Detection Method of Abnormal Text Information in Social Networks Integrating GUIDE and NOTICE

Ying Liu[1] and Jian Liu[2,*]

[1]*Department of Economics and Management, Wenhua College, Wuhan 430074, China*
[2]*Business Academy, Xianning Vocational Technical College, Xianning 437100, China*
E-mail: liuujianj@outlook.com
[*]*Corresponding Author*

Received 16 September 2025; Accepted 13 December 2025

Abstract

The proliferation of abnormal text information in social networks has become an important challenge for digital social governance. Traditional detection methods are unable to cope with increasingly complex semantic camouflage and dissemination strategies due to excessive reliance on one-dimensional analysis. Therefore, this research develops a detection method for abnormal text information in social networks that integrates the Graph-based User Interaction and Diffusion Evaluation (GUIDE) module and the Natural Language Ontology-driven Textual Anomaly Classification Engine (NOTICE) module. The GUIDE module captures anomalous propagation patterns through dynamic propagation tracking and network modeling, while the NOTICE module identifies semantic risks using a multilingual ontology library and deep semantic understanding. By combining structural and semantic analysis through a dual-attention fusion mechanism, the proposed framework simultaneously detects semantic anomalies and

Journal of Cyber Security and Mobility, Vol. 15_1, 95–122.
doi: 10.13052/jcsm2245-1439.1514

propagation topology anomalies, thereby improving detection accuracy and practicality. The experimental results show that the framework achieves F1 score of 91.2%, 89.7%, and 88.3% in detecting fake news, junk advertising, and hate speech tasks, respectively, which is 5.5–17.8 percentage points higher than the optimal baseline model. These evaluations are conducted on a comprehensive dataset from ZN Lab, containing real-world samples from major platforms like Twitter and Weibo. In actual deployment, the system maintains an accuracy rate of 89.4% when processing 230 million pieces of content per day, and reduces manual review by 43%. In terms of resource consumption, the memory usage remains stable at 645 MB and the response time is 76 ms, significantly better than traditional models. The above results indicate that the proposed model has excellent accuracy and applicability in detecting abnormal text information in social networks, effectively solving the problem of lack of accuracy and reliability in current detection methods. It provides an efficient and reliable technical solution for content governance on social platforms, especially in scenarios such as false information prevention and network violence governance, which has important application value.

Keywords: Social network security, anomaly detection, multimodal fusion, graph neural networks, BERT, fake news detection, hate speech, attention mechanism.

1 Background

With the explosive growth of social media, abnormal textual information in social networks has become a major challenge for digital social governance [1]. According to statistics, global social platforms add over 100 million abnormal texts daily, posing a serious threat to the security of the online ecosystem [2]. Effective detection of such information is crucial for maintaining the health and development of social networks. The current mainstream methods primarily fall into two categories. The first is semantic analysis grounded in deep learning, while the second is structural analysis based on graph neural networks. However, both categories exhibit one-dimensional limitations: semantic models struggle to capture anomalies in propagation topology, and structural models fail to deeply comprehend textual risks [3]. The detection of abnormal text information in social networks mainly revolves around two aspects: accurately capturing the structural characteristics of abnormal propagation and effectively identifying the potential risks of text content. By analyzing network models, the

Graph-based User Interaction and Diffusion Evaluation (GUIDE) module can effectively identify abnormal structural patterns in information propagation. The Natural Language Ontology-driven Textual Anomaly Classification Engine (NOTICE) module, based on multilingual ontology libraries and deep semantic understanding, can effectively identify potential risks in text content [4, 5]. The rationale for integrating GUIDE and NOTICE is grounded in their inherent complementary nature. In social networks, abnormal information often exhibits a duality: it may be semantically camouflaged to evade content-based filters, or it may be propagated in a structurally anomalous pattern that appears normal when viewed in isolation. The GUIDE module excels at detecting the latter by modeling user interaction networks and diffusion pathways, but may be fooled by well-disguised text. Conversely, the NOTICE module is adept at piercing through semantic camouflage via deep ontology-driven understanding, but might miss content that spreads through subtle, coordinated behaviors rather than overtly harmful language. By integrating them, the framework achieves cross-domain validation: the structural anomalies identified by GUIDE provide contextual reinforcement to semantic suspicions raised by NOTICE, and vice versa. This synergistic interaction creates a more robust and holistic detection system that is greater than the sum of its parts. To address the identified research gap and leverage the complementary strengths of structural and semantic analysis, this study aims to achieve the following objectives: (1) To design and implement a novel, deeply-integrated GUIDE-NOTICE fusion framework that enables real-time, collaborative analysis of both propagation topology and textual semantics. (2) To develop a dual-attention mechanism within the framework that can dynamically quantify and fuse the evidence of anomalies from both structural and semantic perspectives. (3) To empirically evaluate the proposed framework against state-of-the-art baseline models across multiple social network platforms and types of abnormal content (e.g., fake news, hate speech). (4) To validate the practical applicability and scalability of the system in terms of detection accuracy, computational efficiency, and generalization ability. Therefore, the research aims to address the challenge of detecting increasingly complex and covert abnormal text information in social networks. The research constructs a detection framework that integrates GUIDE and NOTICE. By integrating the propagation structure anomaly analysis capability of the GUIDE module with the deep semantic understanding capability of the NOTICE module, a more comprehensive approach to text analysis can be achieved. This integrated method enables more accurate and robust recognition of abnormal text. Consequently, it improves detection accuracy, reduces

the manual review burden, and enhances the system's usability in practical large-scale environments. While existing hybrid models such as TextGCN-BERT and HAN have demonstrated the value of combining structural and semantic information, the proposed GUIDE-NOTICE framework introduces several key algorithmic and conceptual advancements: (1) Native Co-design vs. Post-hoc Fusion: Unlike approaches that sequentially combine pre-trained GNN and BERT models (e.g., using BERT embeddings as node features in GNNs), GUIDE and NOTICE are architecturally co-designed from the ground up. They feature a dual-attention mechanism that enables real-time, bidirectional information exchange between structural and semantic analysis, rather than merely concatenating their outputs. (2) Dynamic Propagation Modeling: While models like HAN capture static graph structures, GUIDE specifically implements dynamic propagation tracking that models the temporal evolution of information diffusion. This allows for detecting anomalies in dissemination patterns that static graph snapshots would miss. (3) Ontology-Driven Semantic Understanding: NOTICE moves beyond standard BERT embeddings by incorporating a multilingual ontology library and rule-based deep validation. This enables fine-grained classification of text into predefined anomaly categories, with the ontology being dynamically updatable to adapt to emerging threats. (4) Cross-Modal Evidence Reinforcement: The fusion mechanism explicitly models the mutual reinforcement between structural and semantic evidence. Suspicious propagation patterns (detected by GUIDE) can trigger deeper semantic analysis, while semantically risky content (flagged by NOTICE) can prompt re-examination of its dissemination pathway, creating a synergistic detection effect.

2 Literature Review

The detection of abnormal text information in social networks is a core challenge for network security and content governance. With the popularity of social media, abnormal text presents characteristics of multi-modality, strong concealment, and rapid evolution. Domestic and foreign scholars have conducted research on it. For example, Khan W et al. proposed an innovative model that integrates graph convolutional networks, deep residual learning, and residual attention mechanisms to address the limited performance of abnormal node recognition in complex social networks. A dynamic attention mechanism based on residual information was proposed by extracting the structural features of the network through graph convolutional networks. The experimental results showed that the proposed model could effectively

solve the problem of limited performance in identifying abnormal nodes in complex networks [6]. Sufi F K et al. proposed a fully automated network space intelligent application to address the low efficiency of social network monitoring and news reporting analysis worldwide. By integrating multi-source news data and artificial intelligence (AI) technology, a real-time global threat map was constructed to help users remotely perceive dangerous areas. The experimental results showed that this application could greatly improve detection efficiency [7]. Ravichandran B D et al. addressed the issue of insufficient accuracy and efficiency in the classification system of false information on social media. A combination method of neural fuzzy system and neural network was proposed. By combining the logical reasoning ability of neural fuzzy and the representation learning advantage of neural networks, a hybrid model of adaptive neural fuzzy inference system and deep neural network (ANFIS-DNN) was proposed to improve classification accuracy and robustness. The experimental results indicated that the hybrid model exhibited higher potential in theoretical analysis [8]. Madani M et al. proposed a two-stage detection model to address the issues of event dynamism, validation difficulties, and dataset limitations in existing methods for detecting fake news on social networks. Combining natural language processing and machine learning techniques during the process to improve the accuracy and robustness of false news detection. The experimental results showed that the model outperformed benchmark methods, especially in diversity and small data scenarios [9]. Zkik K et al. proposed an integrated AI detection model to address issues such as fraud risks and advanced persistent threats in current social networking systems. They used graph neural networks to analyze the code logic and transaction graph of smart contracts, detected contract layer vulnerabilities such as re-entry attacks and infinite loops, and deployed supervised and unsupervised learning algorithms. The experimental results showed that the proposed model could effectively detect threats [10].

Since the development of GUIDE and NOTICE methods, some of their theories and practical applications have become relatively mature, and scholars from many countries have conducted in-depth research on them. For example, Wu K et al. proposed a GUIDE-based interactive perception multi-modal trajectory prediction framework to address the challenge of vehicle trajectory prediction in congested scenarios on multi lane highways. They proposed a time-varying graph model that abstracts vehicle motion as nodes, quantified the interaction strength between vehicles through a dynamic adjacency matrix, and used a diffusion graph convolutional network to simultaneously capture spatial topological relationships and temporal evolution

features. The experimental results showed that the proposed framework could effectively predict trajectories [11]. Chen X et al. proposed an optimization model for recommendation systems based on group social diffusion to address the issue of low efficiency in information recommendation modules in social recommendation systems such as Weibo. During the process, GUIDE was incorporated to design a heterogeneous ternary graph neural network that captures both binary and ternary relationships between users, items, and groups. The experimental results showed that the proposed model could effectively solve the problem of low module efficiency [12]. Li Y et al. proposed a diversified recommendation model based on graph diffusion to address the problem of difficulty in improving the diversity of recommended content in current graph neural network recommendation systems. During the process, GUIDE was combined and an innovative self gating mechanism was introduced to accurately capture subtle information interactions between users and items. The recommendation accuracy was significantly improved through the information reinforcement module. The experimental results showed that the proposed model could improve the diversity of recommended content [13]. Ci Y et al. proposed an improved ramp control method based on wavelet neural network for optimizing the traffic efficiency of urban expressways. During the process, NOTICE was combined to propose a chicken flock optimization algorithm optimized wavelet neural network, enhancing the accuracy of traffic prediction. The experimental results showed that the proposed method could effectively solve the optimization problem of urban expressway traffic efficiency [14]. Jesi P M et al. proposed an IoT energy efficiency optimization method based on multi expansion convolutional neural networks and composite motion optimization to address the energy efficiency issues of smart city IoT sensor networks. By combining composite NOTICE, an innovative multi expansion convolutional neural network was proposed, which captures multi-scale network features through convolution kernels with different expansion rates. The experimental results showed that the proposed method could solve the energy efficiency problem of sensor networks [15].

Recent studies by domestic and foreign scholars have continued to refine detection methodologies. For example, Yang Y et al. proposed RosGas, a reinforced self-supervised GNN architecture search framework for adaptive social bot detection. This work highlighted the trend of leveraging network structure and automated machine learning to address dynamic network environments [16]. In a different vein, Bacanin N et al. addressed feature selection and model tuning for phishing website detection using a diversity-oriented

social network search algorithm, demonstrating the application of advanced metaheuristics in security tasks [17].

While these studies make progress in their respective focal areas – structural dynamics and feature optimization – they frequently prioritize methodological sophistication in one particular dimension, often at the cost of neglecting integrated analysis. For instance, certain methods excel at capturing topological behavior with remarkable precision, yet they may fall short in achieving a profound understanding of textual semantics. Conversely, other approaches are adept at optimizing feature sets efficiently, but they fail to deeply incorporate the textual semantic layer that is inherently embedded within the content itself. This delineation of structural and semantic analysis pathways reveals a clear research gap: the need for a deeply integrated framework that performs synergistic, rather than parallel, analysis.

Despite these advancements, existing hybrid or multimodal approaches for abnormal text detection often remain suboptimal. Firstly, many models perform a merely superficial fusion of structural and semantic features, such as late-stage score averaging or simple feature concatenation, which fails to capture the deep, synergistic interactions between propagation patterns and linguistic nuances. Secondly, they frequently lack a dedicated mechanism to model the dynamic and temporal evolution of anomaly propagation, treating the network structure as static. Thirdly, their semantic understanding components are often not robust enough against adversarial textual camouflage and struggle with cross-lingual or cross-platform generalization. Consequently, there exists a crucial research void for a profoundly integrated framework capable of simultaneously and adaptively modeling the dynamic propagation topology while also conducting robust and fine-grained semantic analysis in a seamlessly cohesive manner.

In summary, existing research on abnormal text detection in social networks still faces issues of insufficient comprehensiveness and accuracy. To this end, an innovative GUIDE-NOTICE multi-modal deep learning framework is proposed, which integrates the topology analysis of the GUIDE module with the deep semantic understanding of the NOTICE module to achieve comprehensive and accurate detection of abnormal texts. The GUIDE module utilizes dynamic propagation tracking and adaptive cross platform module recognition technology to effectively capture anomalous diffusion patterns. The NOTICE module integrates multilingual semantic parsing and adversarial training strategies to deeply identify text content risks. This framework is expected to provide reliable technical solutions for content security on social platforms.

3 Methods and Materials

3.1 Characteristics Analysis and GUIDE Module Design of Abnormal Text in Social Networks

Abnormal text information in social networks refers to text data that deviates significantly from normal user-generated content in terms of content characteristics, dissemination patterns, or publishing behavior. Such information may cause negative effects, including spam ads, false information, hate speech, online violence, and other illegal content. This type of information has three essential characteristics: non-regularity, potential harm, and deliberate avoidance. In terms of detection dimension, it is manifested as multiple features of content dimension, behavior dimension, and network dimension. With the evolution of technology, detection methods have progressed from early keyword filtering to the current multi-modal fusion detection. However, these methods still confront dual challenges stemming from language complexity and environmental dynamics. Particularly in the context of the AI boom, unusual text detection necessitates key technological breakthroughs, such as multi-modal collaborative analysis, adversarial robustness, interpretable detection, and edge computing, to meet increasingly complex detection demands [18]. The schematic diagram of abnormal text information in social networks is shown in Figure 1.

As shown in Figure 1, abnormal texts in social networks can be classified into five mutually exclusive types based on multi-level ontologies: spam, false information, other violations, inappropriate speech, and online violence. The attribute data and complex interactions in social networks provide key information for identifying abnormal nodes. The fusion of multi-modal features and graph structure analysis for abnormal text node detection aims to accurately identify high-risk nodes with both semantic anomalies and propagation heterogeneity, providing core technical support for the construction of a trustworthy network space. The traditional anomaly detection calculation process is shown in Equation (1).

$$(1-\alpha)\|S-\hat{S}\|_F^2+\alpha\|X-\hat{X}\|_F^2 \tag{1}$$

In Equation (1), F represents the Frobenius norm. α represents weight coefficients. S represents the original matrix. $\hat{S}$ represents estimated values. X represents auxiliary matrices. $\hat{X}$ indicates auxiliary estimation value. After calculating the high-order structure of the network, the loss function of node attributes is calculated as shown in Equation (2).

$$\mathcal{L}=(1-\alpha)R_S+\alpha R_A \tag{2}$$

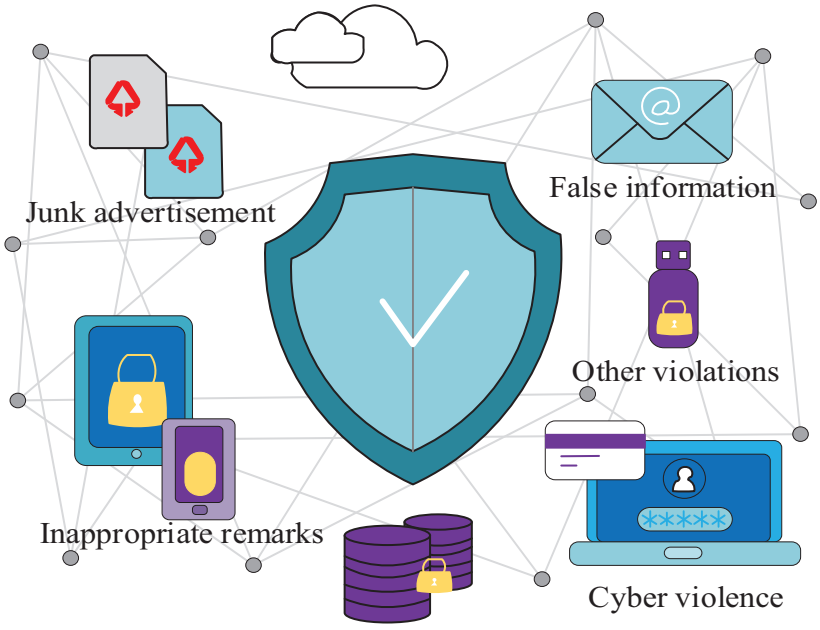


Figure 1 Anomaly text messages in social networks.

In Equation (2), $\mathcal{L}$ represents the loss function. R_S represents the source loss term. R_A indicates auxiliary loss item. Equations (1) and (2) represent the traditional matrix factorizationbased anomaly detection objective, serving as the baseline. The intuition is to learn a compact network representation that minimizes reconstruction error, where poorly reconstructed instances (high error) are flagged as anomalies. Equation (2) enhances robustness by incorporating an auxiliary loss term with additional constraints. Network modules refer to specific subgraph patterns that repeatedly appear in complex networks and have statistical significance, with a frequency significantly higher than expected in random networks [19]. The network module, as the core functional module of the system, carries and analyzes the underlying formation mechanism and functional characteristics of complex networks. The nodal degree of the model is shown in Figure 2.

As shown in Figure 2, the combination of polar coordinate system and heat map visualization is used to demonstrate the distribution characteristics of node degrees in the modular structure of the network. The radial gradient of spacing degree reveals the topological layout from the core to the edge of the network. The difference in connection strength between module boundary

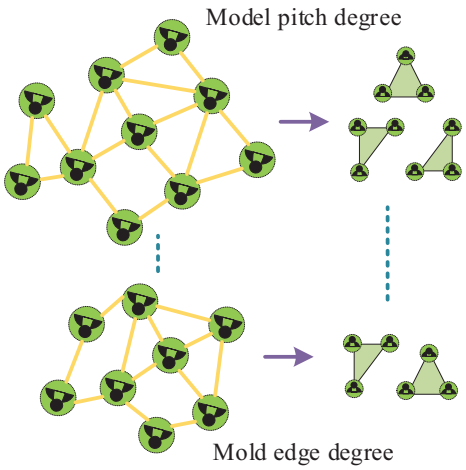


Figure 2 Modular nodes and degrees.

nodes is quantified by edge degree. The statistical significance test of its model is calculated as shown in Equation (3).

$$Z_M = \frac{N_{\text{real}}(M) - \langle N_{\text{rand}}(M) \rangle}{\sigma_{\text{rand}}(M)} \tag{3}$$

In Equation (3), Z_M is the Z fraction of module $M \cdot N_{\text{real}}(M)$ is the number of occurrences of module M in the real network. $\sigma_{\text{rand}}(M)$ is the standard deviation of the number of occurrences of module M in the randomized network. Next, the frequency deviation test is carried out, as shown in Equation (4).

$$\Delta F_M = \frac{F_{\text{abn}}(M) - F_{\text{norm}}(M)}{F_{\text{norm}}(M)} \times 100\% \tag{4}$$

In Equation (4), ΔF_M represents the frequency change rate of the module M. $F_{\text{abn}}(M)$ indicates the frequency of occurrence of the module M in abnormal states. $F_{\text{norm}}(M)$ indicates the frequency of appearance of the module M under normal conditions. Finally, the correlation degree of abnormal nodes is calculated, and the weight of the model node correlation is calculated as shown in Equation (5).

$$w(v, M) = \frac{\text{Deg}_M(v)}{\sum_{u \in V} \text{Deg}_M(u)} \cdot \log(1 + Z_M) \tag{5}$$

In Equation (5), $w(v, M)$ represents the weight related to node v and module M. $\mathrm{Deg}_M(v)$ indicates the degree of node v in the substructure related to module M. $\sum_{u \in V} \mathrm{Deg}_M(u)$ is the sum of degrees of all nodes u (belonging to set V, which is a set of nodes) in the substructure related to the module M. Traditional social network anomaly text detection methods have limitations such as being easily bypassed by adversarial text, difficult to capture dynamic propagation, and weak cross platform generalization ability due to excessive reliance on semantic analysis. The GUIDE module innovatively adopts network module analysis to construct an integrated framework that integrates structure semantic evaluation, dynamic propagation tracking, and transferable module libraries. It can achieve topology based camouflage recognition, real-time discovery of burst propagation chains, and lightweight cross platform adaptive detection. The GUIDE model framework diagram is shown in Figure 3.

As shown in Figure 3, the overall framework diagram of the GUIDE model adopts modular design, presenting the complete process from data input to result output. The framework mainly consists of three parts: The input layer on the left contains raw data of social networks and a prioritized list of nodes. The dual channel encoding layer in the middle is responsible for encoding node attributes and text semantic features to highlight its core role, as well as extracting network topology and propagation mode features. The Attribute Decoder on the right outputs interpretable anomaly detection

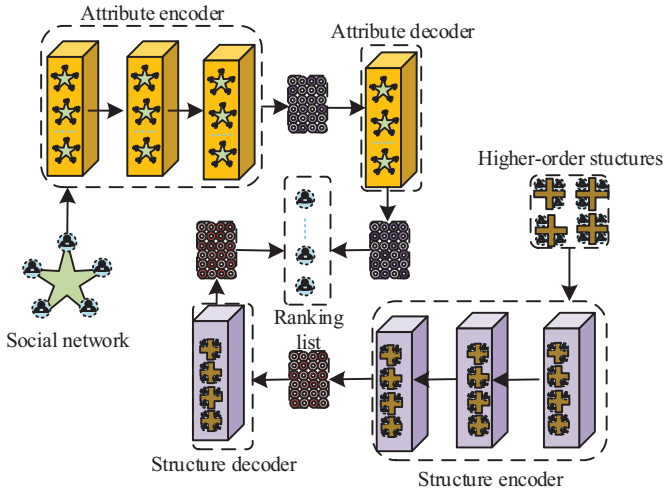


Figure 3 GUIDE model framework diagram.

results. The data flow emphasizes that the GUIDE model achieves end-to-end anomaly detection through a parallel processing mechanism of "semantic structural dual encoding" (SEM-STR Fusion). The study reconstructs node attributes using autoencoders to detect attribute anomalies. An autoencoder network layer implements this function, as shown in Equation (6).

$$H^{(l+1)} = \sigma(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{6}$$

In Equation (6), $H^{(l+1)}$ represents the node feature matrix of the l th layer. σ represents the activation function. D is the degree matrix of the graph, and A is the adjacency matrix of the graph. By using the Relu function as the activation function, the formula for the autoencoder network layer can be simplified, as shown in Equation (7).

$$H^{(l+1)} = f_{\mathrm{Re}lu}(\bar{A}H^{(l)}W^{(l)}) \tag{7}$$

In Equation (7), $f_{\mathrm{Re}lu}$ represents the ReLU activation function. $\bar{A}$ represents the normalized graph adjacency matrix. Equations (6) and (7): Graph Autoencoder Layer. These equations define a graph autoencoder layer in the GUIDE module that reconstructs node attributes through localized, weighted aggregation of neighborhood features. By learning to recreate attributes from their network context, it identifies anomalies as nodes exhibiting high reconstruction error – indicating semantic-structural misalignment. The study innovatively proposes a "semantic structural bimodal fusion" detection framework, which effectively identifies disguised text and abnormal propagation patterns through network module analysis and dynamic propagation tracking technology of GUIDE model. Adopting a dual encoder single decoder architecture, collaborative analysis of node attributes and topological features is achieved, significantly improving the recognition ability of adversarial texts.

3.2 Collaborative Mechanism of GUIDE-NOTICE Fusion Detection Framework

The GUIDE framework proposed in the study overcomes the limitations of traditional single semantic analysis through the "semantic structural bimodal fusion" mechanism, achieving realtime detection of temporal evolution anomalies. However, there are still problems such as insufficient recognition of emerging semantic variations, difficulty in distinguishing fine-grained semantics, and limited cross language adaptability. To this end, the NOTICE

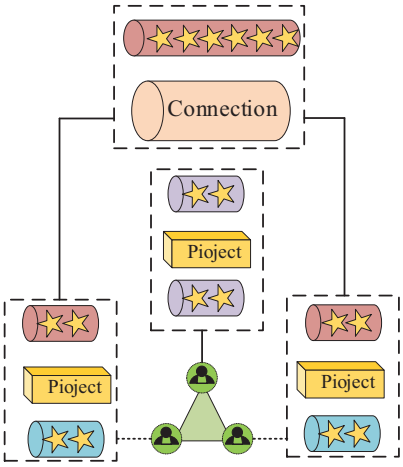


Figure 4 Schematic diagram of node aggregation process in network model.

module is introduced, which is based on deep semantic understanding and a multilingual ontology library. It quantifies the contributions of different modules through fine-grained module feature weighting and dynamically adjusts fusion weights, significantly improving the ability to distinguish semantic anomalies and forming a complementary complete detection system with GUIDE [20]. The node aggregation process in its network model is shown in Figure 4.

As shown in Figure 4, the multi-level architecture of node aggregation process in the network model clearly presents the feature integration mechanism from micro connections to macro project formation through bottom-up visualization. The "Connection" layer at the bottom represents the original node interaction network, while the three parallel "Project" modules in the middle cluster nodes through differentiated weights, ultimately generating compressed high-order feature representations at the top. In social network anomaly detection, there are often significant differences in the contribution of different propagation patterns to the degree of node anomaly. In order to more accurately quantify this differential contribution, the study used a model level attention mechanism based on testing and weight calculation, and derived an abnormal score for attention enhancement, as shown in Equation (8).

$$\mathrm{AbnScore}'(v) = \sum_{M \in \mathcal{M}(v)} \alpha_M \cdot \mathrm{AbnScore}_M(v) \tag{8}$$

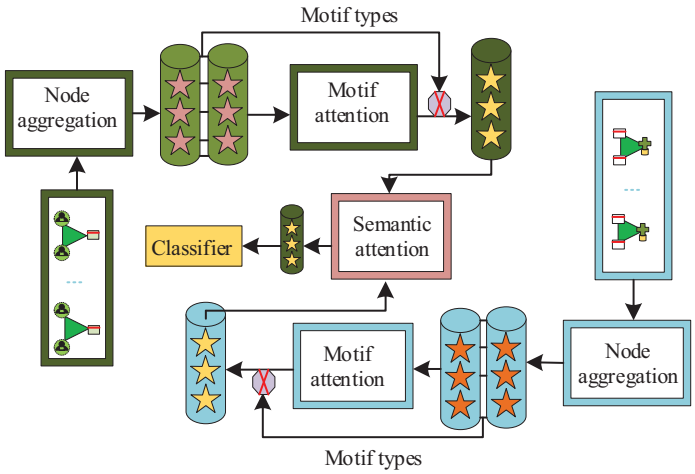


Figure 5 NOTICE framework diagram.

In Equation (8), $\mathrm{AbnScore}'(v)$ represents the final anomaly score of node v; $\mathcal{M}(v)$ represents a set of models related to node v; α_M represents the weight coefficient related to element M; $\mathrm{AbnScore}_M(v)$ represents the anomaly score of node v calculated based on element M. The study performs cross attention calculation with the semantic anomaly signals of the NOTICE module as shown in Equation (9).

$$\beta = \sigma(\alpha_M \cdot S_{\text{semantic}}(v)) \tag{9}$$

In Equation (9), β represents the output value of anomaly detection; α_M represents the weight of M in relevant calculations; $S_{\text{semantic}}(v)$ represents the semantic score of node v. Equations (8) and (9): Attention-based Fusion Mechanism. These equations form the core of our dual-attention fusion mechanism. Equation (8) dynamically weights structural evidence to compute an anomaly score, enhancing detection of diverse patterns. Equation (9) adaptively fuses this structural signal with NOTICE's semantic score, enabling context-aware decisions through bidirectional feedback – a key innovation of our framework. The innovation of the NOTICE framework lies in the construction of a dynamically evolving domain ontology library, which combines expert knowledge with the semantic understanding ability of deep language models, breaking through the bottleneck of traditional methods in fine-grained semantic analysis. The NOTICE framework is shown in Figure 5.

As shown in Figure 5, the NOTICE framework is based on a multi-level analysis architecture of network modules, which provides a complete processing chain from network feature extraction to anomaly classification through modular processes. The core consists of two processing layers: the basic processing layer achieves network topology compression and functional unit extraction through "Node aggregation", and calculates the saliency weights of different modules in conjunction with "module attention". The semantic fusion layer aligns the structural features with the text ontology through "semantic attention", and finally outputs anomaly detection results through "classifier". The innovation of the framework is reflected in the dual attention collaborative mechanism - module attention captures abnormal propagation patterns, semantic attention parses deep semantics of text, and both achieve accurate recognition of adversarial text through iterative optimization. To accurately map text to a predefined anomalous ontology space, NOTICE adopts a semantic projection method based on Bidirectional Encoder Representation from Transformers (BERT), which calculates the similarity between text and ontology concepts to achieve fine-grained classification, as shown in Equation (10).

$$\phi(t) = \sum_{c \in \mathcal{C}} \sin(t, c) \cdot w_c \tag{10}$$

In Equation (10), $\phi(t)$ represents the output result regarding variable t; $\sin(t, c)$ indicates the similarity between variable t and element C; w_c represents weight vector related to element C. Equation (10): Semantic Projection and Ontology Matching. This equation describes how the NOTICE module maps a text snippet to the ontology space for fine-grained classification. The text is classified not by a standard classifier head but by its semantic proximity to predefined conceptual centers in the ontology. This allows for interpretable, fine-grained categorization and enables zero-shot or few-shot learning for new anomaly concepts by simply adding new prototype vectors to the ontology, enhancing the model's adaptability. For multilingual scenarios, NOTICE utilizes multilingual BERT to achieve a unified representation of concept embeddings, eliminating the impact of language differences on detection, as shown in Equation (11).

$$v_c^{mul} = AvgPool(\{mBERT(c)_{en}, mBERT(c)_{zh}, \ldots\}) \tag{11}$$

In Equation (11), v_c^{mul} represent the multilingual representation vectors related to the element c; AvgPool represents average pooling operation;

$mBERT(c)_{en}$ represents the vector representation obtained by encoding element C in English context using a multilingual $BERT$ model; $mBERT(c)_{zh}$ is the vector representation obtained by encoding element c in Chinese context using a multilingual BERT model.For semantic anomalies with strong concealment, NOTICE conducts deep validation through ontology rule sets, as shown in Equation (12).

$$C(t) = 1 - \frac{\sum_{r \in \mathcal{R}} \mathbb{I}(t \vDash r)}{|\mathcal{R}|} \tag{12}$$

In Equation (12), $C(t)$ represents the calculation result regarding variable t; $\mathcal{R}$ represents the number of elements in collection $\mathcal{R}$. For multilingual scenarios, NOTICE adopts an online learning mechanism to dynamically adjust ontology weights to adapt to emerging anomaly patterns, as shown in Equation (13).

$$\Delta w_c = \eta \cdot \frac{\partial \mathcal{L}_{\text{adapt}}}{\partial w_c} \tag{13}$$

In Equation (13), Δw_c represents the update quantity related to parameter w; η represents the learning rate; $\mathcal{L}_{adapt}$ indicates the adaptive loss. By using the GUIDE module to capture propagation structure anomalies, the NOTICE module parses semantic ontology violations. Finally, a GUIDE-NOTICE fusion framework is proposed to achieve comprehensive detection through the following innovative mechanisms. Figure 6 presents the structure of the fusion GUIDE and NOTICE.

As shown in Figure 6, the GUIDE-NOTICE fusion framework constructs a core heterogeneous network structure for abnormal text detection in social networks. The structure establishes a multi-modal anomaly detection system architecture through the complex interactions between three types of nodes: users, information and publishing institutions, and three types of propagation models: “user information user”, “information user information”, and “information publishing institution information”. The framework presents the complex relationship between user behavior, content dissemination, and institutional publishing through differentiated visual coding, intuitively demonstrating the interaction dynamics between users and information, the ownership relationship between information and publishing institutions, and mapping the social connections between users. Through the deep interaction and feature fusion between the NOTICE semantic analysis unit and the GUIDE module detection unit, this framework achieves native collaboration at both the structural and semantic levels. It can synchronously capture

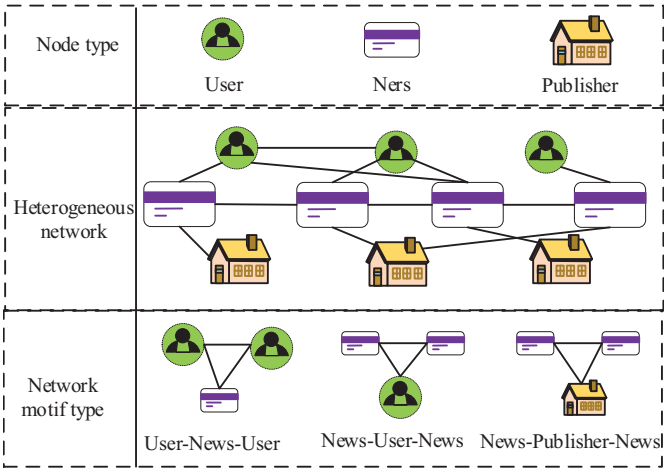


Figure 6 GUIDE and NOTICE fusion structure.

semantic anomalies in information texts and the topological anomalies exhibited during their propagation, thereby achieving accurate recognition and cross validation of abnormal text information in social networks at both the semantic and structural dimensions. By dynamically integrating the propagation anomaly evidence from the GUIDE module with the semantic violation evidence from the NOTICE module, a more robust joint judgment can be achieved, and the comprehensive anomaly probability is shown in Equation (14).

$$P_{abn}(v) = \alpha \cdot AbnScore'(v) + (1 - \alpha) \cdot S_{semantic}(v) \tag{14}$$

In Equation (14), $P_{abn}(v)$ represents the final anomaly probability of node v; $AbnScore'(v)$ represents the fusion based anomaly score of node ν; $S_{semantic}(v)$ represents the semantic score of node v.

4 Validation Analysis of GUIDE-NOTICE Detection Method for Abnormal Text in Social Networks

4.1 Performance Validation of Abnormal Text Information Detection in GUIDE-NOTICE Social Networks

Experiments were conducted on a large-scale, multi-source dataset from ZN Lab, containing 2.35 million posts from platforms like Twitter and Weibo. To address the realistic class imbalance detailed in Table 1, the Synthetic

Table 1 Dataset composition

Category	Size	Proportion (%)	Source Platforms
Fake news	850,000	36.2%	Twitter, Weibo
Junk advertising	750,000	31.9%	Weibo
Hate speech	450,000	19.1%	Twitter
Other violations	200,000	8.5%	Twitter, Weibo
Normal content	100,000	4.3%	Twitter, Weibo
Total	2,350,000	100%	

Table 2 Experimental equipment and parameter configuration

Device Type	Configuration Parameter	Domain/Key Functions
Computing node	2× Intel Xeon Gold 6248R	Large-scale graph data processing
GPU accelerator	4× NVIDIA RTX 3090	Model training/inference
Memory	512GB DDR4 ECC	The graph structure is memory-resident
Storage	8TB NVMe SSD (RAID 0)	High-speed data reading and writing
Deep learning framework	PyTorch 1.10 + CUDA 11.3	Implementation of the GUIDE-NOTICE model
Graph computing library	DGL 0.8 + PyG 2.0	Module analysis and graph convolution
NLP Toolkit	HuggingFace Transformers 4.18	"NOTICE semantic processing"
Distributed scheduling	Ray 1.13	Multi-gpu task parallelism

Minority Over-sampling Technique (SMOTE) was applied. The dataset was split 70/15/15 for training, validation, and testing, with standard text preprocessing (tokenization, lowercasing, removal of user mentions, URLs, and special characters).

To evaluate the detection performance of abnormal text information in social networks that integrate GUIDE and NOTICE, this study compared it with traditional detection models such as Bidirectional Encoder Representation from Transformers Convolutional Neural Network (BERT-CNN), Text Graph Convolutional Network (TextGCN), and Graph Attention Network (GAT). The research experimental data were sourced from the ZN laboratory. The experimental equipment and parameter configuration are shown in Table 2.

Based on the above parameters, this study measured the detection performance of the model by comparing the comprehensive performance of

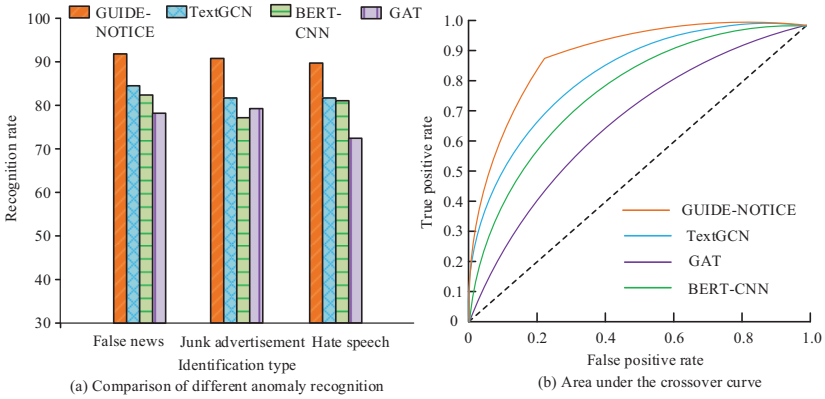
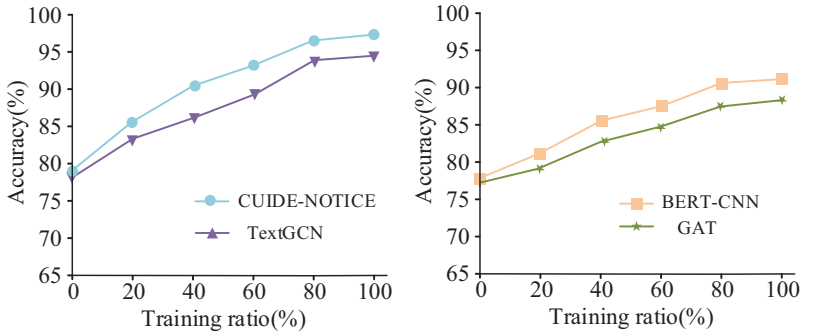


Figure 7 Comprehensive performance of baseline method.

baseline methods, F1 score, detection accuracy, and error rate. Firstly, the comprehensive performance of the baseline method was compared, and the experimental results are shown in Figure 7.

Figure 7(a) indicates the comparative experimental results between GUIDE-NOTICE detection system and mainstream models. The data showed that in the three types of abnormal content detection tasks of fake news, junk advertising, and hate speech, the GUIDE-NOTICE system (91.2%, 89.7%, 88.3%) outperformed the BERT-CNN (82.3%, 76.5%, 79.1%), GAT (78.6%, 81.2%, 73.4%), and TextGCN (85.7%, 83.4%, 82.6%) baseline models comprehensively. Especially in the detection of fake news on Twitter, GUIDE-NOTICE significantly led with an accuracy of 91.2%, an improvement of 5.5 percentage points compared to the optimal baseline model (85.7% for TextGCNs). From Figure 7(b), the GUIDE-NOTICE curve in the cross curve was closest to the upper left corner, with a detection value of 0.886 closest to 1, significantly higher than the comparison models' 0.842, 0.786, and 0.723, demonstrating better detection performance. Overall, these results confirmed that the system achieved a more comprehensive and stable performance improvement in the field of abnormal content recognition through innovative design that integrates multi-modal features and attention mechanisms. To comprehensively evaluate the predictive performance of each model, the study further compared the prediction accuracy of each model, and the results are shown in Figure 8.

As shown in Figure 8(a), the GUIDE-NOTICE model outperformed TextGCN in all training proportion ranges, especially in the low to medium training proportion range of 20%–60%, where the advantage was most



(a) Comparison of accuracy rates between CUIDE-NOTICE and TextGCN

(b) Comparison of accuracy between BERT-CNN and GAT

Figure 8 Comparison of prediction accuracy among models.

significant, and its accuracy improved more significantly with increasing data volume. As shown in Figure 8(b), the overall performance of BERT-CNN exceeded that of GAT. When the training ratio exceeded 40%, the gap between the two gradually widened, while GAT performed well at low training ratios but had limited subsequent improvement. Overall, the performance of each model was GUIDE-NOTICE>TextGCN>BERT-CNN>GAT, and the GUIDE-NOTICE proposed in the study demonstrated the best data efficiency and accuracy stability.

4.2 Practical Application Effect of Abnormal Text Information Detection in GUIDE-NOTICE Social Network

To verify its deployment readiness and scalability, the study conducted an efficiency analysis of the GUIDE-NOTICE model. As shown in Table 3, the proposed model had 125 parameters, which was comparable to BERT-CNN and significantly less than GAT. Each inference required 3.2 GFLOPs. Although the dual-module architecture led to longer one-time training costs, this model performed well in terms of the efficiency of key operations in actual deployment. It achieved a fast inference time of 76 ms and maintained a stable memory usage (∼645 MB), demonstrating a good trade-off between model complexity and runtime performance, and confirming its strong applicability to large-scale real-time scenarios.

The computational resource consumption was further validated by comparing the GUIDE-NOTICE model with TextGCN, BERT-CNN, and GAT models in terms of GPU computation time and peak memory usage, as shown in Figure 9.

Table 3 Model complexity and training efficiency comparison

Model	Parameters (M)	GFLOPs	Training Time (Hours)	Inference Time (ms)	Peak Memory (MB)
GUIDE-NOTICE (Ours)	12.5	3.2	18.5	76	645
TextGCN	8.1	1.5	5.2	138	1250
BERT-CNN	11.8	4.5	12.1	95	980
GAT	25.4	6.8	9.8	155	2100

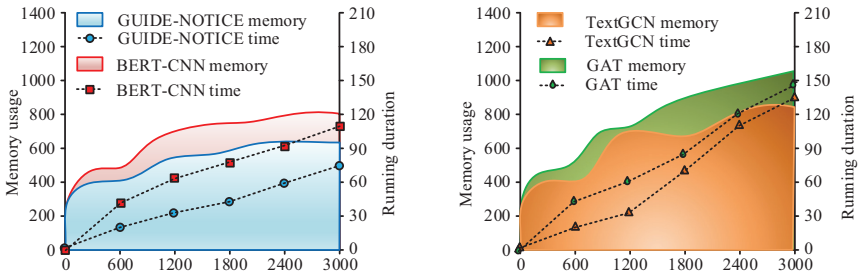


Figure 9 Average GPU operation time per day and peak memory usage.

As shown in Figure 9(a), the GUIDE-NOTICE model experienced a slow increase in memory usage as the data volume increased, ultimately stabilizing at around 645 MB. The GPU computation time was linearly related to the amount of data, and when the amount of data was 3000, the computation time was 76 milliseconds. From Figure 9(b), as the data volume increased, the memory usage of TextGCN and GAT models sharply rose, with a response time of 138 milliseconds. Overall, the GUIDE-NOTICE model had less performance fluctuation when dealing with large amounts of abnormal data, and had more stable and efficient data processing capabilities. Afterwards, a comparative verification of the error rate of the GUIDE-NOTICE model was conducted, and the results are shown in Figure 10.

As shown in Figure 10(a), GUIDE-NOTICE consistently performed the best, with significantly better prediction errors than TextGCN (-1.0 m) and BERT-CNN (-0.8 m) in large data scenarios (-0.6 m). In small data scenarios, the error was further reduced to -0.4 m, which was 33% higher than TextGCN. The research results indicated that GUIDE-NOTICE had the advantage of detection accuracy, especially in resource limited small data scenarios, providing reliable performance guarantees for practical applications.

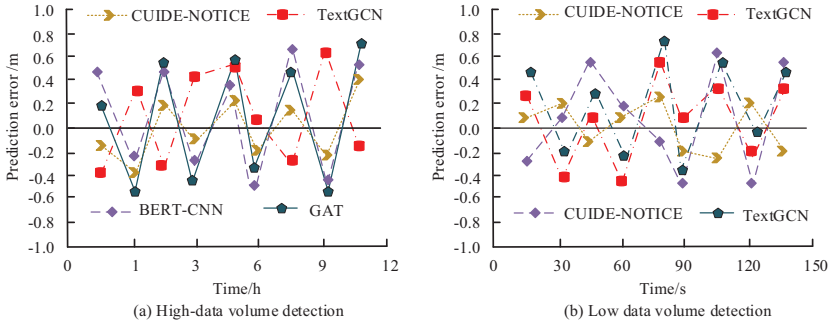


Figure 10 Error rate comparison and verification of GUIDENOTE model.

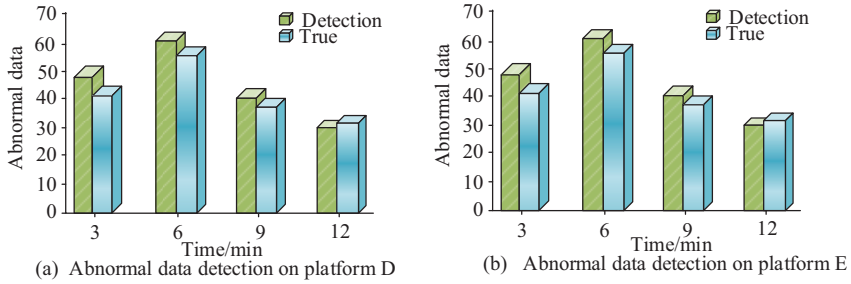


Figure 11 Anomaly data detection of platform D and platform E.

Finally, to validate the generalization ability of the GUIDE-NOTICE model, cross platform testing was conducted to demonstrate the model's adaptability to system heterogeneity. The study selected D platform and E platform for comparative experiments, and the experimental results are shown in Figure 11.

As shown in Figure 11(a), the model detection results of platform D indicated that the error between the detection results and the actual abnormal data was small. As shown in Figure 11(b), the model detection results of the E platform indicated that the error between the detection results and the actual abnormal data was also small. Overall, the actual detection performance of the GUIDE-NOTICE model was excellent. The above results indicated thatthe GUIDE-NOTICE model performed well in terms of GPU computation time and peak memory usage, and had good generalization ability, which can play a good role in social network anomaly detection.

To rigorously evaluate the individual contributions of the GUIDE and NOTICE modules and validate the necessity of their integration, the study conducted a comprehensive ablation study. The study compared

Table 4 The test results of the anomaly detection task on the test set

Model Variant	Fake News	Junk Advertising	Hate Speech	Macro Avg. F1
GUIDE-NOTICE (Full Model)	91.2	89.7	88.3	89.7
w/o NOTICE	83.5	90.1	72.4	82.0
w/o GUIDE	91.0	75.8	86.9	84.6

the performance of the following model variants: (1) GUIDE-only: A model utilizing only the graph-based propagation structure analysis module. (2) NOTICE-only: A model utilizing only the ontology-driven semantic understanding module. (3) GUIDE-NOTICE (Full Model): The complete proposed framework with the dual-attention fusion mechanism. The results on the test set across the three primary anomaly detection tasks are summarized in Table 4.

The ablation study provided key insights into the model's design: the NOTICE-only module excelled in fake news detection by leveraging deep semantic analysis, while the GUIDE-only module performed strongly on junk advertising by identifying anomalous propagation patterns. The fusion of both modules proved particularly critical for detecting hate speech, which required both semantic nuance and structural context, with the full model achieving the highest Macro Average F1 score across all categories. This demonstrated a clear synergistic effect – the integrated framework not only compensated for each module's limitations but also delivered overall performance superior to either module alone, confirming that the dual-attention mechanism successfully unified structural and semantic evidence into a more robust and comprehensive detection system.

5 Conclusion

Aiming at the dual challenges of semantic concealment and complex propagation patterns faced by abnormal text information detection in social networks, an innovative dual-mode analysis framework integrating GUIDE and NOTICE was proposed. This framework achieved multi-dimensional and accurate recognition of abnormal text by utilizing multi-modal deep learning techniques, integrating the network model propagation analysis of GUIDE module and deep semantic understanding of NOTICE module. The experimental results showed that the F1 score of this framework in detecting fake news, junk advertising, and hate speech tasks reached 91.2%, 89.7%, and 88.3%, respectively, which was 5.5–17.8 percentage points higher than the

traditional optimal model. Actual deployment testing showed that the system maintained an accuracy rate of 89.4% under the pressure of processing 230 million pieces of content per day, and reduced manual review workload by 43%. Overall, the fusion of GUIDE and NOTICE bimodal analysis framework had good accuracy in detecting abnormal text information in social networks, and its performance could meet the needs of detecting abnormal text information in social networks. Although the research model performed well in performance and application effect, the lightweight deployment of the model in edge computing scenarios was not fully verified in this research. In the future, further improvements and upgrades will be made to enhance the applicability and practicality of the framework, continuously improving the predictive universality and detection performance of the model.

References

[1] Silva G R S, Canedo E D. Towards user-centric guidelines for chatbot conversational design. International Journal of Human–Computer Interaction, 2024, 40(2): 98–120. DOI: 10.1080/10447318.2022.2118244.

[2] Chefer H, Alaluf Y, Vinker Y, Wolf L, and Cohen-Or D. Attend-and-excite: Attention-based semantic guidance for text-to-image diffusion models. ACM transactions on Graphics (TOG), 2023, 42(4): 1–10. DOI: 10.1145/3592116.

[3] Chung W, Zhang Y, Pan J. A theory-based deep-learning approach to detecting disinformation in financial social media. Information Systems Frontiers, 2023, 25(2): 473–492. DOI: 10.1007/s10796-022-10327-9.

[4] Balshetwar S V, Rs A, R D J. Fake news detection in social media based on sentiment analysis using classifier techniques. Multimedia tools and applications, 2023, 82(23): 35781–35811. DOI: 10.1007/s11042-023-14883-3.

[5] Zhao, J., An, K. and Wang, X. 2024. Research on Fast Early Warning of False Data Injection Attack in CPS of Electric Power Communication Network. Journal of Cyber Security and Mobility. 13, 6 (Nov. 2024), 1331–1356. DOI: 10.13052/jcsm2245-1439.1365.

[6] Khan W, Mohd A, Suaib M, Ishrat M, Shaikh A A, and Faisal S M. Residual-enhanced graph convolutional networks with hypersphere mapping for anomaly detection in attributed networks. Data Science and Management, 2025, 8(2): 137–146. DOI: 10.1016/j.Dsm.2024.09.002.

[7] Sufi F K, Alsulami M, Gutub A. Automating global threat-maps generation via advancements of news sensors and AI. Arabian Journal for Science and Engineering, 2023, 48(2): 2455–2472. DOI: 10.1007/s13369-022-07250-1.

[8] Ravichandran B D, Keikhosrokiani P. Classification of Covid-19 misinformation on social media based on neuro-fuzzy and neural network: A systematic review. Neural Computing and Applications, 2023, 35(1): 699–717. DOI: 10.1007/s00521-022-07797-y.

[9] Madani M, Motameni H, Roshani R. Fake news detection using feature extraction, natural language processing, curriculum learning, and deep learning. International Journal of Information Technology & Decision Making, 2024, 23(03): 1063–1098. DOI: 10.1142/S0219622023500347.

[10] Zkik K, Sebbar A, Fadi O, Kamble S, and Belhadi A. Securing blockchain-based crowdfunding platforms: an integrated graph neural networks and machine learning approach. Electronic Commerce Research, 2024, 24(1): 497–533. DOI: 10.1007/s10660-023-09702-8.

[11] Wu K, Zhou Y, Shi H, Li X & Ran B. Graph-based interaction-aware multi-modal 2D vehicle trajectory prediction using diffusion graph convolutional networks. IEEE Transactions on Intelligent Vehicles, 2023, 9(2): 3630–3643. DOI: 10.1109/TIV.2023.3341071.

[12] Chen X, Xie R, Qiu Z, Cui P, Zhang Z, Liu S,& Lin L. Group-based social diffusion in recommendation. World Wide Web, 2023, 26(4): 1775–1792. DOI: 10.1007/s11280-022-01079-2.

[13] Li Y, Zhao M, Zhang J, Xie Z, Liu Y & Zhan Q. GDDRec: graph neural diffusion model for diversified recommendation. Knowledge and Information Systems, 2025, 67(5): 4401–4430. DOI: 10.1007/s10115-025-02348-y.

[14] Ci Y, Wu H, Sun Y, and Wu L A prediction model with wavelet neural network optimized by the chicken swarm optimization for on-ramps metering of the urban expressway. Journal of Intelligent Transportation Systems, 2022, 26(3): 356–365. DOI: 10.1080/15472450.2021.1890070.

[15] Jesi P M, Antony Asir Daniel V, Rajagopal R, and Femila L Cluster Head Selection Using Multi-Dilation Convolutional Neural Network Optimized with BCMO for IoT Networks. IETE Journal of Research, 2024, 70(8): 6702–6710. DOI: 10.1080/03772063.2024.2315208.

[16] Yang Y, Yang R, Li Y, Cui K, Yang Z, Wang Y, and Xie H. Rosgas: Adaptive social bot detection with reinforced self-supervised gnn

architecture search. ACM Transactions on the Web, 2023, 17(3): 1–31. DOI: 10.1145/3572403.

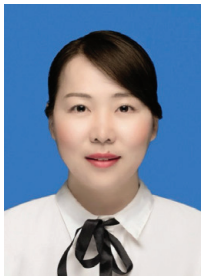
[17] Bacanin N, Zivkovic M, Antonijevic M, Venkatachalam K, Lee J, Nam Y, and Abouhawwash M. Addressing feature selection and extreme learning machine tuning by diversity-oriented social network search: an application for phishing websites detection. Complex & Intelligent Systems, 2023, 9(6): 7269–7304. DOI: 10.1007/s40747-023-01118-z.

[18] Krishnasamy B, Muthaiah L, Kamali Pushparaj J E, and Pandey P S. DIWGAN optimized with Namib Beetle Optimization Algorithm for intrusion detection in mobile ad hoc networks. IETE Journal of Research, 2024, 70(5): 4422–4441. DOI: 10.1080/03772063.2023.2223181.

[19] Bandewad G, Datta K P, Gawali B W, and Pawar, S. N. Review on Discrimination of Hazardous Gases by Smart Sensing Technology. Artificial Intelligence and Applications. 2023, 1(2): 86–97. DOI: 10.47852/bonviewAIA3202434.

[20] M. Hasanvand, M. Nooshyar, E. Moharamkhani, and A. Selyari. "Machine Learning Methodology for Identifying Vehicles Using Image Processing," AIA, vol. 1, no. 3, pp. 170–178, Apr, 2023, DOI: 10.47852/bonviewAIA3202833.

Biographies



Ying Liu obtained her Bachelor's degree in Computer Science and Technology from Central China Normal University in 2008 and earned her Master's degree in Information Science from the same university in 2011. Currently, she serves as a full-time faculty member in the Department of Economics and Management at Wenhua College. Her teaching responsibilities include courses such as Database Systems, New Media Operations, and

Business Data Analysis. She has also passed the Advanced Big Data Analyst examination. Her primary research interests focus on information resource management and digital transformation.

Jian Liu obtained his Bachelor's degree in E-commerce from Central China Normal University in 2004. Currently, he serves as a full-time faculty member at the Business Academy, XianningVocational Technical College. With years of teaching experience in e-commerce, he is dedicated to research on e-commerce platform operations and e-commerce marketing strategies. He has also undertaken various responsibilities related to corporate e-commerce operations and talent development in the field of e-commerce. His main research interests include new media marketing, agricultural product e-commerce, and e-commerce logistics.