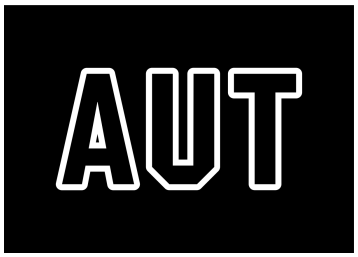


Enhancing Social Network Analysis Through NLP, Knowledge Graphs, and Deep Learning Technologies

Guan Wang

Supervisor: Dr Weihua Li
Prof. Edmund Lai
A/Prof. Quan Bai

School of Engineering, Computer & Mathematical Sciences
Auckland University of Technology

A thesis submitted to Auckland University of Technology
in fulfilment of the requirements for the degree of
Doctor of Philosophy

January 2025

I would like to dedicate this thesis to my loving family and my baby daughter Abigail. Even though Abi is no longer with us, I will never forget how encouraged I was when she was with us. To my supervisors, Dr Weihua Li, Prof. Edmund Lai and A/Prof. Quan Bai, for their wise, patient and invaluable guidance and support without which I would never achieve what I have now.

Copyright

Attestation of Authorship

I hereby declare that except where specific reference is made to the work of others, the contents of this thesis are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This thesis is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgments.

Guan Wang
January 2025

Acknowledgements

It is my great pleasure to express my deepest gratitude to my supervisor, Dr Weihua Li, whose expertise, guidance, and support have been invaluable throughout my PhD journey. Your mentorship has profoundly influenced my academic growth, and this work would not have been possible without your unwavering encouragement. I am also grateful to my secondary supervisor and advisor, Prof. Edmund Lai and A/Prof. Quan Bai, for their valuable suggestions and guidance.

I would like to extend my great appreciation to my colleagues within the School of Engineering, Computer and Mathematical Sciences at the Auckland University of Technology, who have given me tremendous help during my PhD study, William Wong, Verica Rupar, Jinglong Duan, Isabel Wang, Jingli Shi, Xiaodan Wang, Rebecca Frederick, Boshra Talebi Haghighi, Lindsay Zhang, Bumjun Kim, Saide Lo and Terry Brydon.

I am deeply grateful to my family for their unconditional love and support. To my parents, thank you for your sacrifices and for instilling in me the value of education. To my wife, your patience, understanding, and constant encouragement have been crucial to the completion of this thesis. To Abigail, your encouragement was the most powerful strength that pushed me forward. This work is dedicated to you.

Abstract

The rapid evolution of social media and recommendation systems has transformed social interactions and information dissemination, presenting both opportunities and challenges for social analysis. This thesis provides comprehensive methods for analysing social media content, detecting echo chambers, mitigating misinformation, and maximizing influence in social networks.

First, the thesis addresses a fundamental task of social network analysis. Text summarization with knowledge graphs is crucial for condensing and contextualising vast unstructured data, enhancing trend detection and decision-making. Two novel text summarization models, KATSum and AaKOS, are proposed, demonstrating the effectiveness of integrating knowledge graphs. KATSum is a Knowledge-aware Abstractive Text Summarization model that enhances the standard Seq2Seq model, while AaKOS is an Aspect-adaptive Knowledge-based Opinion Summarization model that generates personalized, aspect-oriented summaries from product reviews.

Second, the thesis investigates user behaviour modelling and information diffusion in online social networks. A novel user behaviour model, utilising deep neural networks and knowledge graphs, presents users' personalised actions after receiving influence messages. The proposed model serves as the foundation for the rest of the chapters. Based on this, a novel influence maximisation algorithm is then proposed based on the user behaviour model, which aims to maximise the influence in online social networks by considering the information alteration during the diffusion process.

Third, the focus shifts to investigating solutions for tackling misinformation detection, especially given how information alteration through reframing can lead to the development of misinformation. By adopting frame theories from communication studies, we aim to identify misinformation portrayed from factual information but presented misleadingly. To address this challenge, we propose a deep-learning-based model called FrameTruth, which leverages large language models to extract the framing of information, incorporating this as a crucial feature in misinformation classification. This approach delves into the nuanced manipulation of narrative frames, an under-explored area within the AI community. By leveraging the power of pre-trained large language models and deep neural networks, FrameTruth detects

misinformation originating from accurate facts portrayed under different frames, utilising the various impacts of framing theory elements.

Finally, the thesis tackles the challenge of identification of echo chambers in online social networks by applying deep learning methodologies to model user beliefs based on historical message content and behaviours. A novel content-based framework is proposed, capable of detecting potential echo chambers by creating user belief graphs. This framework also demonstrates the evolution of user belief over time using the user behaviour model.

Publications

Wang, G., Frederic, R., Duan, J., Wong, W., Rupar, V., Li, W. & Bai, Q. (2024). Detecting misinformation through Framing Theory: the Frame Element-based Model. IEEE Transactions on Human-Machine Systems (submitted)

Wang, G., Li, W., Lai, E. & Jiang, J. (2022). KATSum: Knowledge-aware abstractive text summarization. In Proceedings of Principle and Practice of Data and Knowledge Acquisition Workshop, https://doi.org/10.48550/arXiv.2212.03371

Wang, G., Li, W., Bai, Q., & Lai, E. M. (2023). Maximizing Social Influence With Minimum Information Alteration. IEEE Transactions on Emerging Topics in Computing, 12(2):419–431. https://doi.org/10.1109/tetc.2023.3292384

Wang, G., Li, W., Wu, S., Bai, Q., & Lai, E. M. K. (2023, November). BeECD: Belief-Aware Echo Chamber Detection over Twitter Stream. In Pacific Rim International Conference on Artificial Intelligence (pp. 307-319). Singapore: Springer Nature Singapore, https://doi.org/10.1007/978-981-99-7025-4_27

Wang, G., Frederick, R, Haghighi, B. T., Wong, W., Rupar, V., Li, W., Bai, Q. (2024). FramedTruth: A Novel Model Utilising Large Language Models for Frame-based Misinformation Detection. In the 16th Asian Conference on Intelligent Information and Database Systems, accepted in March 2024

Wang, G., Li, W., Lai, E. & Bai, Q. (2024). Aspect-adaptive Knowledge-based Opinion Summarization. In Knowledge Management and Acquisition for Intelligent Systems, pages 29-41, Singapore. Springer Nature Singapore, https://doi.org/10.1007/978-981-96-0026-7_3

Table of contents

List of figures

List of tables

Nomenclature

Acronyms / Abbreviations

AaKOS Aspect-adaptive Knowledge-based Opinion Summarization

ABID Alteration-Based Influence Degree

ABSS Alteration-Based Seed Selection

AI Artificial Intelligence

ALBERT A Lite BERT

AsIC Asynchronous Independent Cascades

AsLT Asynchronous Linear Threshold

BeECD Belief-Aware Echo Chamber Detection

BERT Bidirectional Encoder Representations from Transformers

Bi-LSTM Bidirectional Long Short-Term Memory

CELF Cost Effective Lazy Forward

D-C Diffusion-Containment

DL Deep Learning

DRUC Decaying Reinforced User-Centric

EMLo Embeddings from Language Model

FEM Framed Element-based Model

FN False Negatives

FP False Positives

FTM FrameTruth Model

GAI Generative Artificial Intelligence

GATs Graph Attention Networks

GloVe Global Vectors for Word Representation

GNNs Graph Neural Networks

GPT Generative Pre-trained Transformer

HITS Hyperlink-Induced Topic Search

IC Independent Cascades

IM Influence Maximization

IM-MIA Influence Maximization with the Minimum Information Alteration

IMM Influence Maximization via Martingales

KATSum Knowledge-aware Abstractive Text Summarization

KGs Knowledge Graphs

KIAID Knowledge-aware Information Alteration Influence Diffusion

LLM Large Language Models

LSTM Long Short-Term Memory

LT Linear Threshold

ML Machine Learning

NDM Neural Diffusion Model

NLP Natural Language Processing

OSNs Online Social Networks

LLM Resource Description Framework

RoBERTa Robustly Optimized BERT Pretraining Approach

ROUGE Recall-Oriented Understudy for Gisting Evaluation

SEIR Susceptible-Exposed-Infected-Recovered

SEIRS Susceptible-Exposed-Infected-Recovered-Susceptible

SIR Susceptible-Infected-Recovered

SIRS Susceptible-Infected-Recovered-Susceptible

SI Susceptible-Infected

SIS Susceptible-Infected-Susceptible

SNA Social Network Analysis

TAP Topical Affinity Propagation

TFG Topical Factor Graph

TIC Topic-aware Independent Cascades

TLT Topic-aware Linear Threshold

TN True Negatives

TP True Positives

Chapter 1

Introdcution

In the digital age, online social networks (OSNs), such as X[1] (Twitter as the former name), Facebook[2], Weibo[3], Youtube[4], etc., have become the backbone of modern communication, enabling the rapid dissemination of information [180]. Along with the outbreak of the COVID-19 pandemic, the war between Russia and Ukraine, Japan's discharging nuclear wastewater and other hot topics, the Social Network Analysis (SNA) has entered a new era and attracted great attention of researchers from multiple disciplines. Understanding the dynamics of these networks requires a multi-faceted approach that encompasses global network analysis, individual user behaviour, and specific user interactions.

Based on graph theory, social network analysis examines the structure of a social network, which is structured from relational data with nodes such as individuals, groups or organizations and edges such as some pattern of relationships or interactions between them [124]. Social network analysis is also known as structural analysis since it leverages graph theory, statistics, and computational techniques to understand the dynamics and impacts of social interactions [175]. Social structures typically manifest in two distinct forms, i.e., global and individual perspectives.

Globally, social network analysis provides insights into the overarching structure and connectivity of the network. Techniques such as community detection and centrality measures identify influential nodes and community clusters, which are critical for understanding how information propagates [17, 121]. On the other hand, individually, examining user behaviour reveals patterns of interaction, content consumption, and sharing tendencies. Behavioural analysis can uncover how individuals influence and are influenced by others, which is

[1] https://x.com/
[2] https://www.facebook.com
[3] https://weibo.com
[4] https://www.youtube.com

essential for targeted interventions. For example, researchers explore the significance of interacting with messages, e.g., retweet behaviours, in the information diffusion process and how information cascades occur through the actions of influential users [86, 15].

Nowadays, with the development of Artificial Intelligence (AI), the way social networks are analysed has also changed. More advanced techniques are applied to analyse social networks, such as Machine Learning (ML), Deep Learning (DL), Knowledge Graphs (KGs) and Natural Language Processing (NLP). The integration of these advanced AI techniques into SNA not only provides deeper insights into user behaviours and network structures but also enables the ability to address complex challenges, such as Influence maximization, Misinformation detection, and Echo Chamber effect [4, 16, 39].

In this chapter, we present the research motivation, briefly outline the rationale and significance of the study, and present the research methodology, including research questions, evaluation methods, and advanced techniques applied to social network analysis in this thesis. We also summarize this thesis's contribution.

1.1 Influence Diffusion

Many researchers and practitioners have dedicated significant efforts to modelling and understanding the patterns of influence diffusion in online social networks. Predicting information items or cascades has garnered substantial attention from both academic and business perspectives [80, 101, 190]. Two seminal influence-diffusion models, the Independent Cascade (IC) model and the Linear Threshold (LT) model, are central to much of the existing research, treating influence as a process akin to hopping and infecting [80, 103].

Building on these foundational models, numerous researchers have developed tailored or varied approaches to address new challenges. For instance, Saito et al. introduced the Asynchronous Independent Cascades (AsIC) and Asynchronous Linear Threshold (AsLT) models, which enhance the IC and LT models by incorporating time delays in user interactions [140]. Lagnier et al. proposed the Decaying Reinforced User-Centric (DRUC) model, a variant of the LT model that incorporates user profiles into the influence diffusion process [88]. Barbieri et al. extended both the IC and LT models to create the Topic-aware Independent Cascade (TIC) and Topic-aware Linear Threshold (TLT) models, which account for message content and simulate topic distribution [19]. Liu et al. enhanced the LT model with the Diffusion-Containment (D-C) model, which includes mechanisms for containing competitive influence spread [107].

Distinct from IC and LT-based models, Li et al. proposed agent-based influence diffusion models, where the cascading process is modelled as an evolutionary pattern driven by individual actions [101, 98].

In this section, several influence diffusion issues, i.e., influence maximization, misinformation detection and echo chamber detection, are introduced.

1.1.1 Influence Maximization

Influence maximization is a central problem in the study of influence diffusion. The goal of influence maximization is to spread influence as widely as possible throughout a network [80]. To achieve this, seed selection algorithms identify a set of influential users to initiate the dissemination process. Most existing works are based on two seminal influence diffusion models: the Independent Cascade (IC) model and the Linear Threshold (LT) model. In these models, influence begins with affected users and spreads through the network's topological structure [36, 35]. However, there are two significant limitations in almost all research in this field.

First, influence diffusion is typically treated as a probabilistic-based hopping and infecting process that neglects users' prior knowledge and the detailed content of the influence. Most studies focus on the outcome of influence diffusion, such as estimating the number of users influenced, and traditional influence maximization efforts aim solely at maximizing the spread of influence. They overlook the connections between users' knowledge and the content of the influence. In reality, users perceive and interpret influence messages differently, often associating them with their knowledge and background [42]. For instance, given the news headline "Nearly 165,000 migrants eligible for fast-tracked residency", different groups will focus on different aspects: eligible migrants on the application process, investors on property market impacts, and local business owners on future recruitment plans. These varied interpretations highlight the importance of modelling users' reactions to influence, considering how they understand and disseminate messages.

Second, in many applications, it is crucial to maintain the consistency of messages throughout the diffusion process. However, few studies consider information alteration. Using the same news example, some people may share it with the belief that it will "boost the economy", while others might claim it will cause a "housing crisis". These differing interpretations can lead to complex influence spread chains, producing many variants of the original message and deviating from the initial influence maximization goal. Therefore, it is essential to consider information alteration and take measures to retain the originality of influence messages during the diffusion process.

To address these limitations, our solution focuses on the influence maximization problem, emphasising suppressing information alteration during the diffusion process. We consider users' personalized prior knowledge and potential information alterations. Specifically, we adopt Knowledge Graph (KG) technology to model users' prior knowledge and subconscious biases. Deep learning models are used to simulate users' behaviours, generating responses and comments that imply potential information alterations. Furthermore, we propose a novel seed selection algorithm aimed at maximizing social influence while minimizing information alterations. Extensive experiments demonstrate that our proposed influence-diffusion model effectively captures information alterations and that our novel seeding algorithm outperforms others in maximizing influence with minimal information alteration.

As information alteration is one of the reasons for the generation of misinformation, this approach potentially has significant implications for detecting misinformation, particularly in how content changes and how it is portrayed. By understanding and modelling these alterations, we can better identify and mitigate misinformation as it spreads through social networks.

1.1.2 Misinformation Detection

Misinformation is growing substantially in today's media landscape, with fake news and misleading information spreading through news websites and social media platforms. Generative AI (GAI) models, like ChatGPT, have expedited the process of creating sophisticated articles and posts, making it difficult for readers to distinguish if the content is generated by AI or humans [73, 111, 85].

Automatically identifying false claims is well-researched, especially using keywords for traditional misinformation detection [134, 155]. However, identifying misinformation becomes challenging when accurate facts are manipulated through biased frames to create misleading narratives. Framing involves highlighting certain aspects of information while omitting others, thus distorting the original message. Investigating strategic framing and its role in online information dissemination is crucial [51, 142].

Framing theory, significant in communication fields, explains how the presentation of information can influence perception. It suggests that the way information is framed can manipulate individuals into accepting false information. By emphasizing certain factors and omitting others, communicators can craft narratives that align with their agenda and mislead the audience [144, 55, 52, 51].

Despite extensive research on framing detection and analysis, few studies explore how frames impact misinformation emergence or which frame element has the greatest impact. Classifying misinformation from factual information through framing manipulation remains

challenging. For example, a factual narrative about water reform can be misleadingly reframed using satire or selective emphasis, altering the original message's perception.

Pre-trained Large Language Models (LLM) and deep neural networks are effective in addressing framing classification and misinformation detection due to their ability to learn from unstructured data and identify complex patterns [73]. Our hypothesis is that news on the same topic can be turned into misinformation with different frames.

Misinformation, as one of the serious issues on social networks, can contribute to the formation of echo chambers by reinforcing pre-existing beliefs and biases within a community, leading to selective exposure and confirmation bias. When misinformation spreads within these echo chambers, it is more likely to be accepted and shared without critical scrutiny, as it aligns with the group's established beliefs. This leads to a homogenized information environment where falsehoods can flourish and become more deeply entrenched [45].

1.1.3 Echo Chamber detection

Online social platforms have become key sources for people to perceive information, reshaping how information is searched, filtered, and disseminated [118]. Modelling and analysing the influence and dissemination of information on social networks, including information maximization, misinformation detection, and echo chamber phenomena, have become prominent subjects [96, 147].

Echo chambers are environments where individuals are predominantly exposed to information and opinions that align with their own while opposing views are excluded or minimized. This selective exposure is exacerbated by the algorithmic filtering of social media platforms, which personalize content to match users' preferences, thus reinforcing their existing viewpoints and creating a feedback loop of like-minded information [39].

The prominence of the echo chamber phenomenon has been heightened by the outbreak of Covid-19, a global pandemic and leading topic of discussion. Conversations around this subject vary widely, encompassing themes such as vaccine hesitancy and vaccination-related deaths. Under these conditions, social platforms provide a conducive environment for misinformation propagation due to the lack of editorial supervision. As a result, echo chambers have emerged among users, significantly influencing responses to the Covid-19 pandemic [5]. For example, if members of a community consistently engage with and promote content sceptical of Covid-19 vaccinations, the community can be identified as an echo chamber resistant to prevailing medical advice on vaccines.

1.2 Research Motivation

Online social networks, such as Facebook, X, WeChat, and Weibo, have become indispensable in everyday life. Social influence is everywhere, not only in our daily physical life but also in the virtual Web space. With the global penetration of online and mobile social platforms, people have witnessed the impact of social influence in every field, such as presidential elections, advertising, and innovation adoption [26, 14, 80, 152]. To date, there is no doubt that social influence has become a prevalent yet complex force that drives our social decisions, making a clear need for methodologies to characterize, understand, and quantify the underlying mechanisms and dynamics of social influence.

Social network analysis has been studied extensively, and it has been applied to many research fields, such as influence maximization, misinformation detection, and echo chamber phenomenon [4, 16, 39]. This can bring benefits to various domains, especially those that rely on analysing their customers' behaviours to offer proper services or products.

Influence Maximization

Influence maximization is to identify a limited set of users from online social networks, expecting that they can spread influences and maximize the impact across the entire network [80]. Traditional models like the Independent Cascade (IC) and Linear Threshold (LT) focus on probabilistic influence spread through network structures but have two major limitations [36, 35].

Firstly, these models treat influence diffusion as a simple probabilistic process, neglecting users' prior knowledge and the specific content of the influence. They mainly focus on estimating the number of influenced users without considering how users' knowledge and background affect their interpretation of the influence message [42].

Secondly, they often overlook the importance of maintaining consistent messages throughout the diffusion process. Few studies address the issue of information alteration, which is crucial for retaining the originality of influence messages.

Misinformation detection

In the age of digital communication, the proliferation of misinformation has become a critical issue, posing significant threats to societal well-being, public health, and democratic processes. Identifying misinformation is particularly challenging under the circumstances of information alteration, especially when accurate facts are strategically manipulated through biased frames to construct misleading narratives. These narratives, often carefully crafted, can distort the truth and influence public perception in subtle ways. Moreover, the potential

for Generative AI models to generate misleading narratives underscores the urgency of this problem.

The manipulation of accurate facts involves presenting them in a context that alters their intended meaning, leveraging cognitive biases and emotional triggers to sway opinions. This tactic makes misinformation harder to detect and more convincing to the audience. Traditional fact-checking methods are often insufficient in these scenarios, as they typically focus on verifying the authenticity of individual facts rather than analysing the framing and context in which these facts are presented.

Given the complexity of this issue, advanced methods are urgently needed to detect and mitigate misinformation originating from accurate facts portrayed in different frames.

Echo Chamber Detection

Furthermore, the rise of social media platforms has amplified the spread and impact of such altered information, such as framed misinformation. These platforms facilitate rapid dissemination and reinforcement of biased narratives within echo chambers, where individuals are exposed to information that aligns with their pre-existing beliefs, further entrenching their views and reducing their willingness to consider alternative perspectives.

Most works only consider the content of information, but individual behaviours and content weights demonstrating the meaningful impact of the content on individuals are neglected. For example, reviewing a message does not explicitly show what an individual thinks about this message. However, a thumbs up after reviewing reveals that the individual likes and agrees with this message. Subsequently, such behaviour increases the weight of the information from this message in the corresponding belief graph. Also, most other works mainly focus on identifying echo chambers globally, requiring high-performance computation. Individual-level detection allows personalized interventions that help users break free from echo chambers and require fewer computation resources. By tailoring strategies to each person's unique behaviour and preferences, we can promote exposure to diverse perspectives and encourage critical thinking. This personalized approach is more likely to be effective than broad solutions.

Text Summarization

The exponential growth of user-generated content on social media and online communication platforms, including short/long posts, product/service reviews and comments to lengthy articles and discussions. This huge volume of information demonstrates both opportunities and challenges for social network analysis, where the goal is to understand the structure,

dynamics, and impact of social interactions and information dissemination. One of the key challenges in this field is efficiently extracting meaningful insights from the data. This is where text summarization becomes critically important.

Text summarization can significantly enhance social network analysis by providing concise and relevant user interactions and content summaries. This capability is particularly valuable for several reasons:

- Information Overload: With the large volume of information on social networks, users and researchers face information overload. Summarization techniques can help manage this by providing brief yet comprehensive overviews of large datasets, enabling quicker understanding and decision-making.
- Content Filtering: Social networks often contain a mix of high-quality content and noise. Summarization can help filter out less relevant information, highlighting the most significant and impactful content. This is particularly useful for identifying influential posts, key discussions, and important themes within a network.
- User Behaviour Analysis: Understanding user behaviour and interaction patterns is crucial to social network analysis. Summarized content can reveal the main topics of discussion, sentiment trends, and user engagement levels, providing valuable insights and more concise information fostering the analysis process.
- Resource Efficiency: analysing large volumes of text data can be resource-intensive. Summarization reduces the amount of text that needs to be processed and analysed, making social network analysis more efficient and scalable.

The aforementioned factors motivate the application of text summarization in the context of social network analysis. By leveraging advanced text summarization techniques, researchers can enhance their ability to extract meaningful insights from vast and complex datasets. This, in turn, contributes to a deeper understanding of social dynamics, user behaviour, and information dissemination on digital platforms.

1.3 Research Questions

Research Question 1: How can knowledge graphs and pre-trained language models be integrated to improve the accuracy and adaptability of text summarization models for analysing and summarizing user opinions and behaviours in online social networks?

Research Question 2: What is the most suitable approach for modelling social network user behaviours for mining social influence diffusion patterns?

- **Sub Research Question 2.1:** How can the user behaviour model contribute to the maximization of social influence?
- **Sub Research Question 2.2:** How to maximize influence with minimum information alteration through diffusion?

Research Question 3: With the spread of information with alterations, what is the best approach to detect misinformation that is portrayed from the facts?

- **Sub Research Question 3.1:** Can framing theory provide extra information for detecting the misinformation derived from the facts?
- **Sub Research Question 3.2:** What framing theory elements are providing insightful information?

Research Question 4: How to effectively detect and evaluate the degree of echo chamber effect with knowledge graph and deep neural networks?

1.4 Research Methodology

In this thesis, the research methodology comprises six steps: literature review, research question definition, research objectives, research method design, data collection and analysis, and evaluation of the proposed method, as shown in Figure 1.1.

Fig. 1.1 Steps of applied research methodology.

The research begins with a literature review to identify and define relevant research questions by analysing existing works, drawing inspiration, and pinpointing limitations and research gaps. Based on insights from the literature review, we formulate our research questions and set specific objectives outlining the expected outcomes. Leveraging previous findings, we design our research method to achieve these objectives, including outlining the experimental approach, selecting methodologies, collecting datasets, and planning execution.

Public datasets available online are utilised to test our proposed methods, which are introduced in the following section. We then conduct experiments by implementing the research design, running simulations or analyses, and collecting experimental data. The results are rigorously evaluated and analysed; if outcomes are unsatisfactory, we iterate on our methods, adjusting and refining the approach as necessary.

1.4.1 Graph Representation

Graph representation is a method of presenting data structures where entities are depicted as nodes and their relationships as edges within a graph. This approach enables the visualization and analysis of complex, interconnected data in a clear and structured manner.

In this thesis, we utilise knowledge graphs to represent factual information. A knowledge graph is a specific type of graph that depicts entities and the relationships between them. These graphs are crucial for modelling user behaviours and calculating belief graph similarities. To leverage the full potential of knowledge graphs, we convert them into representation vectors.

Knowledge Graphs

Human knowledge provides a formal understanding of the world. Knowledge graphs representing structural relations between entities have become an increasingly popular research direction towards cognition and human-level intelligence. In recent years, knowledge graphs as a form of structured human knowledge have drawn great research attention from both academia and the industry [48, 122, 172]. A knowledge graph is a structured representation of facts consisting of entities, relationships, and semantic descriptions. Entities can be real-world objects and abstract concepts, relationships represent the relation between entities, and semantic descriptions of entities and their relationships contain types and properties with well-defined meanings. A knowledge graph can be viewed as a graph when considering its graph structure [151].

As shown in Figure 1.2, knowledge can be expressed in a factual triple in the form of (head, relation, tail) or (subject, predicate, object) under the resource description framework (RDF), for example, (Albert Einstein, WinnerOf, Nobel Prize). It can also be represented as a directed graph with nodes as entities and edges as relations.

In this thesis, knowledge graphs are applied with deep learning and NLP technologies to model user behaviours in social networks. In other words, we use the model to predict user behaviour given received messages to see how a user reacts to them. We also construct

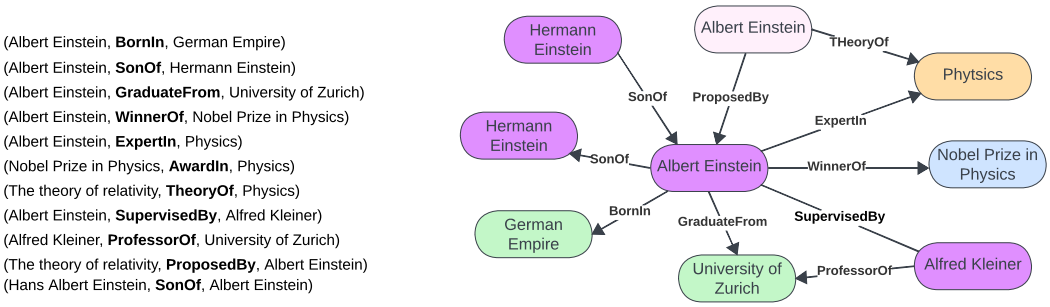


Fig. 1.2 An example of Knowledge Graph.

knowledge graphs from plain text datasets for each user in a social network. These knowledge Graphs will be converted into representation vectors for use in deep learning models.

Graph Attention Networks

In recent years, data in the form of graphs, e.g. social networks, biological networks, recommendation systems, and transportation networks, have attracted much attention, leading to an increasing interest in generating graph representations. Graph attention networks are powerful neural networks based on the classic graph neural networks [141] that operate on graph-structured data, leveraging masked self-attentional layers to compute the hidden representations of each node in the graph by attending to its neighbours using a self-attention mechanism. This attention-based architecture offers several notable advantages: (1) it is efficient because the operation can be parallelized across node-neighbour pairs; (2) it accommodates graph nodes with varying degrees by assigning arbitrary weights to the neighbours; and (3) it is well-suited for inductive learning tasks, enabling the model to generalize to entirely unseen graphs.

In Figure 1.3 (left), a single feed-forward layer α_{ij} represents the attention mechanism as the softmax of attention coefficients e_{ij} which indicates the importance of node $j's$ feature to node i. Similar to the attention mechanism [160], the model allows each node in the graph to attend to every other node. When the graph structure is injected into the attention mechanism, allowing only to compute the attention for node $j \in N_i$ where N_i is some neighbourhood of node i, the attention mechanism is calculated using the softmax function as:

$$\alpha_{ij} = softmax_j(e_{ij}) = \frac{exp(LeakyReLU(e_{ij}))}{\sum_{k \in N_i} exp(LeakyReLU(e_{ik}))}, \tag{1.1}$$

where $LeakyReLU()$ is the activation function and the attention coefficient e_{ij} is as:

$$e_{ij} = a(W\overrightarrow{h}_i, W\overrightarrow{h}_j), \tag{1.2}$$

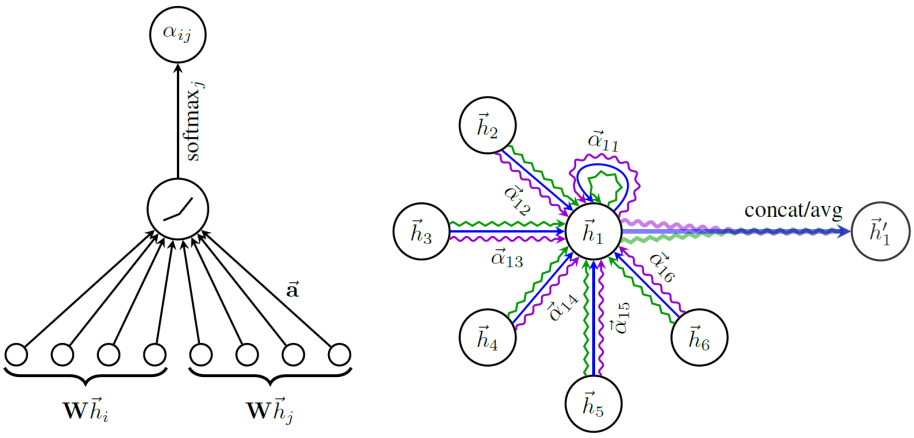


Fig. 1.3 Left: attention mechanism employed in GATs. Right: An illustration of multi-head attention [161].

where $\overrightarrow{h}_i \in \mathbb{R}$ represents the node feature. Finally, the attention mechanism is expressed as:

$$\alpha_{ij} = \frac{exp(LeakyReLU(\overrightarrow{a}^T[W\overrightarrow{h}_i||W\overrightarrow{h}_j]))}{\sum_{k\in N_i} exp(LeakyReLU(\overrightarrow{a}^T[W\overrightarrow{h}_i||W\overrightarrow{h}_k]))} \tag{1.3}$$

Figure 1.3 (right) demonstrates the multi-head attention mechanism by node 1 in its neighbourhood. Multi-head attention is used to stabilize the learning process of self-attention, and the aggregated features from each head are finally concatenated or averaged to update the feature of node 1, as depicted in Figure 1.3 (right). Before this, a nonlinear function, σ, is applied as:

$$\overrightarrow{h}'_i = \sigma(\sum_{j\in N_i} \alpha_{ij} W\overrightarrow{h}_j) \tag{1.4}$$

This thesis uses Graph Attention Networks (GATs) to address various problems. GATs are used to encode graphs extracted from texts. They are extended to encode a user's prior knowledge and subconsciousness in social networks by adding two transformer layers for generating initial embeddings. We convert these topological structure graphs into vector representations by training GATs on each individual's belief graph. Unlike the original GATs, we introduce weighted relation features $R = \{r_{i,j}|0 < i < n, 0 < j < n\}$ as the initial attention coefficients. These weighted relation features are used during the attention calculation as follows:

$$e_{i,j} = a(W\hat{h}_i, W\hat{h}_j, r_{i,j}), \tag{1.5}$$

where $r_{i,j}$ is the weighted relation from node i to node j.

We add a global node to each graph to collect all the graph's features. This global node is linked to every node in the graph, and its representation represents the entire graph every time we feed it to the model.

1.4.2 Language Models

A language model is a statistical model designed to predict the probability of a sequence of words. It can generate, understand, and manipulate natural language text by learning patterns, structures, and relationships within the language data it is trained on. Since the introduction of the transformer architecture [160], language models have advanced to a significantly more sophisticated stage. The advent of the transformer architecture has revolutionized the development of language models, marking a significant leap from traditional recurrent and convolutional neural networks. The proposed self-attention mechanism enables more efficient parallelization and handling of long-range dependencies in text, leading to remarkable improvements in performance and scalability. Building on this foundation, many pre-trained language models like BERT (Bidirectional Encoder Representations from Transformers) and other BERT-like language models have emerged [47, 184, 109, 89].

BERT (Bidirectional Encoder Representations from Transformers)

BERT opens the era of leveraging large-scale unsupervised learning on diverse text corpora to capture rich contextual representations by understanding both the left and right context of words simultaneously. This bidirectional approach differentiates BERT from earlier models like LSTM, ELMo, GloVe, etc., enabling it to achieve state-of-the-art results on various natural language processing tasks [68, 129, 128].

This thesis uses BERT as a text encoder to process news texts for text summarization and product reviews for opinion summarization. Additionally, BERT encodes user posts as part of our proposed user behaviour model. It is also used to generate embeddings, where the embeddings of the special token "[CLS]" are leveraged as features to identify misinformation

BERT incorporates two key concepts that have driven recent advances in the NLP field: (1) the transformer architecture and (2) unsupervised pre-training. The transformer is a network architecture that relies entirely on attention mechanisms, eliminating the need for traditional recurrent and convolutional components in sequence transduction models [160]. BERT's model training involves two main steps: pre-training and fine-tuning. During the pre-training process, BERT is trained on unlabelled data for various tasks, allowing it to learn deep bidirectional representations. In the fine-tuning phase, the BERT model is initialized

by the pre-trained parameters, and these parameters are then fine-tuned with annotated data specific to downstream tasks.

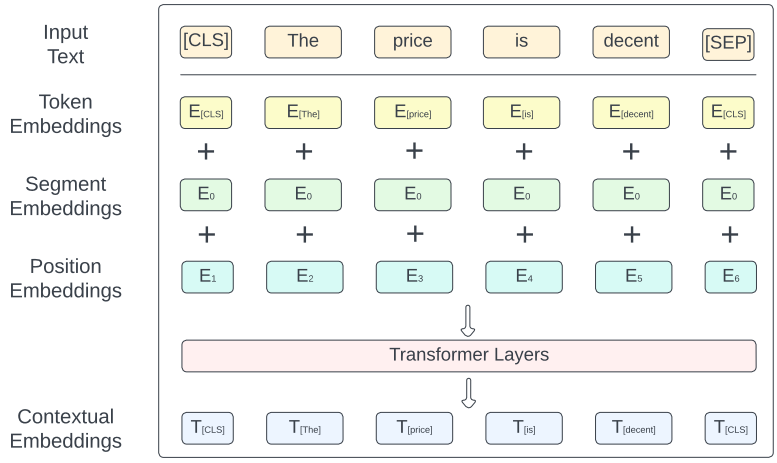


Fig. 1.4 The overview of BERT generating contextual embeddings

Figure 1.4 illustrates how BERT generates contextual embeddings. BERT's input representation can accommodate both individual sentences and pairs of sentences, allowing it to tackle a variety of downstream tasks. Each input sequence begins with a special classification token [CLS] and ends with a [SEP] token to separate distinct sentences. The final input representation is the output of a stack of transformer layers processing the input by summing up three types of embeddings: position embeddings, segment embeddings, and token embeddings.

1.4.3 Evaluation Methods

In this thesis, we apply different evaluation methods to different problems.

ROUGE

In Chapter 3, we apply ROUGE to evaluate the performance of summarization models. ROUGE-1 and ROUGE-2 refer to the overlap of uni-gram and bi-gram between the source text and the generated summary, respectively. ROUGE-L describes the longest common subsequence. ROUGE is one of the most widely used evaluation metrics for text summarization. It focuses on token-level matching but ignores the semantic-level matching between the summary and the source text, which results in the inability to identify factual errors. Formally, ROUGE-N is an n-gram recall between a summary and a golden summary. ROUGE-N is computed as follows[104]:

$$ROUGE-N = \frac{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in \{ReferenceSummaries\}} \sum_{gram_n \in S} Count(gram_n)} \tag{1.6}$$

Alteration-Based Influence Degree (ABID)

In Chapter 4, we propose a novel method to evaluate the influence coverage with a major consideration of information alteration called Alteration-Based Influence Degree (ABID) and calculated as follows:

$$ABID = |f(SC)| \cdot (1 - \frac{\alpha}{|F(SC)|} \sum_{v_i \in F(SC)} Atr(v_i)), \tag{1.7}$$

where $f(SC)$ refers to the influence coverage, i.e., the cardinality of the users who received or spread the influence message, when seed set SC is selected, $|SC| = k$. In contrast, $F(SC)$ is a set of users who spread the influence message, i.e., the users who are activated, $\forall v_i \in F(SC), v_i \in V \wedge s(v_i) = 1$. $\alpha \in [0,1]$ denotes a parameter, balancing the weight of the information alteration degree. If $\alpha = 0$, the alteration is not considered, while only influence coverage contributes to the objective function. $Atr(v_i)$ is a function, which estimates the alteration degree between msg_x and msg'_x, representing originally diffused message and the message delivered by v_i, respectively.

Confusion Matrix

In Chapter 5, as the misinformation detection problem is transformed into a binary classification task, a confusion matrix is used to calculate the Precision, Recall and F1 score by calculating the following terms:

- True Positives (TP): when predicted misinformation is actually labelled as misinformation;
- True Negatives (TN): when predicted information is actually labelled as information;
- False Positives (FP): when predicted information is actually labelled as misinformation;
- False Negatives (FN): when predicted misinformation is actually labelled as information.

Based on the Confusion Matrix, Precision, Recall, F1-score and Accuracy are calculated to assess the model by comparing it with existing baseline models.

- **Accuracy** evaluates the model's performance across all categories.

$$Accuracy = \frac{|TP| + |TN|}{|TP| + |TN| + |FN| + |FP|} \tag{1.8}$$

- **Precision** evaluates the correctness of the positive instances (the correctness of misinformation predicted) that our model has predicted.

$$Precision = \frac{|TP||}{|TP| + |FP|} \tag{1.9}$$

- **Recall** indicates how many of the actual positive instances our model can correctly recognize.

$$Recall = \frac{|TP|}{|TP| + |FN|} \tag{1.10}$$

- **F1-score** is the harmonic mean of precision, and it provides a balanced measure that considers both precision and recall.

$$F1\_score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{1.11}$$

1.5 Contribution

In this section, the key research contributions are summarised below:

- A novel user behaviour model is proposed to predict user interactions and behaviours when receiving a message from other sources. It has the potential to This model consists of two encoders, a Knowledge Graph encoder and a text encoder, and a decoder where the Knowledge Graph embeddings and text embeddings are fused and utilised to generate text. This user behaviour model can help with creating more personalized user experiences by accurately predicting user interactions and behaviours. This can be particularly beneficial for applications like recommendation systems, and targeted advertising. Moreover, this model can be applied to help with echo chamber detection and mitigation by analysing user behaviour patterns and tendencies.

- Two novel summarization models, KATSum and AaKOS, are designed to verify the effectiveness of integrating Knowledge Graphs in text generation used in the user behaviour model. In KATSum, a triplet classifier is also trained to filter key information. Integrating Knowledge Graph and triplet classifier can address the unfaithfulness and

factual error issues in classical text summarization. Relevent results from KATSum are published in [167]. AaKOS utilised the self-supervised manner to train the model to address the issue of a lack of adaptive nature in opinion summarization.

- We first address the influence maximization problem considering the information alteration in the influence diffusion process retaining the influence originality. We also model the complex influence diffusion process by considering users' personalised behaviours and the possibility of information alteration, where the user behaviour model is utilised to produce user reactions. A novel seeding algorithm is also proposed, enabling a trade-off between maximizing the influence diffusion and retaining the originality of spreading influences.

- We propose two deep learning-based models, the FrameTruth Model (FTM) and the Frame Element-based Model (FEM), to tackle the challenge of identifying misinformation portrayed from factual information but presented in a misleading manner, integrating the framing theory. We are also the first to explore the impact of elements of framing theory on identifying misinformation.

- To identify if a user is isolated in an echo chamber, a novel content-based framework is built to model the user's beliefs based on their historical message contents and behaviour. These messages and behaviours are transformed into weighted belief graphs. Each user's belief graph is compared with its neighbour to estimate the echo chamber degree of the ego network. We also utilise the user behaviour model to demonstrate the evolution of the echo chamber showing the alignment between the echo chamber degree and user behaviours.

1.6 Thesis Structure

The following part of this thesis is managed as below:

- **Chapter 2** reviews relevant literature on influence diffusion, influence maximization, misinformation, echo chamber detection and mitigation and text summarization.

- **Chapter 3** introduces two text summarization models based on the enhanced seq2seq model by incorporating NLP and Knowledge Graph technology. This chapter aims to answer the sub-research questions 1.2 and 1.3.

- **Chapter 4** demonstrates a solution for the influence maximization problem by proposing the knowledge-aware model for influence diffusion with information alteration and ABID algorithm for seed selection. This chapter answers the research question 1.
- **Chapter 5** showcases the effectiveness of framing theory in detecting misinformation. Both framing categories and framing elements are explored to tackle research question 2.
- **Chapter 6** presents a method of detecting echo chambers from an individual perspective applying Knowledge Graph and Deep Learning technologies. The results have demonstrated the effectiveness of this method. This chapter addresses the research question 3.
- **Chapter 7** summarises the thesis by concluding the proposed methods and bringing up future works.

Chapter 2

Literature Review

In this chapter, a comprehensive literature review on social network analysis is presented, covering five key aspects: influence diffusion, influence maximization, misinformation detection, echo chamber detection and mitigation, and text summarization.

2.1 Influence Diffusion

Influence diffusion in social networks is an important study area, with applications ranging from influence maximization to misinformation control. Understanding how influence propagates through social networks and developing methods to maximize or limit this spread is crucial for commercial and societal purposes. This section aims to demonstrate key contributions in the field and identifies common limitations.

Influence diffusion models are reviewed and followed by the influence diffusion application, which includes influence maximization, misinformation detection, and echo chamber detection.

2.1.1 Influence Diffusion Models

In the past decade, many researchers and practitioners have dedicated significant efforts to modelling and learning the patterns of influence diffusion in online social networks, revealing that predicting information items or cascades has attracted great attention from both academic and business perspectives [80, 101, 190]. In this section, basic diffusion models and their variations are reviewed and classified into threshold models, epidemic models and user behaviour models.

Threshold Models

Each node in Threshold Models has a specific threshold above which it will be activated and attempt to interact and activate its neighbours if its neighbours include some inactive nodes. Two seminal threshold models, the Independent Cascade (IC) model and the Linear Threshold (LT) model, are utilised by most existing research, where influence is treated as a hopping and infecting process [80, 103].

Given a network $G = (V, E)$ where V represents the nodes in the network or users in a social network, E represents the edges between nodes or relationships between users.

The Independent Cascade (IC) model is initialized with a set of activated nodes $S \subset V$ from which each active node $v_i \in S$ has a single chance to activate its neighbours $v_j \in \Gamma(v_i)$. If v_i successfully activates v_j, then v_j is marked as active and starts to interact and activate its neighbours by disseminating their influence with a specific probability. If v_i is failed to activate v_j, then v_i will not have a second chance to activate v_j. The influence diffusion process ends when all are attempted to be activated, whether successful or not. Once activated, the node will remain active and can not be deactivated.

The Linear Threshold (LT) model is a model in which a node or user is activated by comparing the sum of influence weights from its previous level of neighbours with its threshold. Similar to the IC model, the LT model is also initialized with a set of activated nodes $S \subset V$ from which each active node $v_i \in S$ has a specific influence weight $w_{ij} \in [0, 1]$ to its neighbours $v_j \in \Gamma(v_i)$. When all the influence weights of node V_i's neighbours are added up, it should satisfy that the sum is not larger than 1 as follows:

$$\sum_{v_j \in \Gamma(v_i)} w_{ij} \leq 1, \tag{2.1}$$

Each node in the network has a threshold; the higher the threshold, the harder it is for its neighbours to activate it. When the sum of the influence weights of v_i's neighbours is above the threshold, v_i is activated.

The difference between these two models is that in the IC model, node v_i is activated based on the pre-defined probability from its neighbour, while in the LT model, node v_i is activated based on its own threshold.

Based on these foundational models, many researchers have tailored or varied them to address new problems. For instance, Saito et al. propose novel influence-diffusion models, Asynchronous Independent Cascades (AsIC) and Asynchronous Linear Threshold (AsLT), which improve the IC and LT models, respectively, by considering time delays in user interactions [140]. They address two problems: one is to learn the parameters, probability for the IC model, and influence weight for the LT model, and another one is to align the

asynchronism of the model with the real-world dataset. The probability that node v_i can influence node v_j is:

$$p_{v_i,v_j} = p(x_{v_i,v_j}, \theta) = \frac{1}{1+exp(-\theta^T x_{v_i,v_j})}, \tag{2.2}$$

$$r_{v_i,v_j} = r(x_{v_i,v_j}, \phi) = exp(-\phi^T x_{v_i,v_j}), \tag{2.3}$$

where x_{v_i,v_j} is a $J-$dimentional link attribute vector, $\theta^T = (\theta_0, ..., \theta_J)$ and $\phi^T = (\phi_0, ..., \phi_J)$ are diffusion probability and time-delay parameter, respectively.

Barbieri et al. argue that both IC and LT models are topic-blind, so they extend both the IC and LT models to propose the Topic-aware Independent Cascade (TIC) model and Topic-aware Linear Threshold (TLT) model, which categorize message content and simulate topic distribution [19]. In TIC, they calculate the probability based on the topic. Given a topic $z \in [1, K]$, the corresponding probability $p^z_{v_i,v_j}$ and each disseminated item i, the probability that node v_i succeeds in activating node v_j with the disseminated item i is:

$$p^i_{v_i,v_j} = \sum_{z=1}^{K} \gamma^z_i p^z_{v_i,v_j}, \tag{2.4}$$

where γ^z_i represents the distribution of the topics, with $\sum_{z=1}^{K} \gamma^z_i = 1$.

Similarly, in TLT, a node v_j is activated on item i when the sum of influence weights on topic z from its neighbours is above θ_{v_j} which is a random threshold that uniformly selected from $[0, 1]$. So, at time step t, the sum of influence weights from v_j's neighbours on item i is:

$$W^t_i(v_j) = \sum_{z=1}^{K} \sum_{v \in \Gamma^i(v_j,t)} \gamma^z_i p^z_{v,v_j} \tag{2.5}$$

where $\Gamma^i(v_j, t)$ represents v_j's neighbours at time step t that are active on item i.

Chen et al. is another work extending the IC model with meeting events as IC-M incorporating time-delay [34]. A meeting probability $m(v_i, v_j)$ is defined and calculated as:

$$m : E \to [0,1](if(v_i, v_j) \notin E, m(v_i, v_j) = 0) \tag{2.6}$$

which offers a more realistic probability than the original IC model.

Lagnier et al. introduce a variant of the LT model, the Decaying Reinforced User-Centric (DRUC) model, which incorporates user profiles into the influence diffusion process [88]. Liu et al. enhance the LT model with a diffusion-containment model, the D-C model, by including the containment of competitive influence spread [107]. Unlike the IC and LT-based

models, Li et al. propose agent-based influence diffusion models, where the cascading process is modelled as an evolutionary pattern driven by individual actions [101, 98].

Epidemic Models

Except for the aforementioned traditional influence models, there are also models that mimic the spread of epidemics as the infected people have a chance to infect their surroundings. There are several states in these models, i.e., Susceptible (S), Infected (I), Recovered/Removed (R), and Exposed (E). According to different complexity, there are the Susceptible-Infected (SI) model, Susceptible-Infected-Recovered (SIR) model, Susceptible-Infected-Susceptible (SIS) model, Susceptible-Infected-Recovered-Susceptible (SIRS) model, Susceptible-Exposed-Infected-Recovered (SEIR) model and Susceptible-Exposed-Infected-Recovered-Susceptible (SEIRS) model [60, 178, 126, 77, 12, 95, 24].

In the SI model, there are only two states, and it starts from a set of susceptible nodes. If infected, they will become infectious and have a chance to infect other susceptible nodes. Once nodes are infected, they remain infected and infectious permanently. Different from SI, the SIS model allows infected nodes to recover from infectious and change to susceptible. The recovered nodes can be infected again once their connected nodes are infectious. In SIR, the infected nodes are allowed to be cured and become immune to the disease or influence in social networks. So, after the state of a node is changed to "Recovered", it will never be infected. However, the SIRS model does not agree with that, it argues that the cured nodes can also be susceptible and infected by others. So, in the SIRS model, the cured nodes still have a chance to be infected, and each node in this model has a chance to be infected, cured, and infected again. Moreover, another important state is brought in to make the model closer to real life. The SEIR and the SEIRS models bring the state "Exposed" in as a middle state between susceptible and infected while in SEIR, the node becomes immune after being cured, in SEIRS, each node can be infected again even after being cured.

When it comes to social networks, those states are translated into terms like "infected", meaning "influenced," "susceptible", meaning easy to be influenced, and "exposed", meaning exposed to some specific influence, like advertisements, rumours, etc., and "recovered" meaning getting rid of influence for some reason.

User Behaviour Models

The aforementioned influence models all focus on the network structure in which nodes represent users and edges represent relationships between them, omitting one of the most important elements, which is the user behaviours, like their interests in topics, opinions

to some events, etc, [190]. By considering user behaviours, the methods will have more potential to offer solutions with higher relevance to realistic scenarios. The study on user behaviour in echo chambers is not well researched until recent years with the development of AI.

Cui et al. select users by estimating their historical evidence of user behaviours to predict the cascade outbreak. They aim to leverage historical cascade data to develop a novel data-driven approach for selecting critical users as sensors and predicting outbreaks based on the cascading behaviours of these users. Specifically, they introduce the Orthogonal Sparse Logistic Regression (OSLOR) method, which jointly optimizes user selection and outbreak prediction. In this method, the prediction loss is combined with both an orthogonal regularizer and an L1 regularizer to ensure high prediction accuracy while maintaining the sparsity and low redundancy of the selected sensors [43].

Saito et al. have addressed the challenge of constructing influence from users' historical behaviours, focusing on the independent cascade model. They formally define the likelihood maximization problem and then the Expectation Maximization (EM) algorithm is employed for its resolution. However, their formulation does not scale effectively to large datasets as in social networks, due to the necessity of updating the influence probability in each EM algorithm iteration [139].

Tang et al. investigated topic-based social influence within social networks, where discussion topics are dispersed among users. Their objective was to identify topic-specific sub-networks and determine topic-specific influence weights among members of these sub-networks. To achieve this, they proposed a graphical probabilistic model known as the Topical Factor Graph (TFG) to integrate the information into a single probabilistic framework. Additionally, they introduced the Topical Affinity Propagation (TAP) model, which utilises TFG to infer the influence graph. To address efficiency concerns, they developed a distributed implementation of the TAP model [152].

However, nearly all existing influence-diffusion models overlook the detailed contents of the information. Specifically, while a few studies consider influence topics or general content categories, these are insufficient to capture fine-grained information. Research has shown that users' prior knowledge significantly impacts message acceptance [78, 101]. Therefore, it is crucial to consider the association between users' prior knowledge and detailed influence content in the spreading of influence.

2.1.2 Influence Maximization

Influence maximization is recognized as a popular NP-hard problem, widely investigated by researchers [80]. The objective is to select a finite set of users from a social network,

expecting them to maximize the impact across the entire network [80]. This set of users is called the *seed set*, the process of identifying these influencers is known as *seed selection*, and the algorithms used for seed selection are referred to as *seeding algorithms*. The classic influence maximization problem aims to maximize the overall number of influenced users, thereby reaching a large coverage audience.

Several classic algorithms, such as Random Selection, Greedy Selection, Degree-based Selection, etc., are widely used as baselines to compare with the proposed approaches. The descriptions of these algorithms are as follows.

- **Random Selection:** is a method that follows a randomized rule to activate influential users from the network randomly. The randomized rule can be different distributions which offer different chances to nodes according to different networks and spreading strategies;

- **Greedy Selection:** aims to maximize the spread of influence through a network by iterating each node and selecting the ones with the most coverage.

  It provides a $(1-1/e)$-approximation guarantee for the influence maximization problem under sub-modular influence functions:

  $$\sigma(S) \geq (1-1/e) * \sigma(S^*) \tag{2.7}$$

  where S is the set of nodes the greedy algorithm selects, and S^* is the optimal set of nodes.

  The greedy algorithm is a widely used method in influence maximization tasks. However, it requires a high cost of computation, which motivates researchers to develop more efficient algorithms;

- **Degree-based Selection:** is a heuristic approach based on the degree of nodes in the network. The degree of a node in a network is the number of edges connected to it. So, the idea behind the degree-based selection is that nodes with more connections are more influential.

- **CELF Selection:** retains the approximation guarantee of the greedy algorithm while achieving faster execution by leveraging the sub-modular property of the influence function. it significantly reduces the number of influence spread evaluations, making it more practical for large-scale networks.

Both CELF and its improved variant CELF++ optimize the influence maximization process by reducing redundant computations, making it a practical choice for large-scale social network analysis [92, 62]

- **IMM Selection:** is an algorithm combining the Reverse Influence Sampling technique and martingale theory to identify influential nodes in large-scale social networks efficiently. It is highly efficient compared to traditional greedy algorithms and provides strong probabilistic guarantees on the quality of the solution.

However, this objective cannot be directly applied to scenarios with different goals, such as minimizing rumours and misinformation. Consequently, numerous studies have explored variants of the influence maximization problem. For example, Li et al. model multiple influence diffusion by considering the relationships among various influences [102]. They also study influence maintenance and propose a novel seeding algorithm aimed at sustaining long-term influence [98]. Additionally, Li et al. develop a model to suppress negative social influences by injecting other influences [97]. Bharathi et al. investigate competitive influence diffusion, focusing on scenarios where multiple types of messages compete within a social network [22]. Furthermore, Gershtein et al. introduce the IM-Balanced system, which allows users to explicitly declare the desired balance when multiple objectives are present [59].

Furthermore, almost no research in the field of influence maximization considers the alteration of information within the cascading process. Users often modify influence messages based on their prior knowledge or attach their opinions before spreading them. This critical feature is overlooked by almost all existing studies. In contrast, we aim to maximize influence while suppressing the alteration of influence, ensuring the integrity of the original message as it spreads through the network. By addressing this gap, our approach seeks to maintain the fidelity of the information being disseminated, thereby enhancing the effectiveness and reliability of influence maximization strategies.

Influence Diffusion with Deep Learning

Influence diffusion is a naturally complex process where users spread influences based on numerous factors such as prior commitment level, preferences, neighbours' opinions, and more [101]. To accurately model real-world influence diffusion, it is crucial to account for users' prior knowledge and behaviours. In recent years, deep learning techniques and knowledge graphs (KG) have been recognized as suitable tools for modelling users' complex behaviours and knowledge structures; thus, they are increasingly adopted for predicting influence diffusion.

Yang et al. propose a Neural Diffusion Model (NDM) to predict influence cascades from a microscopic perspective, employing deep learning techniques such as attention mechanisms and convolutional networks for cascade modelling [182]. Wang et al. introduce an attention-based RNN to capture cross-dependencies in influence diffusion and employ a coverage strategy to mitigate the attention misallocation problem [173]. Hu et al. develop a memory-based deep recurrent network, known as Diffusion-LSTM, to recursively predict the entire diffusion path of an image through a social network [69]. Wang et al. explore the use of representation learning to assist in information diffusion prediction on graphs [171].

Qiu et al. utilise users' local network information as input to a graph neural network with an attention mechanism, learning users' latent feature representations to predict social influence, considering both network structures and user-specific features [132]. Keikha et al. employ deep learning technologies to learn network embeddings by extracting local and global structural information from single and multiple interconnected networks to tackle the influence maximization problem [79]. Similarly, Tian et al. propose a deep influence evaluation model to learn node features and address the topic-aware influence maximization problem, using network structural information, topics, and users' interests to identify influential users [157].

Cai et al. propose a holistic influence diffusion model that considers users' cyber and physical interactions, addressing a new holistic influence maximization problem [31]. Teng et al. employ knowledge graphs to model relevant items in a sequence of promotions, considering multiple target products in multiple promotions and users' dynamic perceptions of these product relations [156].

While most existing deep learning-based influence cascade prediction approaches focus on learning sharing behaviours in a social network—namely, "who influenced whom" information is explicitly labelled in the influence-diffusion chain—they often overlook how users perceive information and generate influencing messages based on their existing knowledge. In contrast, we aim to leverage deep learning and knowledge graphs to model users' complex behaviours with a key consideration of information alteration, simulating influence diffusion from a microscopic perspective. By doing so, we hope to capture the nuances of how information is transformed and propagated through social networks, providing a more accurate and comprehensive understanding of influence diffusion.

2.1.3 Misinformation Detection

With the rise of social media, the field of misinformation detection has attracted significant attention from both researchers and practitioners, and the ease with which information can be distributed and consumed has increased, allowing misinformation also to increase [73].

Traditional Misinformation Detection

Traditional misinformation detection focuses on incorporating user-based and content-based approaches. Specifically, user-based or context-based strategies analyse the social environment surrounding misinformation, focusing on user attributes and behaviour, while content-based approaches delve into textual and emotional facets of content [150].

The user-based approaches include extracting explicit and implicit features from user profiles [148] and focusing on status-sensitive users to facilitate early detection [110, 114]. For example, Hamdi et al. propose a novel methodology for evaluating the credibility of information sources on Twitter, utilising node2vec to extract features from Twitter's follower and followee graph and incorporating user-specific attributes [66].

Modern content-based strategies have resulted from combining tensor-based article modelling with semi-supervised learning [2], leveraging transformer-based language models [127], and hybridising Convolutional and Recurrent Neural Networks [120]. While these works provide valuable insights into misinformation detection, they primarily address misinformation through individual user characteristics and isolated content features, often overlooking the broader narrative structure or frames via which the content is portrayed.

Moreover, there is also traditional rule-based misinformation detection for fact-checking, and fake news is focused on detecting misinformation by paying attention to who provided the information or what the content was. Manual fact-checking relied on the author's reputation and/or the source to determine the veracity of the information [65]. Similarly, to detect fake news on social media, the social contexts, such as explicit and implicit features of user's profiles, are evaluated to determine the credibility of the information [150]. In addition to social contexts, fake news detection focuses on the content of the text by extracting linguistic features in order to detect sensational headlines that are frequent in fake news [150]. Moreover, identifying negation keywords, such as 'no,' 'not,' or 'never,' significantly enhances the classification of rumours [87]. Traditional rule-based approaches rely on information specific to the topic to correctly identify misinformation. Therefore, these approaches experience limitations when detecting misinformation about a new topic [163]. These shortcomings are addressed with the introduction of semi-supervised and unsupervised methods [125].

Furthermore, these traditional methods experience limitations in dealing with misinformation derived from factual events but framed to convey alternative implications. This is particularly challenging with lengthy articles that contain a mix of truthful and misleading information.

Deep Learning Based Misinformation Detection

Many researchers have explored the use of deep learning techniques to automate misinformation detection, such as tensor and transformer-based models and convolutional and recurrent neural networks [73, 2, 120, 127, 130]. Latent patterns and spatial context are extracted from tensor-based models to construct k-nearest-neighbour graphs and belief propagation for semi-supervised misinformation detection [2]. Liu and Wu propose a novel deep neural network composed of a status-sensitive crowd response feature extractor, a position-aware attention mechanism, and a multi-region mean-pooling mechanism, addressing the early detection of fake news on social media [110]. A hybrid of convolutional neural networks (CNN) and recurrent neural networks (RNN) leverages the strengths of CNN in extracting local features and of RNN in capturing long-term dependencies to detect fake news [120]. Another RNN model finds that combining sentiment, emotional, irony and hate analysis with bagging, boosting, stacking and voting means, produces a higher accuracy than without the various analyses [130]. An evaluation of transformer-based models Large Language Models, namely, BERT variants to be used as baselines for misinformation detection, can achieve comparable or better performance than more complex state-of-the-art methods [127]. More recently, a transformer-based model, MisRoBÆRTa, utilises RoBERTa and BART to outperform single transformer misinformation detection models [159]. Finally, a hybrid deep learning model integrating features-based models and universal sentence encoding reveals promising results on the PHEME dataset [6].

While these techniques are able to accurately detect misinformation without considering the narrative or frame, their challenge lies in dealing with misinformation stemming from factual events that are skewed to convey a different implication. Furthermore, they also face difficulties handling lengthy news articles that potentially contain both truthful and misleading information.

Framing Theory

Moreover, news frames significantly impact reader interpretation. The frame of a piece of text can increase the salience of specific parts of information, i.e., to make information more meaningful, noticeable, or memorable [52, 74]. An example by Entman shows that a frame can influence how many readers notice, understand, remember, evaluate, or act upon information presented to them [52]. Furthermore, the problem definition, causal interpretation, moral evaluation, and treatment recommendation are also demonstrated as the four identifiable elements of a frame [52]. Multiple methods have been developed to detect frames using different approaches. Liu et al. develop a neural network-based

approach for detecting frames from news based on the article headlines through fine-tuning a Large Language Model, i.e., BERT, where one prominent public affairs issue in the US, i.e., gun violence, is focused [106]. Alternatively, Walter and Ophir leverage computational tools to develop a novel method, the Analysis of Topic Model Networks, for the inductive identification and categorisation of frames, demonstrating its effectiveness across diverse U.S. news corpora, thus offering potential theoretical, methodological, and practical implications for framing research [164]. Similarly, Arendt et al. adopt the reinforcing spiral framework and a mixed-methods approach to explore the underlying mechanism of the news-framing effect [10]. Cabot et al. finetune a joint method based on RoBERTa, including metaphor, emotion, and frames defined by Card et al. to model political discourse [30, 33].

Despite the existing research works on frame detection and analysis, very few studies investigate how frames impact the emergence of misinformation [106, 74]. The key challenge in detecting framing-induced misinformation lies in classifying misinformation stemming from accurate facts but presented under different frames.

2.1.4 Echo Chamber detection and mitigation

Echo Chambers on Social Platforms

Social platforms have become the primary sources of information for many individuals, offering an unprecedented volume of data. This shifts the way people access information and forms echo chambers, isolating them in the process [39]. The detection of echo chambers has been a research focus across various fields [13, 136], serving as the first step towards mitigating this phenomenon.

Social structures typically manifest in two distinct forms, i.e., global and individual perspectives. Most research has studied echo chambers from a global or topological perspective, primarily focusing on user interactions while overlooking the source of these interactions [18, 29]. Cinelli et al. analyse echo chambers by assessing whether the overall network is strongly polarized towards two sides of a controversy, emphasizing user interaction networks [39]. Cossard et al. explore echo chambers within vaccine communities using clustering techniques, demonstrating the existence of echo chambers within real social networks [41].

Analysing extensive topological structure datasets from a global perspective necessitates high-performance computing resources. The ego network centred around a focal user offers a feasible way to model a community, enabling measurement of the echo chamber degree with a focus on that user. Thus, inspired by Li et al. and Valerio et al., we incorporate the concept of the ego network in our study [11, 99]. Li et al. propose agent-based influence diffusion models, where the influence cascading process is modelled as an evolutionary pattern driven

by individuals' actions. Valerio et al. analyse the micro-level structural properties of online social networks and demonstrate that ego networks play a significant role in social networks, impacting information diffusion within the network. Hu et al. investigate the impact of AI recommendation on forming echo chambers from both individual and topological levels [70].

Graph-Based Echo Chamber Detection

Examining social networks can provide valuable insights into echo chambers. Due to homophily, individuals within an echo chamber are well-connected, while those outside these chambers have weaker connections. Beyond connectedness, a common method for detecting echo chambers involves analysing the propagation patterns of discussed topics.

Network-based detection is among the most prevalent methods for identifying echo chambers. With the growing use of social media platforms, these have increasingly become habitats for echo chambers. For instance, Guarino et al. demonstrate the use of community detection to show that high segregation and clustering of communities by political alignment serve as evidence of echo chambers on social media [63].

To analyse the holistic network, several algorithms are used to detect communities in a network or a graph, such as Fast Greedy, Louvain, WalkTrap, and Infomap.

As optimally finding communities is NP-hard, the Fast Greedy algorithm is always utilised as a community detection approach [121]. The Fast Greedy algorithm obtains a network that optimizes on modularity, which is a measure of the quality of a graph containing communities as:

$$Q = \frac{1}{2m}\sum_{i,j}[A_{i,j} - \frac{k_i k_j}{2m}]\sigma(c_i, c_j) \tag{2.8}$$

where A_{ij} is the weight of the edge between node i and j, k_i is the total weight of the edges connected to node i, c_i is the community for node i, $\sigma(c_i, c_j)$ represents an indicator function and m is the number of total edges [41].

Similar to the Fast Greedy, the Louvain algorithm's objective is also the quality of community using the same equation 2.8. This algorithm initializes the graph by assigning a unique community to each node in the graph. Then, these communities are merged if the modularity is increased when a node is compared with its neighbours. Each community is subsequently condensed into a single node, with the new edge weights being the sum of the previous graph's edge weights [25]. Nourbaksh et al. identify communities that indicate groups of users with large intra-group co-linking behaviour, and small inter-group co-linking [123]. Cossard et al. also utilise the Louvain algorithm to identify echo chambers on an Italian Twitter dataset.

The WalkTrap algorithm is a random walk-based method that targets the same end community [131]. The distance between two specific nodes v_i and v_j is defined as:

$$r_{v_i,v_j} = \sqrt{\sum_{k=1}^{n} \frac{(P^t_{v_i,v_k} - P^t_{v_j,v_k})^2}{d(k)}} \tag{2.9}$$

where $P^t_{v_i,v_k}$ is the probability that node v_i will walk to node v_k in t timesteps, $d(k)$ represents the degree of node v_k.

The WalkTrap algorithm also begins by initializing each node in the graph with a unique community. Hierarchical clustering groups similar nodes by merging the closest communities to form a new graph where each node represents a merged community. This process repeats until only one community remains. A Monte Carlo approach can be employed to estimate the probabilities for random walks to address the high computational complexity of identifying optimal communities.

Del Vicario et al. apply the WalkTrap algorithm to identify echo chambers to study the opinions on Brexit on a Facebook dataset. To achieve that, a Bipartite graph is created, and the WalkTrap is used to detect communities as echo chambers [46].

The Infomap algorithm optimizes the map equation to identify communities [137]. The map equation aims to determine the lower bound on the length of the sequence used to represent a walk on the graph. This representation can be minimized using Huffman codes [71], where the most frequently visited nodes are represented in a minimal form. To further reduce the walk's length, the graph can be divided into modules. Each module has its own codebook (module codebook) and a codebook for movements between modules (index codebook). The description length for a module can be represented by the map equation:

$$L(M) = q_{\curvearrowright} H(Q) + \sum_{i=1}^{m} p^i_{\circlearrowright} H(P^i) \tag{2.10}$$

where $H(Q)$ is the frequency-weighted average length of codewords in a codebook, $H(P^i)$ represents the frequency-weighted average length of codewords in module codebook i, $q_{\curvearrowright}$ means the probability of existing from a module while $p^i_{\circlearrowright}$ represents the probability that a module codebook i is used when visiting node v_j with a probability of p_{v_j}. Different from other algorithms, the WalkTrap optimizes the flow of information while others optimize modularity.

Content-Based Echo Chamber Detection

Content-based methods identify echo chambers by analysing the information texts produced by individuals. Several researchers have made significant contributions to this field. Villa et al. propose both a topology-based and content-based approach, analysing the topological structure of the social network and sentiment aspects related to the content [162]. Cinelli et al. conduct a comparative analysis on a large-scale dataset to identify echo chambers through social network homophily. They define "leaning" as the attitude expressed by a piece of content towards a specific topic about the content [39]. Abd-Alrazaq et al. propose a text-mining method on a large dataset, considering information texts but neglecting temporal information, which can provide contextual insights [1]. Lwin et al. and Xue et al. demonstrate that discourses on Twitter about Covid-19 continually evolve, develop, or change over time [113, 181]. Inspired by these studies, we restructure the dataset into chronologically user-specific streams.

Most existing studies solely consider the content of information but overlook individual behaviours and content weights, which demonstrate the significance of content on individuals. For instance, reading a message does not explicitly reveal an individual's thoughts about the message. However, a subsequent "like" or "upvote" implies that the individual agrees with this message, thereby increasing the weight of information from this message in the corresponding belief graph. We argue that beliefs in individuals' minds carry different weights, and not all beliefs hold equal significance. As a result, behaviours offer valuable insights into people's perspectives on related messages.

Therefore, we propose a belief-aware echo chamber detection framework incorporating content and individual behaviours. Our framework constructs belief graphs for each individual in our dataset, considering their behaviours. To measure the degree of echo chambers, we calculate the similarities between the belief graphs of the focal user and their neighbours. With this framework, social platforms can detect communities where members are primarily exposed to reinforcing views, potentially limiting the diversity of thoughts and contributing to polarization.

2.2 Text Summarization

Numerous research efforts have focused on enhancing abstractive generators' paraphrasing capabilities. For example, Rush et al. apply the neural encoder-decoder architecture to text summarization and discuss potential encoder choices [138]. Based on this research work, many researchers investigated approaches to improve encoding and decoding capabilities, addressing issues such as out-of-vocabulary and repetition. See et al., for example, employ

pointer and coverage mechanisms to tackle these problems while developing the pointer-generator network to accurately reproduce source text information [145]. Gehrmann et al. propose a two-step process to address content selection issues: token-level sequence tagging for content selection and bottom-up copy attention to restrict attention over selected source text fragments [58].

In recent years, fact-aware summarization has attracted significant attention, highlighting the challenges associated with generating factually accurate summaries. Previous studies have revealed that abstractive summarization models are prone to hallucinating phenomena, with approximately 30% of summaries from state-of-the-art models exhibiting factual inconsistency [84, 32]. To mitigate this issue, various approaches have been proposed. Cao et al. introduce fact-aware neural abstractive summarization, incorporating extracted facts into the encoder alongside the source text [32]. Kryscinski et al. develop a novel method for verifying factual consistency and identifying conflicts between the source text and summaries [84]. Li et al. treat fact-aware summarization as an entailment-aware process, arguing that summaries should be semantically entailed by the source text [94]. Zhu et al. utilise Knowledge Graphs to integrate factual information into the summarization process [192].

Despite the advancements in abstractive summarization and fact-aware techniques, existing works possess certain limitations that stem from overlooking the adaptive nature of opinion summarization. The ability to adapt summaries based on user-specific preferences, requirements, or interests is essential for capturing a wide range of aspects and accommodating diverse user needs. Furthermore, most existing approaches predominantly focus on generating summaries that preserve the general or most salient information from the source text, often leading to summaries that may not cater to individual user preferences.

Keywords Extraction

Classic abstraction models learn the distribution over the input texts and generate the token-level or sentence-level representations. These vectors are fed into the decoder, producing a summary with the trained model parameters to maximize the output likelihood. However, the output is sometimes difficult to control due to the unconstrained nature of neural-based language models. Thus, many researchers attempt to adopt constraints during the text summary-generating process. For example, Zhou et al. propose a selective gate network to extract key points, leading the summary to be more focused on these extracted keywords [191]. Ercan et al. leverage WordNet in the lexical chain-building algorithm, which can discover the relations between two-word senses [53]. Building lexical chains turns out to be very time-consuming work since it relies on exhaustive algorithms. Litvak et al. [105] propose a graph-based method of both supervised and unsupervised approaches. In the

supervised approach, the novel method first converts the documents into word graphs and trains a classifier to check if the corresponding word is included in the document extractive summary. In the unsupervised approach HITS algorithm [82] is utilised to calculate the weight of each node and select the top ones.

The aforementioned works have achieved promising results. However, the limitations are still presented. Classical abstractive summarization models do not consider the impact of additional references from the source text but generate summaries that only rely on the hidden states of the inputs. Fact-aware models apply a fact-aware approach to guide the generating process, but unrelated information is also used as guidance. Such unrelated information may distract the summarization generator from producing a better summary. Meanwhile, approaches with keyword extraction neglect the relation of these keywords, and the generator may manipulate the facts.

In this thesis, we propose a novel Knowledge-aware Abstractive Text Summarization (KATSum) model that can address the limitations mentioned above. We utilise the Knowledge Graph to extract triplets, eliminate the noise with a trained classifier, and identify useful triplets as auxiliaries to guide the text summarization generation process.

Opinion Summarization

Opinion summarization can also be categorized into abstractive and extractive methods. Mirroring the developmental trajectory of general summarization, extractive methods initially gained widespread use in this domain due to the high cost of creating golden summaries for datasets, particularly for review datasets where such summaries are not mandatory [38]. Early works in this field predominantly treat the task as a sentence or phrase selection problem, employing either ranking or classification approaches. For instance, Wei et al. prioritize sentences that closely correspond to the query [174]. At the same time, Erkan et al. employ a stochastic graph-based method to rank sentences by calculating their importance based on eigenvector centrality within a graph representation of the sentences [54]. Extractive methods remain popular for opinion summarization [40, 90, 56, 7].

Abstractive methods for opinion summarization began to emerge around 2010, with Ganesan et al. introducing a graph-based algorithm for generating abstractive summaries [57]. Although this approach produces abstractive summaries, it selects words, phrases, or sentences from the original text, rendering it more akin to an extractive method. In recent years, a growing number of studies have begun to leverage machine learning and deep learning techniques for this task. Notably, Chu et al. present an unsupervised neural model for multi-document summarization, proposing an end-to-end architecture featuring an

auto-encoder. This approach decodes the mean of the input reviews' representations into a coherent summary review without relying on any review-specific attributes [38].

Self-supervision

In addition to the abstractive and extractive categories, opinion summarization can be classified into supervised and unsupervised categories. Given the scarcity of golden summaries for datasets, unsupervised approaches are more frequently employed to address this challenge. Both [28] and [38] are unsupervised models that employ auto-encoders, suggesting that review representations can encapsulate sentiment, topics, and opinions about products.

Recently, text summarization has shifted toward a self-supervised approach by utilising similar content as pseudo summaries [7, 50]. To achieve this, Amplayo et al. fine-tune a pre-trained model using their synthetic training dataset of (review, summary) pairs and generate aspect-specific summaries by modifying their introduced aspect controllers [7]. Meanwhile, Elsaha et al. tackle multi-document opinion summarization by assuming one of the documents serves as a target summary for a set of similar documents [50].

However, sampling similar content or documents from the entire dataset may result in the loss of salient information, as a subset may not encompass all relevant details. In the context of opinion summarization, this lost information could be significant to different individuals with distinct requirements. Amplayo et al. develop a method for generating aspect-specific summaries and constructing a synthetic dataset composed of (review, summary) pairs. To achieve this, they employ a technique involving the sampling of reviews as pseudo summaries, introducing three distinct aspect controllers at the word, sentence, and document levels [7]. Nevertheless, this approach unavoidably leads to a certain degree of information loss.

2.3 Summary

This chapter conducts a comprehensive literature review in all relevant fields, i.e., influence diffusion, influence maximization, misinformation detection, echo chamber detection and text summarization.

Limitations in these issues are recognized and summarized as follows:

- Classic abstractive summarization models do not consider the impact of additional reference from the source text but generate summaries only relying on the hidden states of the inputs. Fact-aware models apply a fact-aware approach to guide the generating process, but unrelated information is also used as guidance. Such unrelated information

may distract the summarization generator from producing a better summary. Meanwhile, approaches with keyword extraction neglect the relation of these keywords, and the generator may manipulate the facts.

Despite the advancements in abstractive summarization and fact-aware techniques, existing works possess certain limitations stem from overlooking the adaptive nature of opinion summarization. The ability to adapt summaries based on user-specific preferences, requirements, or interests is essential for capturing a wide range of aspects and accommodating diverse user needs. Furthermore, most existing approaches predominantly focus on generating summaries that preserve the general or most salient information from the source text, often leading to summaries that may not cater to individual user preferences.

- Probabilistic Diffusion: Most research treats influence diffusion as a simple probabilistic hopping and infecting process, ignoring users' prior knowledge and the detailed content of influence.

  Information Alteration: Few studies consider the importance of maintaining constant messages throughout the diffusion process. Information alteration can complicate the influence spread chain, producing many variants and deviating from the initial objective. It is crucial to account for information alteration and take measures to retain the originality of influence messages.

- Despite the existing research on frame detection and analysis, very few studies investigate how frames impact the emergence of misinformation [106, 74]. The key challenge in detecting framing-induced misinformation lies in classifying misinformation stemming from accurate facts but presented under different frames. Although both frame and framing element detection are possible, the impact of frames on misinformation detection requires further research.

- For echo chamber detection, most existing studies solely consider the content of information but overlook individual behaviours and content weights, which demonstrate the significance of content on individuals. For instance, reading a message doesn't explicitly reveal an individual's thoughts about the message. However, a subsequent 'like' or 'upvote' implies that the individual agrees with this message, thereby increasing the weight of information from this message in the corresponding belief graph. We argue that beliefs in individuals' minds carry different weights, and not all beliefs hold equal significance. As a result, behaviours offer valuable insights into people's perspectives on related messages.

In Chapter 3, two summarization models, KATSum and AaKOS, are proposed to incorporate knowledge graphs and deep learning to summarize articles and product reviews, respectively. Chapter 4 addresses the issue of information alteration by proposing a novel seed-selecting algorithm that maximizes influence while considering information alterations, maintaining the originality of influence messages. In Chapter 5, two models, the FrameTruth Model (FTM) and Frame Element-based Model (FEM), are introduced to identify misinformation stemming from accurate facts but presented under different frames. Additionally, a belief graph-based method is proposed to detect if a user is isolated in an echo chamber at the individual level.

Chapter 3

Text Summarization with Knowledge Graph

In this chapter, two novel text summarization models, KATSum and AaKOS, are proposed. Both are composed of knowledge graphs and pre-trained language models.

Text Summarization is recognised as one of the NLP downstream tasks, and it has been extensively investigated in recent years. Meanwhile, the exponential increase in online information has led to an overwhelming amount of opinions and comments on various activities, products, and services. This makes it difficult and time-consuming for users to sift and process all the available information when making decisions. Text summarization can assist people with rapidly perceiving information from long or multiple documents, including news articles, social posts, product reviews, etc.

Most existing research works attempt to develop summarization models to produce a better output. However, advent limitations of most existing models emerge, including unfaithfulness and factual errors. In this chapter, we propose a novel model, named KATSum (Knowledge-aware Abstractive Text Summarization), which leverages the advantages offered by Knowledge Graphs to enhance the standard Seq2Seq model. On top of that, the Knowledge Graph triplets are extracted from the source text and utilised to provide keywords with relational information, producing coherent and factually errorless summaries.

Recent advances in pre-trained language models, such as ChatGPT, have demonstrated the potential of Large Language Models (LLMs) in text generation. While LLMs hold immense potential for text summarization, their practical application in real-time contexts is hindered by their substantial data and resource demands. Furthermore, existing text summarization approaches often lack the "adaptive" nature required to capture diverse aspects in opinion summarization, which is particularly detrimental to users with specific requirements or preferences. In this chapter, we also propose an Aspect-adaptive Knowledge-

based Opinion Summarization model for product reviews, which effectively captures the adaptive nature required for opinion summarization. The model generates aspect-oriented summaries given a set of reviews for a particular product, efficiently providing users with useful information on specific aspects they are interested in, ensuring the generated summaries are more personalized and informative.

Extensive experiments use real-world datasets to evaluate the proposed model. The results reveal that the KATSum can effectively utilise the information from the Knowledge Graphs and significantly reduce the factual errors in the summary. In the meantime, AaKOS outperforms state-of-the-art approaches in opinion summarization and is adaptive and efficient in generating summaries that focus on particular aspects, enabling users to make well-informed decisions and catering to their diverse interests and preferences.

3.1 Overview

The rapid development of the Internet generates a massive amount of information daily, leading to challenges in efficiently retrieving valuable information, such as news, online shopping reviews, etc., which are important for people to consume daily information. These reviews are crucial as they support consumers' decision-making and guide businesses in refining their marketing strategies. Individuals often seek insights into specific aspects of products or services, highlighting the need for effective summarization tools [7].

Text Summarization aims to produce short and brief texts for long or multiple documents while keeping the core information. In response to this, text summarization techniques have emerged as powerful solutions, enabling the ability to condense salient information from multiple comments, aiding efficient decision-making [7]. Text summarization is to generate concise summaries while preserving core information. Two main methods are commonly used: extractive and abstractive methods. **Extractive** methods identify the most meaningful phrases and sentences and combine them without modification to form a summary [189, 179]. In contrast, **abstractive** methods employ innovative language to craft summaries, often introducing new expressions [49, 145]. Hybrid approaches have also been proposed, combining both methods to generate coherent text while retaining key information [187, 108].

Based on the above comparisons, it is evident that the extractive methods can produce a summary that closely follows the grammar rules and presents the facts since the summary consists of text chunks copied over from the source text without any modification. In other words, the selected text chunks are supposed to be concatenated to produce a short passage as a summary. However, by leveraging such an approach, the summary appears not coherent since two sentences in a summary may be far from each other in the source text.

Compared with the extractive methods, the abstractive methods can produce cohesive text. However, they suffer from two major issues, i.e., unfaithfulness and factual inconsistency. The former indicates that the content of the generated summary is far from the main idea of the source text, failing to retain the salient information. The latter refers to the issue that the summarization generator produces incorrect facts from the source text due to its unconstrained nature. Studies have shown that abstractive generator tends to distort the facts from the source text [84, 115], and 30% summaries generated through abstractive methods suffer from the factual inconsistency [32]. In this sense, it is vital to mitigate this issue and avoid producing misleading texts.

Many research works employ keywords in the summarization models, ensuring that the key information from the source text is incorporated. For instance, Li et al. develop an extractor to extract keywords as guidance, which can prevent the model from losing key information [93]. Zhou et al. propose a selective gate network to select keywords from the source text [191]. See et al. utilise the pointer mechanism and propose a pointer-generator model to identify the keywords [145]. However, the relations, representing the semantic relationships between the keywords, are neglected, leading to unfaithfulness and factual errors in the summary.

Knowledge Graph has been widely acknowledged as a suitable tool for modelling connected data, where the nodes present entities and links are modelled as corresponding semantic relationships. In this chapter, we aim to propose a novel text summarization framework by leveraging the advantages offered by Knowledge Graph. The proposed framework is capable of producing coherent text and emphasising the facts extracted from the source text, where the Knowledge Graph guides the summarization generation process. Specifically, we utilise the Knowledge Graph triplets, in the form of $\{head, relation, tail\}$, to incorporate keywords with their relationships in the process of text summarization. Different from the existing keyword extraction methods, we present the summarization process as a pipeline, starting from Knowledge Graph construction. Then, the extracted Knowledge Graph will be mapped into a lower-dimensional embedding space, in which each triplet embedding is supposed to be fed into a trained classifier, checking if this triplet should be included in the summary or not. To evaluate the proposed framework, we conduct experiments on the CNN/Daily Mail dataset and calculate the ROUGE scores. The experimental results reveal that our model can yield a better performance by comparing it against the baseline models.

Except for document summarization, opinion summarization is acknowledged as a sub-research domain of text summarization, which aims to generate summaries from a set of reviews [112, 7, 3]. Opinion summarization has been studied using extractive and abstractive approaches, where ranking algorithms, rule-based methods, and machine learning

are employed to identify important phrases, sentences, or paragraphs [135, 40]. Abstractive approaches utilise neural networks and deep learning technologies to generate more coherent summaries [7, 9].

From the training perspective, the field has seen a shift towards self-supervised methods, which employ similar content as pseudo summaries instead of gold-standard summaries [28, 72, 7]. Although these methods facilitate the training of summarization models without the need for labour-intensive and expensive human-generated summaries, they lack the "adaptive" nature that is essential for capturing diverse aspects of opinion summarization. One major concern with using similar content as pseudo summaries is the potential loss of salient information, which is particularly detrimental to users with specific requirements or preferences. This limitation leads to incomplete or inadequate summaries, as not all critical aspects may be represented. Consequently, users may not have access to all the necessary information for making informed decisions. Moreover, when using sampled reviews as pseudo summaries, there is a risk that the generated summaries may not comprehensively cover the diverse aspects present in a larger set of reviews. The lack of adaptability is particularly problematic for users interested in specific aspects or features. This underscores the importance of developing summarization methods that can effectively capture and present a wide range of aspects to cater to users' diverse needs and interests.

Recently, ChatGPT has shown proficiency in generic text summarization, with various methodologies emerging to harness its capabilities for producing concise summaries [23, 153]. Though ChatGPT can generate concise summaries, the performance does not exceed traditional fine-tuned methods, especially when it is aspect or query-based summarization [183].

To address the aforementioned limitations, in this chapter, we also propose the Aspect-adaptive Knowledge-based Opinion Summarization (AaKOS) model, which can effectively capture the adaptive nature required for opinion summarization. By incorporating adaptability in AaKOS, it can dynamically adjust to users' preferences and interests, ensuring the generated summaries are more personalized and informative. On top of that, AaKOS does not rely on datasets with text-golden summary pairs, which are difficult and expensive to create. Instead, AaKOS works with datasets without human-written summaries as labels, allowing for greater dataset flexibility. Specifically, AaKOS transforms plain texts into weighted Knowledge Graphs utilising LLMs as the triplet extractor to extract useful information with fine-grained prompts and encodes both aspect-related triplets in plain text format and corresponding graphs using a text encoder and a graph encoder, respectively.

AaKOS model is trained in a self-supervised approach, which involves utilising the original sentences derived from the aspects and knowledge graph as pseudo summaries.

In contrast to other self-supervised models that rely on synthetic pairs of source texts and pseudo summaries, AaKOS does not sample reviews from the dataset. Instead, it accurately pairs knowledge graphs and aspects with their corresponding content.

AaKOS significantly enhances output precision, particularly for given aspects. When an aspect is determined, relevant content and sub-graphs are extracted from the processed dataset. Consequently, the output contains information only about the specific aspects included in the graph, ensuring that other aspects are not summarized in the output. This adaptability enables the AaKOS model to generate aspect-specific summaries, catering to users with diverse interests and preferences and providing more personalized and informative results.

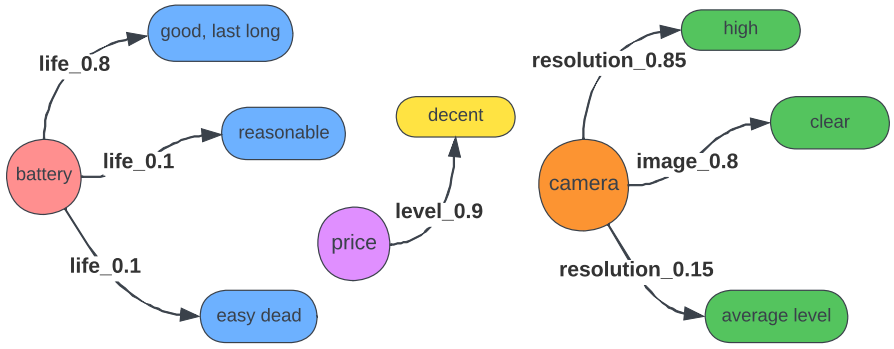


Fig. 3.1 Sample Summary Generated

Figure 3.1 demonstrates the summary generated by our model when given several specific aspects. Specifically, with particular aspects, our model can generate a corresponding summary for a mobile phone from the dataset. Aspects and related contents are coloured in the summary. The value of the weight controller is set to $wc > 0.2$, so triplets with weights higher than 0.2 and corresponding sub-graphs are used to generate the summary, and triplets {battery, life_0.1, easy dead}, {battery, life_0.1, easy dead} and {battery, life_0.1, easy dead} and {camera, resolution_0.15, average level} are not included in the summary.

By leveraging knowledge-based techniques and aspect-adaptive summarization, our proposed model addresses the limitations of existing methods and offers a more efficient way to extract and present relevant information to users. This approach not only helps users make well-informed decisions but also supports enterprises in refining their strategies based on the aspects that matter most to their customers.

Our contributions can be summarized as below:

- We utilise Knowledge Graphs to address the unfaithfulness and factual error problem;
- We imply a triplet classifier to identify the core triplets;
- We utilise LLMs as triplets extractors with fine-grained prompts to facilitate the construction of Knowledge Graphs from the review sets;
- We utilise the self-supervised manner to train a graph-to-text model which also incorporates the embeddings of relevant triplets generated by a pre-trained language model, i.e., BERT;
- We also introduce a weight controller that is able to control the sentiment trends of the output.

The rest of this chapter is organized as follows: In Section 6.2, the related works and limitations are articulated. In Sections 3.3 and 3.4, we elaborate on the proposed KATSum text summarization model, the AaKOS opinion summarization model, and the corresponding experimental setup and results, respectively. Finally, this chapter is concluded in Section 6.5.

3.2 Related works

Text Summarization

Numerous research efforts have focused on enhancing abstractive generators' paraphrasing capabilities. For example, Rush et al. apply the neural encoder-decoder architecture to text summarization and discuss potential encoder choices [138]. Based on this research work, many researchers investigated approaches to improve encoding and decoding capabilities, addressing issues such as out-of-vocabulary and repetition. See et al., for example, employ pointer and coverage mechanisms to tackle these problems while developing the pointer-generator network to reproduce source text information accurately [145]. Gehrmann et al. propose a two-step process to address content selection issues: token-level sequence tagging for content selection and bottom-up copy attention to restrict attention over selected source text fragments [58].

In recent years, fact-aware summarization has attracted significant attention, highlighting the challenges associated with generating factually accurate summaries. Previous studies have revealed that abstractive summarization models are prone to hallucinating phenomena, with approximately 30% of summaries from state-of-the-art models exhibiting factual inconsistency [84, 32]. To mitigate this issue, various approaches have been proposed. Cao et al.

introduce fact-aware neural abstractive summarization, incorporating extracted facts into the encoder alongside the source text [32]. Kryscinski et al. develop a novel method for verifying factual consistency and identifying conflicts between the source text and summaries [84]. Li et al. treat fact-aware summarization as an entailment-aware process, arguing that summaries should be semantically entailed by the source text [94]. Zhu et al. utilise Knowledge Graphs to integrate factual information into the summarization process [192].

Despite the advancements in abstractive summarization and fact-aware techniques, existing works possess certain limitations that stem from overlooking the adaptive nature of opinion summarization. The ability to adapt summaries based on user-specific preferences, requirements, or interests is essential for capturing a wide range of aspects and accommodating diverse user needs. Furthermore, most existing approaches predominantly focus on generating summaries that preserve the general or most salient information from the source text, often leading to summaries that may not cater to individual user preferences.

Keywords Extraction

Classic abstraction models learn the distribution over the input texts and generate the token-level or sentence-level representations. These vectors are fed into the decoder, producing a summary with the trained model parameters to maximize the output likelihood. However, due to the unconstrained nature of neural-based language models, the output is sometimes difficult to control. Thus, many researchers attempt to adopt constraints during the text summary-generating process. For example, Zhou et al. propose a selective gate network to extract key points, leading the summary to be more focused on these extracted keywords [191]. Ercan et al. leverage WordNet in the lexical chain-building algorithm, which can discover the relations between two-word senses [53]. Building lexical chains turns out to be very time-consuming work since it relies on exhaustive algorithms. Litvak et al. [105] propose a graph-based method of both supervised and unsupervised approaches. In the supervised approach, the novel method first converts the documents into word graphs and trains a classifier to check if the corresponding word is included in the document extractive summary. In the unsupervised approach, the HITS algorithm [82] is utilised to calculate the weight of each node and select the top ones.

The aforementioned works have achieved promising results. However, the limitations are still presented. Classical abstractive summarization models do not consider the impact of additional references from the source text but generate summaries that only rely on the hidden states of the inputs. Fact-aware models apply a fact-aware approach to guide the generating process, but unrelated information is also used as guidance. Such unrelated information may distract the summarization generator from producing a better summary. Meanwhile,

approaches with keyword extraction neglect the relation of these keywords, and the generator may manipulate the facts.

In this chapter, we propose a novel Knowledge-aware Abstractive Text Summarization (KATSum) model, which can address the limitations mentioned above. We utilise the Knowledge Graph to extract triplets, eliminate the noise with a trained classifier and identify useful triplets as auxiliaries to guide the text summarization generation process.

Opinion Summarization

Opinion summarization can also be categorized into abstractive and extractive methods. Mirroring the developmental trajectory of general summarization, extractive methods initially gained widespread use in this domain due to the high cost of creating golden summaries for datasets, particularly for review datasets where such summaries are not mandatory [38]. Early works in this field predominantly treated the task as a sentence or phrase selection problem, employing either ranking or classification approaches. For instance, Wei et al. prioritize sentences that closely corresponded to the query [174], while Erkan et al. employed a stochastic graph-based method to rank sentences by calculating their importance based on eigenvector centrality within a graph representation of the sentences [54]. Presently, extractive methods remain popular for opinion summarization [40, 90, 56, 7].

Abstractive methods for opinion summarization began to emerge around 2010, with Ganesan et al. introducing a graph-based algorithm for generating abstractive summaries [57]. Although this approach produced abstractive summaries, it selected words, phrases, or sentences from the original text, rendering it more akin to an extractive method. In recent years, a growing number of studies have begun to leverage machine learning and deep learning techniques for this task. Notably, Chu et al. presented an unsupervised neural model for multi-document summarization, proposing an end-to-end architecture featuring an auto-encoder. This approach decoded the mean of the input reviews' representations into a coherent summary review without relying on any review-specific attributes [38].

Self-supervision

In addition to the abstractive and extractive categories, opinion summarization can also be classified into supervised and unsupervised categories. Given the scarcity of golden summaries for datasets, unsupervised approaches are more frequently employed to address this challenge. Both [28] and [38] are unsupervised models that employ auto-encoders, suggesting that review representations can encapsulate sentiment, topics, and opinions about products.

Recently, text summarization has shifted toward a self-supervised approach by utilising similar content as pseudo summaries [7, 50]. To achieve this, Amplayo et al. fine-tune a pre-trained model using their synthetic training dataset of (review, summary) pairs and generate aspect-specific summaries by modifying their introduced aspect controllers [7]. Meanwhile, Elsaha et al. tackle multi-document opinion summarization by assuming one of the documents serves as a target summary for a set of similar documents [50].

However, sampling similar content or documents from the entire dataset may result in the loss of salient information, as a subset may not encompass all relevant details. In the context of opinion summarization, this lost information could be significant to different individuals with distinct requirements. Amplayo et al. develop a method for generating aspect-specific summaries and constructing a synthetic dataset composed of (review, summary) pairs. To achieve this, they employ a technique involving the sampling of reviews as pseudo summaries, introducing three distinct aspect controllers at the word, sentence, and document levels [7]. Nevertheless, this approach unavoidably leads to a certain degree of information loss.

Differing from the existing works, our approach first extracts aspects using LLMs and then transforms the text dataset into Knowledge Graphs, which consist of numerous triplets with weighted edges. The edges are formatted as *aspect_weight*, as depicted in Figure 3.1. To train our model, we utilise a self-supervised methodology, creating pseudo summaries for the relevant triplets and mapping them as sample pairs. Aspect control is achieved through a Weight Controller α to guide the selection of aspects, enabling the regulation of sentiment trends within these aspects.

3.3 KATSum: Knowledge-aware Abstractive Text Summarization

In this section, we introduce the proposed novel text summarization model, called Knowledge-aware Abstractive Text Summarization (KATSum) model, by explaining the architecture and triple extraction process.

The architecture of KATSum has been presented in Figure 3.2. The proposed KATSum model consists of a knowledge-aware encoder encoding the article and identifying key information, as well as a decoder generating texts. The knowledge-aware encoder incorporates two pipelines, i.e., the classic encoder pipeline and the knowledge-graph pipeline. The former is composed of a pre-trained BERT model [47], which produces the hidden states used as part of the decoder input. The latter extracts key triplets from the source texts and maps them

into an embedding vector space. Before feeding into the decoder, the outputs from the BERT encoder and knowledge graph are fused as the decoder input.

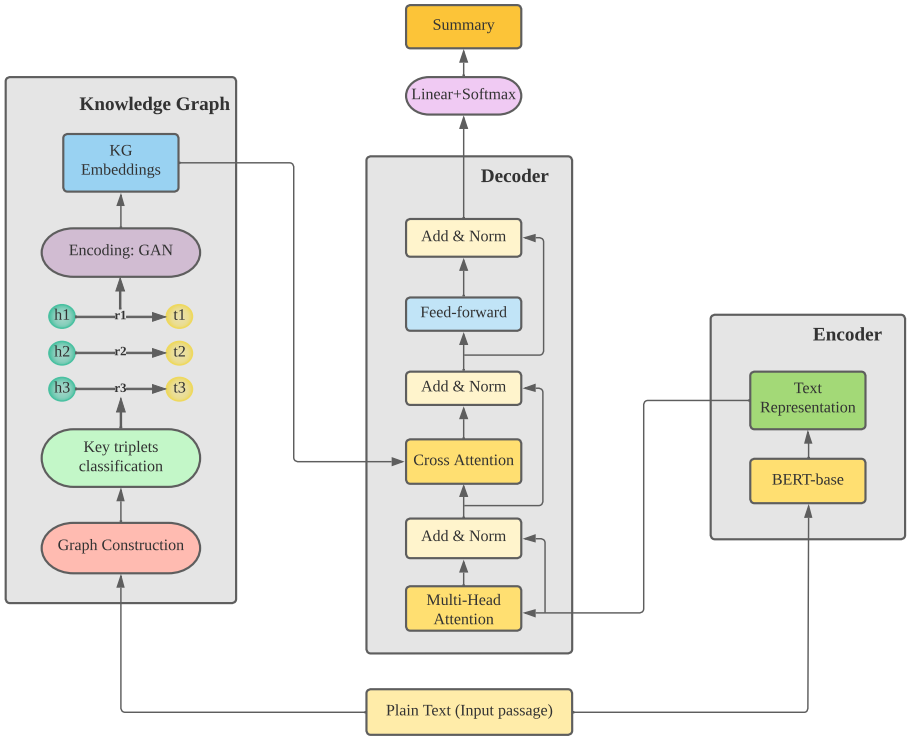


Fig. 3.2 The Architecture of KATSum

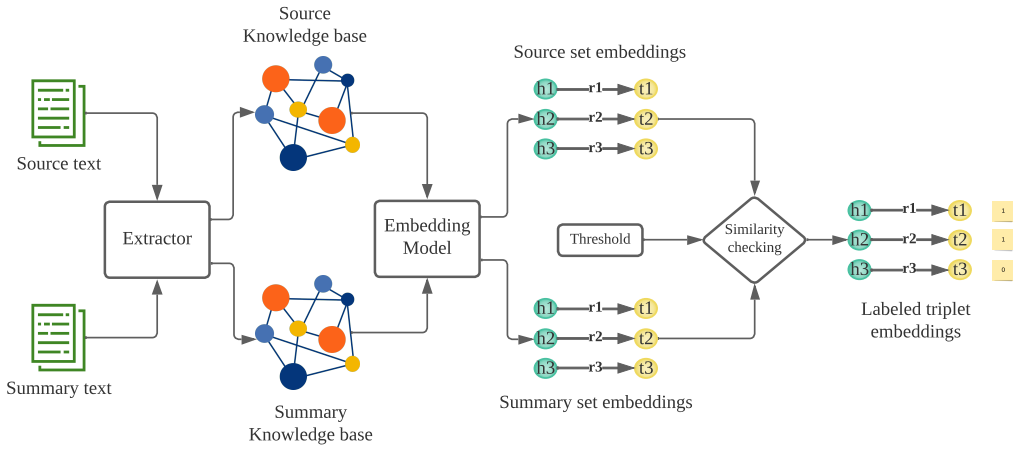


Fig. 3.3 The process of extracting triplets from source text.

On top of that, we employ a Knowledge Graph embedding classifier to identify the knowledge included in the summary from the source text. Triplets are extracted from both

source texts and summaries to train this classifier and presented as source triplet set and summary triplet set, respectively. In order to identify the key triplets, TransE [27] has been adopted to transform the extracted Knowledge Graph into a low-dimensional embedding space so that the semantic-level similarity can be compared. Each triplet from the source set is labelled as 1 if a similar triplet is found in the summary triplet set, and 0 otherwise. We train the Knowledge Graph embedding classifier to identify the key information from the source text by using the labelled data. The information from the source text can be extracted in the form of a triplet, i.e., $\{head, relation, tail\}$. The Sigmoid classifier is formulated in Equation 3.1.

$$\hat{y}_i = \sigma(We_i + b), \tag{3.1}$$

where e_i denotes the vector of Triplet t_i. The identified triplet embedding is fused into the baseline model from where we get the Vector v_i.

The decoder of KATSum derives from the Transformer Decoder with 6 identical layers. Inspired by [108], we follow the training strategy using two Adam optimizers respectively, with different warm-up schedules for the learning rate:

$$lr_e = \widetilde{lr}_e \cdot min(step^{-0.5}, step \cdot warmup_e^{-1.5}) \tag{3.2}$$

$$lr_d = \widetilde{lr}_d \cdot min(step^{-0.5}, step \cdot warmup_d^{-1.5}) \tag{3.3}$$

3.3.1 Experimental Setup

In this section, we first formulate the text summarization problem. Then, the datasets and baselines used in the experiments are introduced. Finally, we explain the parameter selections.

Text Summarization Problem

Given an input sequence, $X = \{x_1, x_2, x_3, ..., x_n\}$, where n refers to the sequence length. The objective is to obtain a text summary, $Y = \{y_1, y_2, y_3, ..., y_m\}$, $m \leq n$, where m refers to the summary length.

Experimental Datasets

Two real-world datasets have been utilised in the experiments, i.e., CNN/Daily Mail [67, 145] and XSum [119]. The statistics of both datasets are listed in Table 3.6.

Table 3.1 The statistics of two datasets

Dataset	Data Type	Sample Pairs			Avg Tokens	
		Training	Validating	Testing	Source Text	Summary
CNN/Daily Mail	News	287,226	13,368	11,490	781	56
XSum	News	204,045	11,332	11,334	431	23

- **CNN/Daily Mail** is a widely used dataset in many research works especially has been used for evaluating text summarization. It consists of news articles (781 tokens on average) from CNN and Daily Mail websites paired with multi-sentence summaries (3.75 sentences or 56 tokens on average). It, in total, contains 287,226 (92%) training pairs, 13,368 (4.3%) validation pairs and 11,490 (3.7%) test pairs.
- **XSum** is a dataset of articles (431 tokens on average) and one-sentence summaries (23 tokens on average) collected from BBC. The official random split includes 204,045 (90%), 11,332 (5%) and 11,334 (5%) documents in training, validation and test sets, respectively.

Evaluation Metrics

ROUGE [104] is selected as the evaluation metric for measuring the quality of the generated summaries. ROUGE_1 and ROUGE_2 refer to the overlap of uni-gram and bi-gram between the source text and the generated summary, respectively. ROUGE_L describes the longest common sub-sequence. ROUGE is one of the most widely used evaluation metrics for text summarization [104]. It focuses on token-level matching but ignores the semantic-level matching between the summary and the source text, which results in the inability to identify factual errors.

Baselines

To evaluate KATSum, we utilise the existing models with encoder-decoder architecture as the baselines. Specifically, the transformer-based pre-trained language models, e.g., BERT [47], are adopted as the encoder. For the decoder, we select the original transformer decoder with six identical layers [81]. To accommodate the BERT input, three types of embeddings, i.e., token embeddings, segment embeddings and position embeddings, are fused as one input vector before being fed into the encoder.

There are many other pre-trained language models. Different combinations of these models, e.g., BERT as encoder and GPT2 as decoder, can yield a different performance. Therefore, we select different combinations of several pre-trained language models as the baselines. In this chapter, the optional encoders include BERT and XLNet [184]. As for the decoder, the original transformer decoder is applied.

Parameters Selection

As the encoder of the KATSum backbone, we leverage "bert-base-uncased" implemented by Hugging Face[1]. While the decoder possesses 768 hidden units, and the hidden size of the feed-forward layer is 2,048.

As the decoder is transformer-based, we train the model using the same strategy as [108] since the transformer decoder needs to be trained from scratch while the BERT encoder is well pre-trained. We fine-tune the encoder and train the decoder separately with different schedules. For encoder, we use $\beta_1 = 0.9$ for Adam optimizer, $\hat{lr}_e = 2e^{-}3$ as the initial learning rate, $warmup_e = 20,000$. For decoder, we use $\beta_2 = 0.999$ for the Adam optimizer, $\hat{lr}_d = 0.1$ as the initial learning rate, $warmup_d = 10,000$. The parameter settings are given according to the empirical values from the existing studies.

To label the Knowledge Graph triplets, we also conduct experiments on the similarity threshold. Given the choices, i.e., $threshold \in \{0.5, 0.8, 0.9\}$, the model yields the best performance when $threshold = 0.8$. To initiate the training, we first train the classifier based on the labelled data for 5 epochs, having 10,000 steps in each epoch. Then, we train the model by using the original data for 200,000 steps. Next, we evaluate and save checkpoints every 2,500 steps. As the GPU memory limits the batch size, we employ the gradient accumulation and calculate the gradient every 5 steps. All the models are trained on 1 Tesla P100 GPU with a memory of 16GB.

3.3.2 Experimental Results

Experiment 1

In this experiment, we calculate ROUGE scores on both CNN/Daily Mail and XSum to compare the performance. The experimental results are illustrated in Tables 3.7 and 3.9. Two baseline models are implemented. One adopts BERT as the encoder, while the other employs XLNet as the encoder.

[1] https://huggingface.co/

The experimental results explicitly show that our model can remarkably outperform the baselines with the assistance of Knowledge Graph, especially when using XLNet as the encoder. Our model's score with XLNet is more than 20% higher than the baseline on CNN/Daily Mail. Using XSum, KATSum with XLNet also outperforms other models, especially the ROUGE_L scores.

Table 3.2 ROUGE results on CNN/Daily Mail over different models

Model	**ROUGE_1**	**ROUGE_2**	**ROUGE_L**
Baseline (BERT - Transformer decoder)	32.5	14.2	29.8
Baseline (XLNet - Transformer decoder)	34.2	15.6	31.5
KATSum (BERT)	41.4	20.3	39.5
KATSum (XLNet)	**42.8**	**20.7**	**40.2**

Table 3.3 ROUGE results on XSum over different models

Model	**ROUGE_1**	**ROUGE_2**	**ROUGE_L**
Baseline (BERT - Transformer decoder)	33.7	14.6	27.2
Baseline (XLNet - Transformer decoder)	34.1	14.3	28.4
KATSum (BERT)	44.8	22.4	38.1
KATSum (XLNet)	**45.6**	**22.9**	**38.4**

Different pre-trained language models are employed as our backbone encoders. We conduct analysis and discuss the results. Intuitively, XLNet is more suitable for working with Knowledge Graph since XLNet can encode the entire article without length restrictions [184]. Specifically, BERT only uses the first 512 tokens of the source text, and the rest are not encoded. In contrast, the Knowledge Graph can offer key information extracted from the entire source text. In this case, the decoder may not effectively utilise the key information. XLNet has been proven to be effective at increasing the performance of a summarizer [170]. Thus, the XLNet has also been applied as the encoder of KATSum.

Experiment 2

In this experiment, we conduct an ablation study to demonstrate the importance of the Knowledge Graph component of KATSum. Specifically, we first remove the classification process from the Knowledge Graph component, directly using the knowledge graph embeddings from the graph neural network, and observe the impact on the final results. Then, we remove the whole Knowledge Graph component from the model and compare the results with the complete model with the Knowledge Graph component and classification process. We use 1/3 of the entire CNN/Daily Mail dataset and conduct experiments using the Bert-base KATSum model.

As can be observed from Table 3.10, the model with the classification process (removing noisy information) can yield a better performance. From Table 3.11, we can observe that the Knowledge Graph component brings contributions to the text summarization, and it can help the model generate better summaries. The Knowledge Graph constructed based on the source text incorporates the complete information. Whereas, some information appears not as important as others, and it may distract the generator from identifying the main idea.

Table 3.4 ROUGE results under the condition of with and without the classification process. The model is BERT-based, where 1/3 of the dataset is utilised to conduct experiments.

Model	ROUGE_1	ROUGE_2	ROUGE_L
KATSum (with classification)	31.7	11.3	28.8
KATSum (without classification)	29.5	10.6	27.4

Table 3.5 ROUGE results under the condition of with and without the Knowledge Graph component. The model is BERT-based, where 1/3 of the dataset is utilised to conduct experiments.

Model	ROUGE_1	ROUGE_2	ROUGE_L
KATSum (with KG component)	31.8	11.2	28.8
KATSum (without KG component)	26.7	9.3	25.2

3.4 AaKOS: Aspect-adaptive Knowledge-based Opinion Summarization

In this section, we present an in-depth description of the proposed Aspect-adaptive Knowledge-based Opinion Summarization (AaKOS) model for tackling the aspect-adaptive opinion summarization task. This task is characterized as a text generation process that utilises inputs in both graph and plain text formats. These inputs are subsequently transformed into graph embeddings via a graph encoder and text embeddings via a text encoder.

In AaKOS, Graph Attention Networks (GATs) [161] are employed as the means to encode these graphs. This is because GATs lie in the ability to effectively capture the complex relationships between nodes in a graph through attention mechanisms by casting the weighted edges into weighted nodes representing the relationships of two connected nodes. In this way, GATs can adaptively focus on more relevant and informative connections while encoding the graph structure.

Data pre-processing is a key step of AaKOS. Given a collection of reviews related to a single product, our objective is to convert the plain text data into Knowledge Graphs and subsequently utilise these relevant graphs for generating summaries. Prior to implementing the proposed model, it is essential to pre-process the dataset through a series of steps, including cleaning the dataset, classifying aspects, and ultimately extracting relevant triplets to construct Knowledge Graphs and recording the corresponding sentences as pseudo summaries. The processed reviews are also used to pre-train a BERT model further, which is used in the subsequent steps.

The architecture of AaKOS is illustrated in Figure 3.4, which comprises two encoders and a decoder. The two encoders consist of a BERT-based text encoder [47] and a GATs-based graph encoder. The former utilises a pre-trained BERT model [47] to generate hidden states from a given set of triplets, which are then employed as part of the decoder input. The latter transforms the filtered sub-graphs into graph embeddings. The decoder is constructed with multi-head attention layers for text and cross-attention layers for integrating text embeddings and graph embeddings. The cross-attention mechanism can be represented by the following equations:

$$Q = W_q E^g, K = W_k E^t, V = W_v E^t, \tag{3.4}$$

$$Attn = softmax(\frac{QK^T}{\sqrt{d_k}})V, \tag{3.5}$$

where W_q, W_k, W_v represent learnable parameters. E^g and E^t denote the embeddings of graph and text, respectively. Additionally, d_k corresponds to the embedding dimension. In the AaKOS model, Graph Attention Networks (GATs) [161] are employed along with an additional global node featuring the product name as the node value. This is mainly due to the outstanding performance in the current setting. Thus, GATs are utilised as the graph encoder in the final design. The attention mechanisms in GATs are computed using Equations 4.3, 4.4, and 4.5.

$$\alpha_{ij} = softmax_j(\beta_{ij}) = \frac{exp(\beta_{ij})}{\sum_{k\in\Gamma(n_i)} exp(\beta_{ik})}, \tag{3.6}$$

$$\beta_{ij} = LeakyReLU(a^T[Wh_i||Wh_j]), \tag{3.7}$$

$$h_i = \sigma(\sum_{j\in N_i} \alpha_{ij} W h_j), \tag{3.8}$$

where $\Gamma(n_i)$ denotes the neighbourhood of node n_i within the graph, while h_i corresponds to the representations of n_i. β_{ij} signifies the importance of n_j with respect to n_i, and α_{ij} represents the normalized attention of n_i across all neighbouring nodes. The final graph embeddings, denoted as $\mathbf{emb} \in \mathbf{R^d}$, are gathered by the additional global node, which connects every node in the graph.

In this model, all reviews are initially converted into knowledge graphs, which are subsequently mapped into a lower-dimension vector space using GATs. The resulting vectors (graph embeddings) are utilised in conjunction with aspect embeddings from a text encoder as input for a decoder to generate summaries.

To regulate the sentiment trend, the Knowledge Graph is designed as a weighted structure, and a parameter called the Weight Controller is introduced. This parameter serves as a threshold for filtering relevant attributes by comparing edge weights to the controller's value. If the edge weight of a triplet is bigger than the controller, this triplet will be selected, and its corresponding text records will be selected as a pseudo summary. Adjusting the weight controller enables the filtering of less-discussed perspectives, thereby highlighting aspects that most people focus on or identifying prevailing trends. For example, in Figure 3.1, the weight controller is set to $wc > 0.2$, excluding attributes with a weight lower than 0.2 from the summary.

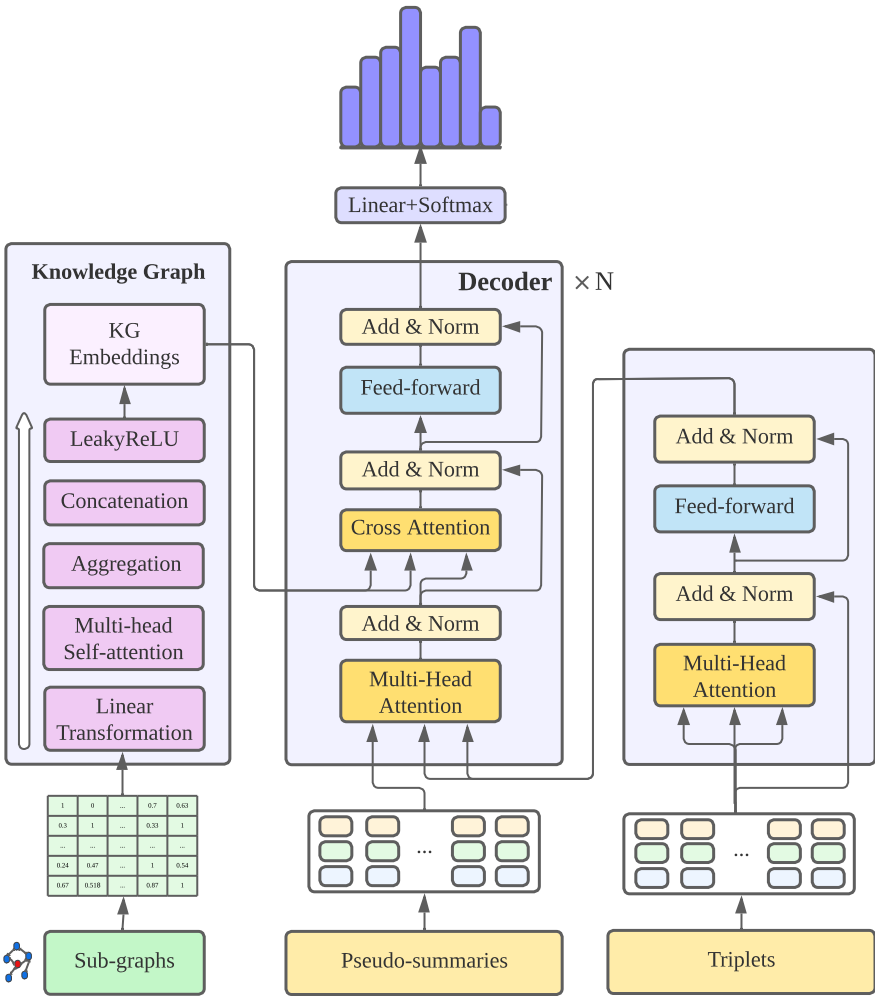


Fig. 3.4 The architecture of the AaKOS model.

3.4.1 Experimental Setup

Datasets

To conduct the experiments, three real-world datasets are employed to evaluate the performance of the proposed AaKOS model, i.e., Amazon Product Review dataset [2], SPACE dataset [8] and YELP dataset [3].

- **Amazon Product Review** dataset is an extensive collection of millions of product reviews from Amazon, encompassing 46 categories. We focus on the top 10 categories with the highest number of reviews. Each review in the dataset includes 15 attributes, such as marketplace, customer_id, star_rating, review_id, and review_date, among

[2] https://jmcauley.ucsd.edu/data/amazon/index_2014.html
[3] https://www.yelp.com/dataset

others. In this work, we retain only the essential features, specifically product_id, review_headline, and review_body.

- **SPACE** (Summaries of Popular and Aspect-specific Customer Experiences) dataset is a comprehensive compilation of "hotel" reviews sourced from TripAdvisor [4]. It features human-written abstractive opinion summaries intended solely for evaluation purposes.

- **YELP** is a dataset composed of businesses, reviews and user data from Yelp. Chu et al. also collect 200 reference summaries from Amazon Mechanical Turk [5] for evaluating and testing [38] their approach on YELP. We evaluate our model using these reference summaries as well.

In addition to the aforementioned datasets, the Amazon Product Review dataset with golden summaries is also used in [28]. Specifically, Bražinskas et al. curated a dataset with gold-standard summaries derived from the **Amazon Product Review** dataset, where 15 products from each of the Amazon review categories, i.e., Electronics, Clothing, Shoes and Jewelry, Home and Kitchen, Health and Personal Care, are sampled. On top of that, 8 reviews from each product are selected to serve as summaries. Three workers were assigned to each product, who were to read the reviews and write a summary text. These summaries are exclusively used for evaluation purposes. We assess our model, trained on our custom pre-processed Amazon dataset, using the corresponding summaries in our experiments.

In order to train our model on these datasets, we first need to pre-process them in accordance with the pipeline detailed in Section 3.4.1. This ensures that our model can be effectively trained using a self-supervised approach.

Data Pre-processing

Figure 3.5 demonstrates the process of data pre-processing. Basically, after extracting aspects from reviews, they are employed in conjunction with the review set to construct Knowledge Graphs and corresponding pseudo-summary sets. Three key stages are elaborated as follows.

Data Cleaning Process. Initially, the dataset undergoes a cleaning procedure, where entries with a review length of less than 100 characters are removed. While this may lead to the loss of some information from the overall review pool, the final performance of our model relies on the processed curated dataset. Any eliminated information

[4] https://www.tripadvisor.com/
[5] https://www.mturk.com

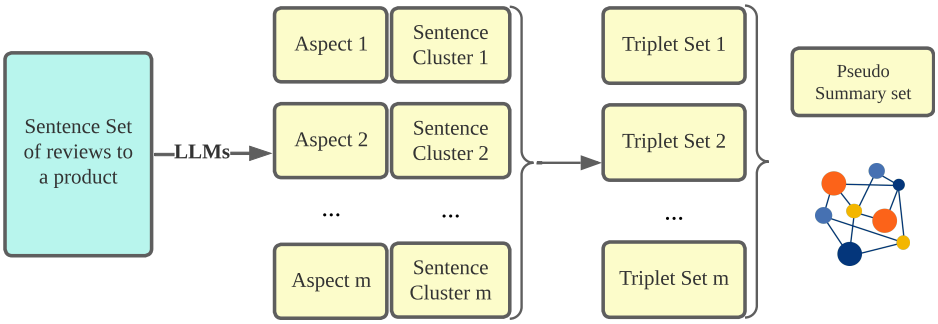


Fig. 3.5 Brief data pre-Processing

is considered non-existent. Subsequently, we leverage product IDs to determine the number of products with more than five reviews, ensuring a sufficient quantity of information for each product. The dataset is then partitioned into smaller subsets based on product categories. The processed review text serves as the foundation for further pre-training a BERT model, which will function as the text encoder in subsequent stages.

Aspect Extraction and Knowledge Graph Construction. The cleaned dataset is comprised of product IDs, product titles, product categories and corresponding review texts. The subsequent steps involve extracting aspects from the dataset and constructing Knowledge Graphs for each product, incorporating aspects and related content. To extract aspects of each product, we design a fine-grained prompt to guide LLMs in extracting key triplets from the texts. Given a set of reviews of a product $R = \{r_1, r_2, ..., r_m\}$, for each review r_i composed of sentences $S_i = \{s_{i1}, s_{i2}, ..., s_{in}\}$, each sentence is classified into an aspect class $A(R) = \{a_x | s_{ij} \in S_i\}$, where $A(R)$ is the final set of various aspects related the given product. For each aspect class a_x, we then cluster all the sentences $S = \{s_{ij} \in \bigcup_{i=1}^{m} S_i\}$ across all reviews $Cluster(a_x) = \{sij | a_x, \forall s_{ij} \in \bigcup_{i=1}^{m} S_i\}$, where $Cluster(a_x)$ is the cluster of sentences related to the given aspect a_x. For each cluster associated with aspect a_x, triplets in the format of $(head, relation, tail)$ are extracted $T(a_x) = \{(head, relation, tail) | a_x\}$ which will subsequently be used to form the weighted knowledge graph $G = \bigcup_{a_x \in A} \{(head, relation, tail) \in T(a_x)\}$ of a product. Edges within these graphs are weighted by calculating the proportion of mentioned attributes, where Figure 3.1 further illustrates the weighted graphs.

Sample Pairs Mapping. Given that these datasets lack gold-standard summaries, it is not feasible to train the model in a supervised manner or evaluate the model using metrics like ROUGE [104]. Therefore, our model is trained using a self-supervised approach. To create the dataset for self-supervised training while extracting triplets from reviews,

we also record the corresponding sentences, which are used as pseudo summaries in the training process. We randomly select k aspects and merge the corresponding triplets into one graph. This graph is subsequently mapped with the pseudo summary, which consists of related sentences. The value of k varies based on the number of aspects of each product.

Evaluation Metrics

To evaluate our model, we use ROUGE [104] to measure the lexical overlap between generated and reference summaries. Rouge_1, Rouge_2, and Rouge_L are reported, representing uni-gram, bi-gram, and longest common subsequence overlaps, respectively.

In addition to ROUGE, we also assess our model using another metric called aspect coverage. This metric employs an aspect-based sentiment analysis (ABSA) task [116] to predict the category and sentiment of extracted opinion phrases from summaries.

Baselines

LexRank [54] is an unsupervised graph-based summarization method. It employs a ranking algorithm to determine node centrality. In LexRank, sentences are treated as nodes to form a graph with weighted edges calculated using TF-IDF. In this chapter, following the settings from [8], we also use BERT [47], and SentiNeuron [133] vectors to calculate the adjacency matrices.

Opinosis [57] is a graph-based summarization framework that generates concise abstractive summaries of highly redundant opinions. It assumes no domain knowledge and leverages mostly the word order in the existing text.

Meansum [38] is an unsupervised neural model for multi-document summarization. It proposes an end-to-end architecture with an auto-encoder, where the mean of input review representations decodes into a reasonable summary review without relying on any review-specific features.

Copycat [28] is a summarization model based on the pointer-generator mechanism [145]. It follows the intuition of controlling the "amount of novelty" during summary generation. With this intuition, they define a hierarchical variational autoencoder model to produce summaries that reflect common opinions.

QT [8] enhances the ability to control the summarization process by leveraging the properties of quantized space to generate aspect-specific summaries.

Table 3.6 The results of the general summarization experiment on the SPACE dataset.

SPACE(general)	Rouge_1	Rouge_2	Rouge_L	Aspect Coverage(F1)
LexRank	29.85	5.87	17.56	0.520
LexRank(SENTI)	30.56	4.75	17.19	0.520
LexRank(BERT)	31.41	5.05	18.12	0.520
Opinosis	28.76	4.57	15.96	0.570
MeanSum	34.95	7.49	19.92	0.610
Copycat	36.66	8.87	20.90	0.676
QT	38.66	10.22	21.90	0.758
AaKOS(ours)	**39.42**	**11.06**	**23.48**	**0.772**
Human Up. Bound	49.80	18.80	29.19	0.845

3.4.2 Experimental Results

In this section, we present the results of our model in comparison with other baselines. We evaluate our method alongside the baselines on both General and Aspect-adaptive Summarization to demonstrate its performance.

Experiment 1: General Summarization

We begin by presenting the results for General Summarization. In this experiment, we employ the full aspect set and do not control the weight, generating a summary that encompasses all aspects of a product. Aspects with low-weight labelled attributes are also included in the summary. A comparison of our model with other baselines on the SPACE dataset is shown in Table 3.6. Considering the Aspect Coverage metric, our model outperforms the other baselines. However, compared to human-written summaries, there remains a significant gap, indicating substantial room for improvement.

The comparison of our model with other baselines on the Amazon dataset is presented in Table 3.7. With the exception of the Rouge_L result, our model surpasses all baselines in performance. The performance comparison between our proposed model and several baseline models on the YELP dataset is detailed in Table 3.8. It is evident that our model demonstrates superior performance, outperforming all other baseline models under comparison.

Although the ROUGE results meet our expectations, the Aspect Coverage does not significantly outperform the other baselines as anticipated. In General Summarization, we apply all extracted aspects, and all relevant content in the Knowledge Graph is utilised, so every aspect should be included in the summary.

Table 3.7 The results of the general summarization experiment on the Amazon dataset.

Amazon(general)	Rouge_1	Rouge_2	Rouge_L	Aspect Coverage(F1)
LexRank(BERT)	31.47	5.07	16.81	0.663
Opinosis	28.42	4.57	15.50	0.614
MeanSum	29.20	4.70	18.15	0.710
Copycat	31.97	5.81	**20.16**	0.731
QT	34.04	7.03	18.08	0.739
AaKOS(ours)	**35.21**	**7.58**	20.04	**0.752**

Table 3.8 The results of the general summarization experiment on the YELP dataset.

YELP(general)	Rouge_1	Rouge_2	Rouge_L	Aspect Coverage(F1)
LexRank(BERT)	26.46	3.00	14.36	0.601
Opinosis	24.88	2.78	14.09	0.672
MeanSum	28.46	3.66	15.57	0.713
Copycat	29.47	5.26	18.09	0.728
QT	28.40	3.97	15.27	0.722
AaKOS(ours)	**30.12**	**5.68**	**20.35**	**0.736**

Two factors may limit the improvement of Aspect Coverage: **1)** aspect extraction: we employ clustering to identify aspects with a pre-defined number of clusters, which is likely to introduce bias. The impact of the number of clusters should also be evaluated; **2)** output length limitation: since we set an output length limit of 256, some information is inevitably omitted after reaching this constraint.

As the first factor is related to data pre-processing, which affects the entire training process and requires a considerable amount of time to complete, we only conducted an experiment on varying the output length to validate our assumption regarding the second factor. Results are displayed in Table 3.9. As the table indicates, our model exhibits a noticeable improvement when the output length is increased from 256 to 512, confirming that our assumption about the second factor is accurate.

Experiment 2: Aspect-adaptive Summarization

In this experiment, we demonstrate the ability of our model to adapt to various aspects of the SPACE dataset. In Table 3.10, we compare six specific aspects individually based on Rouge_L, and average results for Rouge_1, Rouge_2, and Rouge_L are also shown. To

Table 3.9 The results of Aspect Coverage with output length 512.

Models	AC(SPACE)	AC(Amazon)	AC(YELP)
LexRank(BERT)	0.520	0.663	0.601
Opinosis	0.570	0.614	0.672
MeanSum	0.610	0.710	0.713
Copycat	0.676	0.731	0.728
QT	0.758	0.739	0.722
AaKOS(256)	0.772	0.752	0.736
AaKOS(512)	**0.803**	**0.824**	**0.826**
Human Up. Bound	0.845	-	-

ensure that the outputs of other baselines contain only the content of the particular aspect, we use the aspect to filter out the relevant sentences from our pre-processed SPACE dataset, as described in Section 3.4.1. These filtered contents are then used as input for all baseline models. Since our model requires different inputs, we utilise the corresponding graphs and aspects as inputs.

Table 3.10 The results of aspect-adaptive summarization experiment on SPACE dataset.

SPACE(ASP)	Rouge_L						R_1	R_2	R_L
	Building	Cleanliness	Food	Location	Rooms	Service	Average		
LexRank(ASP)	11.56	17.65	11.73	16.64	15.23	18.65	21.46	3.26	15.24
MeanSum(ASP)	15.67	14.03	13.25	19.03	15.67	18.14	21.87	4.48	15.97
Copycat(ASP)	13.28	22.64	15.25	21.59	16.80	24.62	24.35	7.16	19.03
QT(ASP)	15.31	21.38	16.03	22.16	**19.83**	24.38	26.18	7.85	19.85
AaKOS(ASP)	**17.12**	**25.06**	**16.29**	**24.02**	19.58	**26.13**	**28.67**	**8.65**	**21.37**
Human	40.33	38.76	33.63	35.23	29.25	30.31	44.86	18.45	34.58

The result for the “Rooms” aspect falls below the QT’s result, and for “Food”, it only slightly surpasses the best baseline. Our model outperforms all baselines on other aspects and average results, exhibiting significant improvement, particularly on “Cleanliness”. However, a considerable gap between our model and human-written results still exists.

Table 3.11 The general summaries from all models about "Hotel Erwin".

Models	**General Summaries for "Hotel Erwin"**
Human	All staff members were friendly, accommodating, and helpful. The hotel and room were very clean. The room had modern charm and was nicely remodeled. The beds are extremely comfortable. The rooms are quite with wonderful beach views. The food at Hash, the restaurant in lobby, was fabulous. The location is great, very close to the beach. It's a longish walk to Santa Monica. The price is very affordable.
MeanSum	It was a great stay! The food at the hotel is great for the price. I can't believe the noise from the street is very loud and the traffic is not so great, but that is not a problem. The restaurant was great and the food is excellent.
Copycat	This hotel is in a great location, just off the beach. The staff was very friendly and helpful. We had a room with a view of the beach and ocean. The only problem was that our room was on the 4th floor with a view of the ocean. If you are looking for a nice place to sleep then this is the place for you.
QT	Great hotel. We liked our room with an ocean view. The staff were friendly and helpful. There was no balcony. The location is perfect. Our room was very quiet. I would definitely stay here again. You're one block from the beach. So it must be good! Filthy hallways. Unvacuumed room. Pricy, but well worth it.
AaKOS	The Hotel Erwin is an excellent choice for its great and convenient location right on Venice Beach. Rooms were quiet, clean, comfortable and with great ocean views, even though some claimed the noise levels. The rooftop bar provides a great view of the ocean and the experience of watching the sunset. The service of Hotel Erwin is friendly and accommodating. Food and drinks are delicious at fare price.

Case Study

In this section, we present a case study by delving into a detailed comparison of the general and aspect-adaptive summarization capabilities of various models using the reviews of "Hotel Erwin".

Table 3.12 The aspect-adaptive summaries from AakOS about “Hotel Erwin”.

Aspects	**Aspect-adaptive Summaries for Hotel Erwin by AaKOS**
Room and Service	The rooms of Hotel Erwin were quiet, clean, comfortable and with great ocean views. They have a modern and spacious design with comfortable beds. However, bathrooms with sliding glass doors may have privacy concerns. The hotel offers great, friendly and accommodating service with helpful staff.
Building and Location	The Hotel Erwin has a very great and convenient location. It locates right on Venice Beach. The building is decorated newly and very stylish. The rooftop bar provides a great view of the ocean and the experience of watching the sunset.
Cleanliness and Food	The hotel is clean. The rooms are clean, comfortable and well-maintained. The hotel and the restaurant off the lobby provide great food and drinks are delicious with fare price including breakfast, and room service.

In Table 3.11, we demonstrate the general summaries produced by several models, including the human-written summary and the one generated by the proposed AaKOS model. Upon examining the table, it is evident that the summaries provided by the AaKOS model align more closely with the human-written summary in terms of detail and comprehensiveness.

One of the noticeable differences is the perspective from which the summaries are written. Except for the human-written summary and that generated by AaKOS, all other models use a first-person perspective, as indicated by the frequent use of pronouns like “I” and “we”. However, a summary should ideally represent the collective opinion of numerous reviews rather than an individual perspective. This highlights one of the strengths of the AaKOS model, i.e., its text-to-graph transformation process effectively filters out irrelevant nouns, such as personal pronouns, retaining only the ones relevant to the aspects being summarized.

Moving to Table 3.12, we can observe the aspect-adaptive summaries generated by AaKOS for “Hotel Erwin”. These aspect-oriented summaries provide detailed insights into specific features of the hotel, such as “Room and Service”, “Building and Location”, and “Cleanliness and Food”.

A noteworthy observation here is the depth and precision of the aspect-specific summaries produced by AaKOS. They bring out the nuances of different aspects, like the privacy concerns related to bathroom designs under "Room and Service" or the enjoyable experience of watching the sunset from the rooftop bar under "Building and Location." Such focused summaries would be particularly beneficial for potential customers looking for information on specific aspects of the hotel.

This case study shows the effectiveness of the AaKOS model in generating both general and aspect-specific summaries. It effectively condenses broad opinions and provides detailed, relevant summaries. Moreover, its ability to create aspect-oriented summaries proves invaluable in providing potential customers with targeted information, ultimately aiding their decision-making process.

3.5 Conclusions

In this chapter, we present a novel knowledge-aware text summarization model, called KATSum, and a novel approach to conduct aspect-adaptive and knowledge-based opinion summarization, through the development of the Aspect-adaptive Knowledge-based Opinion Summarization (AaKOS) model. The KATSum model employs the Knowledge Graph to improve the quality of the summaries in terms of ROUGE scores. With the knowledge-aware encoder, the input text will be processed by the pre-trained language model and converted into Knowledge Graph embeddings. Such features can help to address unfaithfulness and factual inconsistency. The AaKOS model is self-supervised, training on accurately matched pairs of aspect graphs and pseudo summaries. It proves effective in capturing diverse aspects from reviews and tailoring the summaries to align with the specific requirements of users. To achieve this, reviews are first transformed into knowledge graphs, providing a structured representation of the information contained in the review. If users specify certain aspects they're interested in, the model retrieves the corresponding sub-graphs and leverages these to produce summaries that directly address the desired aspects. The model also introduces a weight controller that aids in accounting for varying sentiment trends, lending a dynamic dimension to the summarization.

Extensive experiments have been conducted to evaluate KATSum using two real-world datasets, i.e., CNN/Daily News and XSum, to evaluate the performance of AaKOS models under both general text summarization and aspect-adaptive summarization tasks, where three real-world datasets are adopted, i.e., Amazon product review, SPACE, and YELP. The experimental results have, respectively, shown that the KATSum significantly outperforms the baselines with pre-trained models, and explicitly demonstrated its superiority over other

baseline models in crafting comprehensive general summaries. These summaries span all aspects of reviews. Additionally, the AaKOS model excels in generating targeted, aspect-specific custom summaries.

The results of KATSum and AaKOS have been published in [167, 168], respectively.

Considering the huge volume of messages generated every day on social networks, text summarization is a powerful tool for distilling key topics, identifying trending themes, and supporting various relevant tasks. By leveraging these capabilities, the methodologies discussed in this chapter provide a foundation for social analysis. This integration can enhance the ability to analyze large-scale social data efficiently, like the Amazon Product Reviews, offering insights into network behaviours, topic trends, and the dynamics of online interactions.

Chapter 4

Maximizing Social Influence With Minimum Information Alteration

With the rapid advancement of the Internet and social platforms, how to maximize the influence across popular online social networks has attracted great attention from both researchers and practitioners. Almost all the existing influence diffusion models assume that influence remains constant in the process of information spreading. However, in the real world, people tend to alternate information by attaching opinions or modifying the contents before spreading it. Namely, the meaning and idea of a message normally mutates in the process of influence diffusion. In this chapter, we investigate how to maximize the influence of online social platforms with a key consideration of suppressing information alteration in the diffusion cascading process. We leverage deep learning models and knowledge graphs to present users' personalised behaviours, i.e., actions after receiving a message. Furthermore, we investigate the information alteration in the process of influence diffusion. A novel seed selection algorithm is proposed to maximize the social influence without causing significant information alteration. Experimental results explicitly show the rationale of the proposed user behaviours deep learning model architecture and demonstrate the novel seeding algorithm's outstanding performance in both maximizing influence and retaining the influence originality.

4.1 Overview

Social media, such as Facebook[1], Twitter[2] and WeChat[3], has been reconsigned as an impartible part of people's daily life. Through social media platforms, people communicate,

[1] https://www.facebook.com/
[2] https://twitter.com/
[3] https://www.wechat.com/

produce and share information arbitrarily without physical and temporal limitations [64]. Meanwhile, social media has also become one of the main sources through which users perceive information. Influences spreading across online social networks exert a significant impact on people's opinions and behaviours [35]. Therefore, a great many research works have been dedicated to a common social phenomenon, i.e., influence diffusion, due to its various applications, e.g., recommendation systems [177, 186], viral marketing [91, 80], epidemics spreading modelling [188], and even in the US congressional election [26].

Influence maximization is to identify a limited set of users from online social networks, expecting that they can spread influences and maximize the impact across the entire network [80]. The initially selected user set is named as *seed set*, and the selecting process is called *seed selection*. Most existing works are developed based on two seminal influence diffusion models, i.e., the Independent Cascade (IC) model and Linear Threshold (LT) model, where the influence is initiated from the affected users, spreading through the network topological structure [36, 35]. However, there are two major limitations for almost all the research works in this field. **First**, influence diffusion is simply treated as a probabilistic-based hopping and infecting process where the users' prior knowledge and the detailed content of influence are ignored. Specifically, almost all the studies focus on the outcome of influence diffusion, i.e., estimating the number of users getting influenced. Likewise, the traditional influence maximization only looks at maximizing the influence spread but ignores the connections between users' knowledge and the influence content. Whereas, in the real world, users tend to perceive and interpret the influence message in different ways, mainly by associating it with their knowledge and background [42]. For example, given the same piece of news entitled *"Nearly 165,000 migrants eligible for fast-tracked residency"*, different people will focus on different aspects. Eligible migrants care about the detailed resident application process; investors could be more interested in the property market due to the prospective increasing demand; local business owners may plan future recruitment. Given such various interpretations and responses, it is important to model the users' reactions to the influence, namely, how they understand the messages and diffuse them. **Second**, in most applications, it is important to remain constant messages to be propagated throughout the entire diffusion process. However, few research works take information alteration into consideration. Using the same example, some people may share the same news and believe this action will "boost the economy", while others claim it may cause a "housing crisis". With such different information alterations, the influence spread chain can be complex, producing many influence variants in the network and deviating from the initial objective of influence maximization. Therefore, it is important to consider the information alteration and take measures to retain the originality of influence messages in the process.

To address the above two limitations, in this chapter, we focus on the influence maximization problem with a key consideration of suppressing the information alteration in the diffusion process, where users' personalised prior knowledge and possible information alteration are taken into consideration. To be more specific, we adopt Knowledge Graphs (KG) to model users' prior knowledge and sub-consciousness. Deep learning models are utilised to model the users' behaviours, which can generate responses/comments that imply potential information alteration. On top of that, A novel seed selection algorithm is proposed, aiming to maximize the social influence without causing significant information alterations. Extensive experiments are conducted, and the results explicitly reveal that the proposed influence-diffusion model can capture information alteration. The proposed novel seeding algorithm outperforms the others in terms of maximizing influence with minimum information alteration.

To summarise, our contributions to this research work can be presented in three folds:

- To the best of our knowledge, this is the first full research work considering the information alteration in the influence diffusion process. We are also the first to aim to achieve influence maximization by retaining the originality of influence.
- We model the complex influence diffusion process by considering users' personalised behaviours and the possibility of information alteration, where Graph Transformer Network is adopted to generate KG embeddings for representing users' prior knowledge, and NLP technologies with deep learning methods are utilised to produce user reactions.
- We propose a novel seeding algorithm, enabling a trade-off between maximizing the influence coverage and retaining the originality of spreading influences.

The rest of this chapter is organised as follows. In Section 4.2, we introduce the related works and briefly describe our proposed information diffusion model with Graph Transformer Network, deep learning methods and NLP models. In Section 4.3, we formulate the problem and explain the equations and some tricks we use during modelling, and we also introduce the dataset, including pre-processing and demonstrating statistics. In Section 4.4, we introduce our diffusion model, including the Graph Transformer Network, which is used to learn the representation of users' knowledge, Multi-head Attention, which is used to combine user knowledge and received messages to generate user feature representations in order to be fed into our NLP language model which is used to generate the response to the received message due to users' knowledge. In Section 6.4, extensive experiments are conducted, and results are demonstrated to evaluate the performance of the proposed algorithm. Conclusion and future works are described in Section 4.6.

4.2 Related Work

4.2.1 Information Diffusion

Many researchers and enterprises have dedicated significant efforts to modelling and learning influence diffusion in online social networks, revealing that predicting information items or cascades has attracted great attention from both academic and business perspectives [80, 101, 190]. Most existing research works have been developed based on two seminal influence-diffusion models, i.e., the IC model and the LT model, where influence is treated as a hopping and infecting process [80, 103]. Based on these two fundamental models, many researchers attempt to tailor them to fit new problems. For example, Saito et al. propose a novel influence-diffusion model, called Asynchronous Independent Cascades and Asynchronous Linear Threshold (AsIC and AsLT), which improves the IC model and the LT model, respectively, by considering time delay on user interactions [140]. Lagnier et al. propose a variant of the LT model, named the Decaying Reinforced User-Centric (DRUC) model, where the information content user's profiles are involved in the process of influence diffusion [88]. Barbieri et al. extend both the IC model and LT model and propose the Topic-aware Independent Cascade (TIC) model and Topic-aware Linear Threshold (TLT) model, which can categorise the content of each message and simulate topic distribution [19]. Liu et al. enhance the LT model and propose a diffusion-containment model, i.e., D-C model, by including the realistic specialities of the containment of the competitive influence spread [107]. Different from IT and LT-based models, Li et al. propose agent-based influence diffusion models, where the influence cascading process is modelled as an evolutionary pattern, driven by individuals' actions [101, 98].

However, almost all the existing influence-diffusion models ignore the detailed contents of the information. Specifically, a few studies only consider the influence topics or general content categories, but these appear insufficient to capture the fine-grained information of an influence. Studies have shown that the users' prior knowledge significantly impacts the acceptance of a message [78, 101]. Therefore, it is important to consider the association between the user's prior knowledge and detailed influence content in the influence spreading.

4.2.2 Influence Maximization

Influence maximization is recognised as a popular NP-hard problem, which has been widely investigated by many researchers [80]. The objective is to select a finite set of users from the networks, expecting they can maximize the impact across the entire network [80]. The identified set of users is called a *seed set*, the process of identifying this set of influencers is

called *seed selection*, and the algorithms used for seed selection are named *seeding algorithms*. The objective of the classic influence maximization problem focuses on maximizing the overall number of influenced users, serving the goal of reaching a large audience, but this cannot be directly applied to some particular situations with a different objective, e.g., how to minimize the rumours. Therefore, numerous research works have been dedicated to the investigation of influence maximization variants. For example, Li et al. model multiple influence diffusion by considering the relationships among various influences [102]. Li et al. study influence maintenance and propose a novel seeding algorithm aiming to sustain long-term influence [98]. Li et al. propose a novel model which aims to suppress negative social influences by injecting other influences [97]. Bharathi et al. investigate competitive influence diffusion by focusing on the scenario when multiple types of messages are competing within a social network [22]. Gershtein et al. propose an IM-Balanced system, allowing users to explicitly declare the desired balance when multiple objectives are presented [59].

Almost no research in the field of influence maximization considers information alternation within the information cascading process. Users always tend to modify the influence messages based on their prior knowledge or attach their opinions before spreading them. Such a critical feature is ignored by almost all the research works. In contrast, in this chapter, we intend to maximize the influence and suppress the alternation of influence.

4.2.3 Influence Diffusion with Deep Learning

Influence diffusion is a naturally complex process, where users spread influences based on numerous factors, e.g., prior commitment level, preference, neighbours' opinions, etc. [101]. Therefore, to model real-world influence diffusion, it is important to model users' prior knowledge and behaviours. Nowadays, deep learning techniques and KG are acknowledged as suitable tools for modelling users' complex behaviours and knowledge structures; thus, they are also adopted to predict the influence diffusion. Yang et al. propose a Neural Diffusion Model (NDM) to generally influence cascade prediction from a microscopic perspective, where deep learning techniques, including attention mechanism and convolutional network for cascade modelling, are employed [182]. Wang et al. propose an attention-based RNN to capture the cross-dependence in the influence diffusion and introduce a coverage strategy to mitigate the attention misallocation problem [173]. Hu et al. propose a memory-based deep recurrent network, i.e., Diffusion-LSTM, to recursively predict the entire diffusion path of an image through a social network [69]. Wang et al. investigate how to use representation learning to assist information diffusion prediction on graphs [171]. Qiu et al. utilise the users' local network as input to a graph neural network with an attention mechanism to learn users' latent feature representation to predict social influence, where the network structures and

user-specific features are taken into consideration [132]. Keikha et al. employ deep learning technologies to learn network embedding by extracting local and global structural information from single and multiple interconnected networks to tackle the influence maximization problem [79]. Similarly, Tian et al. propose a deep influence evaluation model to learn node features and address the topic-aware influence maximization problem, where network structural information, topic and users' interests are used to identify influential users [157]. Cai et al. propose a holistic influence diffusion model by considering users' both cyber and physical interactions and address a new holistic influence maximization problem [31]. Teng et al. employ the Knowledge Graph to model relevant items in a sequence of promotions considering multiple target products in multiple promotions, and users' dynamic perceptions of these product relations [156].

Whereas most existing deep-learning-based influence cascade prediction approaches focused on learning the sharing behaviours in a social network. Namely, "who influenced whom" information is explicitly labelled in the influence-diffusion chain, but ignores how the users perceive the information to produce an influencing message based on users' existing knowledge. Different from these research works, we intend to leverage deep learning and KG to model the users' complex behaviours with a key consideration of information alteration, simulating influence diffusion from a microscopic perspective.

4.3 Formal Definition and Problem Formulation

4.3.1 Formal Definition

Definition 1: Social Network refers a graph $G = (V, E)$, including a set of users $V = \{v_0, v_1, ..., v_n\}$ and a collection of edges among the users $E = \{e_0, e_1, ..., e_m\}$. The cardinality of V and E is represented as $|V|$ and $|E|$, respectively. Meanwhile, user v_i has a set of neighbours $\Gamma(v_i)$, where $\Gamma(v_i) = \{v_j | e_{ij} \in E\}$. If user v_i is identified as a member of the seed set or is activated, v_i's activation state is regarded as active, i.e., $s(v_i) = 1$, and $s(v_i) = 0$ vice versa.

Definition 2: Influence Message is considered one of the concrete representations of social influence, which generally refers to a piece of information sent to or left for a recipient. In the current setting, influence message msg_x denotes a sequence of words, e.g., a tweet posted by the neighbour, which can be delivered from one user to another, potentially affecting the user's opinions or behaviours. $msg_x(v_i \rightarrow v_j)$ describes that user v_i delivers influence message msg_x to user v_j. The representation of msg_x and $msg_x(v_i \rightarrow v_j)$ is denoted as $\mathbf{msg_x}$

and $\mathbf{msg_x}(\mathbf{v_i} \rightarrow \mathbf{v_j})$, respectively. For simplicity, msg_x refers to a message received by v_i and delivered from the neighbour.

Definition 3: Information Alteration, in general, refers to an action from a user, altering the existing information based on the user's knowledge and perception. In this context, information alternation describes user v_i's behaviour, modifying an existing influence message $msg_x(v_j \rightarrow v_i), v_j \in \Gamma(v_i)$ by changing the existing words of msg_x or adding new words to msg_x, producing a new influence message msg'_x.

To explain this further, in the influence-diffusion process, msg_x can be forwarded without any alteration but is also potentially modified by any user before sending it to the neighbours, i.e., msg'_x.

Definition 4: Prior Knowledge is defined as all the knowledge that a person has before learning a topic. Prior knowledge significantly impacts people's perception of influence messages [37]. In other words, given the same message, different users have different understandings due to their prior knowledge [185]. In the current setting, user v_i's prior knowledge K_i is represented as a KG, consisting of a collection of triples extracted from delivering message set $\{msg_x(v_i \rightarrow v_j) | v_i \in \Gamma(v_j)\}$. The embedding of K_i is represented as $\mathbf{K_i}$.

Definition 5: Sub-consciousness refers to the unconscious mind, the knowledge and process that existed well under the surface of conscious awareness, which is theorised to exert an influence on users' behaviours [176]. In this research, sub-consciousness forms a context of how a user perceives the information, user v_i's sub-consciousness S_i is represented as a KG, formed through resolving all the triples extracted from the receiving message set $\{msg_x(v_j \rightarrow v_i) | v_j \in \Gamma(v_i)\}$. The embedding of S_i is denoted as $\mathbf{S_i}$.

Prior Knowledge and **Sub-consciousness** are modelled as Knowledge Graphs in which the knowledge is presented as factual triples (*head, relation, tail*) or (*subject, predicate, object*). Multiple triples form a directed graph. Head/Subject and tail/object are entities extracted from plain texts, which denote the messages given in the dataset. Relation/Predicate refers to a link between two entities, describing the corresponding semantic relationship. For example, given a message *Elbert Einstein was born in German Empire.*, the triplet is represented as (*Elbert Einstein, bornin, German Empire*).

4.3.2 Problem Formulation

Influence Maximization with the Minimum Information Alteration (IM-MIA) problem is defined as the process of maximizing the overall influence coverage, i.e., the total number of influenced users, and minimizing the information alternation in the diffusion process, i.e., attempting to retain the originality of the influence message throughout the cascade. Therefore, we adopt *Alteration-Based Influence Degree (ABID)* as the evaluation metric to measure the influence coverage by considering information alteration. Specifically, the influence coverage presents a positive correlation with ABID, while the alteration positions a negative correlation.

Specifically, in social network $G = (V, E)$, each user v_i has a chance to alter the influence message msg_x received from the neighbours $\Gamma(v_i)$ before delivering it to the neighbours. The information alteration initiated by v_i is based on both prior knowledge K_i and sub-consciousness S_i. Given a finite budget k (seed set size) and social network $G = (V, E)$ with the users' sending and receiving messages, i.e., $\{K_0, K_1, ..., K_n\}$ and $\{S_0, S_1, ..., S_n\}$, the objective is to maximize ABID, which is formulated in Equation 4.1.

$$max\left(|f(SC)| \cdot (1 - \frac{\alpha}{|F(SC)|} \sum_{v_i \in F(SC)} Atr(v_i))\right), \tag{4.1}$$

where $f(SC)$ refers to the influence coverage, i.e., the cardinality of the users who received or spread the influence message, when seed set SC is selected, $|SC| = k$. In contrast, $F(SC)$ is a set of users who spread the influence message, i.e., the users who are activated, $\forall v_i \in F(SC), v_i \in V \wedge s(v_i) = 1$. $\alpha \in [0, 1]$ denotes a parameter, balancing the weight of the information alteration degree. If $\alpha = 0$, the alteration is not considered, while only influence coverage contributes to the objective function. $Atr(v_i)$ is a function, which estimates the alteration degree between msg_x and msg'_x, representing originally diffused message and the message delivered by v_i, respectively.

Furthermore, to measure the alteration degree between two influence messages $\mathbf{msg_x}(\mathbf{v_k} \rightarrow \mathbf{v_i})$ and $\mathbf{msg'_x}(\mathbf{v_i} \rightarrow \mathbf{v_j})$ from user v_i to v_j, the cosine similarity function given in Equation 4.2 is employed.

$$Atr(v_i) = 1 - cos(\mathbf{msg_x}(\mathbf{v_k} \rightarrow \mathbf{v_i}), \mathbf{msg'_x}(\mathbf{v_i} \rightarrow \mathbf{v_j})), \tag{4.2}$$

where $v_i \in \Gamma(v_k), v_j \in \Gamma(v_i)$, and msg_x denotes any influence message received by v_i and delivered from v_k.

4.4 Knowledge-aware Influence Diffusion with Information Alteration

4.4.1 KIAID-based Influence Diffusion Model

To address the aforementioned IM-MIA problem, we propose a novel influence-diffusion model, named Knowledge-aware Information Alteration Influence Diffusion (KIAID) Model. KIAID leverages users' prior knowledge, sub-consciousness and historically interactive behaviours for building a deep-learning-based model, which is capable of predicting the user's behaviours, namely, generating a message or forwarding the message after receiving one from the neighbours.

Figure 4.1 utilises one of the influence diffusion paths as an example, which demonstrates how the KIAID model is applied to the information-diffusion process. In this figure, User v_i is initially activated and successfully activates v_j. As a seed user, v_i delivers the original influence message msg_x to v_j. When msg_x reaches user v_j, if v_j is activated, the KIAID model takes the corresponding prior knowledge K_j, sub-consciousness S_j and msg_x as the input of the KIAID model. Subsequently, influence message msg_j, i.e., the altered influence message produced by v_j, is generated. Thus, the information alteration degree $Atr(v_j)$ can be calculated based on the cosine distance between $\mathbf{msg_x}$ and $\mathbf{msg_j}$ by using Equation 4.2. Similarly, with the input message msg_j, user v_m is activated and produces msg_m by considering both K_m and S_m. The diffusion process of this path ends when the influence message reaches v_n if v_n is not activated successfully or all its neighbours are attempted at least once.

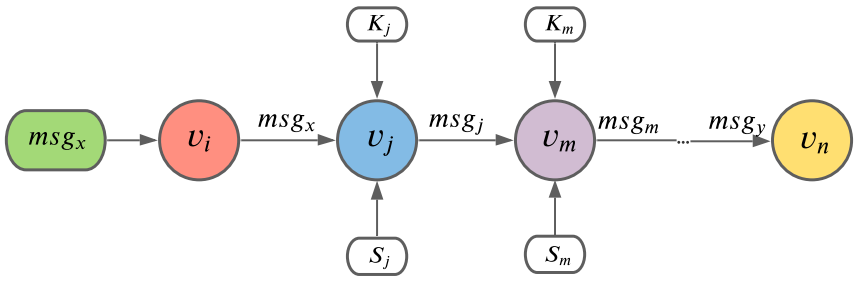


Fig. 4.1 Influence Diffusion with the KIAID Model

The detailed architecture of the KIAID model is presented in Figure 4.2, which consists of three key modules, i.e., User KG (including both prior knowledge and sub-consciousness), Encoder and Decoder. There are two major reasons why KG is applied to model the user's

knowledge. First, KG is capable of presenting the interconnections of the key entities. KG's reliability appears higher than that of plain text. Second, KG can automatically filter redundant data, retaining key information.

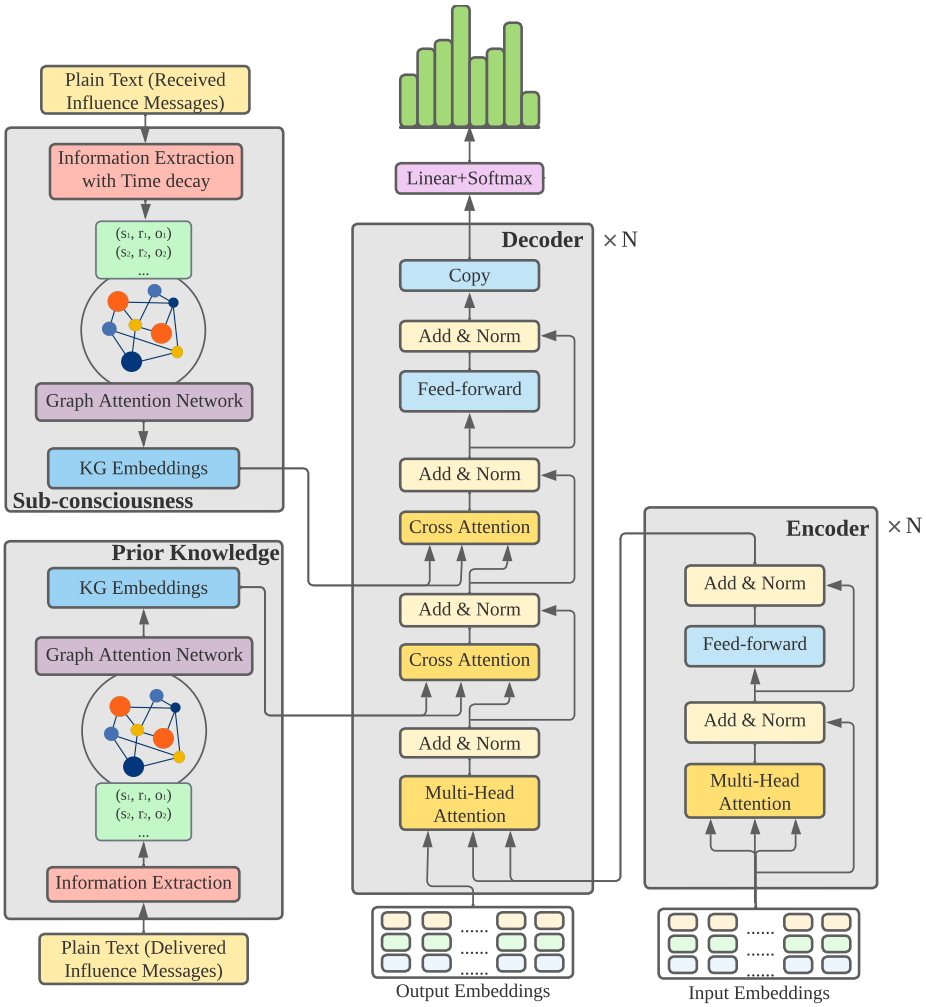


Fig. 4.2 The Architecture of the KIAID Model

Specifically, Sub-consciousness KG and Prior Knowledge KG are generated by following the same approach, i.e., extracting semantic triples and conducting entity resolution, but with different data sources. The former leverages the user's received influence messages, while the latter uses the delivered influence messages. Graph Neural Network (GNN) [141] is adopted to obtain the embeddings of both KGs. Furthermore, we extend the Graph Attention Networks (GATs) [161] by adding two transformer layers, which generate the initial embeddings. The attentions are calculated by using Equations 4.3, 4.4 and 4.5.

$$\alpha_{ij} = softmax_j(\beta_{ij}) = \frac{exp(\beta_{ij})}{\sum_{k \in \Gamma(n_i)} exp(\beta_{ik})}, \tag{4.3}$$

$$\beta_{ij} = LeakyReLU(a^T[Wh_i||Wh_j]), \tag{4.4}$$

$$h_i = \sigma(\sum_{j \in N_i} \alpha_{ij} W h_j), \tag{4.5}$$

where $\Gamma(n_i)$ denotes the neighbourhood of node n_i in the graph; h_i refers to the embeddings of n_i; β_{ij} represents the importance of n_j to n_i; α_{ij} indicates the normalised attention of n_i across all the neighbours. The final graph embeddings, $\mathbf{emb} \in \mathbf{R^d}$, are collected by an inserted global node which connects every node in the KG.

The encoder module has multiple layers, and each layer is composed of multi-head attention, layer normalisation and a feed-forward network, which are formulated in Equations 4.6 and 4.7.

$$\hat{h}^l = LayerNorm(h^{l-1} + MHAtt(h^{l-1})), \tag{4.6}$$

$$h^l = LayerNorm(\hat{h}^l + FFN(\hat{h}^l)), \tag{4.7}$$

where h^0 denotes the first input vector of text; LayerNorm represents the layer normalization operation; MHAtt refers to the multi-head attention; FFN describes the Feed-Forward Network and l here denotes one of the multiple encoder layers.

Inspired by [160], we extend the decoder module by inserting two more attention layers to integrate the text representations and graph embeddings. The output of the decoder is the generated text, i.e., the reaction towards a message based on the incoming influence message and the user's knowledge. There are two scenarios: **(1)** the output is the same as or close to the original text. This will be treated as a forwarding behaviour; **(2)** the output produces different text. We regard the influence message as altered and will be diffused to the neighbours.

To analyse the rationality and validity of the KIAID, we demonstrate the convergence of the model training on two datasets and give a study case in Experiment 1.

4.4.2 Alteration-based Seed Selection Algorithm

There are numerous classic seed selection algorithms for addressing the traditional influence maximization problem, e.g., degree-based, greedy and random selection [80].

To address the IM-MIA problem, we propose a novel seeding algorithm based on the KIAID model, named *Alteration-Based Seed Selection (ABSS)* algorithm, which considers

both node degree and the information alteration degree. The ABSS is described in Algorithm 1.

Algorithm 1 Alteration-Based Seed Selection Algorithm

Input: $G = (V, E)$, k, msg_x, α
Output: Seed set SC, $|SC| = k$

1: Generate a seed set: $SC \leftarrow \emptyset$
2: Generate a candidate set: $C \leftarrow \emptyset$
3: **for** $\forall v_i \in V$ **do**
4: $sum\_Atr := 0$
5: **for** $\forall v_j \in \Gamma(v_i)$ **do**
6: Calculate $Atr(v_j)$ w.r.t. msg_x using Equation 4.2
7: $sum\_Atr := sum\_Atr + Atr(v_j)$
8: **end for**
9: $avg\_Atr(v_j) = \frac{1}{|\Gamma(v_i)|} \cdot sum\_Atr$
10: $C \leftarrow C \cup \{v_i\}$
11: Sort C by avg_Atr in a descending order
12: **if** $|C| > 2k$ **then**
13: Remove the last element from C
14: **end if**
15: **end for**
16: **for** $\forall v_j \in C$ **do**
17: Estimate v_j's ABID using Equation 4.8
18: $SC \leftarrow SC \cup \{v_j\}$
19: Sort SC in terms of ABID in a descending order
20: **if** $|SC| > k$ **then**
21: Remove the last element from SC
22: **end if**
23: **end for**
24: return SC

$$I(v_i, \alpha) = |f(v_i)|(1 - \alpha \frac{\sum_{v_j \in F(v_i)} Atr(v_j)}{|F(v_i)|})/|V| \tag{4.8}$$

In Algorithm 1, the inputs incorporate social network G with users V and relationships E, seed set size k, influence message to be diffused msg_x, and parameter α. The output is the selected seed set SC. Lines 1-2 initialize two sets for carrying seeds and candidate seeds. Lines 3-9 aim to calculate the average information alteration degree of each user's neighbourhood. Line 10 adds the user to the candidate set. Lines 11-15 limit the size of the set by eliminating users who have the lowest average degree of information alteration. Lines

16-23 estimate the ABID of each element in C and identify the top k users as the seeds. Line 24 returns the seed set SC.

Our algorithm measures the alteration degree by considering both the user degree and the average message alteration degree contributed by the neighbours. For each user, only the one-hop neighbour is involved in the computation. Thus, the alteration degree calculation presents a linear time complexity, which is associated with the cardinality of the neighbours, $O(|\Gamma(v_i)|)$. The worst scenario is that a user connects all the others, and the complexity will be $O(|V|)$. The complexity of iterating all the users and corresponding neighbours is $O(\sum_{i=1}^{n} Degree(v_i)) = O(|E|)$, and the worst case is that every node is connected to all other nodes in the graph, which will lead the complexity to $O(|V| * (|V|-1)) = O(|V|^2 - |V|) = O(|V|^2)$. As we select k seeds from $2k$ candidates, the complexity of selection is $O(2k)$. Therefore, the complexity is $O(|V|^2 + 2k)$, where $2k$ is constant, and the final complexity is $O(|V|^2)$.

4.5 Experiments and Analysis

We conducted seven experiments for this research study. Experiment 1 aims to demonstrate the rationality of the KIAID model by analysing the convergence and using case studies. Experiment 2 investigates the information alteration patterns of five different influence messages in the influence-diffusion process. Experiment 3 evaluates the performance of the proposed ABSS algorithm in terms of influence coverage. In Experiment 4, by applying the ABSS algorithm, we further compare the patterns of influence coverage and ABID for different influence messages. Experiment 5 compares the performance of seeding algorithms in terms of ABID. In Experiment 6, we explore the performance difference by varying the parameter α. In the last experiment, we conduct an ablation study to verify the importance and contribution of the Knowledge Graphs.

4.5.1 Dataset Description

In this research, we conduct experiments by using two real-world datasets, i.e., Enron Email[4] and Twitter dataset from Kaggle[5].

- Emails have been proved to be useful for studies in information and social network analysis, especially in detecting hidden organisational structures and selecting influential spreaders [146]. The dataset includes 150 named users and 11,013 edges [83].

[4]https://www.cs.cmu.edu/ enron/

[5]https://www.kaggle.com/datasets/hwassner/TwitterFriends

We assume that email reciprocation implies relationships and that the email contents reflect users' knowledge. The Enron email dataset is pre-processed by eliminating some long advertising emails, and duplicate and short emails. Next, each user's sent and received emails are organised in chronological order.

- The Twitter dataset is widely used as an academic research dataset among researchers. The original dataset contains 40,000 users. Data pre-processing is conducted by eliminating those without friends or followers. Furthermore, users with fewer than 45 posts or fewer than 12 friends are removed. Finally, we construct a Twitter friend graph, having 4,139 nodes and 18,321 edges, with an average of 136 posts.

On both datasets, for each user, we construct two KGs based on their sent/post and received messages, respectively, with Stanford OpenIE [6]. Recall that the KG extracted from a user's sent/post messages is defined as Prior Knowledge, and the KG extracted from received messages is named as sub-consciousness. We assume that the received messages will gradually fade out of the user's memory. Thus, a time-decay function is applied to each user's sub-consciousness KG, which is formulated in Equation 4.9.

$$w_i = 1 - \frac{t_n - t_i}{t_n - t_0 + \beta}, \tag{4.9}$$

where t_i denotes the timestamp of the *i-th* message in the reverse chronological order, t_n describes the timestamp of the last message, t_0 represents the timestamp of the first message, and β refers to a parameter that keeps the weight from being 0.

4.5.2 Experiment Setup

In the experiments, some classic seeding algorithms are used as the counterparts of our proposed ABSS algorithm, including Random selection, Degree-based selection, Greedy algorithm [80], CELF algorithm [92] and IMM algorithm [154]. We compare the overall influence coverage and information alteration of all six algorithms. Meanwhile, we also evaluate each of them by estimating the ABID. Since the diffusion model is IC-based, i.e., probabilistic-based, all experiment results are obtained by calculating the average value over multiple trials.

The weight of the information alteration is controlled by parameter $\alpha \in [0,1]$, aiming to balance the impact of information alteration. $\alpha = 1$ is applied to all the experiments except Experiment 6, in which we showcase the impact of α. The size of the seed set varies from 1 to 21 on the Enron email dataset, and 1 to 101 on the Twitter dataset.

[6] https://nlp.stanford.edu/software/openie.html

To calculate the influence alteration of each user's output message, we adopt a fine-tuned BERT [47] on the dataset to generate a representation vector for each output message and calculate the distance to the original message by using the Cosine similarity function as shown in Equation 4.2.

4.5.3 Experiment 1: KIAID Model Rationality Analysis

The first experiment aims to reveal the validity and rationality of the proposed KIAID model. As mentioned previously, KIAID is a deep-learning-based model that considers the received influence message and user knowledge and generates text as the delivering message. We present the convergence of model training and give a study case.

The model training aims to reduce the loss. Since two datasets are leveraged in this research, the loss and accuracy trends of two models are demonstrated in Figures 4.3 and 4.4. For both Enron Email and Twitter datasets, the KIAID model is trained for 100 epochs. The model evaluation is conducted based on the word prediction task at the end of each epoch using the validation dataset. As we can observe both loss and accuracy trends demonstrate good convergence on both datasets, confirming the model's validity by accurately modelling user behaviour. By achieving a reliable prediction of user reactions, the KIAID model provides a robust foundation for suppressing information alteration during influence diffusion.

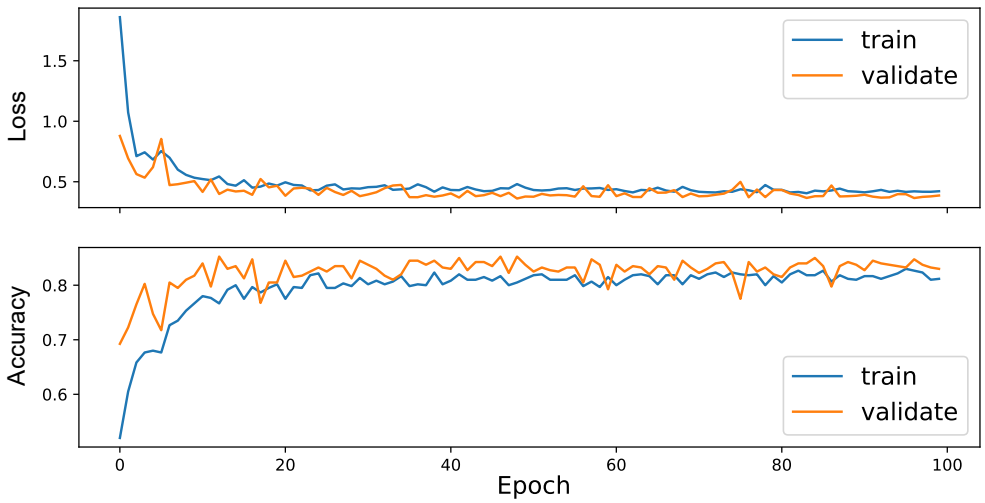


Fig. 4.3 Enron: The loss and accuracy of both training and validating

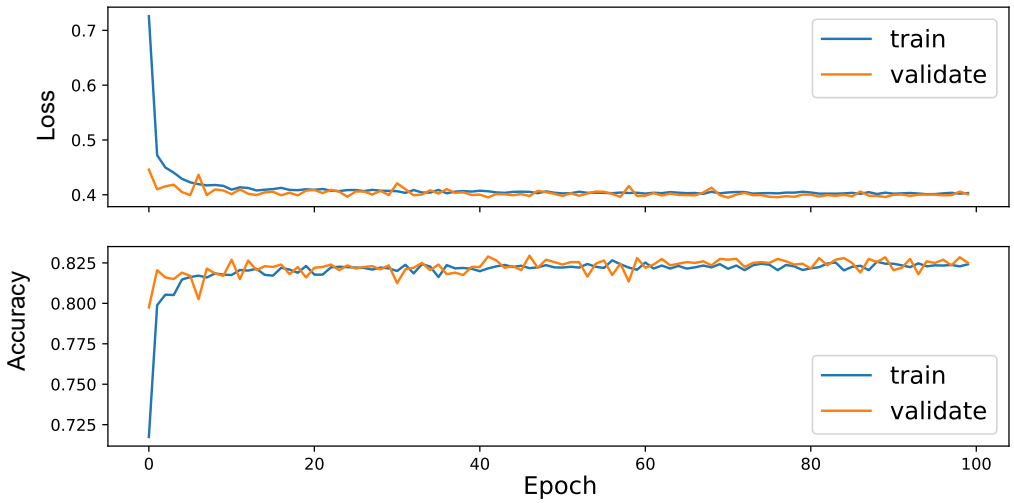


Fig. 4.4 Twitter: The loss and accuracy of both training and validating

Next, we present a study case based on the KIAID model. Figure 4.5 demonstrates an example, i.e., how user Victor reacts when an influence message is received. In this figure, the keywords of the received message, i.e., *extend*, *#465490*, *ENA* and *HPL*, are captured from the received message. Victor provides all the relevant information about entities mentioned in the message. The model not only gives the corresponding response but also corrects the false statement in the received message by indicating the right deal number, *#462490*, between *ENA* and *HPL*. A corresponding part of Victor's KG is also illustrated, and corresponding entities are coloured in blue and red. In this case, this received message will be diffused as Victor's reaction towards the received message.

4.5.4 Experiment 2: Alteration Patterns

Experiment 2 investigates information alteration patterns by employing degree-based and ABSS algorithms on the Enron email dataset. Five different influence messages are picked from the corpus for the influence-diffusion simulation, including meeting arrangement ($msg1$), advertisement ($msg2$), business question ($msg3$), irrelevant message ($msg4$) and personal talking ($msg5$). The details are listed in Table 4.1.

Figures 4.6 and 4.7 show the results after applying the degree-based seed selection algorithm and ABSS algorithm, respectively. As can be seen from both figures, $msg2$ and $msg4$ are altered the least among these five messages, while $msg5$ is altered the most. $msg3$ and $msg1$ sit in the middle, initially showing a rapid upper trend with the increase of k and

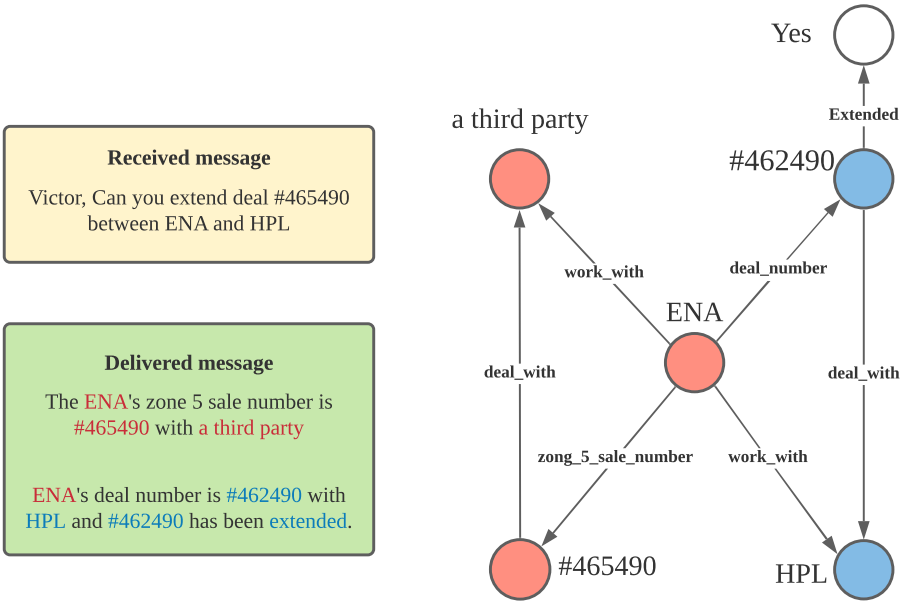


Fig. 4.5 An example of received and delivered messages

then a downside when k reaches 13. This implies that increasing k over a particular threshold can help to retain the originality of the influence message. Meanwhile, the message content impacts the degree of information alteration. People tend to alter the message if it is an open topic, e.g., personal talking but seldom update a fact, e.g., meeting arrangement.

Based on the results in Figures 4.6 and 4.7, we can also observe the proposed ABSS algorithm appears more effective. Given the same influence message, the ABSS algorithm produces a lower average alteration and shows an evident decreasing trend when k increases up to 13. This phenomenon is the most obvious on $msg3$.

These findings provide a better understanding of how different message types influence alteration patterns. Such insights are critical for designing algorithms that effectively suppress information alteration while maintaining a high-influence coverage.

The Twitter dataset does not define a clear type of message, which is different from that of Enron Email. Thus, we randomly select 10 messages for influence propagation and estimate the average alteration degree of these messages using various algorithms. The results are presented in Figure 4.8. As can be seen from the figure, the ABSS algorithm demonstrates strong capabilities of suppressing influence alteration.

Table 4.1 5 messages of different types

	Messages	Types
msg1	Here are the net open Socal border positions we have for Elvis and Cactus. Let's try and set up a conference call with Phillip and John to talk about their offers at the back-end of their curves.	Meeting arrangement
msg2	The legendary Stan Lee, creator of Spiderman and the Incredible Hulk, brings us "Let the Game Begin", episode #1 of 7th Portal – a new series exclusively on shockwave.com.	Advertisement
msg3	Are there behind closed doors discussions being held prior to the meeting? Is there the potential for a surprise announcement of some sort of fixed price gas or power cap once the open meeting finally happens?	Business question
msg4	In this future, you will be able to teleport instantly as a hologram to be at the office without a commute, at a concert with friends, or in your parents' living room to catch up. This will open up more opportunity no matter where you live. You'll be able to spend more time on what matters to you, cut down time in traffic, and reduce your carbon footprint.	Irrelevant content
msg5	I am sorry I have not called to see how you are doing. How are you? How is Olivia and the rest of the family? How is the house coming along? We are doing fine. Meredith is sitting and eating some solid foods. She is at a very fun age. I am going to Chicago tomorrow on business. I am going to have dinner with Esther on Thursday night.	Personal talking

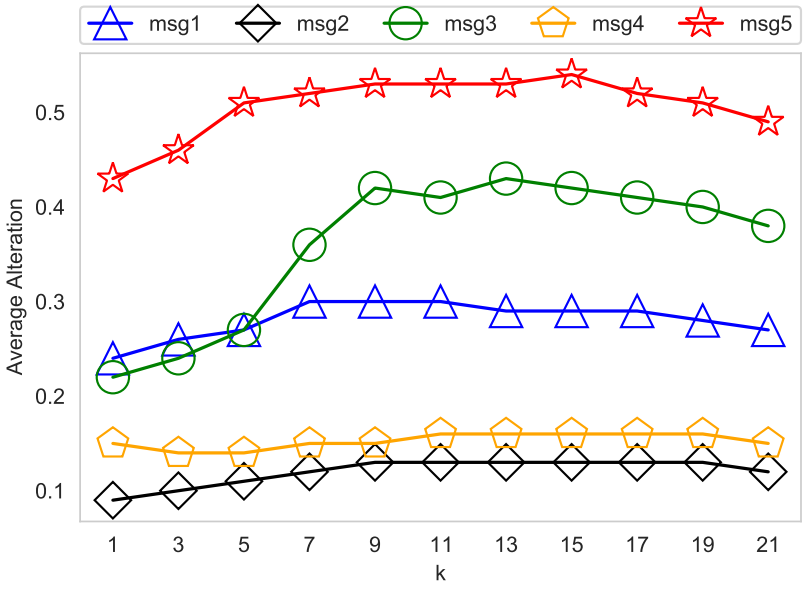


Fig. 4.6 Enron: Average alteration degree of influence messages using degree-based algorithm

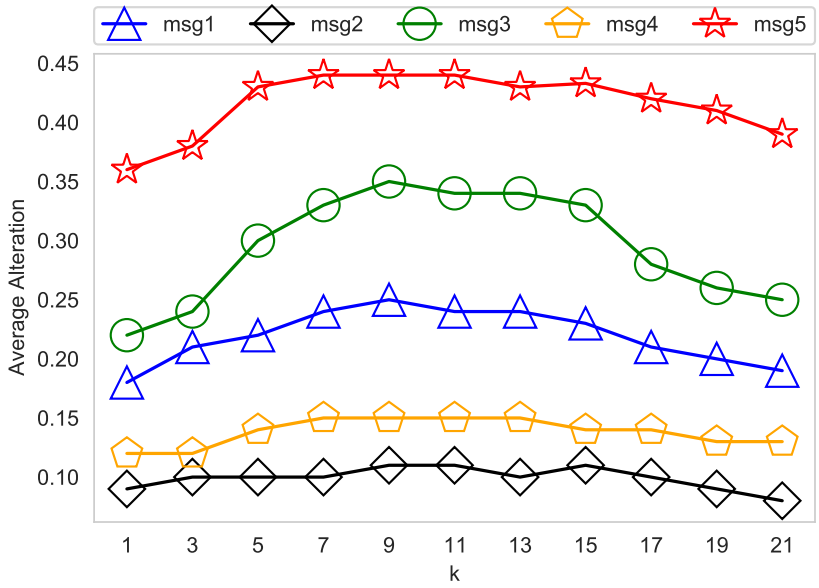


Fig. 4.7 Enron: Average alteration degree of influence messages using ABSS algorithm

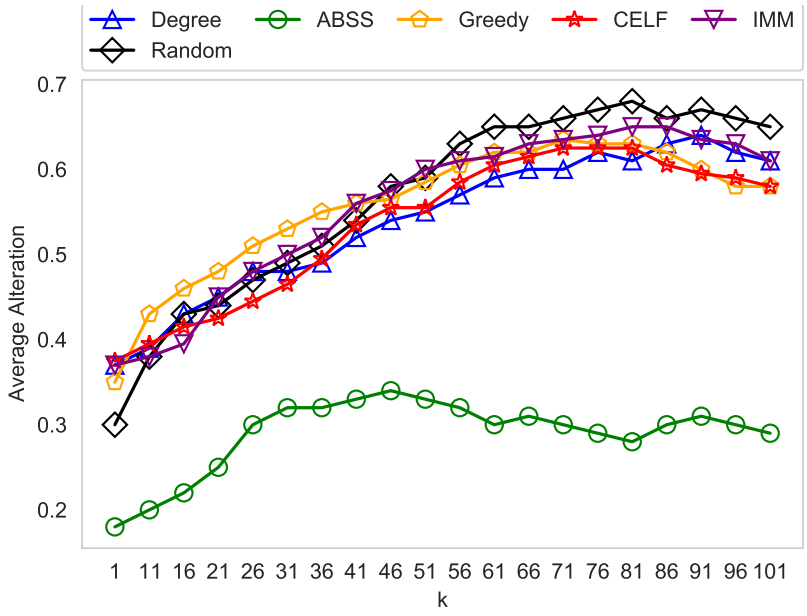


Fig. 4.8 Twitter: Average alteration of six algorithms on 10 messages

4.5.5 Experiment 3: Coverage Comparison

In this experiment, we aim to investigate the performance of the ABSS algorithm in terms of influence coverage and compare it with other classic seed selection algorithms, including Random selection, Degree-based selection, Greedy algorithm [80], CELF algorithm [92] and

IMM algorithm [154]. The experimental results are presented in Figure 4.9 and 4.10. The greedy algorithm performs the best on both datasets, and the proposed ABSS is quite close to it. However, the greedy algorithm is not scalable for large networks [36]. This proves that our algorithm is as stable as the greedy algorithm to select seeds.

One observation is that, when the seed size is small, the influence coverage of ABSS on the Enron email dataset (small dataset) appears a bit lower than most of the other algorithms and converges to the same level as the seed set size k increases. As for the Twitter dataset (large dataset), the influence coverage of ABSS always remains lower than the others at the beginning. However, as seed set size k increases, it also converges to the close level as other algorithms. This is because we also take information alteration into consideration while selecting the seed set. Nodes with high degrees are not always the best choice in our algorithm.

From this experiment, we can see that the ABSS algorithm demonstrated comparable performance to other classic algorithms in terms of influence coverage. This highlights its ability to achieve the dual objectives of maximizing influence while retaining the originality of influence messages, aligning with the goal of optimizing influence propagation in a realistic social network setting.

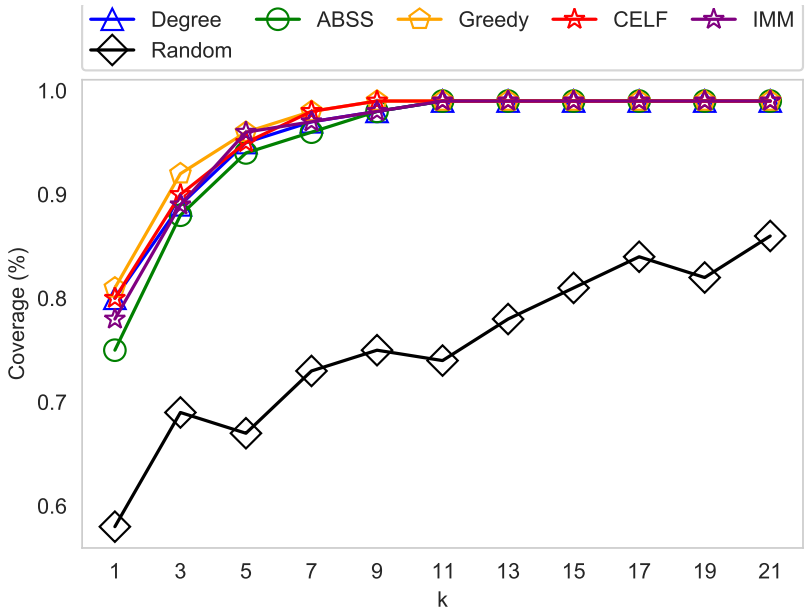


Fig. 4.9 Enron: Average Coverage of six algorithms

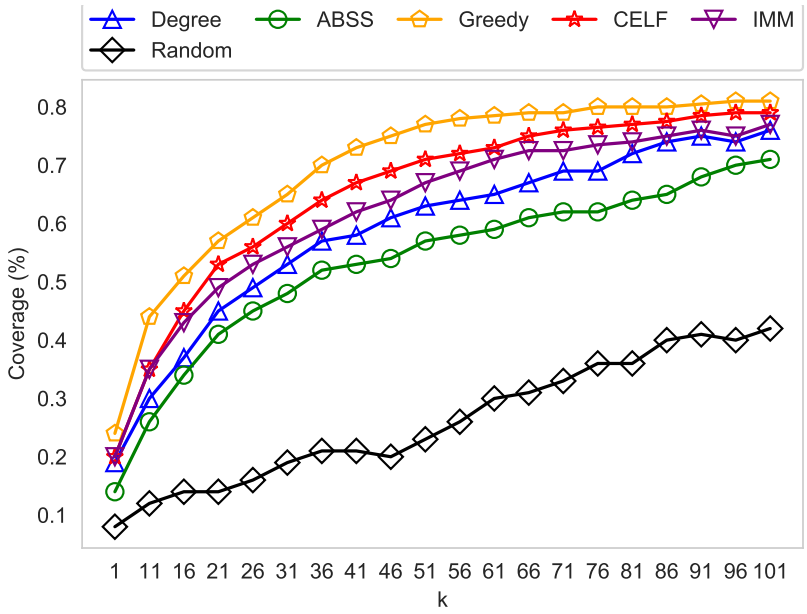


Fig. 4.10 Twitter: Average Coverage of six algorithms

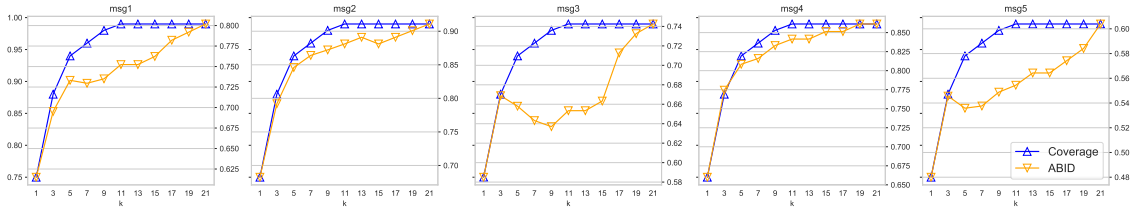


Fig. 4.11 Influence coverage and ABID of 5 influence messages

4.5.6 Experiment 4: Simulation

In this experiment, we simulate the information-diffusion process by using the proposed KIAID model to spread five influence messages. After applying the ABSS algorithm to each influence message, the influence coverage and ABID are calculated and presented.

In Figures 4.11, the ABID values of $msg1$, $msg2$ and $msg4$ rise as k increases. However, the values of $msg3$ and $msg5$ decrease. As can be seen from Figure 4.7 the overall alterations of $msg1$, $msg2$ and $msg4$ appear stable, and there are no significant ups and downs. Based on the observations from these figures, there are two reasons why the ABID declines as the coverage increases. First, the alteration increasing rate appears high while the coverage increasing rate becomes low, as shown in $msg3$ of Figures 4.11 and Figure 4.7, when k increases from 3 to 9, a series of consecutive drops is presented with the rapid increase of alteration. Second, when the influence coverage starts to converge, the ABID diminishes as the alteration rises, even with a minor increase. Based on the observation from these figures,

we can conclude that ABID is very sensitive to information alteration, especially when the coverage is converged, slight information alterations can lead to an obvious reduction in ABID. In Figure 4.11, given $k = 15$, $msg2$ shows an obvious fall when the alteration slightly increases, referring to the corresponding alteration of $msg5$ in Figure 4.7.

However, given a larger seed set size k, the alteration decreases when the influence coverage starts to converge, and the total influence coverage also starts to increase.

This experiment shows that ABID values are highly sensitive to alteration and that slight changes can significantly affect the influence's quality. This reinforces the goal of quantifying the trade-off between influence coverage and information alteration. The ABID metric is a practical tool for assessing this balance, directly addressing the core problem of influence maximization with minimal alteration.

4.5.7 Experiment 5: ABID Comparison

This experiment aims to evaluate the performance of the proposed novel seeding algorithm, i.e., the ABSS algorithm, in terms of ABID. Recall that the evaluation metric ABID measures the influence coverage with a major consideration of information alteration.

Figures 4.12 and 4.13 demonstrate the ABID produced by six algorithms with different seed set sizes on both Enron email and Twitter datasets, respectively. It is evident that the proposed ABSS algorithm outperforms the other classic seeding algorithms.

Given a low budget k, we can see from Figure 4.9 and 4.10, the ABSS algorithm weakens the influence coverage, but appears strong in retaining the originality of the influence message. As the seed set size k increases, our ABSS algorithm performs better and better, and we can observe that when k is bigger than 30, our ABSS algorithm starts to outperform other algorithms.

Meanwhile, as can be observed from Figure 4.12 the greedy algorithm performs even worse than the degree-based algorithm, especially when k increases up to 17. In Figure 4.13, other algorithms also perform very close to the Greedy algorithm. This implies that high influence coverage is not always the best choice for spreading influences since the information can be altered. Whereas, a balance between influence coverage and information alteration becomes the key point.

This experiment shows that the ABSS algorithm outperforms other seed selection algorithms in terms of ABID when seed set sizes are sufficiently large. This demonstrates its efficacy in addressing the Influence Maximization with Minimum Information Alteration (IM-MIA) problem, directly contributing to the goal of balancing coverage and originality in influence propagation.

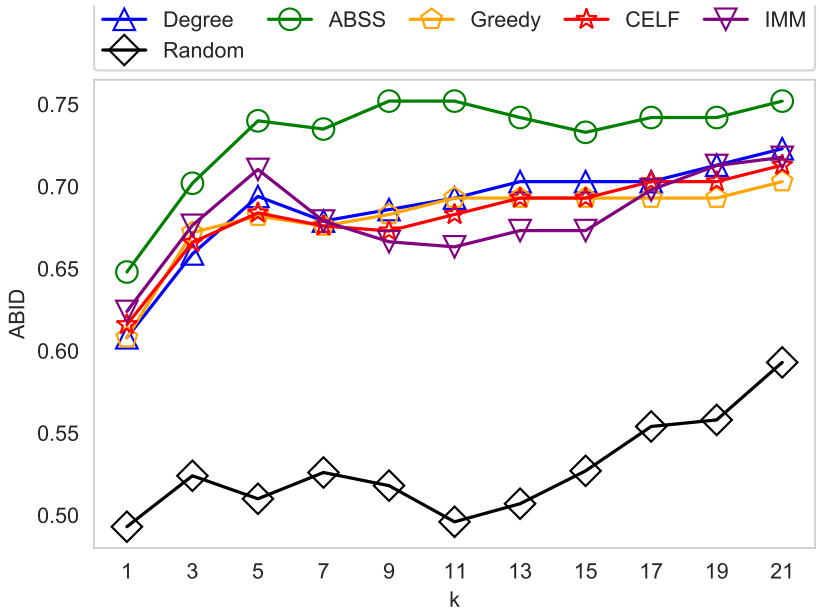


Fig. 4.12 Enron: Average ABID of six algorithms

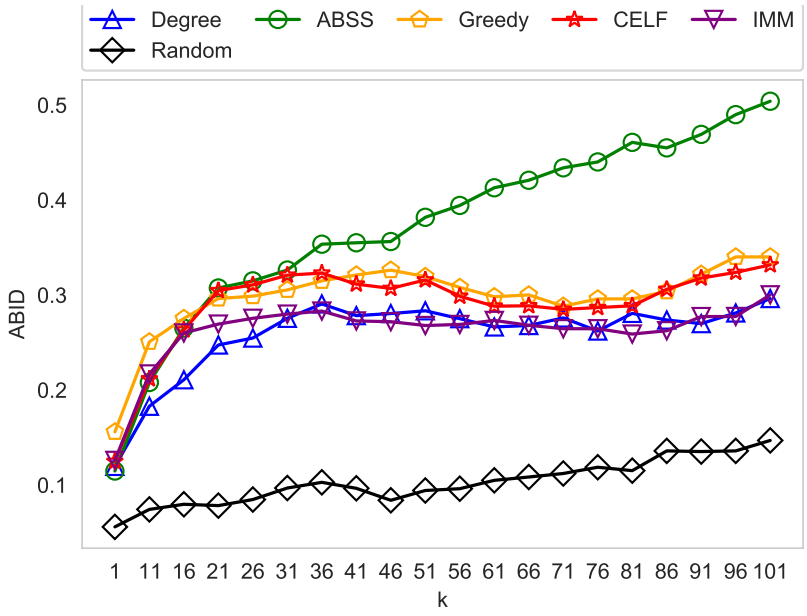


Fig. 4.13 Twitter: Average ABID of six algorithms

4.5.8 Experiment 6: Parameter Analysis

Experiment 6 aims to investigate the impact of the introduced parameter $\alpha \in [0,1]$. We calculate the ABID on all four algorithms with different α values, i.e., $\alpha \in \{0, 0.5, 0.8, 1\}$.

The results on both Enron Email and Twitter datasets are listed in Tables 4.2 and 4.3, respectively.

α is the parameter controlling the strength of alteration while disseminating a message. As can be observed from both tables, a lower α value corresponds to a smaller difference among the algorithms. When $\alpha = 1$, the effect of the alteration is fully considered, having both influence coverage and alteration contributed to the ABID. In this case, our seeding algorithm ABSS outperforms other algorithms, especially when k grows larger. With the increase of α, the ABID appears close to the convergence. When $\alpha = 0$, the alteration is not taken into consideration, i.e., only the influence coverage contributes to the ABID. In this case, the results appear the same as those of coverage. We can also draw the conclusion that a α value corresponds to a higher difference among the seeding algorithms. In this case, our ABSS seeding algorithm performs much better than the others.

This experiment shows that lower values of the parameter α result in reduced sensitivity to information alteration, while higher values amplify its effect. This flexibility underscores the adaptability of the proposed model to varying social network scenarios, showcasing the ability to create a customizable and scalable framework for influence maximization.

Table 4.2 Enron email: ABID of six algorithms with different alpha values

Alpha values	Algorithms	Seed Set Size k										
		1	3	5	7	9	11	13	15	17	19	21
$\alpha = 1$	Greedy	0.608	0.672	0.682	0.676	0.683	0.693	0.693	0.693	0.693	0.693	0.703
	CELF	0.616	0.666	0.684	0.676	0.673	0.683	0.693	0.693	0.703	0.703	0.713
	IMM	0.624	0.676	0.710	0.679	0.666	0.663	0.673	0.673	0.698	0.713	0.718
	Degree	0.608	0.659	0.694	0.679	0.686	0.693	0.703	0.703	0.703	0.713	0.723
	Random	0.493	0.524	0.510	0.526	0.518	0.496	0.507	0.527	0.554	0.558	0.593
	ABSS	**0.648**	**0.702**	**0.740**	**0.735**	**0.752**	**0.752**	**0.742**	**0.733**	**0.742**	**0.742**	**0.752**
$\alpha = 0.8$	Greedy	0.648	0.721	0.737	0.737	0.744	0.752	0.752	0.752	0.752	0.752	0.760
	CELF	0.653	0.713	0.737	0.737	0.737	0.744	0.752	0.752	0.760	0.760	0.768
	IMM	0.655	0.719	0.760	0.737	0.729	0.729	0.737	0.737	0.756	0.768	0.772
	Degree	0.646	0.705	0.745	0.737	0.745	0.752	0.760	0.760	0.760	0.768	0.776
	Random	0.510	0.558	0.541	0.567	0.564	0.545	0.562	0.583	0.612	0.610	0.647
	ABSS	**0.676**	**0.742**	**0.783**	**0.784**	**0.800**	**0.800**	**0.792**	**0.784**	**0.792**	**0.792**	**0.800**
$\alpha = 0.5$	Greedy	0.709	0.796	0.821	0.828	0.832	0.842	0.842	0.842	0.842	0.842	0.846
	CELF	0.708	0.783	0.817	0.828	0.832	0.837	0.842	0.842	0.846	0.844	0.851
	IMM	0.702	0.783	0.835	0.825	0.823	0.827	0.832	0.832	0.844	0.851	0.854
	Degree	0.704	0.774	0.822	0.825	0.833	0.842	0.846	0.846	0.846	0.851	0.856
	Random	0.537	0.607	0.590	0.628	0.634	0.618	0.644	0.668	0.697	0.689	0.727
	ABSS	**0.719**	**0.801**	**0.850**	**0.857**	**0.871**	**0.871**	**0.866**	**0.861**	**0.866**	**0.866**	**0.871**
$\alpha = 0$	Greedy	**0.808**	**0.922**	**0.963**	**0.978**	**0.987**	**0.994**	**0.994**	**0.994**	**0.994**	**0.994**	**0.994**
	CELF	0.803	0.904	0.947	0.976	0.985	0.989	0.993	0.993	0.993	0.993	0.993
	IMM	0.782	0.893	0.962	0.969	0.984	0.987	0.989	0.993	0.993	0.993	0.993
	Degree	0.802	0.887	0.951	0.973	0.982	0.989	0.993	0.993	0.993	0.993	0.993
	Random	0.582	0.694	0.672	0.728	0.752	0.737	0.781	0.814	0.842	0.820	0.859
	ABSS	0.751	0.883	0.942	0.958	0.982	0.989	0.993	0.993	0.993	0.993	0.993

Table 4.3 Twitter: ABID of six algorithms with different alpha values

Alpha values	Algorithms	Seed Set Size k										
		1	11	21	31	41	51	61	71	81	91	101
$\alpha = 1$	Greedy	**0.156**	**0.251**	0.296	0.306	0.321	0.320	0.298	0.288	0.296	0.322	0.340
	CELF	0.125	0.212	0.305	0.321	0.312	0.316	0.288	0.285	0.289	0.318	0.332
	IMM	0.126	0.217	0.270	0.280	0.273	0.268	0.273	0.265	0.259	0.277	0.300
	Degree	0.120	0.183	0.248	0.276	0.278	0.284	0.267	0.276	0.281	0.270	0.296
	Random	0.056	0.074	0.078	0.097	0.097	0.094	0.105	0.112	0.115	0.135	0.147
	ABSS	0.115	0.208	**0.308**	**0.326**	**0.355**	**0.382**	**0.413**	**0.434**	**0.461**	**0.469**	**0.504**
$\alpha = 0.8$	Greedy	**0.173**	**0.289**	**0.351**	0.374	**0.403**	0.410	0.396	0.389	0.397	0.419	0.434
	CELF	0.140	0.239	0.350	**0.377**	0.383	0.395	0.377	0.380	0.385	0.411	0.423
	IMM	0.141	0.244	0.314	0.336	0.342	0.348	0.361	0.357	0.355	0.374	0.394
	Degree	0.134	0.206	0.288	0.326	0.339	0.353	0.343	0.359	0.369	0.366	0.389
	Random	0.061	0.084	0.091	0.116	0.119	0.121	0.144	0.156	0.164	0.190	0.202
	ABSS	0.120	0.218	0.328	0.357	0.390	**0.420**	**0.448**	**0.471**	**0.497**	**0.511**	**0.545**
$\alpha = 0.5$	Greedy	**0.198**	**0.345**	**0.433**	**0.478**	**0.526**	**0.545**	**0.542**	**0.539**	0.548	0.564	0.575
	CELF	0.163	0.281	0.417	0.460	0.491	0.513	0.509	0.523	0.529	0.551	0.561
	IMM	0.163	0.284	0.380	0.420	0.446	0.469	0.492	0.495	0.500	0.519	0.535
	Degree	0.155	0.241	0.349	0.403	0.429	0.457	0.458	0.483	0.500	0.510	0.528
	Random	0.068	0.097	0.109	0.143	0.153	0.162	0.203	0.221	0.238	0.273	0.284
	ABSS	0.127	0.234	0.359	0.403	0.443	0.476	0.502	0.527	**0.550**	**0.575**	**0.607**
$\alpha = 0$	Greedy	**0.243**	**0.442**	**0.568**	**0.652**	**0.728**	**0.767**	**0.785**	**0.793**	**0.804**	**0.806**	**0.812**
	CELF	0.204	0.348	0.532	0.598	0.667	0.713	0.734	0.761	0.772	0.784	0.789
	IMM	0.202	0.348	0.493	0.563	0.620	0.673	0.712	0.725	0.741	0.763	0.770
	Degree	0.192	0.302	0.452	0.532	0.577	0.632	0.652	0.687	0.722	0.748	0.762
	Random	0.083	0.116	0.138	0.192	0.206	0.227	0.301	0.328	0.360	0.412	0.421
	ABSS	0.141	0.264	0.413	0.484	0.532	0.574	0.593	0.624	0.642	0.681	0.714

4.5.9 Experiment 7: Ablation Study

An ablation study to showcase the importance and contribution of the Knowledge Graph components of the model, including prior knowledge and sub-consciousness. Such components impact the quality of the generated output text. For simplicity, we choose three k values to show the results.

Tables 4.4 and 4.5 compare the performance by adding or removing Knowledge Graphs. When they are removed, the ability to suppress the influence alteration of our ABSS algorithm is weakened. On the Twitter dataset, when $k = 21$, the result of ABSS without KG appears lower than Greedy. This shows that our model benefits from the Knowledge Graph rather than from a sole language model.

This ablation study shows that removing the Knowledge Graphs reduce the model's ability to control alteration, showcasing their importance in the model architecture. This aligns with the technical goal of leveraging user prior knowledge and subconsciousness to minimize information alteration. The integration of Knowledge Graphs is a critical component of the proposed KIAID model.

Table 4.4 Abalation study using Twitter dataset

Algorithms	k		
	21	61	101
ABSS (with KG)	**0.308**	**0.413**	**0.504**
ABSS (without KG)	0.287	0.336	0.426
Greedy	0.296	0.298	0.340

Table 4.5 Abalation study on Enron email dataset

Algorithms	k		
	5	11	21
ABSS (with KG)	**0.740**	**0.752**	**0.752**
ABSS (without KG)	0.702	0.723	0.723
Greedy	0.682	0.693	0.703

4.5.10 Discussion

Seven experiments are conducted by leveraging the proposed novel influence-diffusion model, i.e., KIAID, considering the user's knowledge and information alteration. Through these experiments, we demonstrate the advantages of addressing the IM-MIA problem with the proposed ABSS algorithm. The results from these experiments validate the effectiveness of the KIAID model and ABSS algorithm in maximizing social influence while suppressing information alteration.

Based on the experiment results, the following insights can be revealed as follows:

- The results from Experiment 2 prove that our seed selection algorithm considering information alteration can effectively reduce the alteration of the message during the diffusion process.
- From the observation of Experiment 4, we uncover that the alteration-based influence is very sensitive to information alteration. Slight information alteration can cause an obvious reduction in alteration-based influence, especially when influence coverage starts to converge.
- Experiment 5 carries out the result that, the proposed ABSS algorithm can outperform other baseline algorithms in terms of ABID.

By conducting this research, we have come to understand that the scale of social networks raises concerns about the applicability of our model. As the language model is trained on

messages, the sheer size of the network can result in an increase in training data. Consequently, the simulation's runtime may also increase as the network expands. However, by designing the language model training and individual view-based simulation of information diffusion separately, we can mitigate the impact of large-scale networks.

4.6 Conclusion and Future Works

In this chapter, we focused on the problem of Influence Maximization with Minimal Information Alteration. A novel influence maximization approach, which can measure the extent to which the users in a social network can change the messages that they receive, was proposed. In the proposed approach, we use Knowledge Graphs to represent users' knowledge and train a model to generate the users' reactions toward the received messages. In addition, a novel Knowledge-Aware Information Alteration Diffusion Model was proposed. By applying this model to the IC-based diffusion process, we also propose a novel Alteration-Based Seed Selection algorithm to tackle the Information Maximization problem with a key consideration of suppressing information alteration. We further conduct extensive experiments to demonstrate the advantages of our model and algorithms. Experimental results show that compared with other classic seed selection algorithms, our algorithm can effectively increase the influence coverage and retain the originality of the influence message. The results of this chapter have been published in [166].

Chapter 5

Misinformation Detection with Deep Learning and Framing Theory

The proliferation of misinformation jeopardizes social cohesion, distorts the truth, and destabilizes democratic processes. Despite numerous studies on detecting misinformation within online social networks, much of this misinformation is derived from factual information but presented in misleading ways, implying meanings that differ from literal interpretations and thus leading readers astray. Such instances pose a significant challenge for classification since their textual features are very similar to those of truthful information.

In this chapter, We delve into the rapidly evolving challenge of misinformation detection, with a specific focus on the nuanced manipulation of narrative frames — an under-explored area within the AI community, by proposing deep-learning-based models to detect misinformation originating from accurate facts portrayed under different frames. The potential for Generative AI models to generate misleading narratives underscores the urgency of this problem. Drawing from communication and framing theories, we posit that the presentation or 'framing' of accurate information can dramatically alter its interpretation, potentially leading to misinformation. In particular, the intricate user interaction in social networks plays an important role in this process, as these platforms provide an unsupervised environment for the dissemination of misinformation. Through real-world examples, we demonstrate how shifts in narrative frames can transmute fact-based information into misinformation.

Our innovative approach leverages pre-trained Large Language Models and deep neural networks to detect misinformation from accurate facts portrayed under different frames. These advanced AI techniques offer unprecedented capabilities in identifying complex patterns within unstructured data, critical for examining the subtleties of narrative frames. Our objective is to bridge a significant research gap in the AI domain, providing valuable

insights and methodologies for tackling framing-induced misinformation, thus contributing to the advancement of responsible and trustworthy AI technologies.

Two novel models, FrameTruth Model (FTM) and Framed Element-based Model (FEM), are proposed to address the challenges mentioned above. FTM leverages Large Language Models to extract the framing of the information, incorporating this as an important feature in the process of misinformation classification. FEM investigates how framing elements affect misinformation detection, treating each element as a separate feature for the language model to process. Our research systematically examines these elements, offering important insights into the subtle ways information can be skewed using framing.

We evaluate both models by comparing their performance against popular baselines on real-world datasets, demonstrating their superiority in classifying misinformation derived from facts. We also examine the impact of framing theory elements, proving the rationale for applying framing theory to enhance misinformation detection performance.

5.1 Overview

Misinformation in today's media landscape is growing substantively, where fake news and false or misleading information are disseminated through various media channels, e.g., news websites and online social media platforms that most people frequently use and consume information [73]. Artificial Intelligence (AI) has advanced from exclusively understanding language to Generative AI (GAI) models that can automatically generate articles, posts, and narratives with remarkable sophistication[111]. The accessibility of GAI models such as ChatGPT has expedited the process of creating manipulative misinformation. In most cases, it can be difficult for readers to distinguish whether the narrative was written by a GAI model or a human author [85].

Automatically identifying incorrect facts or claim validation is a well-researched task that has achieved high accuracy results, particularly with traditional misinformation detection focused on keywords [134, 155]. However, identifying misinformation becomes challenging when accurate facts are manipulated through biased frames to create misleading narratives. Manipulating through biased frames involves selecting and highlighting certain aspects of information while omitting others, presenting an alternative perspective that can distort the original message [51]. This manipulation of accurate information by changing the perspective and frames can lead to the propagation of misinformation, making framing a critical factor in misinformation detection. Investigating strategic framing and its role in online information dissemination is essential to address this issue [142].

The framing theory has been recognised as an important concept in the communication fields, explaining how the presentation of information can influence an individual's perception and interpretation of that information [144, 55, 52]. It illuminates the process by which communicators strategically highlight specific facets of a perceived reality within a communication text. In the context of misinformation, framing theory suggests that the way information is presented or framed can be used to manipulate individuals into accepting false or misleading information as accurate. Framing involves the selection of some factors about an issue or event, making them salient, or emphasizing these factors over other factors. It is about selecting and deciding which parts of a situation or event to make salient to an audience. Salience can be achieved by bringing the factors to the audiences' attention through their typographic appearance and layout, such as headlines and bold text; or the way their discussion is carefully crafted through the communicating text; or by the deliberate omission of these factors and the skilled manipulation of the logic of the arguments presenting the facts of the situation. It also suggests how information is presented and communicated in a narrative - the story that communicates the facts in a meaningful way - can influence an individual's perception and interpretation of that information and is recognized as an important concept in the communication and social science fields [51, 52, 61, 144, 55].

Despite identifying specific frames, framing theory also suggests that four frame elements contribute to how information is presented: problem definition, causal interpretation, moral evaluation, and treatment recommendation [52]. Specifically, the problem definition defines the problem by determining the actions of a causal agent along with their associated costs and benefits. It is measured by what is culturally acceptable, while the causal interpretation identifies which forces cause the problem. The moral evaluation makes moral judgments on the causal agent and their effects, with the treatment recommendation offering suggestions to solve the problem and the possible effects these might have. When it comes to misinformation, framing theory suggests that the manner in which information is conveyed or framed can be harnessed to persuade readers into embracing inaccurate or misleading information as truthful. By strategically highlighting specific facts or interpretations while purposefully excluding others, individuals or organizations can craft a specific narrative that aligns with their agenda to mislead the audience [51]. By selectively emphasising certain facts or elements, individuals or organisations can frame a particular narrative in a way that suits their agenda and misleads the audience.

There are numerous research studies concerning the detection of framing, the identification of elements within a frame, and the analysis of framing itself [52, 164, 158]. While there are few existing research works on framing detection and framing analysis, very few studies explore how frames impact the emergence of misinformation or which frame element has the

greatest impact on the overall frame. It is challenging to classify misinformation stemming from factual information. Therefore, misleading information created by manipulating the frame of a truthful narrative would be undetected by traditional misinformation detection models. The key challenge is that it is very difficult to classify the misinformation stemming from the correct information by alternating the corresponding frames or framing elements. For example, an excerpt of a factual information narrative with a political frame:

"The proposed three waters reform program harks back to the Havelock North water contamination event in 2016... The government estimates that we'll need a mind-boggling $120 billion to $185 billion over the next 30 years... The government believes that four entities, aggregating all the water services across the country, offer the best and quickest opportunity to achieve the desired improvements... The review was expanded to cover all three waters, and this acknowledges the inter-relationships between the three networks."

Information is presented in a straightforward and factual manner, explaining the motivation behind the reform, the expected costs, and the time frame, describing the government's belief that larger entities can achieve efficiency gains, and understanding why all three water networks were reviewed. However, an excerpt of a misleading narrative with a semantic frame that uses specific terms to associate the statement with other communication contents or features, including irony, lettering, metaphor and so on [143]. The following example shows the satire/irony which suggests the opposite of the original message:

"Because nothing says 'clean water' like shifting responsibility from local government to some fancy-sounding entity, right?... They even established a drinking water regulator to ensure everything meets regulatory standards because we all know how important it is to regulate things, right?... Because who needs small, local councils when you can have these big entities making all the decisions for you? Efficiency gains are just a bonus, my friends!... Because why bother keeping it simple when you can add some unnecessary complexity?"

Satire, oversimplification, and selective framing are used to mislead as it mocks the idea of clean water as a priority, ignoring the serious health concerns that prompted the government to consider these reforms, downplays the significance of regulatory standards by sarcastically framing them as if they are unnecessary. At the same time, the actual cost estimates are not addressed seriously, dismissing the efficiency gains, oversimplifying the government's rationale for proposing larger entities to handle water services and sarcastically dismissing the complexity of reviewing all three waters, suggesting that it is unnecessarily complicated.

Pre-trained Large Language Models (LLM) and deep neural networks have been acknowledged as efficient and effective techniques to address the framing classification and

misinformation detection problem since they can learn from unstructured data and identify complex patterns that are difficult to detect using traditional methods [73].

Our hypothesis is that news or articles on the same topic can be converted into misinformation when given different frames, and in this chapter, we use the frame of a narrative and the frame elements as key considerations in the process of identifying misinformation.

To address this challenging issue, in this chapter, we propose and develop two framing theory-based models, FramedTruth Model (FTM) and Frame Element-based Model (FEM), for identifying misinformation stemming from portrayed facts under different frames. The primary goal of the first model is to detect misinformation that has been intentionally reframed by individuals or entities with the aim of deceiving or misleading their audience. The second model is the first full research work, addressing the framing-based misinformation detection issue, investigating how framing elements affect misinformation detection, and treating each element as a separate feature for the language model to process. Our research systematically examines these elements, offering important insights into the subtle ways information can be skewed using framing. This also enhances the accuracy and effectiveness of detecting misinformation.

To assess the performance of the proposed models, we evaluated them against several well-known baselines using real-world datasets. Experimental results highlight their superior capability in classifying misinformation manipulated from factual content. Additionally, we perform ablation studies to confirm the contributions of key elements and analyse the effects of parameter selection. We also provide case studies for a qualitative evaluation of the proposed models.

Our contributions of this chapter include:

- We formally define misinformation that is portrayed from the facts and formulate the misinformation detection problem in the context of Generative AI.
- We propose a novel model called FrameTruth Model (FTM) and Framed Element-based Model (FEM), which can effectively identify misinformation stemming from portrayed facts under different framing. To the best of our knowledge, this is the first full research work, tackling the framing-based misinformation detection problem.
- We are the first to investigate how framing elements affect misinformation detection, treating each element as a separate feature for the language model to process. Our research systematically examines these elements, offering important insights into the subtle ways information can be skewed using framing. This also enhances the accuracy and effectiveness of detecting misinformation.

5.2 Related Works

With the rise of social media, the ease with which information can be distributed and consumed has increased, allowing misinformation also to increase [73], leading to significantly increased attention from both researchers and practitioners.

5.2.1 Traditional Misinformation Detection

Traditional misinformation detection focuses on incorporating user-based and content-based approaches. Specifically, user-based or context-based strategies analyse the social environment surrounding misinformation, focusing on user attributes and behaviour, while content-based approaches delve into textual and emotional facets of content [150].

The user-based approaches include the extraction of explicit and implicit features from user profiles [148] and the focus on status-sensitive users to facilitate early detection [110, 114]. For example, Hamdi et al. propose a novel methodology for evaluating the credibility of information sources on Twitter, utilising node2vec to extract features from Twitter's follower and followee graph and incorporating user-specific attributes [66].

Modern content-based strategies have resulted from combining tensor-based article modelling with semi-supervised learning [2], leveraging transformer-based language models [127], and hybridising Convolutional and Recurrent Neural Networks [120]. While these works provide valuable insights into misinformation detection, they primarily address misinformation through individual user characteristics and isolated content features, often overlooking the broader narrative structure or frames via which the content is portrayed.

Moreover, there is also traditional rule-based misinformation detection for fact-checking, and fake news is focused on detecting misinformation by paying attention to who provided the information or what the content of the information was. Manual fact-checking relied on the author's reputation and/or the source to determine the veracity of the information [65]. Similarly, to detect fake news on social media, the social contexts, such as explicit and implicit features of user's profiles, are evaluated to determine the credibility of the information [150]. In addition to social contexts, fake news detection focuses on the content of the text by extracting linguistic features in order to detect sensational headlines that are frequent in fake news [150]. Moreover, identifying negation keywords, such as 'no,' 'not,' or 'never,' played a significant role in enhancing the classification of rumours [87]. Traditional rule-based approaches relied on information specific to the topic to correctly identify misinformation. Therefore, these approaches experienced limitations when detecting misinformation about a new topic [163]. These shortcomings were addressed with the introduction of semi-supervised and unsupervised methods [125].

Furthermore, these traditional methods experience limitations in dealing with misinformation derived from factual events but framed to convey alternative implications. This is particularly challenging with lengthy articles that contain a mix of truthful and misleading information. Unlike these conventional approaches, our proposed method advances the field by incorporating 'framing theory' as a critical factor in misinformation detection. FTM analyses not only the content but also the underlying narrative, which is important in understanding how information can be skewed by framing to mislead. This represents a significant advancement, offering a more nuanced and comprehensive approach to tackling the complex nature of misinformation in today's information landscape.

5.2.2 Deep Learning Based Misinformation Detection

Many researchers have explored the use of deep learning techniques to automate misinformation detection, such as tensor and transformer-based models and convolutional and recurrent neural networks [73, 2, 120, 127, 130]. Latent patterns and spatial context were extracted from tensor-based models to construct k-nearest-neighbour graphs and belief propagation for semi-supervised misinformation detection [2]. Liu and Wu propose a novel deep neural network composed of a status-sensitive crowd response feature extractor, a position-aware attention mechanism, and a multi-region mean-pooling mechanism, addressing the early detection of fake news on social media [110]. A hybrid of convolutional neural networks (CNN) and recurrent neural networks (RNN) leverages the strengths of CNN in extracting local features and of RNN in capturing long-term dependencies to detect fake news [120]. Another RNN model found that combining sentiment, emotional, irony and hate analysis with bagging, boosting, stacking and voting means produced a higher accuracy than without the various analyses [130]. An evaluation of transformer-based models Large Language Models, namely, BERT variants to be used as baselines for misinformation detection, can achieve comparable or better performance than more complex state-of-the-art methods [127]. More recently, a transformer-based model, MisRoBÆRTa, utilised RoBERTa and BART to outperform single transformer misinformation detection models [159]. Finally, a hybrid deep learning model integrating features-based models and universal sentence encoding revealed promising results on the PHEME dataset [6].

While these techniques are able to accurately detect misinformation without considering the narrative or frame, their challenge lies in dealing with misinformation stemming from factual events that are skewed to convey a different implication. Furthermore, they also face difficulties handling lengthy news articles that potentially contain both truthful and misleading information.

5.2.3 Framing Theory

Moreover, news frames significantly impact reader interpretation. The frame of a piece of text can increase the salience of specific parts of information, i.e., to make information more meaningful, noticeable, or memorable [52, 74]. An example by Entman showed that a frame can influence how a large portion of readers notice, understand, remember, evaluate, or act upon information presented to them [52]. Furthermore, the problem definition, causal interpretation, moral evaluation, and treatment recommendation are also demonstrated as the four identifiable elements of a frame [52]. Multiple methods have been developed to detect frames using different approaches. Liu et al. develop a neural network-based approach for detecting frames from news based on the article headlines through fine-tuning a Large Language Model, i.e., BERT, where one prominent public affairs issue in the US, i.e., gun violence, is focused [106]. Alternatively, Walter and Ophir leverage computational tools to develop a novel method, the Analysis of Topic Model Networks, for the inductive identification and categorisation of frames, demonstrating its effectiveness across diverse U.S. news corpora, thus offering potential theoretical, methodological, and practical implications for framing research [164]. Similarly, Arendt et al. adopt the reinforcing spiral framework and a mixed-methods approach to explore the underlying mechanism of the news-framing effect [10]. Cabot et al. finetune a joint method based on RoBERTa including metaphor, emotion, and frames defined by Card et al. to model political discourse [30, 33].

Despite the existing research works on frame detection and analysis, very few studies investigate how frames impact the emergence of misinformation [106, 74]. The key challenge in detecting framing-induced misinformation lies in classifying misinformation stemming from accurate facts but presented under different frames. Inspired by the framing theory from communication studies [52], we integrate this theory in the proposed FTM, enabling it to distinguish the misinformation portrayed from the actual facts.

Although both frame and framing element detection are possible, the impact of frames on misinformation detection requires further research. Our proposed FEM explores this impact by incorporating the framing theory presented by Entman to solve the earlier challenge of detecting misinformation stemming from accurate facts that are skewed to be misleading potentially [52]. Additionally, FEM discerns the respective contributions of the four framing elements to the overall accuracy of misinformation detection.

5.3 Preliminary

In this section, we give formal definitions, and the problem of detecting misinformation portrayed from the facts is also formulated.

5.3.1 Formal Definition

Definition 1: Narrative generally refers to a way of sharing stories or information, whether it is spoken, written, or shared online. In the current context, the narrative indicates the news stories and articles that are being disseminated online. Let $N = \{n_1, n_2, \ldots, n_n\}$ denote the set of narratives, a narrative can be information or misinformation in online social networks, where n_i represents a single narrative.

Definition 2: Information refers to the presentation of facts in a way that aims to convey these facts accurately. The narrative of the information is constructed to reflect its true nature and implications without distorting or omitting key elements. Let $I \in A$ refer to the information set:

$$I = \{(fa, n) \mid fa \in FA \wedge n \in N\} \tag{5.1}$$

where $fa \in FA$ refers to a specific fact in a fact set, $n \in \mathcal{N}$ refers to a narrative of a narrative set for information.

Definition 3: Misinformation, converse to information, involves using the same facts but framing them within a narrative that is designed to mislead, deceive, or manipulate the audiences. The key aspect of misinformation in this chapter is not the distorted facts but how they are presented in a misleading narrative. Let $M \in A$ represent a set of misinformation which is composed of the content and the specific narrative:

$$M = \{(fa, n) \mid fa \in FA \wedge n \in N\} \tag{5.2}$$

where $fa \in FA$ refers to the same fact in a fact set as information, $n \in N$ refers to the narrative of a narrative set for misinformation.

Definition 4: Frame suggests how information is structured and presented in a story, including the perspective from which it is told. It suggests how information is presented in a narrative - the story that communicates the facts in a meaningful way - can influence an individual's perception and interpretation of that information and is recognized as an important concept in the communication and social science fields. Mathematically, f represents a frame, and the set of frames is $FR = \{fr_1, fr_2, \ldots, fr_m\}$. The relationship between a frame and a narrative of one article can be represented by $R : N \rightarrow FR$, where $R(n_i) = fr_i$, and R represents the element extractor.

Definition 5: Frame Elements are the specific components used to construct a frame in articles. A frame is generally composed of four elements, and they constitute how information should be displayed in front of the readers and how the readers would perceive the content. Each article has four elements: "problem definition", "causal interpretation", "moral evaluation", and "treatment recommendations". Let e represent one of the elements of a frame in an article and $\mathscr{E}_i = \{\mathrm{e}_1, \mathrm{e}_2, \mathrm{e}_3, \mathrm{e}_4\}$ represents the element set of the article a_i where e_1 represents the "problem definition", e_2 represents the "causal interpretation", e_3 represents the "moral evaluation", and e_4 represents the "treatment recommendations".

5.3.2 Problem Formulation

The Misinformation Detection problem is defined as the process of classifying articles to identify misinformation stemming from portrayed facts under different narratives, thus misleading the audiences. To achieve that, we adopt the Frame Element-based Model (FEM), incorporating the elements of framing theory extracted from the articles. The FEM is trained to understand the semantics and narratives of articles. Having a set of articles $A = \{a_1, a_2, \ldots, a_n\}$, given an article $a_i \in A$, the model first extracts the frame elements $E_i = \{e_1, e_2, e_3, e_4\}$ of the article, and then encode them to get the hidden state h_i of it which is later used to calculate the probability to predict if the article is misinformation or not.

$$P(h_i) = softmax(w \cdot h_i + b), \tag{5.3}$$

where $P(h_i)$ is the probability that an article a_i contains misinformation, and h_i represents the last hidden state of the given article a_i or corresponding element set E_i, .

The object of the last step of predicting is defined as minimizing the loss function L:

$$\mathbf{w}^*, b^* = \arg\min_{\mathbf{w},b} L(\mathbf{w}, b) + \lambda \|\mathbf{w}\|^2, \tag{5.4}$$

where w^* and b^* are the target optimal weights, and the loss function L, which is the cross entropy loss function, is defined as:

$$L = -\frac{1}{N} \sum_{i=1}^{N} \left[y_i \log P(h_i) + (1 - y_i) \log(1 - P(h_i)) \right], \tag{5.5}$$

where N is the number of samples, and y_i is the actual label of the article a_i.

5.4 FramedTruth: A Frame-based Model Utilising Large Language Models for Misinformation Detection

5.4.1 FramedTruth Model

In this section, we elaborate on the proposed FramedTruth Model (FTM) for misinformation detection in the context of news articles. FTM consists of several key components, including news frame identification, frame and article encoding, the fusion of frame and article representations and classification. This also exhibits the impact of frames for detecting misinformation in news content.

To identify the dominant frame in a given news article, we leverage the power of ChatGPT[1] by providing the following prompt: *"Act as an academic researcher. Here are some examples of news article frames: environmental frame, political frame, business frame, race relations and so on. Please identify the most possible dominant frame in the following news article. Remember, only give the frame name without any other information."*

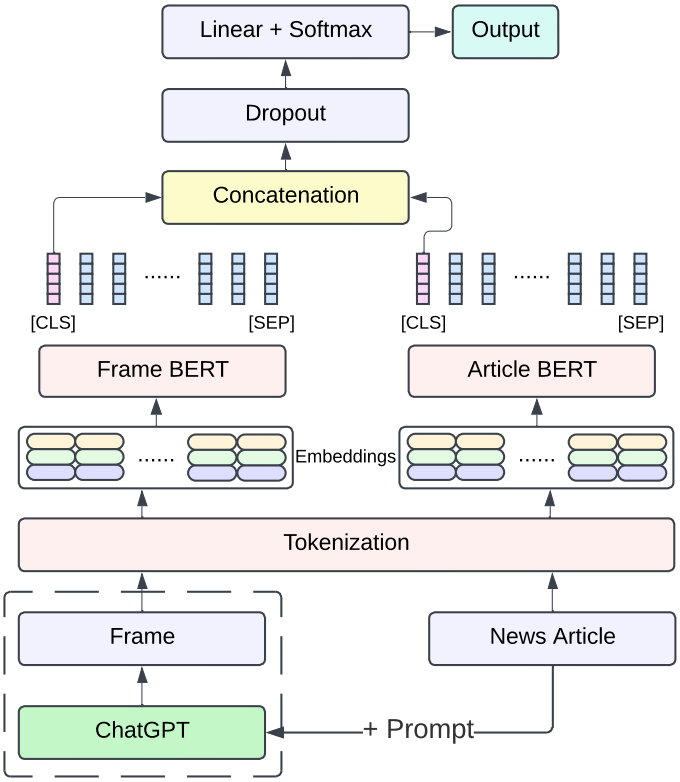


Fig. 5.1 FramedTruth Framework

[1]https://chatgpt.com/

Figure 5.1 demonstrates the overall Framework of the FramedTruth model, which is capable of detecting misinformation by considering frames in news articles. With a Dual-Encoder architecture, the FramedTruth model incorporates two separate BERT layers, each dedicated to handling a different type of input. Subsequent to these BERT layers, a linear layer is implemented. One of these BERT encoders is tasked with the encoding of the news article, while the other is assigned to encode the corresponding frame.

BERT [47] is a bidirectional multi-layer transformer architecture with multi-head self-attention mechanism. Attention is calculated as follows:

$$Attention(Q,K,V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{5.6}$$

$$Q = W_qE, K = W_kE, V = W_vE, \tag{5.7}$$

where Q,K,V refer to the matrices of queries, keys and values, respectively. W_q,W_k,W_v represent learnable parameters. E denotes the embeddings of text. Frame BERT receives the embeddings of frames, while the Article BERT receives the embeddings of the articles. Multi-head attention is defined as follows:

$$MultiHead(Q,K,V) = Concat(head_1,...,head_n)W^O \tag{5.8}$$

The "Article BERT" is dedicated to focusing only on encoding news articles, which can vary in length and cover various topics. It is tasked with transforming these articles into a suitable vector representation that encapsulates their semantic content, preparing it for further stages to identify misinformation.

The "Frame BERT" is exclusively tasked with encoding frames, which serve as additional features within news articles. These frames are instrumental in forming the perspective or viewpoint that a news article conveys, acting like a lens through which readers understand the content. Since frames are distinct from the actual content of the article, they are individually encoded by "Frame BERT". This separate encoding ensures that they do not interfere with the processing of the main article content itself.

Upon identifying the misinformation, both the article and its corresponding frame are fed into their respective encoders. These distinct inputs are processed independently. The output representations of both encoders are then concatenated to form a single vector, merging the semantic information of the article and its frame.

$$h_i = Concat(h_f^i, h_t^i), \tag{5.9}$$

where h_f^i and h_t^i represent the hidden states of the frame and the news article, respectively.

The concatenated output undergoes a dropout layer for regularisation prior to entering the linear layer. After passing through this layer, the output is processed by a softmax function. The resulting output signifies the probabilities associated with the predefined labels, i.e., information and misinformation.

$$P = softmax(Dropout(h_i)W^T + b). \tag{5.10}$$

where P is the probability distribution of the class labels.

5.4.2 Experiments and Results

To evaluate the performance of the proposed FTM, we carry out three experiments. The first experiment involves a comparison of the FTM with various state-of-the-art models in the realm of misinformation classification. In the second experiment, an ablation study is conducted to validate the significance of the frame factor within the model. Lastly, we present a case study in the third experiment to demonstrate the practical validity of the proposed FTM.

Datasets and Generation

In the experiments, we adopt two datasets, each comprising news articles along with their corresponding frames. Each article within these datasets is assigned one of two labels: "0" for information and "1" for misinformation.

- **The Knowledge Basket**[2] is one of New Zealand's news and information archives. We only capture the news focus on the "Three Waters Reform" in New Zealand, a topic of substantial political discourse and interest spanning from 2017 to 2023. This dataset accumulates a total of 1,841 articles. Following the application of our labelling process yields 3,262 articles labelled in concordance with their identified frames.
- **Kaggle Fake News Dataset**[3] contains news articles from multiple sources such as Reuters, the New York Times. For the purpose of our study, we confine our selection to the "TRUE" set. To augment the dataset to fulfil the research objectives, we produce another set of data by varying the frame of existing news, ultimately resulting in 25,718 labelled samples following the augmentation process.

[2]https://www.knowledge-basket.co.nz/

[3]https://www.kaggle.com/datasets/stevenpeutz/misinformation-fake-news-text-dataset-79k

Data Processing and Generation. The methodology we employ to refine our datasets is demonstrated in Figure 5.2. This process begins with news articles from trustworthy platforms and proceeds through several transformative and generative steps. In this process, we employ three prompts for the following purposes: prompt 1) to identify frames from original articles, prompt 2) to generate new articles by altering frames, and prompt 3) to generate new articles while preserving the frames from the original articles.

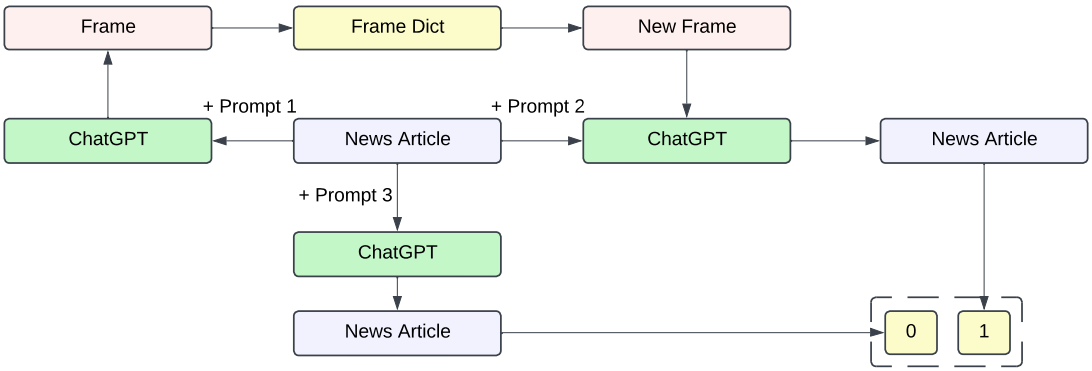


Fig. 5.2 The Flow of Data Processing and Generation

First of all, we employ ChatGPT to identify an article's frame, following the previous approach with prompt 1. The model returns a frame category, in response to which we select a contrasting frame from an expansive frame dictionary. This dictionary comprises an array of frames, each designed in accordance with the selected dataset and proved by domain experts in communication to generate content with the potential to mislead readers. For example, suppose an article is identified as 'Environmental Urgency'. In that case, we might select the contrasting frame 'Economic Impact' to illustrate how different frames can mislead the readers from an environmental issue to an economic issue, thus shaping the readers' opinions on the environmental issue as they should not be. Utilising the newly selected frame in conjunction with the original article, we instruct ChatGPT to produce a new article while emphasising the prescribed frame, using prompt 2. Consequently, the articles generated adhere to the same factual basis as the original ones and maintain semantic similarity. However, they provide different implications that potentially mislead readers. These newly framed articles are ultimately labelled "1", indicating misinformation. To assess the validity of the dataset generated through altering the frame, we have chosen a random sample comprising 10% of the generated news articles. These were then reviewed by domain experts specialising in communication studies with positive feedback. 92% of these reviewed articles are recognised as the given classification. Furthermore, to balance the frame and label distribution, inspired by [20], we also swap a part of the information with misinformation.

Meanwhile, we leverage ChatGPT to re-articulate the original articles, ensuring the preservation of their initial frames, by invoking prompt 3. The key rationale behind this approach is to prevent the model from differentiating between ChatGPT-generated text and human-authored text, thereby enabling it to concentrate solely on classifying misinformation. These rewritten articles are finally labelled as "0", referring to information.

Experiment 1: Performance Comparison

In this experiment, we aim to demonstrate the performance of the FramedTruth model in comparison to several well-established classification models. The models under comparison include BERT [47], RoBERTa [109], ALBERT [89], and XLNet [184]. These models serve as our benchmarks, and we utilise the pre-existing versions from the Hugging Face's transformers library[4] for our experiments. All the models undergo fine-tuning on the training dataset, in accordance with the experimental environment settings specified earlier. In order to assess the performance of our proposed model with the baseline models, we adopt widely accepted evaluation metrics for classification tasks, i.e., Precision, Recall, and the F1-Score.

As can be seen from Tables 5.1 our proposed model, i.e., FramedTruth, demonstrates superior capabilities in classifying misinformation in both "Three Waters" and "Kaggle Fake News" datasets. On the "Three Waters" dataset, FramedTruth substantially outperforms the traditional sequence classification models (BERT, RoBERTa, ALBERT, and XLNet). Specifically, FramedTruth exhibits remarkable accuracy, precision, recall, and F1-Score values of 0.9862, 0.9695, 0.9734, and 0.9715, respectively. These values surpass those of its closest competitor (RoBERTa), which delivers an accuracy of 0.8622, a precision of 0.8784, a recall of 0.7915, and an F1-Score of 0.8327. Similarly, the "Kaggle Fake News" dataset also shows the outstanding performance of FramedTruth over the other models. In this instance, FramedTruth achieves an accuracy of 0.9854, a precision of 0.9625, a recall of 0.9447, and an F1-Score of 0.9535.

Experiment 2: Ablation Study

In this experiment, we conduct an ablation study to assess the performance of the FramedTruth model, both with and without the integration of frame information. The key objective of this experiment is to validate the critical role that frame incorporation plays in enhancing the effectiveness of misinformation detection. This experiment is implemented on both datasets.

The results of the experiment, as illustrated in Tables 5.2 and 5.3, reveal that the model incorporating frame information outperforms its counterpart without this attribute. This finding

[4]https://huggingface.co/docs/transformers/index

Table 5.1 Results on the Three Waters Dataset and the Kaggle Fake News Dataset.

	Three Waters Reform Dataset				Kaggle Fake News Dataset			
Models	Accuracy	Precision	Recall	F1 score	Accuracy	Precision	Recall	F1 score
BERT	0.8469	0.8188	0.8127	0.8364	0.8278	0.8574	0.7806	0.8172
RoBERTa	0.8622	0.8784	0.7915	0.8327	0.8201	0.7862	0.8726	0.8273
ALBERT	0.8086	0.7651	0.8057	0.7849	0.8368	0.8523	0.8095	0.8303
XLNet	0.8545	0.8113	0.8658	0.8376	0.8237	0.8574	0.7709	0.8118
FrameTruth	**0.9862**	**0.9695**	**0.9734**	**0.9715**	**0.9854**	**0.9625**	**0.9447**	**0.9535**

shows the fundamental role that framing plays in the successful detection of misinformation that is portrayed from facts. Framing embeds vital contextual information into the articles, thus contributing significantly to the overall model performance.

Table 5.2 Results of Ablation Study on Three Waters Dataset

Models	Accuracy	Precision	Recall	F1_score
FramedTruth(No frame)	0.8132	0.8015	0.7562	0.7782
FramedTruth(With Frame)	**0.9862**	**0.9695**	**0.9734**	**0.9715**

Table 5.3 Results of Ablation Study on the Kaggle Fake News Dataset

Models	Accuracy	Precision	Recall	F1_score
FramedTruth(No frame)	0.8361	0.8902	0.7564	0.8179
FramedTruth(With Frame)	**0.9854**	**0.9625**	**0.9447**	**0.9535**

On top of that, the training loss comparison between the model with frames and the model without frames is visualised in Figure 5.3. When frame information is considered, the model converges more rapidly, indicating that the frames offer distinct and advantageous information for the detection of misinformation.

Experiment 3: Case Study

In this experiment, we conduct a case study to analyse the similarities between two articles written about the same event where each article has been written with a different frame before testing for misinformation with and without consideration for their frames. The articles focus on a council's decision to withdraw its membership from Local Government New Zealand (LGNZ).

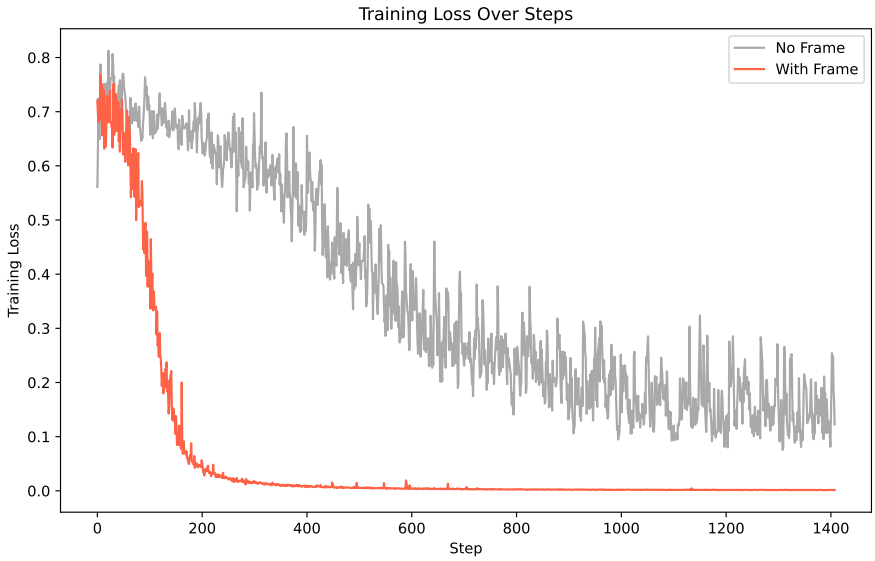


Fig. 5.3 The Training Loss Comparison

The first article has a political frame and writes:

> "...the council felt that LGNZ had not done enough to advocate for councils...", "...councillors expressed their dissatisfaction with LGNZ, with some stating that the organization had become 'mouthpieces for the Department of Internal Affairs'...", "Timaru Mayor Nigel Bowen explained that the council had been concerned about LGNZ's approach...", "LGNZ president Stuart Crosby expressed disappointment..."

Similarly, the second article with a blaming frame writes on the same points:

> "The council believes that LGNZ has failed to adequately advocate for councils...", "Councilors expressed their disappointment with LGNZ, stating that instead of advocating for councils, the organization has become 'mouthpieces for the Department of Internal Affairs'...", "Mayor Nigel Bowen has expressed the council's concerns about LGNZ's approach...", "LGNZ President Stuart Crosby expressed..."

The examples from the two articles show how similar the articles are as they both report that the council decided to leave LGNZ due to their disappointment and dissatisfaction with how poorly LGNZ advocated for councils and had become "mouthpieces for the Department

of Internal Affairs". Additionally, both articles relay the Mayor's concerns and LGNZ's response to the council's withdrawal.

As a result, without the article's frame, both articles were classified as 0 (information), whereas with the FramedTruth model, the blaming frame article was correctly detected as 1 (misinformation). It is hard to detect the differences between the two articles due to their semantic similarities. However, once the frames of the articles were considered, misinformation was identified.

5.5 Detecting Misinformation through Framing Theory

5.5.1 Frame Element-based Misinformation Detection Model

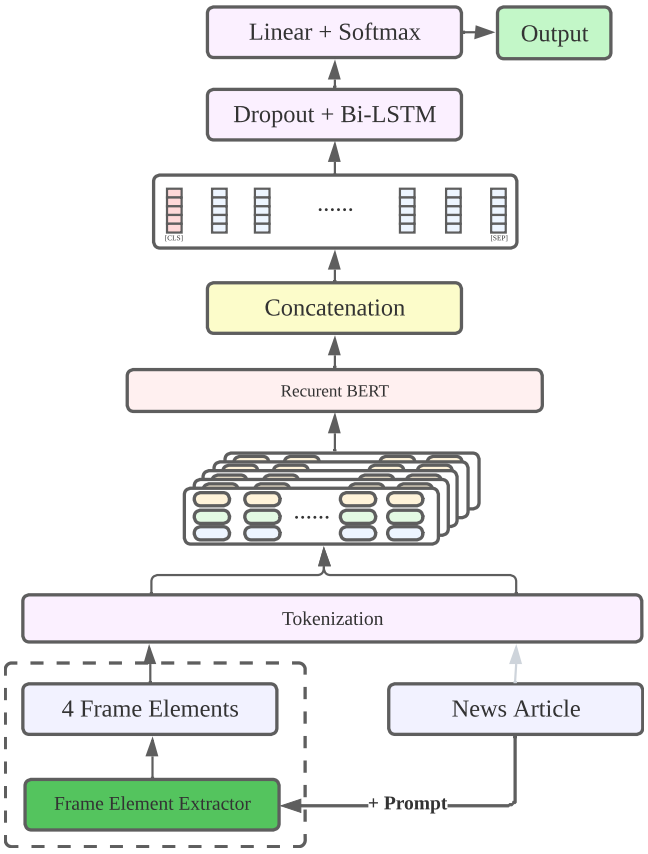


Fig. 5.4 The Architecture of the Frame Element-based Model

In this section, the proposed Frame Element-based Model (FEM) for Misinformation detection is introduced in the context of news articles. Figure 5.4 demonstrates the overall

framework of our proposed model, and in Algorithm 2, we showcase the steps of the whole process.

Initially, the Frame Element Extractor is utilised to process the news article to extract four framing elements: Problem Definition, Causal Interpretation, Moral Evaluation, and Treatment Recommendation. These elements represent the core of how the information is framed. The extracted framing elements, along with the corresponding news article, are then tokenized performing as the fundamental preprocessing step in NLP.

To capture the subtle contextual nuances of each element, following the tokenization, we independently encode each element and the corresponding news article (Lines 6 to 14 in Algorithm 2). This nuanced understanding of different elements is vital as each frame carries different weights and implications for the overall narrative of this news article. The separate encoding also allows us to quantify the impact of each element to reveal the most influential aspects of how the article is framed, thus increasing the chance of identifying misinformation.

Independently encoding each element and article is a strategic choice that can enhance the model's analytical precision, allowing for reliable misinformation detection. Line 6 starts the recurrent process. An empty tensor is created in advance and used to concatenate each embedding from each loop. In the recurrent process, we first encode the article, which performs as the main body of the input, and then each element is encoded.

The embeddings of each element *embE* and the news article *embA* are then concatenated to form a dense vector *emb* as the representation of the whole input followed by a dropout layer to prevent overfitting.

$$emb_t = concat(emb_{t-1}, h_t), \tag{5.11}$$

where emb_t represents the concatenated embeddings of the current time step, and h_t represents the embeddings of an element *embE* or the article *embA*.

The concatenated embeddings *emb* from the previous layer are then fed to a Bi-LSTM layer. The Bi-LSTM layer is applied to capture the holistic context after all elements and article embeddings are concatenated. This allows the model to understand how different elements of the article relate to each other.

$$h_i = Relu(BiLSTM(emb)), \tag{5.12}$$

where h_i is the representation of the input after being processed by a Bi-LSTM layer and a Relu activation function.

A linear layer, including a Dropout, is applied to map the high-dimensional output representation h_i from the Bi-LSTM layer to the target space. The softmax function is used to obtain the probability distribution over the potential classes, which finalizes the prediction process to identify the misinformation.

$$predicts = softmax(Dropout(h_i)W^T + b), \tag{5.13}$$

where $predicts$ is the probability distribution of the class labels, W^T is the learnable weight matrix, b refers to the bias.

Algorithm 2 Frame Element-based Misinformation detection Algorithm

Input: $D = (a,\ E)$
Output: 0 (misinformation) or 1 (information)

$Information \leftarrow Collect(sources)$
$Misinformation \leftarrow ChatGPT(information,\ prompt1)$
$FrameElements \leftarrow ChatGPT(articles,\ prompt2)$
Create BERT, BiLSTM, FC Layer as classifier, Dropout, Relu
$emb := \{\}$
for $a_i, E_i \in D$ **do**
    $embA := BERT(a_i)$
    $emb := concat(emb,\ embA)$
    for $e_j \in E_i$ **do**
        $embE := BERT(e_j)$
        $emb := concat(emb,\ embE)$
    end for
    $emb := Dropout(emb)$
end for
$outputs := BiLSTM(emb)$
$h := Relu(outputs)$
$logits := classifier(Dropout(h))$
$predicts := softmax(logits)$

5.5.2 Experiment Setups

Model Setup

To ensure an efficient training process, we conduct our experiments on the Paperspace[5] platform utilising the following tailored computational and training settings to the unique demands of each dataset:

[5]https://www.paperspace.com/

- GPU Configuration: The model is trained over a span of 100 epochs utilising NVIDIA's A6000 48GB GPU and 45 GB 8 CPU.
- Dropout: To mitigate the risk of overfitting, a dropout rate of 0.3 was applied during training.
- Learning Rate: The training uses an initial learning rate of 1×10^{-5}, and it is modulated following a cosine schedule with a warm-up phase. The warm-up steps vary in accordance with the specificities of each dataset.
- Batch Size: The batch size is determined based on the particular requirements and characteristics of each dataset.
- Frame Element Extractor: ChatGPT, as a powerful generative AI model, is used as the element extractor. Different extractors can be applied for the same purpose.

Datasets

In this section, we introduce 4 datasets used to evaluate our model. To assess the generalization capability of the model, we used three single-topic datasets, which are the Three Waters Reform dataset, the Covid-19 dataset, the Nuclear Pollution dataset, and a mixed-topic dataset, which is the Kaggle Fake News dataset. The statistics of these datasets are displayed in Table 5.4.

- **Three Waters Reform** dataset is collected from The Knowledge Basket[6], it is one of New Zealand's news and information archives, providing researchers, information professionals and library users with access to New Zealand information resources since 1994. We only capture the news focus on the "Three Waters Reform" in New Zealand, a topic of substantial political discourse and interest spanning from 2017 to June 2023. This dataset accumulates a total of 1,841 articles. Following the application of our labelling process yields 3,262 articles labelled in concordance with their identified frames and frame elements.
- **Covid-19** is collected using Newsapi[7], which is an API service that allows developers to retrieve news articles from various sources on a worldwide scale. We use "Covid-19" as keywords to retrieve news articles in the period from 01/12/2019 to 20/08/2023. These articles reflect the in-time attitude to the Covid-19 pandemic. This dataset includes 13,386 articles after the pre-processing.

[6]https://www.knowledge-basket.co.nz/
[7]https://newsapi.org/

- **Nuclear Pollution** dataset is collected using the Newsapi as well and with the keywords "nuclear pollution" over the last 5 years. This dataset provides a comprehensive view of the discourse surrounding nuclear pollution, offering a diverse range of perspectives and information. After the data pre-processing, there are 2,431 articles with an average token length of 482.
- **Kaggle Fake News Dataset**[8] contains news articles from multiple sources such as Reuters and so on. For the purpose of our study, we confine our selection to the "TRUE" set and randomly select 3k articles. To augment the dataset to fulfil the research objectives, we produce another set of data by varying the frame of existing news, ultimately resulting in 5,915 labelled samples following the implementation of our augmentation process.

Table 5.4 The statistics of the datasets after pre-processing.

Dataset	articles	average length
The Three Waters	3,262	823
Covid-19	13,386	537
Nuclear Pollution	2,431	482
Mixed-topic	5,915	469

Data Pre-processing

Our proposed methodology commences with the collection of datasets comprised of news articles from reliable sources. These articles constitute our ground truth, representing information opposite to misinformation. Accordingly, we synthesize misinformation based on the framing theory by altering the frames of our collected news articles. This process augments our datasets in a generative method at the document level and unfolds in three structured phases [20].

- **Frame Identification and Element Extraction**: utilising the capabilities of LLM, we first process the collected news articles to identify their frames and extract four elements of framing theory. These extracted framing elements reflect the news articles' original and unaltered state. They are annotated with the label "1", signifying their category as information. The frames we harness in this chapter are selected and proved by domain experts in communication, including:

[8] https://www.kaggle.com/datasets/stevenpeutz/misinformation-fake-news-text-dataset-79k

Political Frame focuses on shaping the narrative of political topics to influence public opinion, such as elections, policy introduction, etc.

Healthcare Frame centres on healthcare topics, like medical services, medical insurance services, medicines, etc.

Environmental Frame Focusing on environmental issues, this frame is based on climate change and preserving the environment.

Business Frame refers to the deceptive narratives in misinformation. It involves organizational structure, business processes, and strategic planning.

Race Frame involves the perspective through which issues of race are viewed. It encompasses how discussions, policies, and narratives are shaped, influencing societal understanding and perceptions of racial and ethnic issues.

- **Frame Alteration**: The second stage involves the alteration of the frame, utilising ChatGPT to manipulate the article narrative while maintaining the original factual information. This step simulates the process of creating misinformation through narrative manipulation, a common way that preserves factual information but skews the frame to mislead readers. 20% of the altered narratives are verified by the domain experts.
- **Element Extraction**: In the final step, we process the narrative-manipulated articles through ChatGPT to extract the corresponding four elements of framing theory, labelled "0" along with the manipulated articles, signifying their category as misinformation. This eventually establishes the basis for comparison with the information.

This pre-processing procedure is designed to construct binary-category datasets that are comprised of information and misinformation with elements of framing theory, thereby enabling the nuanced training of our model. Through this process, we not only aim to create datasets that serve as the foundation of misinformation detection but also enhance the understanding of how narrative (framing theory in this chapter) can be utilised to generate misinformation.

Evaluation Metrics and Baselines

To evaluate the performance of our proposed model (FEM), we utilise the Confusion Matrix as our primary evaluation measurement. The Confusion Matrix provides a comprehensive visualization of the performance by categorizing predictions into four different classifications [149].

- True Positives (TP): when predicted misinformation is actually labelled as misinformation;
- True Negatives (TN): when predicted information is actually labelled as information;
- False Positives (FP): when predicted information is actually labelled as misinformation;
- False Negatives (FN): when predicted misinformation is actually labelled as information.

Based on the Confusion Matrix, Precision, Recall, F1-score and Accuracy are calculated to assess the model by comparing it with other existing baseline models.

- **Accuracy** is utilised to evaluate the model's performance across all categories.

$$Accuracy = \frac{|TP| + |TN|}{|TP| + |TN| + |FN| + |FP|} \tag{5.14}$$

- **Precision** evaluates the correctness of the positive instances (the correctness of misinformation predicted) that our model has predicted.

$$Precision = \frac{|TP||}{|TP| + |FP|} \tag{5.15}$$

- **Recall** presents an indication of how many of the actual positive instances our model can correctly recognize.

$$Recall = \frac{|TP|}{|TP| + |FN|} \tag{5.16}$$

- **F1-score** is the harmonic mean of precision, and it provides a balanced measure that takes both precision and recall into account.

$$F1\_score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{5.17}$$

As our baselines, we perform the fine-tuning of several highly utilised pre-trained transformer-based language models followed by a feed-forward layer as a classifier for misinformation detection. These baselines include:

- **BERT** [47] is a groundbreaking transformer-based model in the field of NLP. It is known for its deep bidirectional training, meaning it considers the context from both the left and right sides in all layers. This leads to a more nuanced understanding of language context and semantics. BERT has been highly influential in improving the performance of a wide range of NLP tasks.5
- **RoBERTa** [109] is built upon BERT by modifying key hyperparameters, training with more data, and longer training times. These changes help RoBERTa outperform BERT on several benchmark NLP tasks. It is known for its improved robustness and efficiency.
- **ALBERT** [89] is a version of BERT optimized for lower memory consumption and increased speed. It introduces two major modifications: factorized embedding parameterization and cross-layer parameter sharing. These changes reduce the model's size without significantly affecting its performance, making it more scalable and efficient.
- **XLNet** [184] is an extension of the Transformer model. Instead of the standard transformer, XLNet uses transformer-XL [44]. It combines the best of both autoregressive (AR) and autoencoding (AE) models. Unlike BERT, XLNet learns to predict a word at a position in a sequence considering all permutations of the sequence.
- **LongFormer** [21] is designed to handle longer texts. It is an extension of the standard transformer-based model, like BERT, but optimized for lengthy documents. Its key innovation is the introduction of an attention mechanism that scales linearly with sequence length.

5.5.3 Experiments and Analysis

Experiment 1: Model Evaluation - against baselines

In this experiment, we compare the performance of our model with other baseline models on four datasets introduced in Section 5.5.2. The results are displayed in Table 5.5 ,5.6, 5.7 and 5.8 respectively. The results on each dataset consistently show that our model (FEM) incorporating frame elements with the original news article significantly outperforms other models with only articles, presenting the importance of frame elements.

From these results, we can observe that frame elements play a crucial role in understanding and interpreting information. We also demonstrate the results of our model with only frame elements and only texts as input respectively. Compared to the baselines, the results with only frame elements as input are also beyond them.

By analysing the performance of our model with only frame elements compared to the other baselines, we can observe the significance of these elements. Frame elements contribute to the deeper semantic understanding of the content by narrowing it down to the core theme reducing the distracting noise. This enables the model to grasp not just the explicit meaning but also the implicit intentions and nuances thus increasing the probability of precisely detecting misinformation.

Experiment results in Table 5.8 on the Mixed-topic dataset demonstrate that frame elements can also provide a more general representation of information, making the model more adaptable and robust to variations of information.

Table 5.5 Results on the Three Waters Dataset.

Models	Accuracy	Precision	Recall	F1_score
BERT	0.8469	0.8188	0.8127	0.8157
RoBERTa	0.8622	0.8784	0.7915	0.8327
ALBERT	0.8086	0.7651	0.8057	0.7849
XLNet	0.8545	0.8113	0.8657	0.8376
LongFormer	0.8591	0.8283	0.8516	0.8398
FEM (text+frames)	**0.9862**	**0.9695**	**0.9734**	**0.9715**
FEM (only text)	0.8652	0.8316	0.8638	0.8474
FEM (only frames)	0.9278	0.9355	0.9605	0.9478

Experiment 2: Parameter Analysis

In this experiment, we conduct a comparative analysis to explore the contribution of each element to misinformation detection. The experimental framework analyses the composite efficacy of the model, which is equipped with four elements: problem definition, causal interpretation, moral evaluation, and treatment recommendation. Within the area of misinformation detection with frame elements, exploring the individual contribution of distinct frame elements is vital to help us understand how frame elements influence the model's capability to grasp the veracity of information.

Table 5.6 Results on the Covid-19 Dataset.

Models	Accuracy	Precision	Recall	F1_score
BERT	0.8372	0.8052	0.8074	0.8063
RoBERTa	0.8547	0.8539	0.7867	0.8190
ALBERT	0.8104	0.7783	0.8163	0.7968
XLNet	0.8429	0.8207	0.8629	0.8412
LongFormer	0.8546	0.8617	0.8694	0.8655
FEM (text+frames)	**0.9783**	**0.9583**	**0.9708**	**0.9645**
FEM (only text)	0.8865	0.8737	0.8826	0.8781
FEM (only frames)	0.9132	0.9195	0.9361	0.9277

Table 5.7 Results on the Nuclear Pollution Dataset.

Models	Accuracy	Precision	Recall	F1_score
BERT	0.8035	0.7921	0.80167	0.7969
RoBERTa	0.8167	0.8234	0.7826	0.8025
ALBERT	0.8051	0.7568	0.7864	0.7713
XLNet	0.8268	0.8035	0.8284	0.8158
LongFormer	0.8462	0.8254	0.8316	0.8285
FEM (text+frames)	**0.9538**	**0.9429**	**0.9531**	**0.9480**
FEM (only text)	0.8491	0.8365	0.8537	0.8450
FEM (only frames)	0.9035	0.9216	0.9268	0.9242

The model with all 4 elements serves as the benchmark for optimal performance, showing a high degree of accuracy, precision, recall, and F1-score. This provides a holistic frame element-based analysis of information, thus enhancing the probability of identifying misinformation.

Then, we remove each frame element from all four elements, keeping the other three elements remaining. Figure 5.5 displays all performance metrics while Figure 5.6 demonstrates the trend of F1-Scores during the training process.

We can observe from all these figures that when the element of Problem Definition is removed from the model, a pronounced decrement in all measurements is demonstrated. This suggests that the recognition of Problem Definitions is instrumental in the precise detection of misinformation, potentially due to its role in pinpointing the core theme within the

Table 5.8 Results on the Mixed-topic Dataset.

Models	Accuracy	Precision	Recall	F1_score
BERT	0.8354	0.8127	0.8165	0.8146
RoBERTa	0.8497	0.8503	0.7902	0.8191
ALBERT	0.8126	0.7816	0.8257	0.8030
XLNet	0.8528	0.8320	0.8783	0.8545
LongFormer	0.8736	0.8542	0.8867	0.8701
FEM (text+frames)	**0.9696**	**0.9582**	**0.9683**	**0.9632**
FEM (only text)	0.8823	0.8574	0.8929	0.8748
FEM (only frames)	0.9158	0.9207	0.9319	0.9263

narrative that may be manipulated. Without incorporating the element of Problem Definition, the model's capability to differentiate between true and misleading content is significantly compromised.

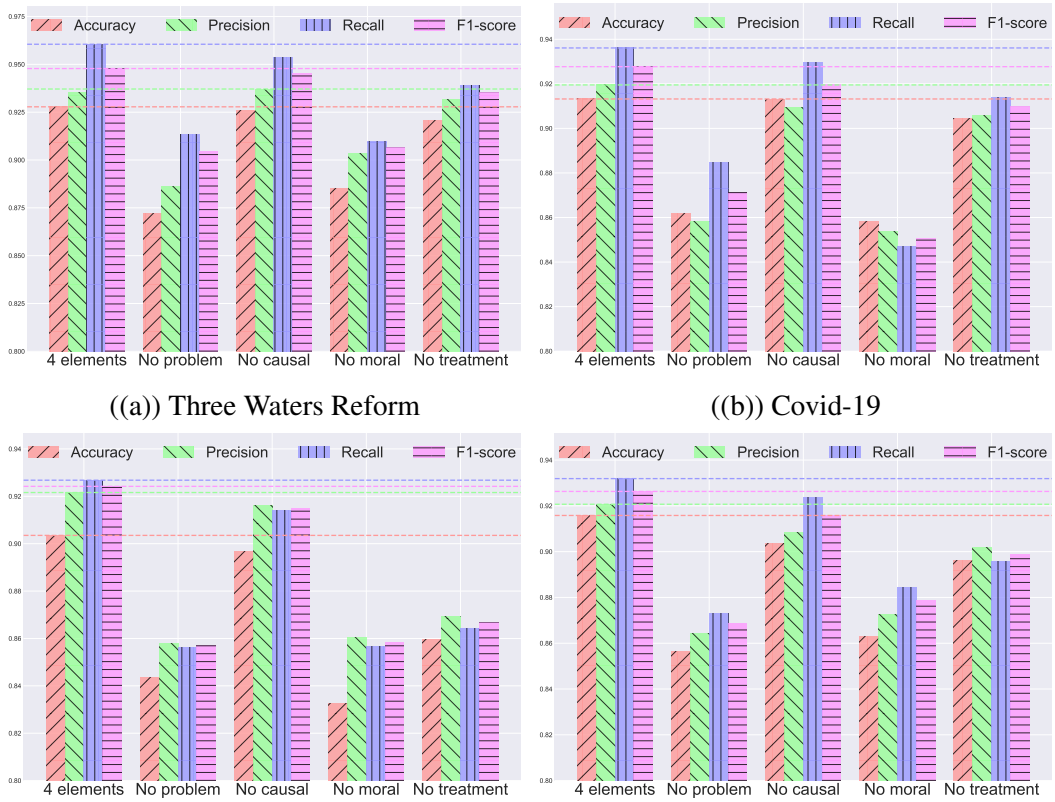
Meanwhile, the absence of the Moral Evaluation frame also results in a noticeable decline in all performance metrics. It appears to be an important factor in the framing of information, indicating that it is often manipulated through misinformation to obtain emotional biases or ethical stances.

The model, which lacks the frame of problem definition or moral evaluation, demonstrates a noticeable drop in precision and accuracy, indicating a higher rate of false positives. This implies that while the model may still identify genuine instances of misinformation, it is also more likely to incorrectly classify accurate information as misinformation.

On the contrary, a lack of the frame of Causal Interpretation or Treatment Recommendation does not show a substantial decline in performance metrics compared to the benchmark. This observation implies that while they have a role in the misinformation detection process, however, their absence does not critically influence the capability of the model to identify misinformation.

One noticeable difference in the results demonstrated in Figure 5.5(c), and Figure 5.6(c) on the Nuclear Pollution dataset is the lack of the frame of Treatment Recommendation. Removing the Treatment Recommendation element also results in a lower performance across all metrics indicating that in the context of nuclear pollution, the treatment recommendation is likely to be a key indicator of the news articles. A lack of this element in this area could also allow misinformation, proposing ineffective or misleading responses to harness the

readers. The decrement in performance of missing the frame of treatment recommendation also implies that the contribution of each element can vary depending on the subject.



((a)) Three Waters Reform

((b)) Covid-19

((c)) Nuclear Pollution

((d)) Mixed-topic

Fig. 5.5 Measure the performance of removing one of the elements on all four datasets.

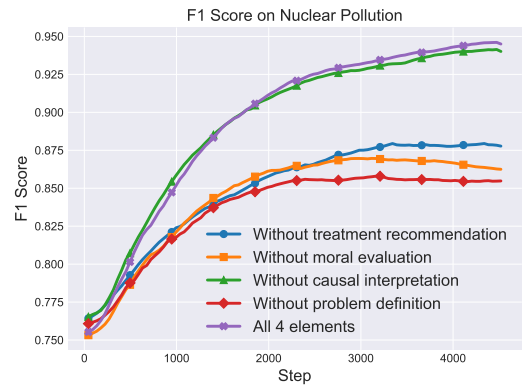
Experiment 3: Similarity Comparison

Table 5.9 One single pair similarity and similarities removing one of the elements.

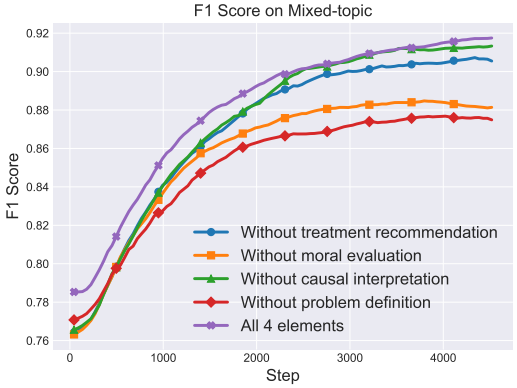
Info vs Mis-info	Similarity	F1-score
Article Similarity	0.86	0.8474
Elements Similarity(all 4 elements)	**0.61**	**0.9478**
Elements Similarity(without problem)	0.79	0.9046
Elements Similarity(without causal)	0.62	0.9454
Elements Similarity(without moral)	0.81	0.9065
Elements Similarity(without treatment)	0.64	0.9354

((a)) Three Waters Reform

((b)) Covid-19



((c)) Nuclear Pollution



((d)) Mixed-topic

Fig. 5.6 The F1-scores during the training process on all four datasets.

Table 5.10 Compare article average similarity with average similarity calculated using 4 elements.

Info vs Mis-info	three water	covid	nuclear	mixed
Article Similarity	0.86	0.82	0.83	0.85
Elements Similarity(all 4 elements)	0.58	0.62	0.59	0.61
Elements Similarity(without problem)	0.83	0.81	0.82	0.83
Elements Similarity(without causal)	0.59	0.62	0.60	0.63
Elements Similarity(without moral)	0.81	0.79	0.81	0.80
Elements Similarity(without treatment)	0.60	0.64	0.78	0.63

In this experiment, we analyse the relationship between the similarity between information and misinformation under different conditions relating to the presence or absence of specific frame elements within our model. This experiment represents how closely misin-

formation mirrors authentic information in terms of framing. The cosine function is used to calculate their similarities:

$$sim(h_i, h_j) = \frac{h_i \cdot h_j}{\| h_i \| \| h_j \|}, \tag{5.18}$$

where h_i and h_j represent the final hidden states of two articles or elements from two articles.

Table 5.9 shows the similarities of one randomly selected article from the Three Waters Reform dataset and the overall performance (F1-score) of the model on this dataset. The similarity between the information article and the misinformation article is used as the benchmark for further analysis. The similarity of 0.86 shows that, without any specialized modification, misinformation is quite successful at resembling a genuine news article.

From Table 5.9, we can observe that the similarity between the information article and the misinformation article is the highest at 0.85. However, the F1 score of the model with only text on this dataset is 0.8474, which is lower than for other conditions, especially when utilising only all 4 elements, which holds the lowest similarity of 0.61 and the highest F1 score of 0.9478.

This experiment indicates the alignment of similarity with the model performance as an inverse relationship. The lower the similarity, the higher the performance is in detecting misinformation. This underscores the importance of elements of framing theory in the detection process and points out their potential impact on improving detection accuracy.

We also calculate the average similarity scores between information and misinformation across four distinct datasets under different conditions. Results are displayed in Table 5.10. The pattern of similarity scores is relatively consistent across different topics, indicating that the manipulation of framing in misinformation follows a similar pattern. However, on the nuclear dataset, when omitting the frame of treatment recommendation, the similarity still remains at a high level, indicating the importance of treatment recommendation in this dataset.

Overall, by comparing the average similarities across all datasets, we can observe the alignment of similarities with omitting distinct elements except the impact when omitting the frame of treatment recommendation on the nuclear dataset. This indicates a unique pattern in the context of a specific topic, showing that different elements of framing theory have varying levels of impact depending on the topic. This also underlines the importance of topic-sensitive approaches in misinformation detection.

Experiment 4: Case study

In this experiment, we conduct a case study to analyse the similarities between two articles written about the same topic where each article has a different frame and frame elements. The articles focus on the government's proposed water reforms.

The first article has a political frame:

"There's a lot of change being proposed by the government... Fundamentally, they're considering shifting responsibility for our three waters: water supply, wastewater, and stormwater, from local government into four large entities... The government now believes that costs of between $120 billion and $185b will be required: between $4 and $6b per year on average... The proposed three waters reform program harks back to the Havelock North water contamination event in 2016... It's on this basis that the government has concluded that four entities, aggregating all the water services across the country, offer the best and quickest opportunity to achieve the desired improvements to the three-waters networks."

While the second article has a semantic frame to show their satire:

"Oh, boy! The government is proposing some exciting changes, folks. Brace yourselves because they're considering taking control of our beloved three waters. You know, the precious water supply, wastewater, and stormwater that our local government has been responsible for?... The government estimates that we'll need a mind-boggling $120 billion to $185 billion over the next 30 years... Well, now they want to hand it over to these big entities. What a brilliant idea, right?... And get this – the government thinks it would be cheaper if larger entities took over the water services. Apparently, they can borrow more, with the government's backing, of course. I mean, who needs small, local councils when you can have these big entities making all the decisions for you?"

The similarities calculated in Table 5.9 of 0.86 indicate that the articles are highly similar as they both share details about the reform. However, once the frame elements are considered, the article similarity decreases to 0.61.

The political frame is informative and objective, presenting a detailed overview using formal language and statistics to support its claims, while the semantic frame is emotional and opinionated, using colloquial language and employing vivid imagery to engage readers emotionally, which may risk oversimplification and bias.

To determine the classification without frame elements, our model only encodes the news article. In comparison, for classifying with the frame elements, the elements and news articles are encoded independently with their embeddings concatenated into one vector prior to classification. The inclusion of these extra features enhances the model's performance.

For example, given the problem definition for the political frame:

"The proposed shift of responsibility for three waters from local government to four large entities known as water supply entities."

As well as the problem definition for the semantic frame:

"The proposed government takeover of three waters"

The problem definition of each article highlights their differences in framing. The politically framed article's problem definition is detailed with a neutral tone, while the semantic frame's problem definition is short and has a negative perspective.

Both articles are classified as information without frame elements. However, once the frame elements are considered, the semantic frame is correctly classified as misinformation. This suggests that including the frame elements in our model contributes to the successful classification of misinformation by reinforcing the differences between the two semantically similar articles.

Discussions

We proposed a Frame Element-based Model (FEM) to distinguish misinformation from information. Several experiments are conducted, providing crucial insights into the importance of elements of framing theory in misinformation detection. Results are evaluated by comparing them with baseline models. Furthermore, we analysed the contribution of each element, demonstrating the different roles of the elements. By comparing the article similarities and element similarities, we obtained insights into how the elements improve the performance of detecting misinformation. Based on the experimental results, we have the following insights observed:

- The results of Experiment 1 on all datasets consistently proved that incorporating the elements of framing theory while detecting misinformation stemming from portrayed facts under different narratives can help improve the performance.
- The parameter analysis experiment revealed that the absence of certain framing elements, particularly Problem Definition and Moral Evaluation, leads to a significant decrease in the model's accuracy and precision. This underscores the critical role these elements play in the accurate detection of misinformation.
- Experiment 2 also demonstrates a finding that the specific element plays a different role on different topics, highlighting the potential impact of elements and underscoring the necessity for topic-sensitive approaches in misinformation detection, as different elements of framing theory have varying levels of impact depending on the topic.

- The similarity comparison experiment further illustrated how misinformation closely mirrors authentic information in terms of framing. The results indicated an inverse relationship between similarity and model performance: lower similarity between the information and misinformation articles led to higher performance in detecting misinformation.
- The case study shows that not applying the framing theory to the articles can lead to the result of the article with a semantic frame incorrectly classified as information, increasing the potential for misleading interpretations. It also shows that while articles may be semantically similar, the choice of framing can greatly impact the narrative of content being misleading or misinterpreted.

5.6 Conclusion and Future Work

In this chapter, we introduce the FramedTruth model (FTM) and the Framed Element-based Model (FEM) to identify misinformation in the context of news articles incorporating the elements of framing theory.

For FTM, we explore the use of LLM in the process of identifying frames and augmenting data samples. On top of that, we conducted extensive experiments to evaluate the performance of the FramedTruth model. The experiments on two real-world datasets demonstrated the FramedTruth model's superior performance in identifying misinformation compared to several well-known transformer-based models. The ablation studies further validated the significant contributions of framing, highlighting its crucial role in misinformation detection. The results of FTM have been published in [165].

The FEM leverages ChatGPT and deep neural networks to detect misinformation originating from accurately portrayed facts under different frames. The efficacy of FEM is demonstrated through comprehensive performance comparisons with other methods, highlighting the effectiveness of Framed Element-based approach against traditional misinformation detection models. The contribution of each element is also evaluated and analysed along with the similarities under different conditions, indicating the importance of the specific element and showcasing how the narrative of an article is framed.

This chapter also provides a foundational understanding of how frames and elements of framing theory influence the perception and interpretation of information. Building upon the insights obtained, there are several future directions. Future studies can delve into the impact of specific elements across various topics, such as the frame of treatment recommendation on the nuclear dataset, which shows more impact than on the other three datasets, raising questions that can be explored in the future. Besides the contribution of each element we

explored, relations among these elements can also be explored as a future direction that can help us understand and enhance the ability to detect misinformation under more complex scopes.

Chapter 6

Echo chamber detection and mitigation

The phenomenon known as the "echo chamber" has been widely acknowledged as a significant force affecting society. This has been particularly evident during the Covid-19 pandemic, wherein the echo chamber effect has significantly influenced public responses. Therefore, detecting echo chambers and mitigating their adverse impacts has become crucial to facilitating a more diverse exchange of ideas, and fostering a more understanding and empathetic society. In response, we use deep learning methodologies to model each user's beliefs based on their historical message contents and behaviours. As such, we propose a novel, content-based framework built on the foundation of weighted beliefs. This framework is capable of detecting potential echo chambers by creating user belief graphs, utilising their historical messages and behaviours. To demonstrate the practicality of this approach, we conducted experiments using the Twitter dataset on Covid-19. These experiments illustrate the potential for individuals to be isolated within echo chambers. Furthermore, our in-depth analysis of the results reveals patterns of echo chamber evolution and highlights the importance of weighted relations. Understanding these patterns can be instrumental in the development of tools and strategies to combat misinformation, encourage the sharing of diverse perspectives, and enhance the collective well-being and social good of our digital society.

6.1 Overview

Nowadays, online social platforms have become one of the key sources for people to perceive information. It also reshapes the way of searching, filtering and disseminating information [118]. Modelling and analysing the influence and dissemination of information on social networks, including information maximization, social sentiment analysis, concern detection, etc. [96, 147], have become prominent subjects [76, 100]. One of the main perspectives of

social media is to expose users to like-minded peers, which may result in echo chambers that could reinforce users' pre-existing viewpoints and drive the communities to be more polarized [39]. Individuals in these communities are easily affected by their surroundings. Estimating the extent to which an individual is isolated in an echo chamber is helpful in breaking the isolation.

Jamieson et al. were the first to define the phenomenon occurring on social media platforms where information within a community is amplified and metaphorically term it as the "echo chamber" [75]. To determine whether an individual is isolated within an echo chamber, we focus on their surroundings and behaviours. Therefore, we define the echo chamber from an individual perspective. This phenomenon is characterized by an individual who:

- Resides in a community that echoes their opinion, wherein most neighbours share similar views.
- Inhabits this community, where the individual's perspectives are repeatedly reinforced by community members. This reinforcement is achieved through reading related messages, sending messages that align with the community's views, or responding to messages from like-minded neighbours.
- Self-reinforces by engaging in communication that aligns with their viewpoint. This could involve sending relevant messages to neighbours, responding affirmatively to messages with the same opinion, or viewing related messages posted by their neighbours or recommended by the platform.

The prominence of the echo chamber phenomenon has been heightened by the outbreak of Covid-19. Recognised globally as a pandemic, Covid-19 has become a leading topic of discussion. Conversations around this subject vary widely, encompassing themes such as vaccine hesitancy and vaccination-related deaths. Under these conditions, social platforms provide a conducive environment for misinformation propagation due to the lack of editorial supervision. As a result, echo chambers have emerged among users, significantly influencing responses to the Covid-19 pandemic [5]. For example, if members of a community consistently engage with and promote content sceptical of Covid-19 vaccinations, the community can be highly identified as an echo chamber resistant to the prevailing medical advice on vaccines. Exposure to such misinformation can lead to adverse societal consequences.

To address this challenging issue, we propose a novel approach to detect the echo chamber phenomenon and estimate the corresponding degree. In pursuing this aim, we explore the utility of knowledge graphs in identifying echo chambers from an individual perspective and

propose a content-based framework that constructs belief graphs for each individual. During this construction process, we extract triplets from the individual's related Twitter content, including tweets, retweets, and replies. To assess the impact of individual behaviours, we incorporate them as distinct parameters to calculate weights for these triplets. Extensive experiments are conducted, and the results explicitly show that modelling an individual's belief graphs with weighted relations can effectively reveal an individual's trends on a specific topic and identify the echo chambers to which the individual belongs.

The rest of this chapter is organised as follows. In Section 6.2, the related works and limitations are articulated. Section 6.3 gives formal definitions and elaborates on the proposed content-based belief-aware echo chamber detection framework. Section 6.4 introduces the data pre-processing and the experimental results. Finally, the chapter has been concluded in Section 6.5.

6.2 Related Work

6.2.1 Echo Chambers on Social Platforms

Social platforms have become the primary sources of information for many individuals, offering an unprecedented volume of data. This shifts the way people access information and forms echo chambers, isolating them in the process [39]. The detection of echo chambers has been a research focus across various fields [13, 136], serving as the first step towards mitigating this phenomenon.

Social structures typically manifest in two distinct forms, i.e., global perspective and individual perspective. Most research has studied echo chambers from a global perspective or topological viewpoint, primarily focusing on user interactions while overlooking the source of these interactions [18, 29]. Cinelli et al. analyse echo chambers by assessing whether the overall network is strongly polarized towards two sides of a controversy, emphasizing user interaction networks [39]. Cossard et al. explore echo chambers within vaccine communities using clustering techniques, demonstrating the existence of echo chambers within real social networks [41].

Analysing extensive topological structure datasets from a global perspective necessitates high-performance computing resources. The ego network centred around a focal user offers a feasible way to model a community, enabling measurement of the echo chamber degree with a focus on that user. Thus, inspired by Li et al. and Valerio et al., we incorporate the concept of the ego network in our study [11, 99]. Li et al. propose agent-based influence diffusion models, where the influence cascading process is modelled as an evolutionary pattern driven

by individuals' actions. Valerio et al. analyse the micro-level structural properties of online social networks and demonstrate that ego networks play a significant role in social networks, impacting information diffusion within the network. Hu et al. investigate the impact of AI recommendation on forming echo chambers from both individual and topological levels [70].

6.2.2 Content-Based Echo Chamber Detection

Content-based methods identify echo chambers by analysing the information texts produced by individuals. Villa et al. propose both a topology-based and content-based approach, analysing the topological structure of the social network and sentiment aspects related to the content [162]. Cinelli et al. conduct a comparative analysis on a large-scale dataset to identify echo chambers through social network homophily. They define "leaning" as the attitude expressed by a piece of content towards a specific topic about the content [39]. Abd-Alrazaq et al. propose a text-mining method on a large dataset, considering information texts but neglecting temporal information, which can provide contextual insights [1]. Lwin et al. and Xue et al. demonstrate that discourses on Twitter about Covid-19 continually evolve, develop, or change over time [113, 181]. Inspired by these studies, we restructure the dataset into chronologically user-specific streams.

Most existing studies solely consider the content of information but overlook individual behaviours and content weights, which demonstrate the significance of content on individuals. For instance, reading a message doesn't explicitly reveal an individual's thoughts about the message. However, a subsequent 'like' or 'upvote' implies that the individual agrees with this message, thereby increasing the weight of information from this message in the corresponding belief graph. We argue that beliefs in individuals' minds carry different weights, and not all beliefs hold equal significance. As a result, behaviours offer valuable insights into people's perspectives on related messages.

Therefore, we propose a belief-aware echo chamber detection framework incorporating content and individual behaviours. Our framework constructs belief graphs for each individual in our dataset, considering their behaviours. To measure the degree of echo chambers, we calculate the similarities between the belief graphs of the focal user and their neighbours. With this framework, social platforms can detect communities where members are primarily exposed to reinforcing views, potentially limiting the diversity of thoughts and contributing to polarization.

6.3 Belief-Aware Echo Chamber Detection

In this section, we formally define related terms and explain the proposed belief-aware echo chamber detection framework.

6.3.1 Formal Definitions

Two types of graph structures are utilised in this chapter: one is the ego network, a directed graph $G =< U, E >$ that includes a focal user and its neighbours, and the other is the user belief graph $BG =< H, WR, T >$.

An ego network consists of a focal user u_f and their neighboring users, denoted as $U = \{u_0, ..., u_n\}$. Each user, represented by $u_i \in U$, corresponds to a node in this directed network. Each $edge(u_i, u_j)$ between nodes is directed as the flow of information, indicating that user u_i follows, replies to, or mentions user u_j. Each user in an ego network also has a unique belief graph, representing their personal network of beliefs, which is clarified in Definition 2.

A belief graph is a unique graph containing multiple triplets $BG = \{H, WR, T\}$, where H and T refer to nodes in belief graphs, and $WR = \{wr_i | 0 < i < m\}$ represents relations between nodes, defined as weighted relations. The belief graph is constructed by extracting triplets from users' historical messages and behaviours on corresponding messages across various topics.

Similarity $sim(v_i, v_j)$ refers to the distance between two vectors in a low-dimensional space. The similarity is a value between $[0, 1]$, where 0 implies completely contrary viewpoints, while 1 signifies identical viewpoints.

Echo chamber degree $p(u, k)$ is a measure that evaluates the likelihood of an echo chamber. A higher degree suggests a higher probability that an individual experiences an echo chamber related to a specific topic k.

Topic k refers to the label of each message, e.g., m_k. The topic set is denoted as $T = \{T_0, T_1, ..., T_n\}$. Messages with the same topics express similar discourse. One message is assigned only one topic. In our framework, topics are used for graph partitioning.

Each user in an ego network is represented by their sub-graphs. The Graph encoder receives each sub-graph and generates representations representing each user in the ego network. Subsequently, the graph representations are utilised for the calculation of the echo chamber probability of this ego network.

As illustrated in Fig. 6.1, we determine the echo chamber degree in 5 phases:

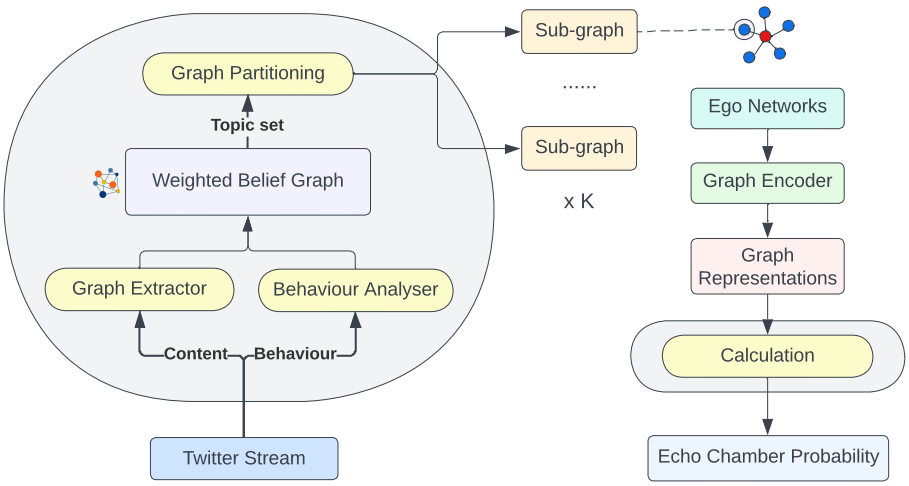


Fig. 6.1 The brief overall process of the framework.

- 1) Construct a belief graph for each individual, considering both their Twitter stream and their user behaviours. Based on each corresponding behaviour to a message, we increase or decrease the weights of triplets extracted from this message.
- 2) Partition the corresponding part of an individual's belief graph into sub-graphs according to different topics.
- 3) Select an individual and its neighbours to create an ego network based on their followee/follower relations and mentioning behaviour.
- 4) Calculate the similarities of sub-graphs on a given topic between the focal user and its neighbours to assess the closeness of their beliefs.
- 5) Quantify the echo chamber degree by evaluating their average similarity and information entropy.

A sub-graph is a graph partitioned from an individual's complete belief graph given a specific topic. Sub-graphs are used to compare users. Messages refer to texts that users receive and post, including tweets, retweets, and replies. User behaviours refer to user operations on a social platform, including:

- Viewing: Users view messages posted by their neighbours (followees) or recommended by the platform.
- Liking: Users like a message by clicking the blank heart symbol.

- Disliking: Users express dislike for a message by cancelling their liking behaviour, i.e., clicking the solid heart symbol.
- Reposting: Users repost viewed messages.
- Sending: Users post a message or reply to someone in their own words.

We presume each behaviour reflects a different perspective on corresponding messages and aids in modelling changes in user beliefs. For example, when a user likes a message, we increase the weight of triplets extracted from this message by assigning a changing rate to the weight. The changing rates are defined for different behaviours as shown in Table 6.1:

Table 6.1 Changing rate of each behaviour on corresponding information.

	Reviewing	**Liking**	**Disliking**	**Reposting**	**Sending**
Changing Rate(r)	0.5	2	-2	1	2

6.3.2 Belief Graph Construction

The first step in this work is to construct belief graphs for each individual in the ego network. We extract triplets (i.e., *{head, relation, tail}*) from the content and calculate weights for these triplets by analysing the individual's behaviours. We then attach the weights to the relations, resulting in weighted relations. A belief graph that reflects an individual's belief consists of multiple triplets with weighted relations. We use Stanford OpenIE[1] to extract triplets from texts.

To calculate the weights, we employ a logarithmic function to prevent the weights from reaching extremely high or low. This logarithmic transformation helps maintain a balanced range of weights. The function is defined as follows:

$$w_i = ln(w_i^{'} + r_i) + 1, \tag{6.1}$$

where $w_i^{'}$ is the previous weight of the same triplet and r_i is the changing rate as shown in Table 6.1.

During extraction, the same triplets may be extracted multiple times. For each new triplet, the initial weight is defined as 0, and its current weight is calculated based on the changing rate of the corresponding behaviour. When we encounter the identical triplet, we add the change in weight according to the current behaviour to its previous weight (i.e., $w_i^{'}$ in Equation 6.1).

[1] https://nlp.stanford.edu/software/openie.html

6.3.3 Belief Graph Partitioning

A complete belief graph of an individual encompasses information from several topics. Comparing complete graphs may allow irrelevant information to affect performance on the given topic. Hence, we perform a graph partitioning step before transforming graphs into graph representations.

In this step, we utilise word embeddings from a word2vec model [117] to identify nodes within the belief graph that have similar words to the given topic and keywords. The cosine similarity function, as shown below, is utilised to measure the similarity between word embeddings. Nodes and relations in both directions are subsequently used to form the sub-graph.

$$sim(v_i, v_j) = \frac{v_i \cdot v_j}{\| v_i \| \| v_j \|}, \tag{6.2}$$

where v_i and v_j denote word embeddings obtained from the word2vec model.

6.3.4 Echo Chamber Detection

To compare the similarities among belief graphs, we convert these topological structure graphs into vector representations. This is achieved through training Graph Attention Networks (GATs) on each individual's belief graph to generate graph representations. Different from the original GATs [**?**], we introduce the weighted relation features $R = \{r_{i,j} | 0 < i < n, 0 < j < n\}$ as the initial attention coefficient. The weighted relation features are used during the attention calculation as follows:

$$e_{i,j} = a(W\hat{h}_i, W\hat{h}_j, r_{i,j}) \tag{6.3}$$

Equation 6.3 represents the importance of node j's features to node i. W denotes a weight matrix used to parameterize a shared linear transformation, $\hat{h}_i$ represents the features of node i, and $r_{i,j}$ is the weighted relation from node i to node j. To collect all features of the whole graph, we add a global node to each graph. This global node is linked to every node in the graph, and its representation represents the entire graph.

We hypothesize that similar graphs express similar beliefs on relevant topics. To test this, we compute the similarity between individuals' belief graphs. We apply the graph representations generated by the trained GATs to a cosine similarity function to calculate these similarities:

$$sim(h_i^k, h_u^k) = \frac{h_i^k \cdot h_u^k}{\| h_i^k \| \| h_u^k \|}, \tag{6.4}$$

where h_i^k denotes the representation of user i's sub-belief graph on topic k, and $\| h_i^k \|$ represents the Euclidean norm of h_i^k. h_u^k refers to the representation of the focal user's sub-graph on topic k. The average similarities are then calculated as follows:

$$avg(h_u^k) = 1/n \sum_{i=1}^{n} sim(h_i^k, h_u^k), \tag{6.5}$$

where n describes the number of the focal user's neighbours.

In addition to similarity, inspired by [70], we also consider information entropy from information theory and statistical mechanics to calculate the probability of an individual being isolated in an echo chamber. The equation is as follows:

$$H(g_u^k) = -\sum_{k \subseteq K} p_k \cdot ln(p_k), \tag{6.6}$$

where p_k is the percentage of a user's sub-graph on topic k, and g_u^k refers to the belief graph of the focal user of an ego network on topic k. Finally, we use both average similarity and information entropy to measure the echo chamber using the following equation:

$$p(u,k) = avg(h_u^k) \cdot H(g_u^k), \tag{6.7}$$

where u represents a focal user. A higher $p(u,k)$ indicates a greater likelihood that u is isolated in an echo chamber. In such a case, the ego network centred around user u is a $p(u,k)$ possibility echo chamber on topic k.

6.4 Experiments and Analysis

This section provides details of two experiments conducted to validate the efficacy of the proposed Belief-based Echo Chamber Detection model. The first experiment evaluates the similarities in responses of echo chamber members to multiple related messages. The second experiment implements an ablation study to elucidate the progression of the Belief Graph module within the BeECD framework and to investigate the impact of weighted relations on belief graphs.

6.4.1 Data Collection and Organisation

The experiments utilise a dataset gathered from Twitter related to COVID-19. This dataset, part of the continually updated COVID-19 Twitter chatter dataset maintained by Georgia State University's Panacea Lab, spans a crucial six-month period from December 2020 to May 2021. This time frame is particularly significant as it encompasses a period when several candidate vaccines displayed safety and the ability to generate immune responses. The proposed BeECD can be applied to any dataset. In this chapter, we leverage Covid-19 as the dataset to validate this approach.

Each individual's content and behaviours are organised into a chronological stream, including the user's tweets, tweets from the user's neighbours, retweets, replies, corresponding tweets, likes, and liked tweets. We limit our focus solely to English tweets, replies, and retweets, and uniquely, we include retweets in the streams of each individual, allowing us to process retweeting behaviour and corresponding content concurrently.

To facilitate computation, we extracted a sub-graph from the total dataset, comprising 285 users, 3,587 interconnections, and 42,478 posts, which include tweets, retweets, and replies.

6.4.2 Experiment 1: Response Analysis

In this experiment, we hypothesize that each user within an ego network can respond appropriately to one or multiple similar messages, anticipating that responses from like-minded users will exhibit greater similarity than those from dissimilar users. The implications of this phenomenon in real-world contexts are substantial. Consider, for instance, an ego network exhibiting an 80% echo chamber probability. If the average response probabilities within this network align closely with this percentage, it will signify that most users within the network are engaged in disseminating and consuming similar information. In a practical sense, this may translate to a reinforcement of a specific narrative or perspective. The resulting lack of engagement with diverse viewpoints could amplify polarization. This may create a self-reinforcing cycle in which users are confined to information confirming their pre-existing beliefs, thereby becoming increasingly resistant to alternative viewpoints or evidence contradicting their established convictions.

In addition to calculating the echo chamber probabilities, we subsequently train an encoder-decoder structured language model on the entire dataset, feeding 20 random messages from the test set into each ego network based on the same topic that the ego network inclines towards. The language model is used to assess the similarity of the users' responses.

By comparing the echo chamber probabilities and response similarities, we assess the efficacy of the proposed framework. The results depicted in Fig 6.2 corroborate our hypothesis.

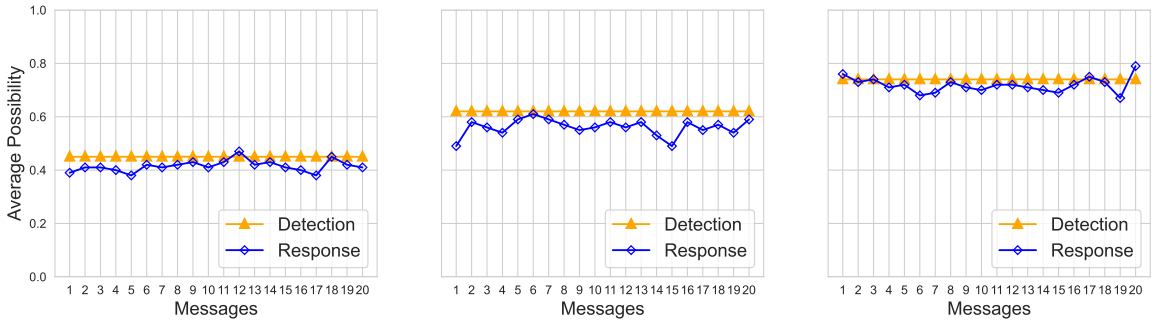


Fig. 6.2 The probabilities of detected echo chamber and user response similarities.

By presenting the outcomes from three distinct ego networks with varying degrees of echo chamber probabilities, it's clear that the average response probabilities align with the calculated echo chamber probabilities.

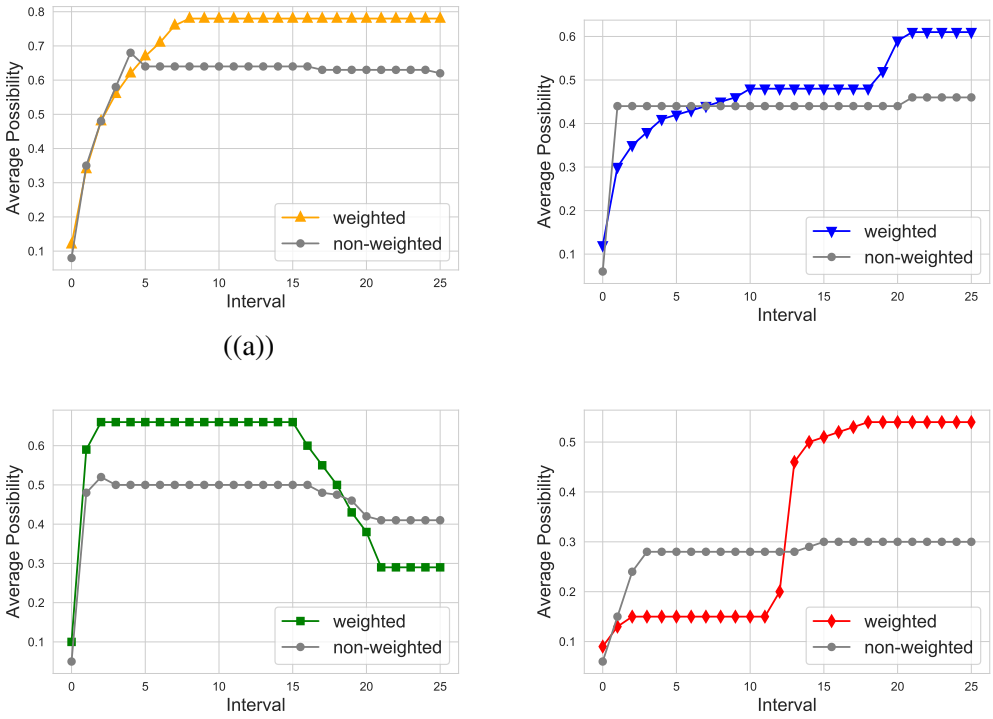
This experiment sheds light on the intricate relationship between online interactions and the formation of echo chambers. The observed alignment between the echo chamber probability and user response patterns underscores the role digital platforms play in shaping real-world perspectives, emphasizing the need for further research and interventions in this domain.

6.4.3 Experiment 2: Belief Graph Impact Analysis

The second experiment seeks to understand the influence of weighted relations on the evolution of echo chambers. We additionally train Graph Attention Networks (GATs) that do not account for the properties of relations during the training process. Belief graphs are initiated using data from the first two months of six, following which the remaining data is partitioned into 25 unique time intervals. From each interval of the users' Twitter streams, users' beliefs and behaviours are extracted and used to update their corresponding belief graphs.

By partitioning data into distinct time intervals, we highlight the significance of temporal evolution in shaping users' beliefs. This process shows the importance of identifying the beliefs and understanding how they transform and develop over time. Such an approach corresponds to real-world scenarios, where individuals often undergo phases or shifts in their perspectives. These changes may be influenced by various factors, such as past experiences, exposure to new information, or personal growth, reflecting the complexity and dynamism of human belief systems.

We anticipate that our framework, which incorporates weighted relations, is capable of detecting variations in user beliefs, including instances where these beliefs intensify before subsequently diminishing. This approach offers insight into the dynamic nature of belief changes. On the other hand, in the absence of such weighted relations, a user's beliefs appear to remain unaltered and static. This lack of dynamism obscures the potential to observe the evolutionary patterns of echo chambers, even in cases where users undergo significant shifts in their perspectives. We selected four representative curves from our proposed framework with weighted relations and corresponding curves from the framework lacking weighted relations for comparison. The results depicted in Fig 6.3 effectively showcase different patterns of evolution of echo chambers in our proposed framework with weighted relations.



((a))

Fig. 6.3 The evolution of echo chambers.

From Fig 6.3, it is clear that our proposed framework can effectively represent the evolution of echo chambers or users' shifting perspectives. The result also highlights a crucial difference between the two models. The use of weighted relations, as opposed to non-weighted ones, allows for a more nuanced representation of the complexities inherent in human interactions and belief systems. In real-world terms, not all interactions influence our beliefs equally. Some might have a significant impact due to the trustworthiness of the source or the emotional resonance of the content, while others might be casually scrolled past

without much thought. Thus, incorporating weighted relations can more accurately model how real people might be influenced by their digital interactions.

Understanding the evolution of echo chambers using advanced models like GATs with weighted relations is crucial in today's digital age. Such insights provide a clearer picture of how beliefs change over time on platforms like Twitter, emphasizing the need for digital platforms to prioritize diverse content exposure and critical thinking among their users.

6.5 Conclusion and Prospective Research Directions

In this study, we introduce a novel content-based methodology for echo chamber detection on social networks, coined as the belief-aware echo chamber detection approach shedding light on the intricate relationship between online interactions and the formation of echo chambers. We leverage Knowledge Graph technology to construct user belief graphs, taking into account both message content and user behaviour. Additionally, we train modified Graph Attention Networks, incorporating weighted relations into the computation process. Similarities between user belief graphs are then computed. The experimental results indicate promising effectiveness and demonstrate real-world implications of our approach in analysing echo chambers on social platforms. The final results have been published in [169].

However, the detection of echo chambers represents a seminal work, and addressing the subsequent effects presents significant challenges. Future research endeavours will focus on strategies for mitigating the echo chambers, further advancing the understanding and management of social network dynamics.

Chapter 7

Conclusion

7.1 Research Contributions

7.1.1 Text Summarization

- I propose a novel knowledge-aware text summarization model, KATSum, which employs the Knowledge Graph to improve the quality of the summaries in terms of ROUGE scores. With the knowledge-aware encoder, the input text will be processed by the pre-trained language model and converted into Knowledge Graph embeddings. Such features can help to address unfaithfulness and factual inconsistency.
- I propose a novel model, Aspect-adaptive and Knowledge-based Opinion Summarization (AsKOS), which is self-supervised training on an unlabeled dataset. This dataset contains accurately matched pairs of aspect graphs and pseudo summaries without requiring a time-consuming labelling process.
- I improve the effectiveness in capturing diverse aspects from reviews and tailoring the summaries to align with the specific requirements of users.

7.1.2 Influence Maximization

- I first proposed suppressing the information alteration while maximizing the influence in the dissemination process.
- I propose a user behaviour model to take the users' personalised prior knowledge and subconsciousness to predict user behaviour and generate user responses given input information.

- I take the possible information alteration into consideration while addressing the influence maximization problem to keep the originality of the disseminated information.
- I adopt Knowledge Graph (KG) to model users' prior knowledge and sub-consciousness. Deep learning models are utilised to model the users' behaviours, which can generate responses/comments that imply potential information alteration.
- I propose a novel seed selection algorithm aiming to maximize the social influence without causing significant information alterations.

7.1.3 Misinformation Detection

- I formulate the misinformation is portrayed from factual information but presented in a misleading manner, implying meanings that differ from literal interpretations and thus leading readers astray.
- I explore the application of framing theory in detecting misinformation with factual information but presented in a misleading way.
- I propose the FrameTruth Model (FTM) incorporating frame categories, the Large Language Model and Deep Learning to address misinformation detection.
- I propose the Frame Element-based Model (FEM), taking advantage of frame elements to precisely identify misinformation originating from accurately portrayed facts under different frames.
- I explore the separate contribution of each frame element to demonstrate how misinformation is framed.

7.1.4 Echo Chamber Detection

- I define the echo chamber detection from the individual level as estimating the echo chamber degree to see if an individual is isolated in an echo chamber.
- I model each user by constructing their belief graphs according to their historical behaviours and posts.
- I propose a novel method to calculate the echo chamber degree incorporating graph similarities between the belief graphs of the focal user and its neighbours and cross-entropy.

7.2 Limitations and future directions

In this thesis, I addressed several problems in the field of social network analysis incorporating knowledge graph technique and Deep Learning.

Future research directions in the fields of text summarization, information diffusion, misinformation detection, and echo chamber detection are summarised as follows:

Current existing summarization models rely solely on language models which could generate biased content which aligns with the training data. Future summarization models should leverage knowledge graphs to enhance the accuracy and coherence of generated summaries. Knowledge graphs can provide structured context and relationships between entities, improving the relevance and informativeness of the summaries. Combining knowledge graphs with self-supervised training techniques can customise summaries based on specific informational needs. This approach allows models to learn from large amounts of unlabeled data, tailoring summaries to user preferences and enhancing the adaptability of summarization systems.

Meanwhile, given the limitations of ROUGE as an evaluation metric, it is essential to explore new methodologies for evaluating text summarization. These metrics should effectively measure coherence, readability, context dependency, and even reliability, aligning more closely with the advancements in large language models.

In the information diffusion field, the proposed approach can be applied to various scenarios where maintaining the integrity of the influence is essential. This includes marketing campaigns, public health messaging, and political communications, where the fidelity of the original message is critical to achieving desired outcomes. In the future, we intend to explore two potential avenues for the future expansion of this work.

First, a more accurate knowledge representation model will be investigated to capture information alteration.

Second, though the proposed model showed strong performance in small- to medium-scale networks, its computational efficiency remains challenging in large-scale and dynamic networks. This limitation was most noticeable where processing delays increased significantly with larger seed sets. We intend to address the same issue in a more challenging environment, i.e., large-scale and dynamic social networks. Agent-based modelling will be applied to distribute the computation cost of the algorithm and improve its efficiency, following the brief steps below:

- Understand the Dynamics of Large-Scale Networks by performing network analysis to identify key features such as scale, density, modularity, dynamic changes and temporal behaviours.

- Define agent characteristics and roles with attributes such as knowledge, preferences, influence thresholds, behaviour patterns and their relationships.
- Develop a scalable agent-based simulation model for scenarios with varying message types, influence thresholds, and information alterations to test robustness.
- Integrate real-time updates for dynamic networks by developing mechanisms to update the agent-based model dynamically as nodes join, quit, or modify behaviours.
- Incorporate learning mechanisms such as deep learning and reinforcement learning for agents.

For misinformation detection, further investigation into framing theory is still needed to understand its role and impact on frame-based misinformation detection, like how frames convey ideas through different portraits, and how frames are utilized on different platforms, e.g. Facebook, X, etc. To that end, more data from different sources will be collected and analysed thoroughly. Based on the analysis, a model that can precisely predict frames will be developed utilizing advanced technologies.

The detection of echo chambers represents a seminal work, and addressing the subsequent effects presents significant challenges. Future research endeavours can focus on strategies for mitigating the echo chambers, further advancing the understanding and management of social network dynamics.

By addressing these research directions, future studies can contribute to more robust and adaptive models in text summarization, influence maximization, misinformation detection, and echo chamber detection, ultimately enhancing the effectiveness and reliability of online social networks.

References

[1] Abd-Alrazaq, A., Alhuwail, D., Househ, M., Hamdi, M., Shah, Z., et al. (2020). Top concerns of tweeters during the covid-19 pandemic: infoveillance study. *Journal of medical Internet research*, 22(4):e19016.

[2] Abdali, S., Bastidas, G. G., Shah, N., and Papalexakis, E. E. (2020). Tensor embeddings for content-based misinformation detection with limited supervision. *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities*, pages 117–140.

[3] Ahuja, O., Xu, J., Gupta, A., Horecka, K., and Durrett, G. (2022). ASPECTNEWS: Aspect-oriented summarization of news documents. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6494–6506, Dublin, Ireland. Association for Computational Linguistics.

[4] Aïmeur, E., Amri, S., and Brassard, G. (2023). Fake news, disinformation and misinformation in social media: a review. *Social Network Analysis and Mining*, 13(1):30.

[5] Alatawi, F., Cheng, L., Tahir, A., Karami, M., Jiang, B., Black, T., and Liu, H. (2021). A survey on echo chambers on social media: Description, detection and mitigation. *arXiv preprint arXiv:2112.05084*.

[6] Alzahrani, A., Baabdullah, T., Almotairi, A., and Rawat, D. B. (2023). A hybrid deep learning architecture for misinformation detection on social media. In *2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science (IRI)*, pages 199–204.

[7] Amplayo, R. K., Angelidis, S., and Lapata, M. (2021). Aspect-controllable opinion summarization. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 6578–6593, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

[8] Angelidis, S., Amplayo, R. K., Suhara, Y., Wang, X., and Lapata, M. (2021). Extractive opinion summarization in quantized transformer spaces. *Transactions of the Association for Computational Linguistics*, 9:277–293.

[9] Angelidis, S. and Lapata, M. (2018). Summarizing opinions: Aspect extraction meets sentiment prediction and they are both weakly supervised. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3675–3686, Brussels, Belgium. Association for Computational Linguistics.

[10] Arendt, F., Forrai, M., and Mestas, M. (2023). News framing and preference-based reinforcement: Evidence from a real framing environment during the covid-19 pandemic. *Communication Research*, 50(2):179–204.

[11] Arnaboldi, V., Conti, M., La Gala, M., Passarella, A., and Pezzoni, F. (2016). Ego network structure in online social networks and its impact on information diffusion. *Computer Communications*, 76:26–41.

[12] Aron, J. L. and Schwartz, I. B. (1984). Seasonality and period-doubling bifurcations in an epidemic model. *Journal of theoretical biology*, 110(4):665–679.

[13] Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout, F., and Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37):9216–9221.

[14] Bakshy, E., Eckles, D., Yan, R., and Rosenn, I. (2012). Social influence in social advertising: evidence from field experiments. In *Proceedings of the 13th ACM conference on electronic commerce*, pages 146–161.

[15] Bakshy, E., Hofman, J. M., Mason, W. A., and Watts, D. J. (2011). Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 65–74.

[16] Banerjee, S., Jenamani, M., and Pratihar, D. K. (2020). A survey on influence maximization in a social network. *Knowledge and Information Systems*, 62:3417–3455.

[17] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *science*, 286(5439):509–512.

[18] Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., and Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science*, 26(10):1531–1542.

[19] Barbieri, N., Bonchi, F., and Manco, G. (2013). Topic-aware social influence propagation models. *Knowledge and information systems*, 37(3):555–584.

[20] Bayer, M., Kaufhold, M.-A., and Reuter, C. (2022). A survey on data augmentation for text classification. *ACM Computing Surveys*, 55(7):1–39.

[21] Beltagy, I., Peters, M. E., and Cohan, A. (2020). Longformer: The long-document transformer. *arXiv:2004.05150*.

[22] Bharathi, S., Kempe, D., and Salek, M. (2007). Competitive influence maximization in social networks. In *International workshop on web and internet economics*, pages 306–311. Springer.

[23] Bhaskar, A., Fabbri, A., and Durrett, G. (2023). Prompted opinion summarization with gpt-3.5. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 9282–9300.

[24] Bjørnstad, O. N., Shea, K., Krzywinski, M., and Altman, N. (2020). The seirs model for infectious disease dynamics. *Nature methods*, 17(6):557–559.

[25] Blondel, V. D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

[26] Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., and Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295–298.

[27] Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. *Advances in neural information processing systems*, 26.

[28] Bražinskas, A., Lapata, M., and Titov, I. (2020). Unsupervised opinion summarization as copycat-review generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5151–5169, Online. Association for Computational Linguistics.

[29] Bruns, A. (2017). Echo chamber? what echo chamber? reviewing the evidence. In *6th Biennial Future of Journalism Conference (FOJ17)*.

[30] Cabot, P.-L. H., Dankers, V., Abadi, D., Fischer, A., and Shutova, E. (2020). The pragmatics behind politics: Modelling metaphor, framing and emotion in political discourse. In *Findings of the association for computational linguistics: emnlp 2020*, pages 4479–4488.

[31] Cai, T., Li, J., Mian, A. S., Sellis, T., Yu, J. X., et al. (2020). Target-aware holistic influence maximization in spatial social networks. *IEEE Transactions on Knowledge and Data Engineering*.

[32] Cao, Z., Wei, F., Li, W., and Li, S. (2018). Faithful to the original: fact-aware neural abstractive summarization. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, pages 4784–4791.

[33] Card, D., Boydstun, A., Gross, J. H., Resnik, P., and Smith, N. A. (2015). The media frames corpus: Annotations of frames across issues. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 438–444.

[34] Chen, W., Lu, W., and Zhang, N. (2012). Time-critical influence maximization in social networks with time-delayed diffusion process. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 26, pages 591–598.

[35] Chen, W., Wang, C., and Wang, Y. (2010). Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1029–1038.

[36] Chen, W., Wang, Y., and Yang, S. (2009). Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 199–208.

[37] Cheung, C. M. and Thadani, D. R. (2012). The impact of electronic word-of-mouth communication: A literature analysis and integrative model. *Decision support systems*, 54(1):461–470.

[38] Chu, E. and Liu, P. (2019). Meansum: A neural model for unsupervised multi-document abstractive summarization. In *International Conference on Machine Learning*, pages 1223–1232. PMLR.

[39] Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., and Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9).

[40] Condori, R. E. L. and Pardo, T. A. S. (2017). Opinion summarization methods: Comparing and extending extractive and abstractive approaches. *Expert Systems with Applications*, 78:124–134.

[41] Cossard, A., Morales, G. D. F., Kalimeri, K., Mejova, Y., Paolotti, D., and Starnini, M. (2020). Falling into the echo chamber: the italian vaccination debate on twitter. In *Proceedings of the International AAAI conference on web and social media*, volume 14, pages 130–140.

[42] Crawford, K. (2009). Following you: Disciplines of listening in social media. *Continuum*, 23(4):525–535.

[43] Cui, P., Jin, S., Yu, L., Wang, F., Zhu, W., and Yang, S. (2013). Cascading outbreak prediction in networks: a data-driven approach. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 901–909.

[44] Dai, Z., Yang, Z., Yang, Y., Carbonell, J. G., Le, Q., and Salakhutdinov, R. (2019). Transformer-xl: Attentive language models beyond a fixed-length context. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 2978–2988.

[45] Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., and Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the national academy of Sciences*, 113(3):554–559.

[46] Del Vicario, M., Zollo, F., Caldarelli, G., Scala, A., and Quattrociocchi, W. (2017). Mapping social dynamics on facebook: The brexit debate. *Social Networks*, 50:6–16.

[47] Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

[48] Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., and Zhang, W. (2014). Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 601–610.

[49] Dou, Z.-Y., Liu, P., Hayashi, H., Jiang, Z., and Neubig, G. (2021). GSum: A general framework for guided neural abstractive summarization. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 4830–4842, Online. Association for Computational Linguistics.

[50] Elsahar, H., Coavoux, M., Rozen, J., and Gallé, M. (2021). Self-supervised and controlled multi-document opinion summarization. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1646–1662, Online. Association for Computational Linguistics.

[51] Entman, R. (2004). *Projections of Power: Framing News, Public Opinion, and U.S. Foreign Policy*. Projections of Power: Framing News, Public Opinion, and U.S. Foreign Policy. University of Chicago Press.

[52] Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of communication*, 43(4):51–58.

[53] Ercan, G. and Cicekli, I. (2007). Using lexical chains for keyword extraction. *Information Processing & Management*, 43(6):1705–1714.

[54] Erkan, G. and Radev, D. R. (2004). Lexrank: Graph-based lexical centrality as salience in text summarization. *Journal of artificial intelligence research*, 22:457–479.

[55] Fairhurst, G. and Sarr, R. (1996). *The art of framing*. San Francisco: Jossey-Bass.

[56] Gamzu, I., Gonen, H., Kutiel, G., Levy, R., and Agichtein, E. (2021). Identifying helpful sentences in product reviews. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 678–691, Online. Association for Computational Linguistics.

[57] Ganesan, K., Zhai, C., and Han, J. (2010). Opinosis: A graph based approach to abstractive summarization of highly redundant opinions. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 340–348, Beijing, China. Coling 2010 Organizing Committee.

[58] Gehrmann, S., Deng, Y., and Rush, A. (2018). Bottom-up abstractive summarization. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4098–4109, Brussels, Belgium. Association for Computational Linguistics.

[59] Gershtein, S., Milo, T., and Youngmann, B. (2021). Multi-objective influence maximization. *algorithms*, 20:33.

[60] Giles, P. (1977). The mathematical theory of infectious diseases and its applications. *Journal of the Operational Research Society*, 28(2):479–480.

[61] Goffman, E. (1974). *Frame analysis: An essay on the organization of experience.* Harvard University Press.

[62] Goyal, A., Lu, W., and Lakshmanan, L. V. (2011). Celf++ optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th international conference companion on World wide web*, pages 47–48.

[63] Guarino, S., Trino, N., Chessa, A., and Riotta, G. (2020). Beyond fact-checking: Network analysis tools for monitoring disinformation in social media. In *Complex Networks and Their Applications VIII: Volume 1 Proceedings of the Eighth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2019 8*, pages 436–447. Springer.

[64] Guille, A., Hacid, H., Favre, C., and Zighed, D. A. (2013). Information diffusion in online social networks: A survey. *ACM Sigmod Record*, 42(2):17–28.

[65] Guo, Z., Schlichtkrull, M., and Vlachos, A. (2022). A Survey on Automated Fact-Checking. *Transactions of the Association for Computational Linguistics*, 10:178–206.

[66] Hamdi, T., Slimi, H., Bounhas, I., and Slimani, Y. (2020). A hybrid approach for fake news detection in twitter based on user features and graph embedding. In *Distributed Computing and Internet Technology: 16th International Conference, ICDCIT 2020, Bhubaneswar, India, January 9–12, 2020, Proceedings 16*, pages 266–280. Springer.

[67] Hermann, K. M., Kocisky, T., Grefenstette, E., Espeholt, L., Kay, W., Suleyman, M., and Blunsom, P. (2015). Teaching machines to read and comprehend. *Advances in neural information processing systems*, 28.

[68] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

[69] Hu, W., Singh, K. K., Xiao, F., Han, J., Chuah, C.-N., and Lee, Y. J. (2018). Who will share my image? predicting the content diffusion path in online social networks. In *Proceedings of the eleventh ACM international conference on web search and data mining*, pages 252–260.

[70] Hu, Y., Wu, S., Jiang, C., Li, W., Bai, Q., and Roehrer, E. (2022). Ai facilitated isolations? the impact of recommendation-based influence diffusion in human society. In Raedt, L. D., editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 5080–5086. International Joint Conferences on Artificial Intelligence Organization.

[71] Huffman, D. A. (1952). A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9):1098–1101.

[72] Im, J., Kim, M., Lee, H., Cho, H., and Chung, S. (2021). Self-supervised multimodal opinion summarization. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 388–403, Online. Association for Computational Linguistics.

[73] Islam, M. R., Liu, S., Wang, X., and Xu, G. (2020). Deep learning for misinformation detection on online social networks: a survey and new perspectives. *Social Network Analysis and Mining*, 10:1–20.

[74] Jacoby, W. G. (2000). Issue framing and public opinion on government spending. *American Journal of Political Science*, pages 750–767.

[75] Jamieson, K. H. and Cappella, J. N. (2008). *Echo chamber: Rush Limbaugh and the conservative media establishment*. Oxford University Press.

[76] Jiang, C., D'Arienzo, A., Li, W., Wu, S., and Bai, Q. (2021). An operator-based approach for modeling influence diffusion in complex social networks. *Journal of Social Computing*, 2(2):166–182.

[77] Jin, Y., Wang, W., and Xiao, S. (2007). An sirs model with a nonlinear incidence rate. *Chaos, Solitons & Fractals*, 34(5):1482–1497.

[78] Karimi, S., Papamichail, K. N., and Holland, C. P. (2015). The effect of prior knowledge and decision-making style on the online purchase decision-making process: A typology of consumer shopping behaviour. *Decision Support Systems*, 77:137–147.

[79] Keikha, M. M., Rahgozar, M., Asadpour, M., and Abdollahi, M. F. (2020). Influence maximization across heterogeneous interconnected networks based on deep learning. *Expert Systems with Applications*, 140:112905.

[80] Kempe, D., Kleinberg, J., and Tardos, É. (2003). Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146.

[81] Klein, G., Kim, Y., Deng, Y., Nguyen, V., Senellart, J., and Rush, A. M. (2018). Opennmt: Neural machine translation toolkit. In *Proceedings of the 13th Conference of the Association for Machine Translation in the Americas (Volume 1: Research Track)*, pages 177–184.

[82] Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632.

[83] Klimt, B. and Yang, Y. (2004). Introducing the enron corpus. In *CEAS*, volume 45, pages 92–96.

[84] Kryscinski, W., McCann, B., Xiong, C., and Socher, R. (2020). Evaluating the factual consistency of abstractive text summarization. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9332–9346, Online. Association for Computational Linguistics.

[85] Kshetri, N. (2023). Chatgpt in developing economies. *IT Professional*, 25(2):16–19.

[86] Kwak, H., Lee, C., Park, H., and Moon, S. (2010). What is twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600.

[87] Kwon, S., Cha, M., Jung, K., Chen, W., and Wang, Y. (2013). Prominent features of rumor propagation in online social media. In *2013 IEEE 13th International Conference on Data Mining*, pages 1103–1108.

[88] Lagnier, C., Denoyer, L., Gaussier, E., and Gallinari, P. (2013). Predicting information diffusion in social networks using content and user's profiles. In *European conference on information retrieval*, pages 74–85. Springer.

[89] Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., and Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*.

[90] Lavanya, A. and Rajeswari, K. (2020). Aspect based extractive summarization of online product reviews. *International Journal of Scientific & Technology Research*, 9(3):3588–3591.

[91] Leskovec, J., Adamic, L. A., and Huberman, B. A. (2007a). The dynamics of viral marketing. *ACM Transactions on the Web (TWEB)*, 1(1):5–es.

[92] Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., and Glance, N. (2007b). Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 420–429.

[93] Li, C., Xu, W., Li, S., and Gao, S. (2018a). Guiding generation for abstractive text summarization based on key information guide network. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 55–60.

[94] Li, H., Zhu, J., Zhang, J., and Zong, C. (2018b). Ensure the correctness of the summary: Incorporate entailment knowledge into abstractive sentence summarization. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1430–1441, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

[95] Li, M. Y., Graef, J. R., Wang, L., and Karsai, J. (1999). Global dynamics of a seir model with varying total population size. *Mathematical biosciences*, 160(2):191–213.

[96] Li, W., Bai, Q., Jiang, C., and Zhang, M. (2016a). Stigmergy-based influence maximization in social networks. In *PRICAI 2016: Trends in Artificial Intelligence: 14th Pacific Rim International Conference on Artificial Intelligence, Phuket, Thailand, August 22-26, 2016, Proceedings 14*, pages 750–762. Springer.

[97] Li, W., Bai, Q., Liang, L., Yang, Y., Hu, Y., and Zhang, M. (2021). Social influence minimization based on context-aware multiple influences diffusion model. *Knowledge-Based Systems*, page 107233.

[98] Li, W., Bai, Q., Nguyen, D. T., and Zhang, M. (2017). Agent-based influence maintenance in social networks. In *Proceedings of the Sixteenth International Conference on Autonomous Agents and Multiagent Systems (Extended Abstract)(AAMAS 2017)*.

[99] Li, W., Bai, Q., and Zhang, M. (2016b). Agent-based influence propagation in social networks. In *2016 IEEE International Conference on Agents (ICA)*, pages 51–56. IEEE.

[100] Li, W., Bai, Q., and Zhang, M. (2018c). Siminer: a stigmergy-based model for mining influential nodes in dynamic social networks. *IEEE Transactions on Big Data*, 5(2):223–237.

[101] Li, W., Bai, Q., and Zhang, M. (2019). A multi-agent system for modelling preference-based complex influence diffusion in social networks. *The Computer Journal*, 62(3):430–447.

[102] Li, W., Bai, Q., Zhang, M., and Nguyen, T. D. (2018d). Modelling multiple influences diffusion in on-line social networks. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, pages 1053–1061.

[103] Li, Y., Fan, J., Wang, Y., and Tan, K.-L. (2018e). Influence maximization on social graphs: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 30(10):1852–1872.

[104] Lin, C.-Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

[105] Litvak, M. and Last, M. (2008). Graph-based keyword extraction for single-document summarization. In *Coling 2008: Proceedings of the workshop multi-source multilingual information extraction and summarization*, pages 17–24.

[106] Liu, S., Guo, L., Mays, K., Betke, M., and Wijaya, D. T. (2019a). Detecting frames in news headlines and its application to analyzing news framing trends surrounding us gun violence. In *Proceedings of the 23rd conference on computational natural language learning (CoNLL)*, pages 504–514.

[107] Liu, W., Yue, K., Wu, H., Li, J., Liu, D., and Tang, D. (2016). Containment of competitive influence spread in social networks. *Knowledge-Based Systems*, 109:266–275.

[108] Liu, Y. and Lapata, M. (2019). Text summarization with pretrained encoders. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 3730–3740, Hong Kong, China. Association for Computational Linguistics.

[109] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019b). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

[110] Liu, Y. and Wu, Y.-F. B. (2020). Fned: a deep network for fake news early detection on social media. *ACM Transactions on Information Systems (TOIS)*, 38(3):1–33.

[111] Longoni, C., Fradkin, A., Cian, L., and Pennycook, G. (2022). News from generative artificial intelligence is believed less. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 97–106.

[112] Lu, Y., Zhai, C., and Sundaresan, N. (2009). Rated aspect summarization of short comments. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, page 131–140, New York, NY, USA. Association for Computing Machinery.

[113] Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., and Yang, Y. (2020). Global sentiments surrounding the covid-19 pandemic on twitter: analysis of twitter trends. *JMIR public health and surveillance*, 6(2):e19447.

[114] Ma, J., Gao, W., and Wong, K.-F. (2018). Rumor detection on twitter with tree-structured recursive neural networks. Association for Computational Linguistics.

[115] Maynez, J., Narayan, S., Bohnet, B., and McDonald, R. (2020). On faithfulness and factuality in abstractive summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1906–1919.

[116] Miao, Z., Li, Y., Wang, X., and Tan, W.-C. (2020). Snippext: Semi-supervised opinion mining with augmentednbsp;data. In *Proceedings of The Web Conference 2020*, WWW '20, page 617–628, New York, NY, USA. Association for Computing Machinery.

[117] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space.

[118] Morini, V., Pollacci, L., and Rossetti, G. (2021). Toward a standard approach for echo chamber detection: Reddit case study. *Applied Sciences*, 11(12):5390.

[119] Narayan, S., Cohen, S., and Lapata, M. (2018). Don't give me the details, just the summary! topic-aware convolutional neural networks for extreme summarization. In *2018 Conference on Empirical Methods in Natural Language Processing*, pages 1797–1807. Association for Computational Linguistics.

[120] Nasir, J. A., Khan, O. S., and Varlamis, I. (2021). Fake news detection: A hybrid cnn-rnn based deep learning approach. *International Journal of Information Management Data Insights*, 1(1):100007.

[121] Newman, M. E. (2006). Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582.

[122] Nickel, M., Murphy, K., Tresp, V., and Gabrilovich, E. (2015). A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33.

[123] Nourbakhsh, A., Liu, X., Li, Q., and Shah, S. (2017). Mapping the echo-chamber: Detecting and characterizing partisan networks on twitter. In *Proceedings of the 2017 International Conference on Social Computing, Behavioral-Cultural Modeling, & Prediction and Behavior Representation in Modeling and Simulation*, pages 5–8.

[124] Oliveira, M. and Gama, J. (2012). An overview of social network analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(2):99–115.

[125] Oshikawa, R., Qian, J., and Wang, W. Y. (2018). A survey on natural language processing for fake news detection. *CoRR*, abs/1811.00770.

[126] Pastor-Satorras, R. and Vespignani, A. (2001). Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200.

[127] Pelrine, K., Danovitch, J., and Rabbany, R. (2021). The surprising performance of simple baselines for misinformation detection. In *Proceedings of the Web Conference 2021*, pages 3432–3441.

[128] Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543.

[129] Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018a). Deep contextualized word representations. *NAACL*.

[130] Pillai, S. E. V. S. and Hu, W.-C. (2023). Misinformation detection using an ensemble method with emphasis on sentiment and emotional analyses. In *2023 IEEE/ACIS 21st International Conference on Software Engineering Research, Management and Applications (SERA)*, pages 295–300.

[131] Pons, P. and Latapy, M. (2005). Computing communities in large networks using random walks. In *Computer and Information Sciences-ISCIS 2005: 20th International Symposium, Istanbul, Turkey, October 26-28, 2005. Proceedings 20*, pages 284–293. Springer.

[132] Qiu, J., Tang, J., Ma, H., Dong, Y., Wang, K., and Tang, J. (2018). Deepinf: Social influence prediction with deep learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2110–2119.

[133] Radford, A., Jozefowicz, R., and Sutskever, I. (2017). Learning to generate reviews and discovering sentiment. *arXiv preprint arXiv:1704.01444*.

[134] Rashkin, H., Choi, E., Jang, J. Y., Volkova, S., and Choi, Y. (2017). Truth of varying shades: Analyzing language in fake news and political fact-checking. In *Proceedings of the 2017 conference on empirical methods in natural language processing*, pages 2931–2937.

[135] Raut, V. B. and Londhe, D. (2014). Opinion mining and summarization of hotel reviews. In *2014 International Conference on Computational Intelligence and Communication Networks*, pages 556–559. IEEE.

[136] Romer, D. and Jamieson, K. H. (2021). Patterns of media use, strength of belief in covid-19 conspiracy theories, and the prevention of covid-19 from march to july 2020 in the united states: Survey study. *Journal of medical Internet research*, 23(4):e25215.

[137] Rosvall, M., Axelsson, D., and Bergstrom, C. T. (2009). The map equation. *The European Physical Journal Special Topics*, 178(1):13–23.

[138] Rush, A. M., Chopra, S., and Weston, J. (2015). A neural attention model for abstractive sentence summarization. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 379–389, Lisbon, Portugal. Association for Computational Linguistics.

[139] Saito, K., Nakano, R., and Kimura, M. (2008). Prediction of information diffusion probabilities for independent cascade model. In *International conference on knowledge-based and intelligent information and engineering systems*, pages 67–75. Springer.

[140] Saito, K., Ohara, K., Yamagishi, Y., Kimura, M., and Motoda, H. (2011). Learning diffusion probability based on node attributes in social networks. In *International Symposium on Methodologies for Intelligent Systems*, pages 153–162. Springer.

[141] Scarselli, F., Gori, M., Tsoi, A. C., Hagenbuchner, M., and Monfardini, G. (2008). The graph neural network model. *IEEE transactions on neural networks*, 20(1):61–80.

[142] Scheibenzuber, C., Neagu, L.-M., Ruseti, S., Artmann, B., Bartsch, C., Kubik, M., Dascalu, M., Trausan-Matu, S., and Nistor, N. (2023a). Dialog in the echo chamber: Fake news framing predicts emotion, argumentation and dialogic social knowledge building in subsequent online discussions. *Computers in Human Behavior*, 140:107587.

[143] Scheibenzuber, C., Neagu, L.-M., Ruseti, S., Artmann, B., Bartsch, C., Kubik, M., Dascalu, M., Trausan-Matu, S., and Nistor, N. (2023b). Dialog in the echo chamber: Fake news framing predicts emotion, argumentation and dialogic social knowledge building in subsequent online discussions. *Computers in Human Behavior*, 140:107587.

[144] Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of communication*, 49(1):103–122.

[145] See, A., Liu, P. J., and Manning, C. D. (2017). Get to the point: Summarization with pointer-generator networks. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1073–1083, Vancouver, Canada. Association for Computational Linguistics.

[146] Shetty, J. and Adibi, J. (2005). Discovering important nodes through graph entropy the case of enron email database. In *Proceedings of the 3rd international workshop on Link discovery*, pages 74–81.

[147] Shi, J., Li, W., Yang, Y., Yao, N., Bai, Q., Yongchareon, S., and Yu, J. (2021). Automated concern exploration in pandemic situations-covid-19 as a use case. In *Knowledge Management and Acquisition for Intelligent Systems: 17th Pacific Rim Knowledge Acquisition Workshop, PKAW 2020, Yokohama, Japan, January 7–8, 2021, Proceedings 17*, pages 178–185. Springer.

[148] Shu, K., Cui, L., Wang, S., Lee, D., and Liu, H. (2019). defend: Explainable fake news detection. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 395–405.

[149] Shu, K., Sliva, A., Wang, S., Tang, J., and Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36.

[150] Shu, K., Wang, S., and Liu, H. (2018). Understanding user profiles on social media for fake news detection. In *2018 IEEE conference on multimedia information processing and retrieval (MIPR)*, pages 430–435. IEEE.

[151] Stokman, F. N. and de Vries, P. H. (1988). Structuring knowledge in a graph. In *Human-Computer Interaction: Psychonomic Aspects*, pages 186–206. Springer.

[152] Tang, J., Sun, J., Wang, C., and Yang, Z. (2009). Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 807–816.

[153] Tang, L., Sun, Z., Idnay, B., Nestor, J. G., Soroush, A., Elias, P. A., Xu, Z., Ding, Y., Durrett, G., Rousseau, J. F., et al. (2023). Evaluating large language models on medical evidence summarization. *npj Digital Medicine*, 6(1):158.

[154] Tang, Y., Shi, Y., and Xiao, X. (2015). Influence maximization in near-linear time: A martingale approach. In *Proceedings of the 2015 ACM SIGMOD international conference on management of data*, pages 1539–1554.

[155] Tchechmedjiev, A., Fafalios, P., Boland, K., Gasquet, M., Zloch, M., Zapilko, B., Dietze, S., and Todorov, K. (2019). Claimskg: A knowledge graph of fact-checked claims. In *The Semantic Web–ISWC 2019: 18th International Semantic Web Conference, Auckland, New Zealand, October 26–30, 2019, Proceedings, Part II 18*, pages 309–324. Springer.

[156] Teng, Y.-W., Shi, Y., Tai, C.-H., Yang, D.-N., Lee, W.-C., and Chen, M.-S. (2021). Influence maximization based on dynamic personal perception in knowledge graph. In *2021 IEEE 37th International Conference on Data Engineering (ICDE)*, pages 1488–1499. IEEE.

[157] Tian, S., Mo, S., Wang, L., and Peng, Z. (2020). Deep reinforcement learning-based approach to tackle topic-aware influence maximization. *Data Science and Engineering*, 5(1):1–11.

[158] Touri, M. and Koteyko, N. (2015). Using corpus linguistic software in the extraction of news frames: towards a dynamic process of frame analysis in journalistic texts. *International Journal of Social Research Methodology*, 18(6):601–616.

[159] Truică, C.-O. and Apostol, E.-S. (2022). MisrobÆrta: Transformers versus misinformation. *Mathematics*, 10(4).

[160] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008.

[161] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., and Bengio, Y. (2018). Graph attention networks. In *International Conference on Learning Representations*, volume 1050, page 4.

[162] Villa, G., Pasi, G., and Viviani, M. (2021). Echo chamber detection and analysis. *Social Network Analysis and Mining*, 11(1):1–17.

[163] Vlachos, A. and Riedel, S. (2014). Fact checking: Task definition and dataset construction. In Danescu-Niculescu-Mizil, C., Eisenstein, J., McKeown, K., and Smith, N. A.,

editors, *Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science*, pages 18–22, Baltimore, MD, USA. Association for Computational Linguistics.

[164] Walter, D. and Ophir, Y. (2019). News frame analysis: An inductive mixed-method computational approach. *Communication Methods and Measures*, 13(4):248–266.

[165] Wang, G., Frederick, R., Haghighi, B. T., Wong, W., Rupar, V., Li, W., and Bai, Q. (2024). Framedtruth: A frame-based model utilising large language models for misinformation detection. In *Asian Conference on Intelligent Information and Database Systems*. Springer.

[166] Wang, G., Li, W., Bai, Q., and Lai, E. M.-K. (2023a). Maximizing social influence with minimum information alteration. *IEEE Transactions on Emerging Topics in Computing*, 12(2):419–431.

[167] Wang, G., Li, W., Lai, E., and Jiang, J. (2022). KATSum: Knowledge-aware abstractive text summarization. In *Proceedings of Principle and practice of data and Knowledge Acquisition Workshop 2022*, Shanghai, China.

[168] Wang, G., Li, W., Lai, E. M.-K., and Bai, Q. (2025). Aspect-adaptive knowledge-based opinion summarization. In *Knowledge Management and Acquisition for Intelligent Systems*, pages 29–41, Singapore. Springer Nature Singapore.

[169] Wang, G., Li, W., Wu, S., Bai, Q., and Lai, E. M.-K. (2023b). Beecd: Belief-aware echo chamber detection over twitter stream. In *Pacific Rim International Conference on Artificial Intelligence*, pages 307–319. Springer.

[170] Wang, G., Smetannikov, I., and Man, T. (2020). Survey on automatic text summarization and transformer models applicability. In *2020 International Conference on Control, Robotics and Intelligent System*, pages 176–184.

[171] Wang, J., Zheng, V. W., Liu, Z., and Chang, K. C.-C. (2017a). Topological recurrent neural network for diffusion prediction. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 475–484. IEEE.

[172] Wang, Q., Mao, Z., Wang, B., and Guo, L. (2017b). Knowledge graph embedding: A survey of approaches and applications. *IEEE transactions on knowledge and data engineering*, 29(12):2724–2743.

[173] Wang, Y., Shen, H., Liu, S., Gao, J., and Cheng, X. (2017c). Cascade dynamics modeling with attention-based recurrent neural network. In *IJCAI*, pages 2985–2991.

[174] Wei, F., Li, W., Lu, Q., and He, Y. (2008). Query-sensitive mutual reinforcement chain and its application in query-oriented multi-document summarization. In *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '08, page 283–290, New York, NY, USA. Association for Computing Machinery.

[175] Wellman, B. and Berkowitz, S. D. (1988). *Social structures: A network approach*, volume 15. CUP Archive.

[176] Westen, D. (1999). The scientific status of unconscious processes: Is freud really dead? *Journal of the American Psychoanalytic Association*, 47(4):1061–1106.

[177] Wu, Q., Gao, Y., Gao, X., Weng, P., and Chen, G. (2019). Dual sequential prediction models linking sequential recommendation and information dissemination. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 447–457.

[178] Xiong, F., Liu, Y., Zhang, Z.-j., Zhu, J., and Zhang, Y. (2012). An information diffusion model based on retweeting mechanism for online social media. *Physics letters A*, 376(30-31):2103–2108.

[179] Xu, J., Gan, Z., Cheng, Y., and Liu, J. (2020). Discourse-aware neural extractive text summarization. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5021–5031, Online. Association for Computational Linguistics.

[180] Xu, W. and Wu, W. (2020). *Optimal social influence*. Springer.

[181] Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., and Zhu, T. (2020). Public discourse and sentiment during the covid 19 pandemic: Using latent dirichlet allocation for topic modeling on twitter. *PloS one*, 15(9):e0239441.

[182] Yang, C., Sun, M., Liu, H., Han, S., Liu, Z., and Luan, H. (2018). Neural diffusion model for microscopic cascade prediction. *arXiv preprint arXiv:1812.08933*.

[183] Yang, X., Li, Y., Zhang, X., Chen, H., and Cheng, W. (2023). Exploring the limits of chatgpt for query or aspect-based text summarization. *arXiv preprint arXiv:2302.08081*.

[184] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., and Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, 32.

[185] Yao, J., Dou, Z., and Wen, J.-R. (2020). Employing personal word embeddings for personalized search. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1359–1368.

[186] Zhang, D., Hsu, C.-H., Chen, M., Chen, Q., Xiong, N., and Lloret, J. (2013). Cold-start recommendation using bi-clustering and fusion for large-scale social recommender systems. *IEEE Transactions on Emerging Topics in Computing*, 2(2):239–250.

[187] Zhang, J., Zhao, Y., Saleh, M., and Liu, P. (2020). Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. In *International Conference on Machine Learning*, pages 11328–11339. PMLR.

[188] Zhang, Z., Wang, H., Wang, C., and Fang, H. (2015). Modeling epidemics spreading on social contact networks. *IEEE transactions on emerging topics in computing*, 3(3):410–419.

[189] Zhong, M., Liu, P., Chen, Y., Wang, D., Qiu, X., and Huang, X. (2020). Extractive summarization as text matching. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6197–6208, Online. Association for Computational Linguistics.

[190] Zhou, F., Xu, X., Trajcevski, G., and Zhang, K. (2021). A survey of information cascade analysis: Models, predictions, and recent advances. *ACM Computing Surveys (CSUR)*, 54(2):1–36.

[191] Zhou, Q., Yang, N., Wei, F., and Zhou, M. (2017). Selective encoding for abstractive sentence summarization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1095–1104.

[192] Zhu, C., Hinthorn, W., Xu, R., Zeng, Q., Zeng, M., Huang, X., and Jiang, M. (2021). Enhancing factual consistency of abstractive summarization. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 718–733, Online. Association for Computational Linguistics.