JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Comprehensive Survey on Suppressing the Endogenous Negative Influence through Node Intervention in Social Networks

RESHMA B
Research Scholar,
Department of Computer Science
Sri Krishna Adithya College of Arts and Science, Coimbatore.
Dr.D.Shanmuga Priyaa
Assistant Professor,Department of Computer Science
Sri Krishna Adithya College of Arts and Science,Coimbatore

Abstract : Endogenous negative influence in social networks results from rumours, fake news, harmful opinions, negative emotions and aggressive behavior that are spread among connected users of the network. It is not so simple to suppress, however, because of the lack of integration with existing harmful content detection, need for intervention, the need for labelled data and the need for heavy models and complete network information for modelling of diffusion and identification of critical nodes. From 2024 to 2026 and 25 studies were reviewed. It included methods for misinformation detection machine-learning, deep-learning, transformer, graph-neural-network, reinforcement-learning, epidemic-diffusion and evolutionary-optimization; opinion modeling emotion recognition, influential-node selection, proactive defence, budget-constrained intervention. The comparative analysis showed high detection performance, but low interpretability and scalability, as well as generalisation and adaptability across platforms and in real time. The survey revealed that future systems must integrate semantic, structural, temporal, emotional and behavioural information into a coherent, explainable, adaptive system to track critical-spreaders, predict the spread and intervene with resources to prevent the propagation.

***Keywords* - Critical Node Identification, Endogenous Negative Influence, Graph Neural Networks, Misinformation Suppression, Node Intervention, Social Networks.**

I.INTRODUCTION

The online social networks found on the Internet have transformed the way information, opinions and emotions get created, transmitted and magnified. It encourage communication and information sharing, but it also help to spread rumours and fake news, false narratives and negative public opinion. Negative information can be created through user interactions and spread through complex network patterns, leading to endogenously generated negative influence, which can influence individual and collective decisions, public trust, social stability and social behavior. Hence, the identification of harmful content, modelling harmful content diffusion, identification of influential nodes and design of effective intervention strategies have become important research challenges.

Towards detecting rumours, attention graph adversarial dual contrast learning was proposed, which is a fusion of textual semantics and propagation structures. It reduced noisy-comment interference and acquired powerful representations from the content and graph relationships for improved identification of misinformation in social networks [1]. To enhance the fake-news detection in social media environments, a graph-augmented transformer ensemble framework was created. In the case of complex data sets, it used the contextual representation of text, graph information and ensemble learning to enhance the robustness, scalability and classification accuracy [2]. To detect fake-news, the authors employ transformer models to create a deep learning model for contextual language representation. A series of optimized BERT models were able to achieve semantic dependency in the news article, as well as to distinguish between misleading information and legitimate information better than those classical text-learning methods [3].To address the problems of opinion mining, a hybrid deep neural network model was proposed combining the BGRU model and LSTM model. The method trained on a set of reviews to learn the sequential and contextual patterns to enhance the recognition of negative sentiment [4].To model public-opinion dynamics in social networks, a GAN-SEIR model is introduced. It used a generative adversarial learning (GAN) framework and an epidemic-inspired diffusion modelling framework to model the temporal evolution of opinions, user interactions and user propagation patterns [5]. A text-based graph neural network-based rumour-detection method based on dual embeddings was developed. This approach was leveraged to interpret the content similarity and structural dependency for rumour propagation, using both semantic characteristics of text and the relationships between text and graph [6].To solve the influence maximization problem in large social networks, a new deep reinforcement learning approach is presented. In the method, the influential node selection was formulated as a sequential decision making problem and sequences of policies were learned for influential node selection to efficiently spread information [7].An influence-maximization approach based on deep reinforcement learning and graph attention networks (GANs) was introduced, which is suitable for any topic. It chose influential users based on the topical relevance and network structure for targeted and context-sensitive propagation of information [8].To solve the issue of influence maximization in social networks adaptively a graph convolution neural network was designed. The graph structures were changed and the model was trained on and then the high impact spreaders that can cover the information in the graph drastically were identified [9].A Rumour Detection Temporal Tree Transformer was proposed, which used the textual information, structural information and temporal information. It modelled propagation trees and changing interaction patterns to make a more accurate identification of rumours as it evolved in time [10].

The approaches considered the following studies as examples helped in the detection of harmful contents, public opinion modelling, temporal diffusion analysis and selection of influential nodes, but most of the approaches treated the separately. There was little focus on any one model which could be used to capture the endogenous negative impact, identify critical spreader nodes, predict the propagation dynamics and then offer actionable intervention strategies at limited resources. However, to block the endogenous negative influence in the dynamic networks, there is a need for a scalable approach, which is adaptive and explainable for integrating semantic, structural, temporal and behavioural information.

II. BACKGROUND STUDY

Roumeliotis et al. (2025) [11] did a comparative study on the usage of convolutional neural networks, BERT and GPT-based large language models for fake-news detection and classification. Text preprocessing, text model fine-tuning and performance comparison were performed on the WELFake dataset. But the structure of social networks, propagation pathways and influential users identification were not taken into account. The research identified a gap in the literature investigating the relationship between content accuracy and node level intervention. The results indicated that the large language models and BERT models were superior to the existing CNN model method.

To tackle the challenge of detecting fake news across news articles, social media posts and user-generated content, Mouratidis et al. (2025) [12] created a comprehensive machine-learning framework for the task. The study experimented with different machine-learning models and neural models using different text representations and datasets. But it is more about classification than harmful-information diffusion. The research gap focused on the usability of the results from the detection for the identification of critical nodes and the suppression action. The results showed that transformer-based models outperformed some of the standard approaches with regards to classification.

Papageorgiou et al. (2025) [13] proposed a state-of-the-art fake-news detection framework that is based on large language models (LLMs), deep neural networks (DNNs), fact-based segment extraction and graph-based representations of textual data. The approach used was: multiple datasets, contextual embeddings and factual features. However, these were not included for user interactions and for temporal propagation and intervention policies. The research gap was the relationship of the two representations (semantic and graph) with the control of the diffusion of social networks. The results demonstrated that the proposed fake-news detection method outperformed the existing feature-based methods in terms of accuracy and exhibited high interpretability.

Yin et al. (2026) [14] proposed a spatio-temporal deep reinforcement learning solution to pre-emptive misinformation defense on heterogeneous social network. This model has been designed to incorporate a temporal encoding to predict nodes prior to large propagation events that are at risk on a graph. It was only a success, however, with correct temporal activity pattern and network information. The first research gap was the ability to adapt to real world (non fully connected, rapidly changing) graphs. The findings showed that the misinformation containment was higher than static intervention tactics.

Aslam et al., (2025) [15] proposed a multi-view machine-learning framework that integrates textual, linguistic, semantic and metadata-based features, to tackle the issue of fake news. The Logistic Regression, Random Forest, LightGBM and ensembling were employed for the classification. However, the structure was primarily aimed at detecting spammy content with no examples provided for node propagation and even no node manipulation. The missing part from the study was the combination of multi-view detection and diffusion-aware suppression mechanism. The results highlighted the excellent performance of the ensemble method in fake news classification.

Furutani et al. 2024 [16] provided an endogenous negative influence suppression problem as a node intervention optimization problem in social networks. The idea that users are chosen, when there is a limited budget of interventions, for persuading, nudging or warning the to reduce negative opinion diffusion. But in large, dynamic networks, the challenge of the combinatorial optimization process in the computation arose. Research gaps identified were the development of real-time, adaptive learning based intervention policies that are scalable. The results showed that intervention nodes selected strategically are more effective in restricting the negative influence compared with conventional node-selection methods.

Xu et al. (2025) [17] introduced a proactive rumor-control model with graph neural network, greedy node selection and evolutionary optimization. The approach learned the graph representation and selected the nodes that were going to protect to minimize the spreading of the rumour, taking multiple exposure by users into account. The method, however, needed to have the structure and propagation information of the network, which may not be available in practical environments. The research gap was the ability to robustly control under partially observed networks, with uncertainty. The results showed improved rumor control and computing efficiencies.

Gu et al. (2025) [18] used a difference-in-differences model and a pre-trained BERT-based sentiment classifier to explore the spillover effect of auditor sanctions on investors' online communication. The method focused on the regulated Chinese interactive platforms and whether the investor questions and negative sentiments were analysed. This study, however, was limited to financial-auditing and did not consider general social network diffusion or node intervention. There was a research gap regarding transfer of sentiment-based influence analysis to broader networks. The results indicated that the negative response of investors to the action on the Internet was greatly influenced by the sanctions issued by the auditors.

TABLE I. COMPARATIVE ANALYSIS OF MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

Author, Year and Ref. No.	Concept	Method Used	Limitations	Results in Values
Anthony and Zhou (2025) [19]	Multi-label emotion recognition in crisis-related social-media posts	Twitter-RoBERTa transformer with two self-attention blocks and label-wise threshold tuning	The study used a relatively small and older SemEval-2018 dataset, focused mainly on English tweets and did not model user-to-user emotion diffusion or node intervention	Improved multi-label emotion recognition in crisis-related tweets.
Hashmi et al. (2024) [20]	Explainable fake-news detection in social-media environments	FastText embeddings with hybrid Bi-GRU–Bi-LSTM models and explainable artificial intelligence techniques	The framework depended on labelled datasets, language-specific preprocessing and computationally intensive hybrid architectures; cross-platform generalisation remained limited	Achieved highly accurate and explainable fake-news detection.
Zhang et al. (2026) [21]	Effect of mandatory IP-location disclosure on users' shared emotions on Weibo	Sentiment-dictionary analysis combined with regression discontinuity in time using 193,761 Weibo posts	The analysis focused on one platform, one policy event and depression-related posts; unobserved temporal factors could also have influenced emotional expression	IP-location disclosure reduced emotional expression and posting activity.
Ding et al. (2025) [22]	Comparison of machine-learning and deep-learning models for mental-illness detection on social media	Logistic Regression, Random Forest, LightGBM, GRU and ALBERT with binary and multiclass classification	The models had difficulty interpreting sarcasm, irony and implicit emotional meaning; deep-learning models required greater computational resources and offered lower interpretability	ALBERT achieved the strongest mental-illness detection performance.
Yan et al. (2025) [23]	Emotion recognition and emotion-propagation modelling in social-media networks	Emotion-RGC Net combining RoBERTa, Graph Neural Network and Conditional Random Field	The model depended on large labelled datasets, had high computational complexity and showed difficulty distinguishing overlapping fine-grained emotion categories	Sentiment140: 89.70% accuracy, 90% recall, 90% F1-score and 0.95 AUC; Emotion: 88.50% accuracy, 89% recall, 89% F1-score and 0.93 AUC
Al-Tarawneh et al. (2024) [24]	Fake-news detection using different word-embedding and classification combinations	TF-IDF, Word2Vec and FastText with SVM, Logistic Regression, Random Forest, Decision Tree, Naïve Bayes, KNN and MLP	Performance varied across embedding–classifier combinations; semantic embeddings did not consistently outperform TF-IDF and generalisation beyond the selected dataset was not established	TF-IDF–SVM method achieved high accuracy
Han et al. (2024) [25]	Detection of cyber-aggressive and trolling behaviour in social-media posts	Random Forest, LightGBM, Logistic Regression, SVM and Naïve Bayes using the Cyber-Troll dataset	The study reported identical confusion matrices for some classifiers, indicating possible experimental or data-processing limitations; advanced neural models were not fully evaluated	Random Forest: TP = 1,802, TN = 2,933, FP = 114 and FN = 152; derived accuracy: 94.68%; precision: 94.05%; recall: 92.22%; F1-score: 93.12%

The table I shows the comparison of some of the recent techniques for fake news, negative emotion, mental health and aggressive behaviour detection in social media. The general models such as transformer-based, recurrent, graph-based and existing machine-learning models generally yielded good performance. The majorities of the studies was labelled studies or data from a platform, but were not generalizable and were not easily interpretable. It did directly cover critical node identification or real-time suppression of negative influence.

A. Research Gap

The reviewed studies were effective on detecting fake news, recognising emotions, classification of mental health and analysing aggressive behaviours. Most approaches however, did not focus on the type of information, but considered harmful information as a classification problem, neglecting user relations, information propagation over time, user feelings and network topology. Current approaches used labelled and platform-specific data sets which could not be applied in general across domains and real-time environments. Some research studies discussed misinformation protection, critical-node selection and node intervention but most of the demanded much data from the network and high computational complexity. There was therefore a clear research gap to develop a scalable, explainable and adaptive framework to model endogenous negative influence, to identify critical spreader nodes, to simulate dynamic diffusion and to implement budget constrained interventions in incomplete and changing social networks.

III. Endogenous Negative Influence Suppression and Node Intervention Methods

To stop the spread of negative influence in social networks, a combination of the analysis of the formation of negative influence, connectivity of users, propagation behavior and possibilities of intervention are required. Recently, some approaches have focused on patterns of semantics, structure, time and behaviour to detect negative influence and to pinpoint users that have a strong influence on its propagation. Negative-influence modelling, critical node identification, diffusion control and adaptive intervention with limited resources are presented below. These are discussed in the following subsections: negative-influence modelling, critical-node identification, diffusion control and adaptive intervention with limited resources.

A. Endogenous Negative Influence Modeling in Social Networks

Rumour detection can therefore lead to endogenous negative influence that learns modelling of the process of creation and propagation of bad information by users connected in the network. Attention-based graph learning can learn important relations and adversarial dual contrast learning can provide better separation between the true information and the negative patterns of the social network that are linked to rumors. [1]

Graph-augmented transformer models: Endogenous negative influence through textual information along with the graph structure between users, posts and interactions. It might be helpful in understanding how false information may be fabricated within social media platforms and how network connections can be used to share false and problematic information. [2]

Transformer based deep learning can model the endogenous negative influence by identifying fake news using patterns of social media content in terms of its context, semantics and linguistic features. This allows identification of harmful information early on and provides the foundation for future identification, intervention and diffusion-control strategies for nodes. [3]

Endogenous negative influence is modeled using hybrid deep neural networks, based on opinions, sentiments and emotional expressions of users. The incorporation of multiple mechanisms of neural learning contributes to better recognition of negative attitudes and negative public reactions, which may partially explain how negative attitudes can emerge and spread in social-network communities. [4]

B. Critical Node Identification and Intervention Strategy

Deep reinforcement learning algorithms are able to learn which social media users can affect the most, in order to identify important nodes within the social network. The model is able to analyze the network structure, propagation characteristics and long-term rewards to determine nodes with high influence and optimize intervention under resource constraints and to effectively control the influence of the network. [7]

Topic-aware influence maximization is a graph attention network-based deep reinforcement learning method to find influential users using both network connectivity and topics of discussion. This approach can help identify nodes that have a greater influence in specific communities and information domains to help inform more targeted interventions. [8]

In order to identify critical nodes, adaptive graph convolution neural networks learn the structural and neighborhood features from the social graph directly. The model is also able to determine the influence potential of each node and which users can influence a large portion of the network, thus supporting targeted and efficient interventions in terms of computation. [9]

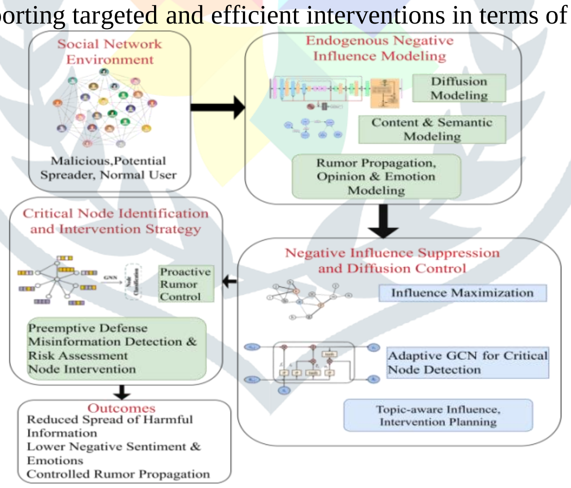

Fig. 1 Integrated Framework for Endogenous Negative Influence Suppression

Fig. 1 is a general model for harmful content, identifying critical spreader nodes and then using targeted intervention strategies. Semantic, emotional, structural and diffusion features are analyzed before selecting the node level control actions. The framework is to curb propagation of the negative information and enhance the stability of the social network.

C. Negative Influence Suppression and Diffusion Control

Predicting harmful diffusion before it starts spreading is used to suppress negative influence in a preemptive deep reinforcement learning approach. The framework acquires the right defensive reaction to the evolving network conditions and intervenes at the right time, improving the capacity to "contain" misinformation and reducing delays because of "reactive" interventions. [14]

Machine-learning approaches to combating fake news may also be useful in controlling the diffusion of misinformation by identifying misinformation and flagging its potential for harm before it is disseminated. If there are harmful pieces of information, the system can take corrective actions such as alert, verification, limitation of information and release of information etc. to minimize the negative social effect of the information. [15]

The node intervention paradigm directly suppresses endogenous negative influence by selecting some key nodes and changing the roles of these nodes in the diffusion process. When the approach is used, it can block or disrupt the transmission of harmful opinions at key junctures and stop the spread of negativity throughout the network. [16]

The main challenge of proactive rumor control is to adopt the graph neural networks and evolutionary optimization to find effective targets to control and to choose appropriate rumors. The method introduced in this survey strives to suppress the rumours in the network by limiting the spread of rumors early in the process, ensuring that the suppression is effective but with fewer nodes being needed in the network and available for the intervention. [17]

D. Adaptive and Budget-Constrained Node Intervention

When the node lacks resources, endogenous negative influence is controlled by choosing a small subset of strategic nodes for persuading, warning, nudging or controlling the participation in the game. The aim is to minimize unwanted diffusion as much as possible, while also minimizing the reduction from the helping effect of intervention resources available [16].

Scalable intervention-target selection in large social networks is possible by the use of graph neural networks along with greedy and evolutionary optimization. Learning the characteristics of these nodes and optimising the placement of protector nodes, the approach is able to enhance the containment of rumors and computational efficiency without affecting the entire network [17].

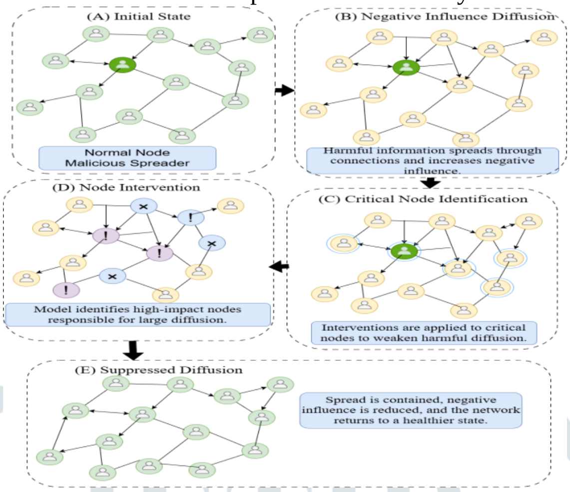

Fig. 2. Diffusion Process and Node Intervention

Fig. 2, the malicious node transmits the bad information by using the connected users. High impact spreaders and/or bridge nodes are then identified and controlled by issuing warnings, restrictions, blocking and/or providing counter-information. These interventions help to minimize negative exposure and put the network back to a healthier state.

Table II. Comparison of Critical Node Identification and Negative Influence Control Methods

Ref. No.	**Method / Model**	**Target Problem**	**Critical Node Strategy**	**Intervention Type**	**Suppression Objective**	**Strength**
[7]	MaDGNN: GNN with DRL	Large-scale influence maximization	Sequential seed-node selection using learned embeddings and rewards	Seed-node activation	Maximize expected influence spread	Scalable across large networks
[8]	DRL with Attentive GNNs	Topic-aware influence maximization	Selects topic-relevant seed nodes using graph attention	Topic-aware seed selection	Maximize topic-specific influence spread	Integrates topic and structural information
[9]	Adaptive Graph Convolution Neural Network	Influence maximization	Predicts node importance and selects top-(K) seed nodes	Seed-node selection	Increase influence spread	Considers influence overlap
[14]	Spatio-Temporal DRL	Misinformation defence	Uses spatial and temporal diffusion states to select defence actions	Pre-emptive defence	Reduce misinformation propagation	Addresses delayed and misplaced intervention
[16]	Node-Intervention Optimization	Endogenous negative influence	Selects intervention nodes under a limited budget	Direct node intervention	Minimize negative influence spread	Exact match to the topic
[17]	GNN with Greedy and Evolutionary Optimization	Proactive rumor control	Learns graph features and optimizes control targets	Proactive rumor intervention	Reduce rumor diffusion	Scalable ML-based rumor control

The techniques for identifying critical nodes, defending against misinformation, controlling rumors and endogenous negative influence in social network are compared in Table II. It provides information about the models, target problems, node selection strategies, types of intervention, suppression goals and most important features of each intervention method.

IV. Machine Learning Techniques for Negative Influence Suppression

Three of the most important and widely discussed applications of machine-learning techniques in social networks are negative influence detection and suppression, which have been assigned to this task. These techniques aid in the timely detection of fake news, misinformation, negative emotions and opinion. Classifiers, ensemble learning techniques, transformers and word embeddings help to improve the accuracy of detection and reduce manual monitoring tasks. These are evaluated against each other in terms of the support for node intervention and diffusion control in Table III.

Table III. Comparative Analysis of Machine Learning Techniques for Negative Influence Suppression

Ref. No.	Method / Model	Dataset	Process	Novelty / Key Contribution	Results
[11]	CNN, BERT, GPT-4o	WELFake: 72,134 articles	Text preprocessing, fine-tuning and comparison	Compares CNN, BERT and LLMs for fake-news detection	GPT-4o achieved the highest accuracy, followed by BERT.
[12]	SVM, RF, CNN, LSTM, BERT	Four fake-news datasets	Feature extraction, balancing and classification	Compares ML and DL across multiple datasets	BERT: 0.99 F1; Passive-Aggressive: 0.992 F1; SVM: 0.962 F1
[15]	LR, RF and LightGBM ensemble	Kaggle: 44,898 articles; LIAR: 12,791 statements	Multi-view feature extraction and stacked ensemble	Combines textual, linguistic and semantic feature views	The ensemble achieved the highest detection accuracy.
[22]	LR, SVM, RF and LightGBM	52,681 social-media statements	TF-IDF, grid search, binary and multiclass classification	Compares predictive performance, interpretability and efficiency	RF binary: 0.9347 F1, 0.9764 AUC; LR multiclass: 0.7474 F1; RF: 0.7454 F1
[24]	SVM, LR, NB, RF and MLP	TruthSeeker: 180,000+ tweets	TF-IDF, Word2Vec and FastText classification	Compares embedding–classifier combinations	SVM with TF-IDF achieved the highest accuracy.

A comparison of selected machine learning methods employed in fake-news detection, misinformation analysis and identification of negative content in social networks is presented in Table III. The comparison is based on the models and datasets used, the processing techniques, the key contributions and the numerical performance. The results indicate that ensemble, transformer and embedding-based methods result in high detection accuracy, enabling the identification of harmful information at early stages before applying node-level intervention and diffusion-control strategies.

V. Deep Learning Techniques for Negative Influence Suppression

The detection and analysis of harmful information, negative opinions and emotional content in social networks is increasingly vital for deep learning techniques. Complex semantic, structural, temporal and relational patterns can be captured by models based on transformers, recurrent neural networks, graph neural networks and explainable learning. These methods enhance the automatic detection and decrease the workload for mass social media surveillance. The relative effectiveness of early node intervention and diffusion control is summarized in Table IV.

Table IV. Comparison of Deep Learning Techniques for Negative Influence Suppression

Ref. No.	Method / Model	Dataset	Process	Novelty / Key Contribution	Results
[1]	AGAD: BERT, GAT and adversarial dual contrast learning	Twitter15, Twitter16 and PHEME	Text–graph encoding, attention filtering, adversarial training and rumor classification	Integrates semantic and propagation features while reducing the effect of noisy comments	Accuracy improvement over GACL: 4.1% on Twitter15, 3.0% on Twitter16 and 3.2% on PHEME
[3]	Fine-tuned BERT transfer-learning model	WELFake: 72,134 articles	Tokenization, contextual feature learning, fine-tuning and binary classification	Uses contextual transformer representations to identify fake and misleading news	Accuracy: 95.3%; precision: 95.2%; recall: 95.4%; F1-score: 95.3%
[4]	Hybrid BGRU–LSTM	IMDB and Amazon review datasets	Class balancing, bidirectional encoding, sequential	Combines BGRU and LSTM to improve negative-	Accuracy: 95%; negative-class recall: 96%; misclassification loss: 13.3%

			learning and sentiment classification	opinion recognition	
[10]	Temporal Tree Transformer with GRU	PHEME	Text encoding, propagation-tree construction, temporal-window learning and rumor classification	Jointly captures semantic, structural and temporal rumor characteristics	Accuracy: 75.84%; Macro F1-score: 71.98%
[20]	FastText with Bi-GRU–Bi-LSTM and XAI	AFND and ArabicFakeTweets	FastText embedding, bidirectional sequence learning, classification and explainability analysis	Combines semantic embeddings, hybrid recurrent learning and interpretable fake-news detection	Achieved high accuracy on both datasets AFND and ArabicFakeTweets
[23]	Emotion-RGC Net: RoBERTa, GNN and CRF	Sentiment140 and Emotion	Context extraction, relational graph learning, sequence modelling and emotion classification	Integrates semantic context and user relationships for social-media emotion recognition	Sentiment140: 89.70% accuracy, 90% F1; Emotion: 88.50% accuracy, 89% F1

Table IV shows the comparison of deep learning techniques applied for rumor, fake-news and sentiment and emotion detection in social networks. It presents an overview of the models, datasets, processing steps, important contributions and numerical performance. These approaches help detect early signs of problematic information and negative feelings before focusing on intervening with specific nodes and diffusion-control strategies.

VI. DISCUSSION

This survey showed that suppressing endogenous negative influence in social networks requires the coordinated integration of harmful-content detection, diffusion modelling, critical-node identification and targeted intervention. Existing methods achieved strong performance in recognising rumours, fake news, negative emotions and aggressive behaviour, but most approaches treated these tasks independently. The practical use was further limited by dependence on labelled and platform-specific datasets, incomplete network information, high computational complexity, limited interpretability and weak real-time adaptability. The survey therefore highlighted the need for a unified, scalable and explainable framework that can analyse semantic, structural, temporal, emotional and behavioural information together. Such a framework should identify influential spreaders early, predict evolving diffusion patterns and select cost-effective intervention actions under limited resources.

VII. CONCLUSION

In this survey, some strategies for suppressing the negative endogenous influence that exists in social networks by detecting harmful contents, modelling the diffusion of harmful contents, identifying the critical nodes and intervening on them are discussed. Machine-learning, deep-learning, transformer, graph-neural-network, reinforcement-learning and optimization methods were able to identify rumours, fake news, negative emotions, aggressive behaviour and influential spreaders. Most of the studies however focused on these tasks separately and used labeled platform specific datasets, full network information and models that require a lot of computational resources.It were not very interpretable, lacked cross-domain generalisation and were not very adaptable in real time for use in dynamic social settings. From the evidence reviewed, a unified framework of semantic, structural, temporal, emotional and behavioural information is necessary for effective suppression. Future work should thus be directed toward scalable, explainable models that are applicable when observing only a subset of nodes in the network, towards predicting changing diffusion patterns, determines the critical spreaders and towards allocating budgets for interventions in an efficient manner. These fully integrated systems can enhance early warning, proactively fight misinformation, implement node intervention with limited resources and manage harmful influence in complex social networks.

VIII. REFERENCES

[1] Zhang, B., Liu, T., Ke, Z., Li, Y., & Silamu, W. (2024). Rumor detection based on attention graph adversarial dual contrast learning. PLOS ONE, 19(4), e0290291. DOI: 10.1371/journal.pone.0290291

[2] Kumar, C., Bansal, M., Khan, M. A., Kaushik, V., Arquam, M., & Alabdultif, A. (2025). Graph-augmented transformer ensemble framework for robust and scalable fake news detection in social media ecosystems. Scientific Reports, 15, 19341. DOI: 10.1038/s41598-025-31653-3

[3] Raza, N., Abdulkadir, S. J., Abid, Y. A., Albouq, S. S., Alwadain, A., Rehman, A. U., et al. (2025). Enhancing fake news detection with transformer-based deep learning: A multidisciplinary approach. PLOS ONE, 20(9), e0330954. DOI: https://doi.org/10.1371/journal.pone.0330954

[4] Hidri, A., Alsaif, S. A., Alahmari, M., AlShehri, E., & Sassi Hidri, M. (2025). Opinion mining and analysis using hybrid deep neural networks. Technologies, 13(5), 175. DOI: 10.3390/technologies13050175

[5] Wang, J., Yin, Y., & Wei, L. (2025). Modeling public opinion dynamics in social networks using a GAN-SEIR framework. Social Network Analysis and Mining, 15(1), 40. DOI: 10.1007/s13278-025-01426-x

[6] Pattanaik, B., Kumar, P., & others. (2024). Rumor detection using dual embeddings and text-based graph neural networks. Discover Computing. DOI: https://doi.org/10.1007/s44163-024-00193-6

[7] Yang, F., Wang, Y., Shu, N., et al. (2026). A deep reinforcement learning framework for influence maximization problem on large-scale social networks. Scientific Reports, 16, 11515. DOI: 10.1038/s41598-026-41731-9

[8] Halal, T., Cautis, B., Groz, B., & Gao, R. (2025). Topic-aware influence maximization with deep reinforcement learning and graph attention networks. Data Mining and Knowledge Discovery, 39, 71. DOI: https://doi.org/10.1007/s10618-025-01133-3

[9] Zhao, X., Zhang, X., Huang, X., & Zhang, H. (2024). Influence maximization based on adaptive graph convolution neural network in social networks. Electronics, 13(16), 3110. DOI: 10.3390/electronics13163110

[10] Wu, S., Deng, Y., Liu, J., Luo, X., & Sun, G. (2025). Rumor detection on social networks based on Temporal Tree Transformer. PLOS ONE, 20(4), e0320333. DOI: https://doi.org/10.1371/journal.pone.0320333

[11] Roumeliotis, K. I., Tselikas, N. D., & Nasiopoulos, D. K. (2025). Fake news detection and classification: A comparative study of convolutional neural networks, large language models and natural language processing models. *Future Internet*, *17*(1), 28. DOI: https://doi.org/10.3390/fi17010028

[12] Mouratidis, D., Kanavos, A., & Kermanidis, K. (2025). From misinformation to insight: Machine learning strategies for fake news detection. *Information*, *16*(3), 189. DOI: https://doi.org/10.3390/info16030189

[13] Papageorgiou, E., Varlamis, I., & Chronis, C. (2025). Harnessing large language models and deep neural networks for fake news detection. *Information*, *16*(4), 297. DOI: https://doi.org/10.3390/info16040297

[14] Yin, F., Zhang, Z., Yu, Z., Wu, C., Chen, J., & Wu, Y. (2026). Breaking the Spatio-Temporal Mismatch: A Preemptive Deep Reinforcement Learning Framework for Misinformation Defense. Information, 17(1), 67. DOI link https://doi.org/10.3390/info17010067

[15] Aslam, Z., Khan, A., & Ahmad, M. (2025). Advancing fake news combating using machine learning. *Knowledge and Information Systems*. DOI: 10.1007/s10115-025-02588-y.

[16] Furutani, S., Aoshima, T., Shibahara, T., Akiyama, M., & Aida, M. (2024). Suppressing the Endogenous Negative Influence Through Node Intervention in Social Networks. IEEE Access, 13, 9290-9302. DOI : https://doi.org/10.1109/ACCESS.2024.3523036

[17] Xu, P., Wang, L., & Peng, Z. (2025). Proactive rumor control using graph neural networks and evolutionary optimization. Data Science and Engineering, 10, 753–763. DOI: 10.1007/s41019-025-00304-y

[18] Gu, X., Lau, Y. W., & Saidin, S. F. B. (2025). The Spillover Effects of Auditor Sanctions on Client Firms: Evidence From Deep Learning on Investors' Online Voice in China. IEEE Access, 13, 88952-88971.DOI: https://doi.org/10.1109/ACCESS.2025.3568536

[19] Anthony, P., & Zhou, J. (2025). Leveraging Transformer with Self-Attention for Multi-Label Emotion Classification in Crisis Tweets. *Informatics*, *12*(4), 114.DOI Link: https://doi.org/10.3390/informatics12040114

[20] Hashmi, E., Yayilgan, S. Y., Yamin, M. M., Ali, S., & Abomhara, M. (2024). Advancing fake news detection: Hybrid deep learning with fasttext and explainable ai. IEEE access, 12, 44462-44480.https://doi.org/10.1109/ACCESS.2024.3381038

[21] Zhang, H., Gao, A., Chen, Z., & Lu, X. (2026). Unveiling the Impact of Mandatory IP Location Disclosure on Social Media Users' Shared Emotions: A Regression Discontinuity Analysis Based on Weibo Data. Information, 17(1), 63. DOI https://doi.org/10.3390/info17010063

[22] Ding, Z., Wang, Z., Zhang, Y., Cao, Y., Liu, Y., Shen, X., ... & Dai, J. (2025). Trade-offs between machine learning and deep learning for mental illness detection on social media. Scientific Reports, 15(1), 14497. DOI : https://doi.org/10.1038/s41598-025-99167-6

[23] Yan, J., Pu, P., & Jiang, L. (2025). Emotion-RGC net: A novel approach for emotion recognition in social media using RoBERTa and Graph Neural Networks. Plos one, 20(3), e0318524. DOI : https://doi.org/10.1371/journal.pone.0318524

[24] Al-Tarawneh, M. A., Al-irr, O., Al-Maaitah, K. S., Kanj, H., & Aly, W. H. F. (2024). Enhancing fake news detection with word embedding: A machine learning and deep learning approach. Computers, 13(9), 239. DOI: https://doi.org/10.3390/computers13090239

[25] Han, H., Asif, M., Awwad, E. M., Sarhan, N., Ghadi, Y. Y., & Xu, B. (2024). Innovative deep learning techniques for monitoring aggressive behavior in social media posts. Journal of Cloud Computing, 13(1), 19. DOI: https://doi.org/10.1186/s13677-023-00577-6