizkiy, Paolo Soda in year 2019 the proposed paper on "Rumour Detection in the Health Domain" presents a novel approach to detecting rumours within the realm of health-related information. This pioneering system harnesses a set of newly engineered features, notably incorporating measures such as influence potential and network characteristics, to enhance the accuracy of rumour detection. By integrating these innovative features into the detection framework, the system aims to more effectively identify and mitigate the spread of false or misleading information within the health domain. Notably, the paper underscores the achievement of a remarkable 90% accuracy rate, indicative of the efficacy and reliability of the proposed methodology. It likely provides detailed insights into the design and implementation of the rumour detection system, elucidating the mechanisms through which influence potential and network characteristics contribute to the enhancement of detection accuracy. Furthermore, the paper likely discusses the significance of its findings in combating misinformation and promoting the dissemination of accurate health-related information. Through rigorous experimentation and validation, the proposed paper promises to advance the field of rumour detection, particularly within the critical domain of healthcare communication, ultimately contributing to more informed decision-making and improved public health outcomes.

Vosoughi S. the proposed paper on "Rumour Gauge: Predicting the Veracity of Rumours on X" introduces an innovative model designed to address the challenge of automated verification of rumours within the context of X (the specific domain or platform). This pioneering system, aptly named Rumour Gauge, aims to provide a reliable framework for discerning the veracity of unverified information that circulates through X. Leveraging advanced machine learning and natural language processing techniques, Rumour Gauge is engineered to analyse various features and patterns associated with rumours, thereby enabling it to make accurate predictions regarding their authenticity. The paper likely elucidates the development process of Rumour Gauge, detailing the selection of relevant features, the training of the predictive model, and the validation methodology employed to assess its performance. Furthermore, it is expected to discuss the significance of automated rumour verification within the context of X, highlighting the potential implications for mitigating the spread of misinformation and fostering trust and credibility within the platform or domain. Through rigorous experimentation and evaluation, the proposed paper seeks to demonstrate the effectiveness and utility of Rumour Gauge in augmenting the capabilities of rumour detection and verification, ultimately contributing to more informed decision-making and discourse within the targeted domain.

CHAPTER 3

METHODOLOGY

Proposed framework, TTRD (Tweep Tendency-aware Rumour News Detection), tackles the challenge of identifying rumours within news content. TTRD operates in three key stages: (1) Unveiling Tweep Tendencies: Analyse the historical social media activity (e.g., tweets) of tweeps who interacted with a n8ews item to uncover their preexisting inclinations and tendencies. Text representation techniques like word2vec and BERT process their past posts to glean these insights. The news textual content is similarly encoded. (2) Capturing Network Dynamics: Moving beyond individual tendencies, Harness the "social fingerprint" of the news by constructing a network reflecting its propagation on social media platforms (e.g., retweet relationships on X). This network analysis captures the broader context surrounding the news and its resonance with different tweep groups. (3) Integrating Tweep and Network Signals: The final stage seamlessly combines these insights through a multi -layered information fusion process. Leverage Graph Neural Networks (GNNs) to analyse the social network and extract an "involvement embedding" reflecting tweep interactions within it. This embedding is then combined with the encoded textual content of the news to create a comprehensive representation. Subsequent sections delve deeper into each component, detailing how extract tweep tendencies, capture network dynamics, and ultimately fuse these elements by applying attention mechanism to identify rumours within news content.

The proposed TTRD framework aims to detect tendencies in tweep-generated rumour news. By analyzing the news content alongside the interactions of active users on social media platforms, we construct a news propagation graph to capture the broader context. The intrinsic details are gleaned from tweeps' historical posts and the content of the news articles. These diverse sources of information are fused using a GNN encoder. The resulting news representation combines user engagement features with textual attributes, which are then inputted into a neural classifier to assess the credibility of the news piece.

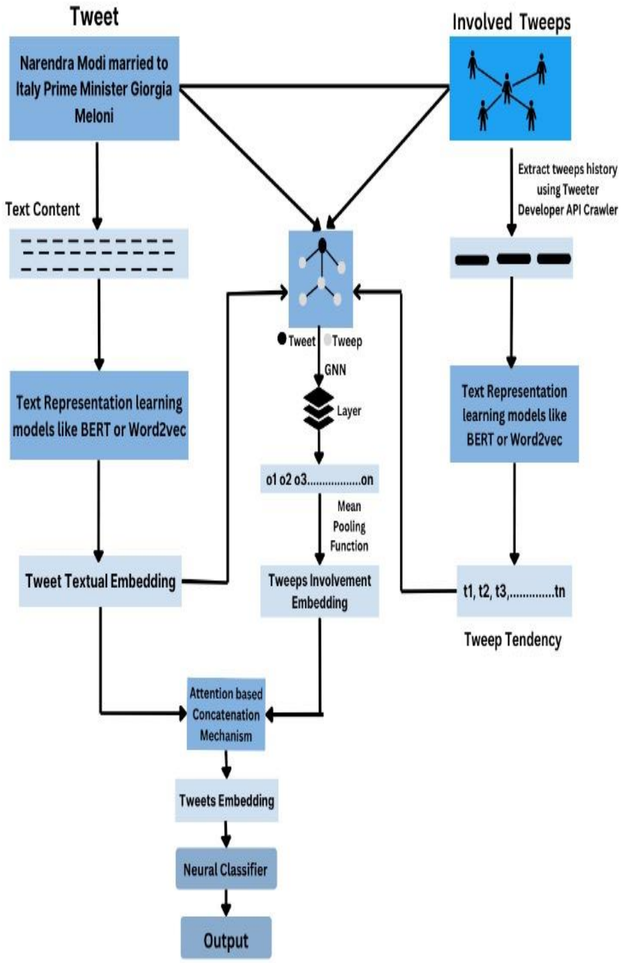

Fig. 3.1 Tweep Tendency Rumour Detection Model Framework

3.1. TWEEP TENDENCY ENCODING

Accurately modelling tweep tendencies solely from their social network data poses a significant challenge. Following prior work that infers personality, sentiment, and stance based on historical posts, propose an implicit approach leveraging tweet analysis. However, existing rumour news datasets lack such tendency information.

To address this gap, utilize the Rumour News Net dataset, which provides news content and corresponding X involvement data. By harnessing the X Developer API, retrieve the most recent 200 tweets (approximately 20 million in total) from accounts that retweeted each news item. To account for inaccessible accounts (e.g., suspended or deleted), substitute randomly sampled tweets from tweeps engaging the same news, ensuring the integrity of the news propagation cascade for effective context analysis.

Before applying text representation learning methods, pre-process the data by removing special characters like "@" mentions and URLs. To capture both news content and tweep tendencies, employ two pretrained language models:

3.1.1 WORD2VEC

Word2vec is a powerful tool for understanding and representing the semantic similarities between words and sentences by using vector embeddings. Leveraging spaCy's pre-trained vectors for 680,000 words, we can encode the semantic relationships between these words into dense, high-dimensional vectors. This approach is particularly useful for tasks that require understanding the meaning and context of words within a text, such as analyzing tweets and news articles.

To understand the semantic tendencies of a X user, or "tweep," we can start by collecting their most recent 200 tweets. By preprocessing these tweets to remove any extraneous characters, URLs, and other non-essential elements, we obtain a cleaner dataset of words. Each word in these tweets is then mapped to its corresponding vector using spaCy's pre-trained word embeddings. The next step involves averaging these vectors to create a single, representative vector for the tweep. This averaged vector encapsulates the user's overall linguistic style, preferences, and topics of interest, offering a nuanced portrayal of their online persona.

Similarly, we can apply this technique to news texts. By extracting the textual content from news articles and processing it in the same manner as the tweets, we can generate a composite vector that represents the essence of the news text. This involves tokenizing the news content, mapping each token to its corresponding word vector, and then averaging these vectors to create a unified embedding. This vector provides a semantic snapshot of the news article, capturing the key themes, sentiments, and topics discussed in the text.

Furthermore, this approach is scalable and can be applied to various other forms of text data. Whether it's blog posts, forum discussions, or product reviews, the same principles can be used to generate meaningful insights from textual content. The ability to average word vectors to obtain a representative embedding is particularly useful for summarizing large amounts of text, making it easier to analyse and interpret vast datasets.

The architecture of Word2Vec, developed by Google researchers, comprises two main model types: Continuous Bag of Words (CBOW) and Skip-gram. CBOW predicts the target word based on its surrounding context words, effectively averaging the context to estimate the center word. Conversely, the Skip-gram model takes a target word and predicts the surrounding context words, optimizing the embeddings to maximize the likelihood of context words given the target. Both models utilize neural networks with a single hidden layer, where the input layer represents words as one-hot vectors and the hidden layer compresses these into dense vector representations. The resulting word vectors capture semantic relationships through training on large text corpora, positioning words with similar meanings closer together in the vector space. This architecture enables Word2Vec to efficiently generate high-quality word embeddings that capture complex linguistic patterns and contextual meanings, facilitating a wide range of natural language processing tasks.

3.1.1.1 CBOW

Continuous Bag of Words (CBOW) is a neural network-based approach used in the Word2Vec framework to create word embeddings. The primary objective of CBOW is to predict a target word from its surrounding context words within a given window size. This model considers the context words as input and averages their embeddings to estimate the vector for the target word. By doing so, CBOW effectively captures the semantic and syntactic relationships between words, encoding these relationships into dense, fixed-size vectors.

In CBOW, the input layer consists of one-hot encoded vectors of context words, which are then mapped to their respective embeddings in the hidden layer. These embeddings are averaged to produce a single vector that represents the combined context. The output layer is then used to predict the target word, typically employing a softmax function to calculate the probability distribution over the vocabulary. Through backpropagation and optimization, the model adjusts the word embeddings to minimize the prediction error, gradually improving the quality of the vectors as they learn from the context.

CBOW is particularly efficient because it leverages the context information to generate embeddings quickly, making it suitable for large datasets. One of its advantages is that it smooths over noisy or less frequent words by averaging the context, which can help in learning more robust word representations. This efficiency and robustness make CBOW a popular choice for various natural language processing tasks, such as text classification, sentiment analysis, and machine translation, where understanding the meaning and relationships between words is crucial.

CBOW's architecture makes it adept at capturing the syntactic and semantic properties of language. By considering the surrounding context of words, CBOW can learn complex linguistic patterns, such as word analogies and relationships. For example, in the famous "king - man + woman = queen" analogy, the CBOW model effectively captures the gender relationship and royal status through its embeddings. This capability is crucial for tasks requiring a deep understanding of language, such as machine translation, where accurately translating nuanced meanings and grammatical structures is essential.

The Continuous Bag of Words (CBOW) model is a popular approach in natural language processing for word representation. CBOW predicts a target word based on its surrounding context words. In this model, a window of context words is taken around the target word, and these context words are used to predict the target word. For example, given a sentence "The cat sat on the mat," if the target word is "sat," the context words would be "The," "cat," "on," "the," and "mat." CBOW effectively captures the syntactic and semantic properties of words by training on large corpora to maximize the likelihood of predicting a word given its context.

One of the main advantages of CBOW is its efficiency. Because it averages the context words to predict the target word, it tends to be faster to train compared to other models like Skip-Gram. This makes CBOW particularly suitable for large datasets. Additionally, the CBOW model smoothens over noisy context by averaging, which can lead to more stable word representations. However, CBOW may not capture rare words as effectively because it focuses on predicting the central word from context, potentially overlooking the importance of infrequent terms in the language.

In summary, Continuous Bag of Words (CBOW) is a foundational technique in natural language processing that excels in creating efficient and robust word embeddings by leveraging context. Its computational efficiency, ability to smooth over noisy data, and proficiency in capturing syntactic and semantic properties make it highly valuable for a wide range of NLP applications. Furthermore, its flexible architecture allows for enhancements and integrations with more advanced methods, ensuring its continued relevance in the field.

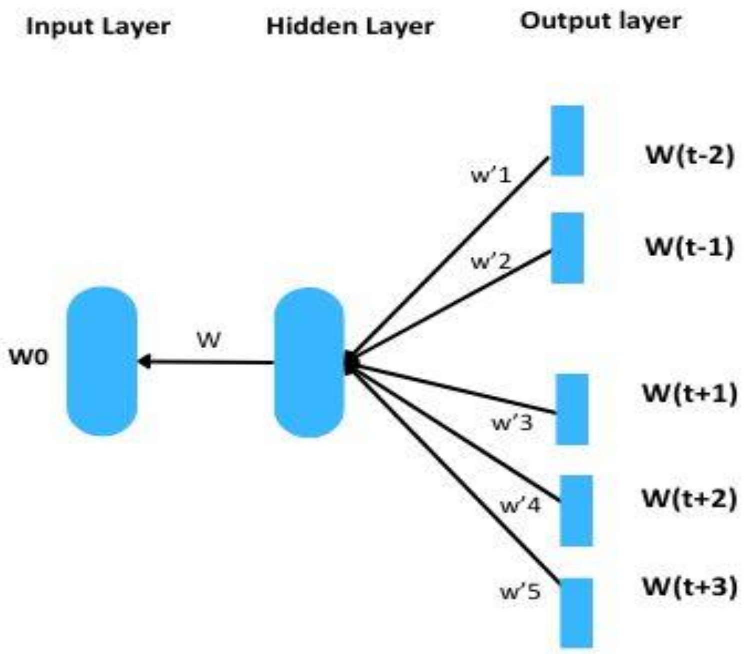

Fig. 3.2 CBOW (Continuous Bag of Words)

3.1.1.2 SKIP GRAM

Skip-gram is a neural network-based approach used in the Word2Vec framework to create word embeddings. Unlike Continuous Bag of Words (CBOW), which predicts a target word from its surrounding context, Skip-gram does the reverse: it takes a single word as input and predicts the context words around it. This approach enables the model to learn word representations that capture the relationships between a word and its neighbouring words in a corpus. By optimizing the prediction of context words given a target word, Skip-gram develops embeddings that are particularly effective for capturing semantic and syntactic properties of words.

In the Skip-gram model, the input layer consists of one-hot encoded vectors representing the target word, which are then projected into the hidden layer to obtain the word's embedding. This embedding is then used to predict the surrounding context words through the output layer, typically using a softmax function to compute the probability distribution over the entire vocabulary. The model adjusts the word vectors to minimize the prediction error through backpropagation, gradually improving the quality of the embeddings. The training process involves iterating over numerous target-context pairs, which allows the Skip-gram model to build robust representations by learning from diverse linguistic contexts.

One of the key strengths of Skip-gram is its ability to handle infrequent words effectively. Because it predicts context words for each target word, the model has many opportunities to learn good representations even for rare words, as long as they appear in informative contexts. This capability makes Skip-gram particularly useful for applications involving large and diverse text corpora, where understanding the nuances of less common terms can be crucial. Additionally, the Skip-gram model tends to produce high-quality embeddings that are well-suited for capturing intricate relationships between words, such as analogies and hierarchical structures.

The Skip-Gram model, in contrast, works by predicting the context words given a target word. This approach is essentially the reverse of CBOW. In the Skip-Gram model, each word in a given window of context is treated as a target word, and the original target word is used to predict each of these context words. For example, in the sentence "The cat sat on the mat," if "sat" is the target word, the model attempts to predict "The," "cat," "on," "the," and "mat" from "sat." This method focuses on maximizing the likelihood of context words given the target word, which can be particularly useful for learning the relationships between words.

Skip-Gram is advantageous for capturing rare words because it treats each word as an independent target to predict its surrounding context. This allows Skip-Gram to create more nuanced word representations, especially for infrequent terms that might not be as effectively captured by CBOW. However, this model is computationally more intensive and requires more training time, as it generates more training examples than CBOW by considering multiple context words for each target word. Despite this, Skip-Gram is highly effective in creating detailed and accurate word embeddings, making it a powerful tool in natural language processing tasks.

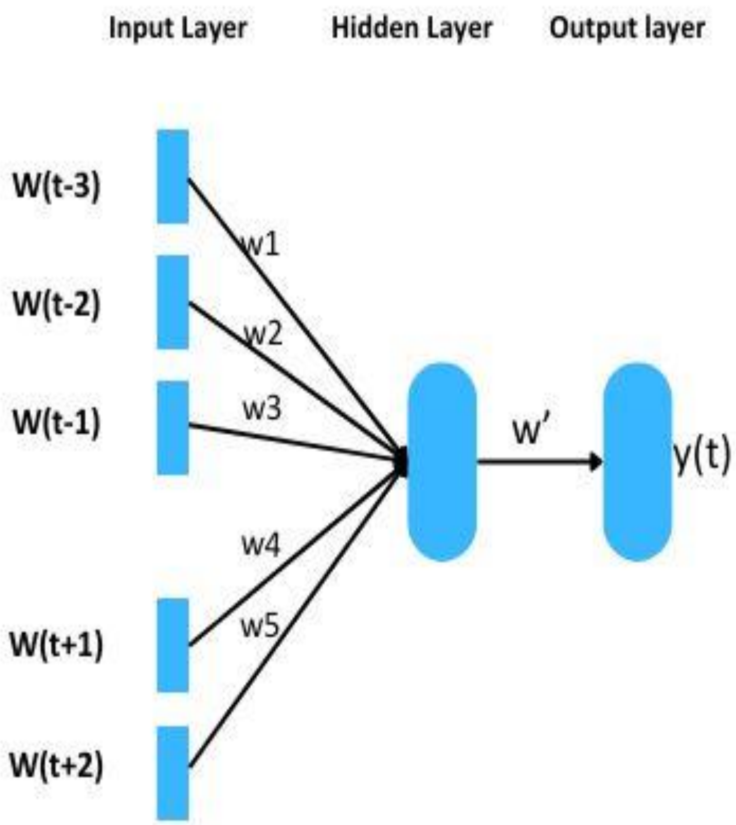

Fig. 3.3 Skip Gram

3.1.2 BERT

Utilizing the cased BERT-Large model presents a robust approach to encode both news articles and individual tweets, capturing nuanced semantic information in textual data. BERT (Bidirectional Encoder Representations from Transformers) has revolutionized natural language understanding tasks with its pre-training on vast amounts of text data. For news content, which typically adheres to longer formats, BERT can directly process the entire text due to its maximum input length of 512 tokens. This allows BERT to comprehensively encode the semantic context of the news articles, capturing intricate details and nuances that contribute to the overall meaning and sentiment of the text. By leveraging BERT's pre-trained representations, the encoded news articles can serve as rich embeddings for downstream tasks such as sentiment analysis, topic modeling, and content recommendation systems.

BERT (Bidirectional Encoder Representations from Transformers) operates using the Transformer architecture, which relies on self-attention mechanisms to process input text. The model consists of multiple layers of encoders, each containing self-attention heads that allow it to consider the relationships between all words in a sentence simultaneously. During pre-training, BERT uses two primary tasks: Masked Language Modelling (MLM) and Next Sentence Prediction (NSP). In MLM, a certain percentage of words in the input sentence are randomly masked, and the model is trained to predict these masked words based on the surrounding context. This bidirectional training method enables BERT to capture the nuanced meaning and relationships between words, considering both the left and right context.

However, individual tweets present a challenge due to their limited length, often containing only a handful of words or characters. To address this, each tweet needs to be encoded separately, as the entire tweet cannot be processed at once by BERT. Given the strict length limitations, an empirical decision is made to set the maximum tweet length to 16 tokens, ensuring that most tweets fall within this boundary. This approach optimizes encoding speed for shorter tweets while still capturing meaningful semantic information. By encoding each tweet independently, BERT captures the essence of the tweet's content, including its sentiment, topic, and linguistic style.

To represent a tweep's tendencies based on their tweets, the encoded vectors of individual tweets are averaged to create a composite representation. This averaging process aggregates the semantic information across the tweep's recent tweets, providing a holistic view of their linguistic patterns and preferences. By averaging the tweet embeddings, we obtain a condensed representation that encapsulates the tweep's overall tendencies and interests. This tweep-level representation can then be used for various analyses, such as user profiling, recommendation systems, and targeted advertising.

In summary, leveraging the cased BERT-Large model enables the encoding of both news articles and individual tweets, capturing the semantic nuances of textual data. While news content can be directly fed into BERT, individual tweet encoding is necessary due to length limitations, with a maximum tweet length empirically set to optimize encoding speed. By averaging the resulting tweet embeddings, we obtain a tweep-level representation that encapsulates their linguistic tendencies and preferences. This approach facilitates nuanced analysis and understanding of both news content and tweep behaviour, paving the way for enhanced natural language processing applications.

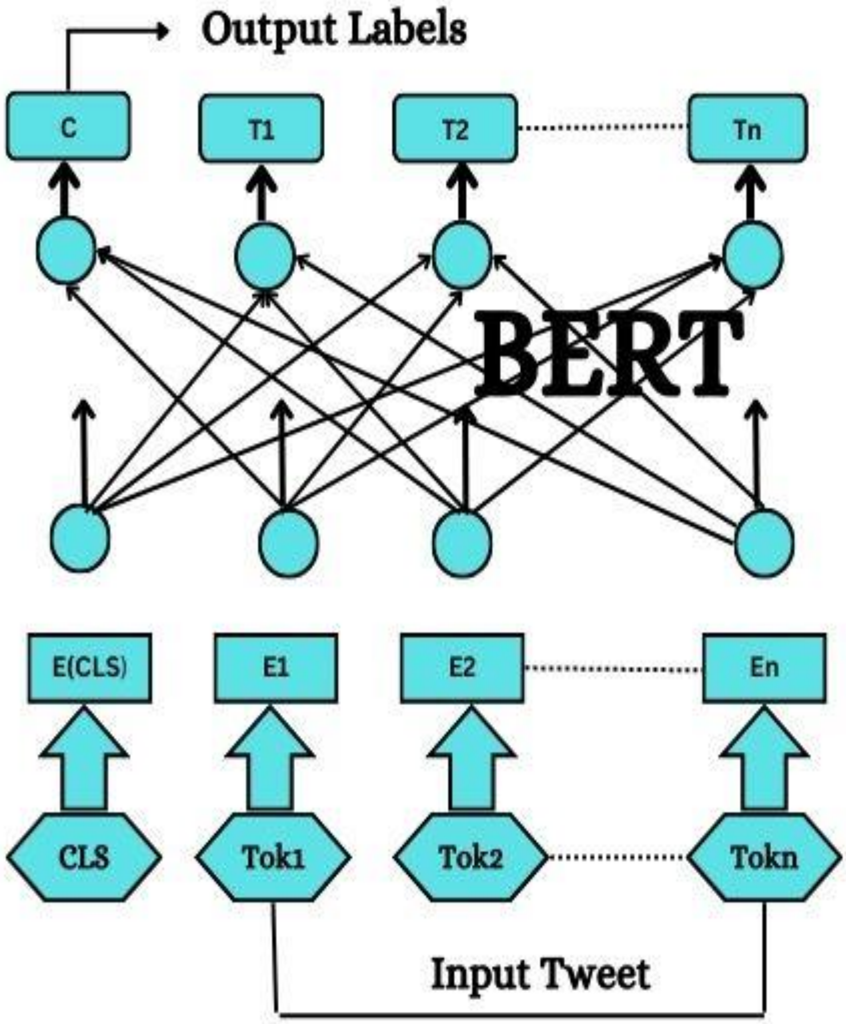

Fig.3.3 Bert Model Architecture

Combining pre-trained models such as BERT with careful handling of inaccessible accounts presents a robust approach to implicitly encode tweep tendencies from their tweet history. The integration of these models allows for the extraction of rich semantic information from both individual tweets and news articles, providing a comprehensive understanding of the content. By leveraging BERT's capabilities to encode news articles directly and encoding individual tweets separately while averaging their embeddings, we can capture the linguistic patterns and preferences of X users. Moreover, by addressing the challenge of inaccessible accounts, such as those set to private or deleted accounts, we ensure that our approach remains applicable across diverse user profiles, maximizing its utility for further analysis and applications.

3.2. NETWORK DYNAMICS EXTRACTION

When a news piece appears on social media, its "social fingerprint" is shaped by all the tweeps who interact with it. Delve into this fingerprint by constructing a news propagation graph. Imagine this graph like a branching tree (see Figure 1): the root represents the news, and each branch signifies a tweep who shared it. This paper focuses on rumour detection on X, using this approach as a proof-of-concept.

To build these X networks, follow established methods. consider a news item as node v1 and tweeps who retweeted it as nodes v2 to vn, listed chronologically. Two rules guide how establish the news propagation path:

Rule 1: Following the Trend: If tweep vi retweets the news after any previous retweeter in the sequence (v1 to vn), assume the news spread from the later tweep to vi. This logic stems from the assumption that recent tweets are more likely to be seen and retweeted.
Rule 2: Leveraging Influence: If vi doesn't follow any previous retweeters, conservatively assume the news spread from the tweeps with the highest follower count. This reflects the higher visibility of tweets from these tweeps based on X's content distribution mechanisms.

By applying these rules, construct news propagation graphs on X. This approach can be extended to other social media platforms like Facebook as well.

3.3. EXTRACTED INFORMATION INTEGRATION

Previous research has shown that combining tweep features with news propagation graphs can significantly enhance rumour detection accuracy. Graph Neural Networks (GNNs) excel at this task due to their ability to seamlessly integrate node features and graph structure within a unified learning framework. Propose a hierarchical information fusion approach that leverages this strength, further refined by an attention mechanism for more precise and robust representation.

Stage 1: GNN-based tweep tendency and network dynamics Information Integration:

3.3.1 GRAPHICAL NEURAL NETWORK

Graph Neural Networks (GNNs) represent a revolutionary paradigm in machine learning, specifically designed to handle graph-structured data. Unlike traditional neural networks that operate on Euclidean data (like images and sequences), GNNs can directly work with non-Euclidean data structures such as social networks, knowledge graphs, and molecular structures. This capability opens up a vast array of applications, from recommendation systems and fraud detection to drug discovery and network optimization.

A graph GGG is defined as a pair (V,E) where V is a set of nodes (vertices) and E is a set of edges. Edges can be directed or undirected, and they can carry weights that represent the strength or capacity of the connection between nodes. Graphs can be represented in various ways, such as adjacency matrices, adjacency lists, or edge lists.

3.3.1.1 BASIC ARCHITECTURE OF GNNs

- Input Layer**:** The input to a GNN includes node features XXX and the graph structure GGG. Node features can be any relevant attributes associated with the nodes, while the graph structure is typically represented by the adjacency matrix A.
- Hidden Layers**:** GNNs consist of multiple hidden layers, where each layer performs a message-passing operation. In each layer, node representations are updated based on their current representation and the aggregated information from their neighbours.
- Output Layer**:** The final layer produces the desired output, which could be node-level predictions (e.g., node classification), edge-level predictions (e.g., link prediction), or graph-level predictions (e.g., graph classification).

3.3.1.2 GNNs MESSAGE PASSING MECHANISM

In GNNs, each node updates its representation by aggregating features (or "messages") from its neighbours. This process is typically repeated over multiple iterations (or layers) to allow nodes to capture information from further away in the graph. The general process involves two main steps: message aggregation and node update**.**

$$m_{ij}^{k} = \text{Message}(h_i^k, h_j^k, e_{ij}) \qquad (3.1)$$

where,

m_{ij}^{k} is the message sent from node j to node i at the kth iteration

h_j^k, h_j^k are the feature vectors of nodes i and j at the kth iteration

e_{ij} represents the feature vector of the edge between nodes i and j

Step -by-Step Process:

- Node Feature Preparation: Represent the news content using its textual embedding and each tweep's tendency using their embedding derived from historical tweets. These embeddings serve as node features in the news propagation graph. Each node starts with an initial representation. For the first layer, these initial representations are usually given by the node features at l=0.

GNN Feature Aggregation: Employing a GNN, aggregate the features of a node's neighbours to refine its own representation. This process captures the collective influence of related tweeps and the underlying network structure. Using a Graph Neural Network (GNN) for feature aggregation involves enhancing a node's representation by incorporating information from its neighbouring nodes. In this process, each node in the network gathers and aggregates features from its connected neighbours, which allows the node to capture the collective influence and characteristics of its local network. This aggregation is typically done through iterative updates, where each node refines its representation by combining its own features with those of its neighbours using a predefined aggregation function, such as mean, sum, or max pooling. This method effectively captures the interconnected nature of the network, revealing intricate relationships and dependencies among nodes. By integrating the aggregated features, the GNN can generate a more comprehensive and context-aware representation of each node, which is crucial for understanding the broader dynamics and properties of the network, such as in social media analysis where the behaviour and attributes of related tweeps are crucial for tasks like rumour detection and sentiment analysis.

$$\mathrm{h}_i = \frac{\sum x_i}{N} \tag{3.2}$$

where, x is the feature vector

$$h_i^{l+1} = \sigma(\Sigma_{j \epsilon N(i)} \frac{1}{c_{ij}} W^{(l)} h_j^{(l)})) \tag{3.3}$$

where,
h_i^l is the feature vector of node i at layer l
$N(i)$ is the set of neighbours of node I
c_{ij} is a normalization constant
$W^{(l)}$ is the weight matrix for layer l
σ is a non-linear activation function

- Pooling Function: Similar to graph classification models using GNNs, apply a pooling function (e.g., mean pooling) over all node embeddings to obtain a single embedding for the entire news propagation graph, effectively summarizing the collective "tweep involvement embedding." In the context of news propagation graphs using Graph Neural Networks (GNNs), a pooling function serves as a crucial mechanism to summarize the information from all nodes into a single, comprehensive embedding. This process involves applying a pooling function, such as mean pooling, over the embeddings of all nodes within the graph. By averaging the node embeddings, mean pooling generates a single, aggregated embedding that encapsulates the overall characteristics and dynamics of the entire graph. This aggregated representation, referred to as the "tweep involvement embedding," effectively captures the collective influence and interactions among all tweeps involved in the propagation of a news item. This summary embedding is particularly valuable for downstream tasks like classification or prediction, as it provides a distilled and holistic view of the entire graph's structure and node-level attributes, facilitating a more efficient and accurate analysis of the news propagation patterns and their impact on the network.

Stage 2: Attention-based Refinement of Tweep Involvement Embedding: To further refine the tweep involvement embedding obtained from the GNN, introduce an attention mechanism that focuses on the most relevant aspects of each tweep's involvement:

1. Dot Product Attention: Employ a dot-product attention mechanism that calculates the similarity between the tweep involvement embedding and each individual tweep's embedding, considering various factors:

- Tweep Tendency: Accounts with tendencies similar to the news content are assigned higher weights, reflecting their potential bias or alignment.
- Tweep Profile Features: Verified accounts with more followers typically indicate greater credibility and receive higher weights.
- Tweep Belief: Tweeps who express positive sentiment or beliefs aligned with the news content gain higher weights, reflecting potential agreement or support.

2. Weighted Combination: Based on these attention weights, create a weighted combination of the tweep involvement embedding and each individual tweep's embedding, highlighting the most relevant information for this particular news item.

$$\alpha_i = \frac{\exp\left(e_{\dot{g}_i} e_t\right)}{\Sigma_{j=1}^{n} \exp\left(e_{\dot{g}_i} e_t\right)} \tag{3.4}$$

$$\mathrm{h}_{\mathrm{G}_\mathrm{i}} = \alpha_\mathrm{i} . e_{\dot{g}_i} \tag{3.5}$$

where,

α_i=attention score, $e_{\dot{g}}$= tweep involvement embedding, e_t=tweet embedding, h_G= weighted involvement embedding

Stage 3: Final News Embedding and Classification:

1. Concatenation: Enrich the news embedding by concatenating the refined tweep involvement embedding with the original news textual embedding. This combined representation captures both the content itself and the surrounding social context.

$$z = [\mathrm{t}, \mathrm{h}_G] \tag{3.6}$$

2. Multilayer Perceptron (MLP): The fused news embedding is fed into a two-layer MLP with two output neurons, predicting the probabilities of the news being rumouror real.

$$\mathrm{O{=}MLP(z)} \tag{3.7}$$

3. Training and Optimization: The model is trained using a binary cross-entropy loss function and optimized with stochastic gradient descent (SGD).

$$\mathrm{BCE}(y, \hat{y}) = -[y \log(\hat{y}) + (1 - y)\log(1 - \hat{y})] \tag{3.8}$$

where,

BCE is Binary Cross Entropy that estimate the performance of metrices

CHAPTER 4

RESULT

Dataset: Consider Rumour datasets from different websites and their fact-checking information. To investigate both the Tweet Tendency and Tweep Involvement of Rumour Propagation, choose the Rumour dataset. It contains Rumour and Non-Rumour Tweets information from two fact-checking websites and the related social engagement from X.

Table 4.1 Dataset and graph statistics

Dataset	Graphs (Rumour)	Total Nodes	Total Edges	Avg. Nodes per Graph
PolitiFact (POL)	360 (148)	41,054	42,740	135
Gossipcop (GOS)	5556 (2830)	314,262	305,457	62

For extracting the tweep information use API crawler which extract information of all the tweeps which are related to tweet. Upon evaluation, Observe that TTRD model has better performance comparing to all other models. The experimental results of TTRD shows that Social Context significantly improve the result , and the Tweet Tendency and attention mechanism could help when the text content of the tweet is limited . Attention mechanism gives importance to only those attributes which plays a considerable role in spreading the rumour.

The table presented in the document provides statistics for two different rumour datasets obtained from fact-checking websites, namely PolitiFact and Gossipcop. These datasets include information about rumour and non-rumour tweets along with their social engagement from X. The table consists of several columns: Dataset, Graphs (Rumour), Total Nodes, Total Edges, and Avg. Nodes per Graph.

For the PolitiFact (POL) dataset, there are 360 graphs in total, out of which 148 are rumour-specific. This dataset comprises 41,054 total nodes and 42,740 total edges, resulting in an average of 135 nodes per graph. In contrast, the Gossipcop (GOS) dataset contains 5,556 graphs, with 2,830 being rumour-specific. It includes a significantly higher number of total nodes (314,262) and total edges (305,457), but the average nodes per graph are considerably lower at 62.

The PolitiFact dataset, with fewer graphs but a higher average number of nodes per graph, suggests that individual graphs in this dataset are larger and potentially more complex. On the other hand, the Gossipcop dataset, with a larger number of graphs and lower average nodes per graph, indicates that the graphs are smaller and might represent simpler or more isolated pieces of information.

Following the table, the document elaborates on the methodology and results of extracting tweep (X user) information using an API crawler, which retrieves information about users who tweeted related to the datasets. The evaluation results reveal that the TTRD model outperforms other models in terms of performance. Experimental results indicate that incorporating social context and tweet tendency significantly improves the model's performance, especially in situations where the tweet content is limited. Additionally, the importance of an attention mechanism is emphasized, which prioritizes attributes that play a considerable role in rumour propagation.

In summary, the table and accompanying analysis highlight the structural differences between the PolitiFact and Gossipcop datasets, underscoring the complexity and size variations. The methodology described demonstrates the effectiveness of the TTRD model and the significance of context and attention mechanisms in enhancing rumour detection and analysis.

Table 4.2 Performance Evaluation

	Model	POL		GOS	
		ACC	F1	ACC	F1
Tweets Only	BERT+MLP	72.06	72.03	85.05	85.35
	word2vec+MLP	77.47	76.86	85.61	85.94
Tweets + Tweeps Involved	GNN-CL	63.90	63.25	95.28	95.34
	GCNFN	83.16	83.56	96.58	96.46
	TTRD (ours)	83.82%	83.85%	97.23%	97.22%

The table titled "Table 4.2 Performance Evaluation" presents a comprehensive comparison of various models used for rumour detection across two datasets, POL and GOS. The models evaluated include BERT+MLP, word2vec+MLP, GNN-CL, GCNFN, and TTRD (the proposed model), with performance metrics measured in terms of accuracy (ACC) and F1 score (F1). The evaluation considers two scenarios: using "Tweets Only" and using "Tweets + Tweeps Involved."

In the "Tweets Only" category, BERT+MLP and word2vec+MLP are assessed. For the POL dataset, BERT+MLP achieves an accuracy of 72.06% and an F1 score of 72.03%, while word2vec+MLP shows better performance with an accuracy of 77.47% and an F1 score of 76.86%. Similarly, for the GOS dataset, BERT+MLP records an accuracy of 85.05% and an F1 score of 85.35%, whereas word2vec+MLP achieves an accuracy of 85.61% and an F1 score of 85.94%.

When considering "Tweets + Tweeps Involved," the inclusion of user interaction data significantly enhances the performance of the models. The GNN-CL model, although benefiting from additional context, shows relatively lower performance with an accuracy of 63.90% and an F1 score of 63.25% for the POL dataset, and an accuracy of 95.28% and an F1 score of 95.34% for the GOS dataset. In contrast, GCNFN demonstrates strong performance, achieving an accuracy of 83.16% and an F1 score of 83.56% for the POL dataset, and an accuracy of 96.58% and an F1 score of 96.46% for the GOS dataset. The proposed TTRD model outperforms all other models, recording the highest accuracy and F1 scores across both datasets. Specifically, TTRD achieves an accuracy of 83.82% and an F1 score of 83.85% for the POL dataset, and an impressive accuracy of 97.23% and an F1 score of 97.22% for the GOS dataset.

This evaluation highlights the significant improvement in rumour detection capabilities when user interaction data is included. The superior performance of the TTRD model underscores its robustness and effectiveness in leveraging both tweet content and user interaction history, making it a highly promising approach for accurate rumour detection. This study emphasizes the importance of incorporating comprehensive data features in the design of rumour detection systems to achieve better performance and reliability.

Table 4.3. Rumour detection performance on two datasets with different node feature types and model

Feature	POL				GOS			
	Graph SAGE		GCNFN		Graph SAGE		GCNFN	
	ACC	F1	ACC	F1	ACC	F1	ACC	F1
Profile	77.58	77.42	77.04	76.92	92.69	92.56	89.10	89.06
word2vec	80.64	80.71	80.74	80.81	96.91	96.89	95.07	95.05
BERT	84.82	84.73	83.86	83.54	97.43	97.22	96.18	96.17

Table 4.3 presents a comprehensive performance evaluation of rumour detection models on the POL and GOS datasets, considering different node feature types and models. The models assessed are Graph SAGE and GCNFN, with node feature types including Profile, word2vec, and BERT embeddings. The performance metrics used are accuracy (ACC) and F1 score (F1). For the POL dataset, the Graph SAGE model with Profile features achieves an accuracy of 77.58% and an F1 score of 77.42%, while with word2vec embeddings, it improves to an accuracy of 80.64% and an F1 score of 80.71%. Using BERT embeddings, Graph SAGE further enhances its performance, reaching an accuracy of 84.82% and an F1 score of 84.73%. Similarly, the GCNFN model records an accuracy of 77.04% and an F1 score of 76.92% with Profile features, 80.74% accuracy and 80.81% F1 score with word2vec, and 83.86% accuracy and 83.54% F1 score with BERT embeddings.

For the GOS dataset, Graph SAGE with Profile features attains an accuracy of 92.69% and an F1 score of 92.56%. This performance increases significantly with word2vec embeddings, achieving an accuracy of 96.91% and an F1 score of 96.89%. The highest performance is observed with BERT embeddings, where Graph SAGE achieves an accuracy of 97.43% and an F1 score of 97.22%. The GCNFN model on the GOS dataset shows an accuracy of 89.10% and an F1 score of 89.06% with Profile features, 95.07% accuracy and 95.05% F1 score with word2vec embeddings, and 96.18% accuracy and 96.17% F1 score with BERT embeddings. These results indicate that BERT embeddings significantly enhance the performance of both models across datasets, with Graph SAGE consistently showing slightly higher performance than GCNFN. Additionally, the GOS dataset yields higher performance metrics compared to the POL dataset, suggesting it may have more distinct or clearer features beneficial for rumour detection. This analysis underscores the importance of advanced text embeddings and graph-based models in improving rumour detection systems, highlighting that leveraging rich semantic features and sophisticated models can substantially enhance detection accuracy and reliability.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1. CONCLUSION

In this project, we argues that user tweep tendency plays a vital role in the rumour detection problem. To verify this argument, we collect the user historical posts to implicitly model the tweep tendency and leverage the news propagation graph on social media of users. An end-to-end rumour detection framework named TTRD is proposed to fuse tweep tendency and network dynamics and predict the news' credibility on social media. Experimental results demonstrate the advantage of modeling the tweep tendency.

5.2. FUTURE WORK

For future work, we will investigate whether, to some extent, the catastrophic forgetting phenomenon in this case can be mitigated by the choices of features—include more features, or find "universal" features that work well despite the different graph structures.

REFERENCES

[1] Junjie Cen 1 and Yongbo Li. “Rumour Detection Method from Social Network Based on Deep Learning in Big Data Environment” 2022

[2] Shaswat Patela, Prince Bansala, Preeti Kaura. “Rumour detection using graph neural network and oversampling in benchmark X dataset” 2022

[3] Li Tan1, Ge Wang1, Feiyang Jia1, Xiaofeng Lian. “Research status of deep learning methods for rumour detection” 2022

[4] Shouzhi Xu1, Xiaodi Liu1, Kai Ma1, Fangmin Dong1, Basheer Riskhan1, Shunzhi Xiang1, Changsong Bing1. ”Detection on social media using hierarchically aggregated feature via graph neural networks” 2022

[5] Qi Huangy, Junshuai Yuy, Jia Wuz, and Bin Wang. “Graph Attention Networks for Early Detection of Rumours on X” 2021

[6] Yi Han, Shanika Karunasekeran, Christopher Leckie. “Graph Neural Networks with Continual Learning for Rumour Detection from Social Media” 2021

[7] Zhiyuan Wu, Dechang Pi, Junfu Chen, Meng Xie, and Jianjun Caob. “Rumour detection based on propagation graph neural network with attention mechanism” 2020

[8] Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J. Jansen, Kam-Fai Wong, Meeyoung Cha. “Detecting Rumours from Microblogs with Recurrent Neural Networks” 2020

[9] Derry Jatnikaa,¬, Moch Arif Bijaksanaa, Arie Ardiyanti Suryani. “Word2Vec Model Analysis for Semantic Similarities in English Words” 2019

[10] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding” 2019

[11] Rosa Sicilia, Stella LoGiudice, Yulong Pei, Mykola Pechenizkiy, Paolo Soda. “X rumour detection in the health domain” 2019

[12] Vosoughi S. “Rumour Gauge: Predicting the Veracity of Rumours on X” 2019