Semantic Reshuffling with LLM and Heterogeneous Graph Auto-Encoder for Enhanced Rumor Detection

Guoyi Li[1, 2], Die Hu[1,2], Zongzhen Liu[1,2], Xiaodan Zhang[*,1,2], Honglei Lyu[1,2]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China

[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

{liguoyi, hudie, liuzongzhen, zhangxiaodan, lvhonglei}@iie.ac.cn

Abstract

Social media is crucial for information spread, necessitating effective rumor detection to curb misinformation's societal effects. Current methods struggle against complex propagation influenced by bots, coordinated accounts, and echo chambers, which fragment information and increase risks of misjudgments and model vulnerability. To counteract these issues, we introduce a new rumor detection framework, the Narrative-Integrated Metapath Graph Auto-Encoder (NIMGA). This model consists of two core components: (1) Metapath-based Heterogeneous Graph Reconstruction. (2) Narrative Reordering and Perspective Fusion. The first component dynamically reconstructs propagation structures to capture complex interactions and hidden pathways within social networks, enhancing accuracy and robustness. The second implements a dual-agent mechanism for viewpoint distillation and comment narrative reordering, using LLMs to refine diverse perspectives and semantic evolution, revealing patterns of information propagation and latent semantic correlations among comments. Extensive testing confirms our model outperforms existing methods, demonstrating its effectiveness and robustness in enhancing rumor representation through graph reconstruction and narrative reordering.

1 Introduction

The proliferation of the internet and social media has accelerated news dissemination and facilitated real-time discussions. Yet, this progress also introduces risks like rapid rumor spread, leading to information fragmentation and group polarization, distorting public perception, influencing responses, and undermining the credibility of online information, thus affecting personal and societal stability (Chen et al., 2022a; Zhou et al., 2022).

With the rapid evolution of Large Language Models (LLMs) (Chiang et al., 2023; Touvron et al., 2023a; Brown et al., 2020; Chowdhery et al., 2023), LLM-based approaches harness their expansive semantic capabilities to analyze rumors (Guan et al., 2023; Wu et al., 2023; Yang et al., 2023). For instance, CICAN (Yang et al., 2023) integrates crowdsourced intelligence with ChatGPT to enhance authenticity identification through textual analysis and knowledge augmentation by LLMs. However, LLMs struggle with the subtleties of information spread within social networks, heavily reliant on text features and failing to capture the structural features of propagation (Hu et al., 2024), making them susceptible to biases from prevailing opinions in propagation threads.

Some strategies employ Graph Neural Networks (GNNs) to exploit the complex topological features of propagation for rumor detection (Bian et al., 2020; Tao et al., 2024; Zhang et al., 2023; Sun et al., 2022). While GNN-based models excel in leveraging propagation graph structures, they of-

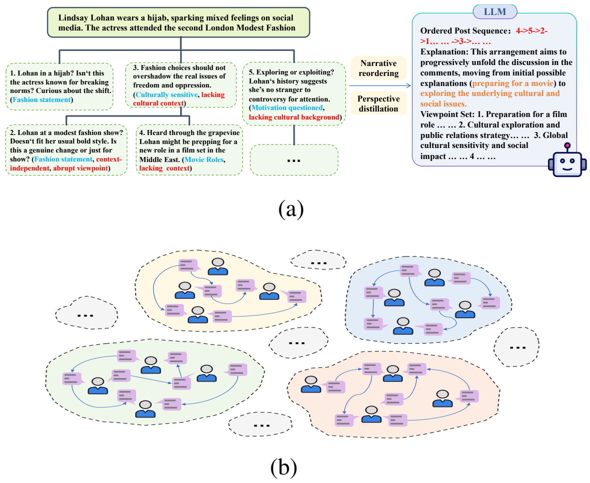


Figure 1: Two examples illustrate our motivation: In (a), blue text represents viewpoint categories, and red text highlights features of information fragmentation. In (b), disconnected propagation chains are prevalent in the data.

Proceedings of the 31st International Conference on Computational Linguistics, pages 8557–8572
January 19–24, 2025.

ten fail to discern latent post relationships, such as semantic shifts and contextual consistency, due to their dependence on supervised learning. This limitation hampers their generalization and performance in complex scenarios.

Existing rumor detection methods face two main challenges: (1) They often assume a complete and static propagation structure, ignoring the "information island" effect (disconnected graphs) (Iandoli et al., 2021; Min et al., 2022), caused by incomplete data collection due to technical or privacy limitations (Shao et al., 2018; Wei et al., 2022) or malicious attacks (Zhu et al., 2024; Wang et al., 2023; Yang et al., 2021), as shown in Figure 1(b). This leads to difficulties in detecting changes in propagation patterns due to missing data, complicating rumor detection and mitigation (Azzimonti and Fernandes, 2023). Models using supervised learning struggle with dynamic propagation structures, and while some use additional links or contrastive self-supervised learning to improve generalization (Sun et al., 2022; Tao et al., 2024; Wei et al., 2022), this can introduce noise. (2) The rapid rise of fragmented, colloquial, and diverse online information, often lacking context, complicates communication and information propagation, fostering viewpoint biases and challenging rumor detection (Bakshy et al., 2015; Zubiaga et al., 2016). Figure 1(a) left illustrates how comments often reflect multiple perspectives without coherent background, making it harder for models to capture contextual nuances (Wang et al., 2022).

To address these challenges, our model is motivated by: (1) Using graph reconstruction based on heterogeneous propagation graphs to capture complex interactions without adding extra links, reducing noise. This approach forces the model to explore deeper structural dependencies, uncovering hidden relationships in propagation beyond static links, enhancing generalization.enhancing generalization. Gradually learning higher-order connections and target node attributes enables inference even with incomplete data, improving robustness. (2) Framing theory suggests that media and individuals present events from different perspectives, enriching information despite increasing fragmentation. We refer to the semantic relevance of understanding all comments through a certain narrative logic as the "semantic narrativity" of information propagation. As shown in Figure 1(a) right, LLMs leverage strong semantic capabilities to extract this narrativity from fragmented information, integrating comments that evolve from initial ideas like "preparing for a movie" to "exploring cultural and social issues." This progression raises further questions about motivations, guiding the model in assessing news authenticity. Capturing this narrativity uncovers global semantic changes, reducing reliance on graph data and enabling diverse, high-level semantic representations.

Thus, we introduce the Narrative-Integrated Metapath Graph Auto-Encoder (NIMGA) for rumor detection. NIMGA incorporates a Metapath-based Heterogeneous Graph Reconstruction module using metapath masks and adaptive attribute masking to rebuild high-order relationships. The Narrative Reordering and Perspective Fusion module leverages Large Language Models (LLMs) with a sliding window technique to refine and reorder comment narratives, capturing evolving semantic features and viewpoints.

Key contributions of this paper include: (1) We introduce the NIMGA for rumor detection, the first to apply generative self-supervised learning on heterogeneous graphs. (2) Our metapath-based masking and adaptive attribute encoding effectively capture advanced structural features and node representations in rumor propagation. (3) The dual-agent mechanism utilizes LLMs for semantic representation of rumor propagation, exploring diverse viewpoints, while narrative reordering provides multiple perspectives on rumor evolution. (4) Experiments on three real-world datasets show our method surpasses baselines in both effectiveness and robustness.

2 Related Work

2.1 Rumor Detection

Text-based Methods. Early rumor detection focused on linguistic features. Potthast et al. (Potthast et al., 2017) assessed rumor veracity using stylistic similarities, while Ma et al. (Ma et al., 2015) applied sequence-based models to capture rumor dynamics. Li et al. (Li et al., 2019a) combined user and text features, training an LSTM for detection.

Propagation-based Methods. Unlike text-based approaches, these focus on propagation patterns between truth and deception (Khoo et al., 2020; Ma and Gao, 2020; Ma et al., 2018; Kumar and Carley, 2019). Graph Convolutional Networks (GCNs) have become popular for effectively capturing propagation features (Sun et al., 2022; Wei et al., 2022; Lin et al., 2021; Zhang et al., 2023), with techniques like adversarial training in GACL (Sun et al., 2022) and deep GCNs in UPSR (Wei et al., 2022) enhancing rumor detection.

Graph encoder-assisted Techniques for Rumor Detection. Graph Autoencoders (GAEs) (Kipf and Welling, 2016; Berg et al., 2017; Hou et al., 2022; Tian et al., 2023) have shown excellent performance in various domains, particularly in link prediction (Berg et al., 2017) and anomaly detection (Li et al., 2019b). GAEs have been applied to rumor detection by uncovering latent graph structures (Hou et al., 2024; Yin et al., 2024; Tao et al., 2024). For instance, (Yin et al., 2024) uses unsupervised encoders with masking and contrastive learning, while (Tao et al., 2024) captures semantic shifts to improve rumor representations. However, challenges arise in handling heterogeneous social graphs, where random masking risks missing key propagation features.

LLM-assisted Techniques for Rumor Detection. Large Language Models (LLMs) (OpenAI, 2022; Touvron et al., 2023b) have impacted rumor detection with their ability to process complex semantic relationships (Yue et al., 2024; Wu et al., 2023; Yang et al., 2023). For example, RARG proposes training and aligning LLMs to generate credible explanations for detected misinformation (Yue et al., 2024). However, despite their effectiveness, these methods often require extensive training data for fine-tuning and may suffer from catastrophic forgetting when domain or concept shifts occur, leading to performance degradation (Gu et al., 2023; Shang et al., 2024; Nan et al., 2022).

Building on these methods, our approach captures high-level propagation structures via heterogeneous graph reconstruction while leveraging LLMs' semantic strengths to explore diverse comment perspectives.

3 Problem Formulation

We define the set of rumor events as $\mathbf{T} = \{T_1, T_2, \ldots, T_M\}$, where M is the total number of events. Each rumor event T_i is modeled as a heterogeneous graph $G = (\mathcal{V}, \mathcal{A}, T_{\mathcal{V}}, T_{\mathcal{E}}, X, \Phi)$, capturing event interactions. $\mathcal{V}$ denotes the nodes, with $T_{\mathcal{V}}$ representing node types, including posts and users, forming the social network. $T_{\mathcal{E}}$ includes post-to-post, user-to-user, and user-to-post edges, reflecting diverse interaction dynamics. The adjacency matrix $\mathcal{A} \in \{0,1\}^{|\mathcal{V}|\times|\mathcal{V}|}$ encodes node connectivity. X denotes the attributes and Φ represents metapaths illustrating interaction sequences between node and edge types, capturing complex information flows. A metapath is expressed as: $T_{\mathcal{V}_1} \xrightarrow{T_{\mathcal{E}_1}} T_{\mathcal{V}_2} \xrightarrow{T_{\mathcal{E}_2}} \cdots \xrightarrow{T_{\mathcal{E}_l}} T_{\mathcal{V}_{l+1}}$, where l is the path length. This framework captures the structural and semantic relationships critical for understanding rumor propagation through the network.

Each event T_i is represented by a propagation graph, where $T_{\mathcal{V}_i} = P_i \cup U_i$, with $P_i = \{p_1^i, p_2^i, \ldots, p_{M_i}^i\}$ as posts and $U_i = \{u_1^i, u_2^i, \ldots, u_{M_i}^i\}$ as users. Posts include original tweets and retweets. The edge set $T_{\mathcal{E}_i}$ consists of $\mathcal{E}_{pp}^i$ (post-to-post), $\mathcal{E}_{uu}^i$ (user-to-user), and $\mathcal{E}_{up}^i$ (user-to-post), where $T_{\mathcal{E}_i} = \mathcal{E}_{pp}^i \cup \mathcal{E}_{uu}^i \cup \mathcal{E}_{up}^i$. We omit the superscript i in subsequent sections. This structure captures both information spread and endorsement within each event. Rumor detection is framed as a binary classification task, where $y \in \{0,1\}$ denotes the class label, with $y = 1$ representing a rumor and $y = 0$ a non-rumor.

4 Methodology

This section presents the Narrative-Integrated Metapath Graph Autoencoder (NIMGA) for rumor detection, depicted in Figure 2. NIMGA boosts detection capabilities through two primary components: (1) **Metapath-based Heterogeneous Graph Reconstruction**: This component focuses on edge reconstruction and attribute reconfiguration along rumor propagation metapaths. It strategically captures critical propagation pathways and context-aware attributes, effectively mapping interaction patterns between various node types and yielding highly indicative rumor representations. (2) **Narrative Reordering and Perspective Fusion**: Utilizing a dual-agent mechanism, this module leverages LLMs for iterative viewpoint distillation and narrative reordering. By integrating diverse viewpoint features with narrative characteristics of comments, the cognitive evolution is finely mapped, enabling precise identification of rumor propagation patterns and structures.

4.1 Metapath-based Heterogeneous Graph Reconstruction

This module consists of three parts: (1) Embedding Layer, (2) Metapath-based Edge Reconstruction, and (3) Dynamic Contextual Attribute Recovery.

Embedding Layer. To enhance contextual fea-

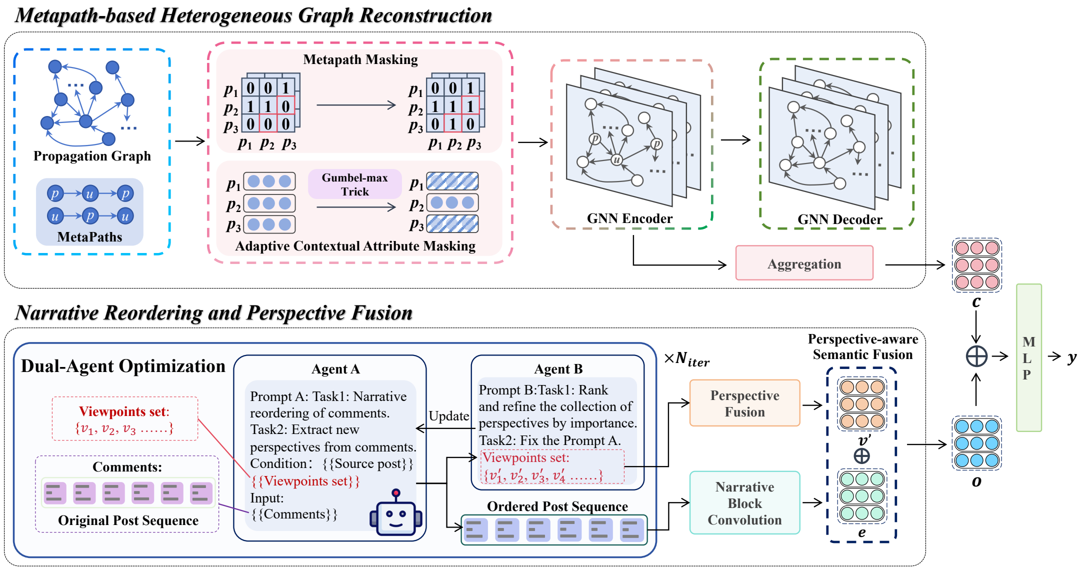


Figure 2: The overall architecture of NIMGA.

ture capture in rumor propagation, we employ TweetBERT (Qudar and Mago, 2020), fine-tuned for tweets, to generate contextualized embeddings for each post p_i from the set $P = \{p_1, p_2, ..., p_M\}$. This process effectively captures the nuances of social media language, providing a robust base for feature representation. We produce refined contextual post representations $X_p = \{h_1^p, h_2^p, ..., h_M^p\}$ and integrate user features such as verified status, follower, and following counts into user representations $X_u = \{h_1^u, h_2^u, ..., h_M^u\}$ essential for evaluating user credibility and influence. The combined node representations are denoted as $X = X_p \cup X_u$.

Metapath-based Edge Reconstruction. To address complex structural and semantic relationships in rumor propagation, we implement a metapath-based edge reconstruction strategy to uncover high-order relationships among key nodes and edges. By masking metapath-based edges and disrupting short-range links, the model infers masked relationships through alternative paths, promoting long-range semantic exploration and improving accuracy. Using a neural attention mechanism, we dynamically weight each metapath Φ, assessing its significance during rumor spread, enabling the model to focus on critical metapaths and refine its predictive abilities.

Specifically, given a heterogeneous graph $G = (\mathcal{V}, \mathcal{A}, T_{\mathcal{V}}, T_{\mathcal{E}}, X, \Phi)$, we create the metapath-based adjacency matrix $\mathcal{A}^{\Phi}$ for each metapath $\psi \in \Phi$ via metapath sampling (Dong et al., 2017) and employ a context-aware attention mechanism to dynamically weigh each metapath ψ, reflecting its relevance:

$$w_{\psi}{=}softmax(q^T \tanh(W^{\psi} h_{\psi}{+}b^{\psi})), \tag{1}$$

where h_{ψ} denotes latent node features related to the metapath, and W^{ψ}, q, and b^{ψ} are adjustable attention parameters. The resulting attention weight w_{ψ} generates a context-sensitive masking matrix:

$$\mathcal{A}_{\mathcal{M}}^{\psi}{=}\mathcal{M}^{\psi}(w_{\psi})\odot\mathcal{A}^{\psi}, \quad \mathcal{M}_{ij}^{\psi}{=}w_{\psi}{\cdot}\mathcal{A}_{ij}^{\psi} \tag{2}$$

adapted dynamically to current rumor dynamics.

During encoding, we input the masked adjacency matrix $\mathcal{A}_m^{\psi}$ and node attributes X into encoder f_{enc1} to produce latent node embeddings, and in decoding, the model focuses on edge prediction relevant to rumor propagation and reconstructs the adjacency matrix via decoder f_{dec1}:

$$H^{\psi}{=}f_{enc1}(\mathcal{A}_{\mathcal{M}}^{\psi}{,}X), \quad \tilde{H}^{\psi}{=}f_{dec1}(\mathcal{A}_{\mathcal{M}}^{\psi}{,}H^{\psi}). \tag{3}$$

We evaluate the reconstruction error of each metapath using an enhanced loss function that integrates these errors into the total loss. We compare the target adjacency matrix $\mathcal{A}^{\psi}$ with the reconstructed matrix $\tilde{\mathcal{A}}^{\psi} = \sigma((\tilde{H}^{\psi})^T \cdot \tilde{H}^{\psi})$ and employ scaled cosine error to assess node-level reconstruction, ensuring semantic consistency within the enhanced propagation graph:

$$C^{\psi}{=}\frac{1}{|\tilde{\mathcal{A}}^{\psi}|}\sum_{(i,j)\in\tilde{\mathcal{A}}^{\psi}}\left(1{-}\frac{\mathcal{A}_{ij}^{\psi}{\cdot}\tilde{\mathcal{A}}_{ij}^{\psi}}{\|\mathcal{A}_{ij}^{\psi}\|{\cdot}\|\tilde{\mathcal{A}}_{ij}^{\psi}\|}\right)^{\gamma_{\psi}}, \tag{4}$$

where C^{ψ} denotes the metapath-specific loss and γ_{ψ} is the scaling factor. Metapath importance is determined using a learnable semantic-level attention vector v, normalized via softmax to produce metapath weights α^{ψ}:

$$\alpha^{\psi}=\frac{\exp(softmax(v^T \tanh(W^r h_{\psi}+b^r)))}{\sum_{\psi} \exp(softmax(v^T \tanh(W^r h_{\psi}+b^r)))}. \tag{5}$$

These weighted losses are aggregated to calculate the total metapath reconstruction loss $\mathcal{L}_{mer}$: $\mathcal{L}_{mer} = \sum_{\psi} \alpha^{\psi} \cdot C^{\psi}$.

Adaptive Contextual Attribute Recovery. To enhance rumor detection in complex propagation graphs, we employ an adaptive context attribute recovery strategy to reconstruct node attributes critical to rumor dynamics, improving detection efficacy. By masking key node attributes (e.g., influential users or widely shared posts), the model learns representations sensitive to attributes influencing rumor spread.

Our approach adopts an adaptive masking strategy tailored to the context of rumor propagation, instead of a static masking rate. Using an attention mechanism, we assign importance weights to node attributes based on their relevance to rumor dynamics, with more significant attributes masked more frequently, forcing the model to focus on reconstructing these crucial attributes.

Formally, for each node v, its importance score s_v is calculated through a trainable attention network:

$$s_v=q^{\top} \tanh(W^v x_v+b^v), \tag{6}$$

where x_v is the attribute vector of node v, and W^v, q, and b^v are trainable parameters. This score is transformed into a masking probability distribution using the Gumbel-Softmax (Gumbel-max Trick) technique (Gumbel, 1954), which allows for differentiable approximation of categorical sampling. This method enables smooth sampling of discrete attributes while supporting backpropagation. Following (Jang et al., 2022), we use the Gumbel-Softmax approach to calculate the probability $\hat{p}_v$ as follows:

$$\hat{p}_v=\frac{\exp((\log s_v+G_v)/\tau)}{\sum_{j\in\mathcal{V}} \exp((\log s_j+G_j)/\tau)}, \tag{7}$$

where G_v and G_j are Gumbel noise $Gumbel(0, 1)$, and τ is the temperature parameter. This mechanism allows for attribute-weight-based sampling, optimized during training. A subset of nodes $\tilde{\mathcal{V}} \in \tilde{\mathcal{V}}_t$ of type t is then sampled and their attributes masked with a learnable token $[\mathcal{M}]$. For each node $v \in \mathcal{V}_t$, the attribute in the masked matrix $\tilde{X}$ is defined as $\tilde{x}_v = x_{[\mathcal{M}]}$ if $v \in \tilde{\mathcal{V}}$, otherwise $\tilde{x}_v = x_v$.

Similar to GraphMAE (Hou et al., 2022; Tao et al., 2024), we input masked node attributes $\tilde{X}$ and the graph adjacency matrix $\mathcal{A}$ into the encoder f_{enc2} to derive latent node embeddings $\hat{H}$, and then input $\mathcal{A}$ and $\hat{H}$ into the decoder f_{dec2} to reconstruct node attributes Z:

$$\hat{H}=f_{enc2}(\mathcal{A},\tilde{X}), \quad Z=f_{dec2}(\mathcal{A},\hat{H}). \tag{8}$$

For context-aware attribute recovery, we define the loss function of target attribute restoration $\mathcal{L}_{car}$ by comparing the masked attribute attributes $\tilde{X}$ and Z with scaling factor γ_2. The loss function is described as follows:

$$\mathcal{L}_{car}=\frac{1}{|\tilde{\mathcal{V}}|} \sum_{v\in\tilde{\mathcal{V}}} \left(1-\frac{\tilde{X}_v \cdot Z_v}{\|\tilde{X}_v\| \times \|Z_v\|}\right)^{\gamma_2}. \tag{9}$$

Representation of Propagation Graph. In order to leverage label information, we also calculate a supervised loss function for optimizing the model. Specifically, given a heterogeneous graph $G = (\mathcal{V}, \mathcal{A}, T_{\mathcal{V}}, T_{\mathcal{E}}, X, \Phi)$, we input the data into encoder f_{enc1} and f_{enc2} to obtain latent representations, respectively. Then, we use mean-pooling operators to aggregate the information of the set of node representations. Finally, we concatenate them to merge the information. Formally, it can be written as follow:

$$H_1=f_{enc1}(\mathcal{A}^{\psi},X), \quad H_2=f_{enc2}(\mathcal{A},X), \tag{10}$$

$$h_1=Mean(H_1), \quad h_2=Mean(H_2), \tag{11}$$

Finally, we obtain two pooled vectors h_1 and h_2, and concatenate them to form the final propagation graph representation c, where $c = concat(h_1, h_2)$.

4.2 Narrative Reordering and Perspective Fusion

Mining narrative correlations from global and local perspectives enhances rumor detection by providing deep semantic insights. Recent Large Language Models (LLMs) offer strong semantic understanding, viewpoint extraction, and paragraph ordering capabilities (Zhao et al., 2023; Chen and Si, 2024).

Two agents, Agent A and Agent B, collaborate: Agent A rearranges narratives and extracts potential new viewpoints from the comments; Agent B ranks the viewpoints by importance, refines the set, and updates Agent A's prompts. A sliding window technique processes large volumes of comments systematically. After multiple iterations, we obtain a refined viewpoint set and

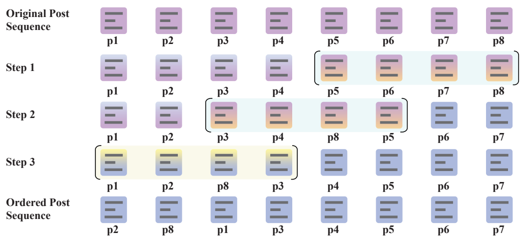


Figure 3: An illustration shows reordering eight comments using sliding windows with a size of 4 and a step of 2. Windows are applied in reverse, with the first two comments from the previous window contributing to the next window's reordering.

a reordered post sequence. We then apply self-attention mechanisms and multi-scale hierarchical convolution to derive multi-perspective fusion features, capturing semantic changes in the narrative. Viewpoint-aware attention fusion generates multi-dimensional features with advanced semantic layers and cross-perspective links, revealing covert manipulation strategies and identifying complex information propagation patterns.

Agent A and Agent B. (1) Agent A - Narrative Reordering Agent: Starting with a basic prompt (Figure 7 (a)), Agent A uses the source post and an empty viewpoint set to reorder comments based on semantic narrative sequences and distill emerging viewpoints. (2) Agent B - Viewpoint Integration Agent: Agent B ranks the extracted viewpoints by importance, retains the top-k, and updates Agent A's prompt (Figure 7 (b)).

Sliding Window Method. Due to the text length constraints of large models affecting LLMs' efficacy in processing text order, we reorder passages in a back-to-first sequence using a sliding window, as shown in Figure 3. This strategy utilizes two hyperparameters: window size w and step size ss. Initially, LLMs reorder posts from the $(M-w)$-th to the m-th. We then adjust the window by ss steps and reorder posts from the $(m-w-ss)$-th to the $(m-ss)$-th. This process is repeated until all posts have been reordered.

Narrative Block Convolution. Narrative block convolution captures interactions and transitions between adjacent posts by applying a one-dimensional convolutional neural network to generate continuous representations of text, as shown in Figure 4. After sorting, the post ordered set $S_{p_i} = \{p_{ij} \,|_{j=1}^{|p_i|}\}$ is obtained. Post embeddings generated by the embedding layer are $g = \{g_j \,|_{j=1}^{|p_i|}\}$. To capture interactions and transitions between

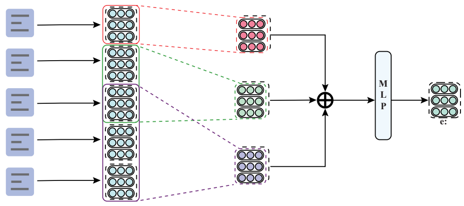


Figure 4: Narrative block convolution.

adjacent posts, tracking the evolution of rumors and mitigating fragmented information, we employ an narrative block convolution to derive continuous representations of the text as shown in Figure 4. A one-dimensional convolutional neural network propagates these features to capture hidden sequential local contexts. The convolution layer generates feature maps $F = \{f_j\}_{j=1}^{|p_i|-d+1}$ from the continuous input sequence $\{g_{j:(j+d-1)}\}_{j=1}^{|p_i|-d+1}$ using filter w^s, where each local input comprises a group of d consecutive posts, represented as: $f_j = \sigma(w^s \cdot g_{j:(j+d-1)} + b^s)$, where $g_{j:(j+d-1)} = concat(g_j, g_{j+1}, \ldots, g_{j+d-1})$, b^s is the bias, and σ is the ReLU activation function.

To track the most significant local changes in rumor propagation, we apply max-pooling to the convolution outputs, reducing dimensionality and emphasizing key local features that may determine propagation direction: $\hat{f} = \max\{f_j\}_{j=1}^{|p_i|-d+1}$. We then derive text representations through $s = W_s\hat{f} + b_s$, where W_s, b_s are trainable parameters. Specifically, $n \in \{1, 2, 3\}$ denotes three window sizes capturing post-level information—single, double, and triple posts. Finally, passing s through a fully connected layer with L2 regularization, we obtain the narrative representation e of the text.

Perspective-aware Semantic Fusion. From Agent A's final results, we extract the top-k significant viewpoints related to the source news text. Given the viewpoint set $S_v = \{v_i\}_{i=1}^{k}$, we fuse the representations of the source post $source$ and the viewpoints S_v to calculate the viewpoint fusion feature v': $v' = ATTN(source, S_v)$, where $ATTN$ maps the source post representation and the viewpoint set to a weighted sum of elements in S_v (Vaswani et al., 2017). We then learn the attention coefficients between the viewpoint features v' and the narrative comment features e:

$$\alpha = \sigma(W^{v'}v' + W^e e + b^f), \quad o = \alpha v' + (1-\alpha)e, \tag{12}$$

where $\sigma(\cdot)$ is the sigmoid activation function, and $W^{v'}, W^e, b^f$ are parameters of the fusion gate.

4.3 Classifier

We derive the final rumor representation by combining the representation of propagation graph c with viewpoint-comment integrated representation o, yielding $\hat{x} = c \oplus o$. This representation is input into a fully connected layer for classification:

$$\hat{y}=softmax(w^c\hat{x}+b^c), \tag{13}$$

where $\hat{y}$ is the predicted label distribution, and w^c, b^c are trainable parameters. The model minimizes the cross-entropy loss:

$$\mathcal{L}_y=-(y\log(\hat{y})+(1-y)\log(1-\hat{y})), \tag{14}$$

where y represents the true label.

The total NIMGA loss includes metapath-based edge reconstruction, adaptive contextual attribute recovery, and classification loss:

$$\mathcal{L}_{total}=\mathcal{L}_y+\lambda\mathcal{L}_{mer}+\mu\mathcal{L}_{car}, \tag{15}$$

where λ, μ are balancing coefficients.

5 Experiments

Datasets. To evaluate our model, we used three representative real-world rumor datasets: MC-Fake (Shu et al., 2020), Weibo (Ma et al., 2016), and MM-COVID (Li et al., 2020). Further details are in Appendix .1.

Baselines. We compare our proposed model with the following baseline on all three datasets: **TweetBert** (Qudar and Mago, 2020), **LLaMA2**$_{text}$ (Touvron et al., 2023b), **LLaMA2**$_{text}$ **+ SPT** (Lester et al., 2021), **ChatGPT**$_{text}$ (Ouyang et al., 2022), **ChatGPT**$_{full}$, **CICAN** (Yang et al., 2023), **GACL** (Sun et al., 2022), **GARD** (Tao et al., 2024), **PSIN** (Min et al., 2022), **HAN** (Yang et al., 2016), **HG-SL** (Sun et al., 2023). These models represent a range of approaches, including traditional text-only methods (TweetBert), LLM-assisted methods (LLaMA2$_{text}$, LLaMA2$_{text}$ + SPT, ChatGPT$_{text}$, ChatGPT$_{full}$, CICAN), graph-based methods (GACL, GARD), and heterogeneous graph-based models (PSIN, HAN, HG-SL).

In the NMGIA's heterogeneous graph propagation enhancement module, **HAN** (Wang et al., 2019) served as both encoder and decoder. **ChatGPT** (gpt-3.5-turbo) was utilized as the LLM. Further details can be found in the parameter analysis .2 and Appendix .1.

Main Results. The comparative performance of baseline models and our proposed NMGIA is detailed in Table 1. NMGIA excels, achieving higher accuracy, precision, recall, and macro F1 scores across three datasets: 89.6% on MC-Fake, 93.4% on Weibo, and 93.1% on MM-COVID, surpassing the state-of-the-art by 2.8%, 3.0%, and 4.7%, respectively.

LLM-based approaches generally underperform compared to supervised models due to limited domain-specific knowledge in tasks like rumor detection. This can result in over-reliance on majority opinions, ignoring differences in propagation threads. However, fine-tuned LLaMA2$_{text}$ + SPT and CICAN, which enhance rumor semantics with LLMs, perform better, indicating LLMs' improved deep comprehension of rumor texts. Propagation-structure-based models also show superior results, with heterogeneous graph-based approaches benefiting from complex relationships and node-type distinctions, leading to higher accuracy and F1 scores.

Key advantages of NMGIA include: (1) In propagation structure enhancement, the adaptive masking mechanism effectively captures key features of heterogeneous propagation graphs. By disrupting short-range semantic links, long-distance interactions are better explored, enhancing the characterization of rumor propagation structures.(2) Leveraging the strong semantic understanding of LLMs, the Narrative Reordering and Perspective Fusion Module introduces a unique learning approach for rumor propagation, refining narrative perspectives to capture deeper semantic representations. This multi-perspective refinement enables the model to learn more comprehensive information.

Ablation Study. We assessed the contributions of various NMGIA model components by excluding each individually. The variants tested included the removal of the narrative reordering and perspective module **(w/o** o**)**, node sampling strategy Gumbel-max Trick **(w/o GT)**, metapath-based edge reconstruction loss **(w/o** $\mathcal{L}_{mer}$**)**, contextual attribute recovery loss **(w/o** $\mathcal{L}_{car}$**)**, viewpoint-based post ranking optimization **(w/o vo)**, viewpoint feature representation **(w/o** v'**)**, narrative comment features **(w/o** e**)**, sliding window (random selection of 30 posts) **(w/o sw)**, narrative block convolution (replaced with CNN) **(w CNN)**, and substituting the LLM for LLaMA2 (13b version) **(w LLaMA2)**.

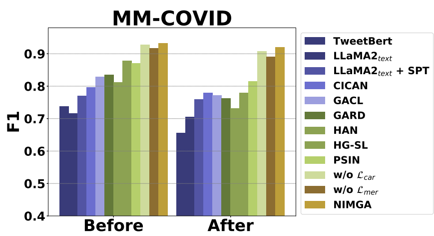
Results from Table 2 reveal several points: (1) Each module is crucial for the model's rumor detection ability. The removal of the semantic enhancement module (-w/o c) impacted accuracy on MC-Fake, Weibo, and MM-COVID, underscoring

Table 1: Experimental results on MC-Fake, Weibo, and MM-COVID. And with * indicate the significance test over HG-SL and PSIN presents a statistically significant improvement (with p < 0.05).

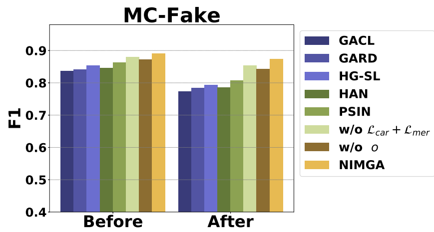
		MC-Fake				Weibo				MM-COVID			
Type	Method	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1	Acc.	Prec.	Rec.	F1
Text-only	TweetBert	0.792	0.795	0.774	0.780	0.825	0.842	0.813	0.823	0.739	0.757	0.717	0.738
Large Language Model-assisted	LLaMA2_{text}	0.735	0.729	0.753	0.735	0.788	0.766	0.795	0.792	0.726	0.717	0.774	0.716
	LLaMA2_{text} + SPT	0.811	0.792	0.829	0.806	0.844	0.846	0.835	0.842	0.778	0.777	0.744	0.770
	ChatGPT_{text}	0.747	0.794	0.715	0.741	0.793	0.829	0.754	0.782	0.716	0.782	0.683	0.708
	ChatGPT_{full}	0.764	0.813	0.748	0.758	0.813	0.828	0.734	0.803	0.725	0.785	0.703	0.721
	CICAN	0.828	0.845	0.803	0.814	0.862	0.859	0.847	0.865	0.792	0.809	0.814	0.796
Graph-based	GACL	0.840	0.849	0.839	0.837	0.875	0.881	0.867	0.877	0.830	0.824	0.834	0.829
	GARD	0.843	0.853	0.852	0.841	0.881	0.873	0.892	0.872	0.834	0.827	0.826	0.835
Heterogeneous Graph-based	PSIN	0.868	0.858	0.867	0.863	0.902	0.893	0.901	0.897	0.870	0.833	0.895	0.871
	HAN	0.850	0.834	0.830	0.846	0.894	0.892	0.899	0.889	0.832	0.825	0.763	0.812
	HG-SL	0.862	0.859	0.864	0.854	0.904	0.903	0.903	0.900	0.875	0.884	0.834	0.878
	NIMGA	**0.896***	**0.897***	**0.888**	**0.890***	**0.934***	**0.932**	**0.931***	**0.933**	**0.931***	**0.934**	**0.928**	**0.932***

Table 2: Ablation study on the architecture design of NIMGA across three datasets.

Method	MC-Fake		Weibo		MM-COVID	
	Accuracy	F1 score	Accuracy	F1 score	Accuracy	F1 score
NIMGA	**0.896**	0.890	**0.934**	**0.933**	**0.931**	**0.932**
-w/o o	0.873	0.872	0.906	0.908	0.905	0.905
-w/o GT	0.890	0.888	0.930	0.928	0.926	0.927
-w/o $\mathcal{L}_{mer}$	0.880	0.881	0.921	0.920	0.917	0.917
-w/o $\mathcal{L}_{car}$	0.884	0.886	0.924	0.922	0.925	0.924
-w/o vo	0.889	0.888	0.928	0.929	0.928	0.927
-w/o v'	0.885	0.883	0.925	0.922	0.924	0.927
-w/o e	0.881	0.880	0.912	0.912	0.914	0.914
-w/o sw	0.885	0.885	0.920	0.920	0.924	0.924
-w CNN	0.882	0.884	0.914	0.917	0.918	0.916
-w LLaMA2	0.888	**0.892**	0.930	0.930	0.925	0.926



(a)



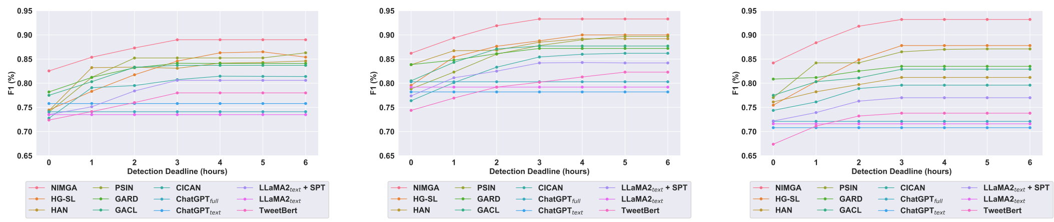
(b)

Figure 5: Performance of NIMGA and baseline models in robustness tests.

the importance of heterogeneous graph reconstruction and viewpoint-comment integration. (2) The results from -w/o $\mathcal{L}_{mer}$ and -w/o $\mathcal{L}_{car}$ show edge reconstruction as more critical, likely due to its role in capturing complex propagation structures crucial for detecting covert rumors. (3) The strong performance of NMGIA with LLaMA2 underscores the model's compatibility with different LLMs, effectively leveraging their semantic understanding for viewpoint distillation and global semantic capture. The impacts of the sliding window technique and narrative block convolution suggest that proper application of these methods can better unearth potential narrative connections in global semantics.

Robustness Testing. We conducted two adversarial tests to assess our method's robustness: (1) We added Gaussian noise to node features and randomly perturbed edges to simulate adversarial attacks (Chang et al., 2020; Chen et al., 2022b; Gu et al., 2020), with results in Figure 5 (a). (2) Introducing a Multi-Agent Reinforcement Learning (MARL) attack (Wang et al., 2023) based on rumor propagation graphs, where agents (e.g., bots, semi-bots, crowdsourced workers) add meaningless comments and new retweet edges to disrupt social engagement, with results shown in Figure 5 (b).

Our model showed high robustness across scenarios. In (a), the LLM-Assisted variant had less performance degradation due to LLMs' resistance to node and content interference, thanks to extensive pretraining. The variant analysis reveals that our model excels by effectively leveraging LLMs' semantic robustness. Additionally, the results for NMGIA w/o $\mathcal{L}_{mer}$ confirm the critical role of metapath-based edge reconstruction in enhancing graph representation. This is likely because adversarial attacks lead to many disconnected subgraphs, while our metapath-based heterogeneous graph reconstruction uncovers latent relationships



(a) MC-Fake (b) Weibo (c) MM-COVID

Figure 6: Results of rumor early detection on three datasets.

effectively. In (b), MARL introduces high-fidelity noise into the propagation structure, significantly impacting graph-based methods. The larger drop in the variant (-w/o e) compared to w/o $\mathcal{L}_{car+mer}$ suggests that capturing global semantic relationships may help mitigate the effects of propagation-structure-based attacks.

Early Detection. Early rumor detection is critical for evaluating model performance, focusing on identifying rumors at their inception (Liu and Wu, 2018). For this, detection deadlines were set at 0, 2, 4, and 6 hours. The F1 scores of the NMGIA model were then compared against a baseline using only pre-deadline posts. Figure 6 displays the performance of various methods across datasets, showing that NMGIA and GARD perform well early on, likely due to the generalization of graph autoencoder-based models with limited data.

Case Study. Our analysis explores the roles of Agent A and Agent B, with the propagation graph displayed in Figure 9. Agent A's sorting of comments, shown in Table 5, organizes fragmented comments into a coherent format that facilitates relationship exploration and viewpoint synthesis. The initial two viewpoints are pre-existing, while the latter two are newly extracted. Agent B further refines these inputs, enhancing task relevance by supplementing, improving, and reordering viewpoints, as detailed in Table 6, providing a clearer and more substantive basis for judgment.

6 Conclusion

In summary, our NIMGA enhances rumor detection on social media through Metapath-based Heterogeneous Graph Reconstruction, capturing complex interactions and hidden pathways within social networks. A dual-agent mechanism, driven by LLMs, refines semantic narratives and viewpoints, making learning more intuitive and comprehensive. Extensive real-world dataset testing confirms NIMGA's superiority over existing models, improving robustness and effectiveness in news event representation and providing a more reliable tool for rumor detection.

7 Acknowledgements

This work was supported by Grant 2022YFB3103704 from the National Key R&D Program of China.

Limitations

In this section, we discuss the limitations of our work and propose corresponding solutions. First, to address common issues such as data missingness and graph network attacks in social networks, we designed a heterogeneous graph reconstruction to disrupt short-range semantic links and uncover long-range potential propagation structural associations. Although this design is simple and effective, it may still not be able to handle attacks on more complex and advanced rumor detection models. In the future, we will continue to focus on the robustness of models based on rumor propagation graphs. On the other hand, we considered the semantic narrativity features in social propagation, providing a new, directly investigable perspective for information dissemination. In the future, we hope to explore more effective social propagation patterns to provide multi-perspective learning for rumor detection models.

Ethics Statement

Our research is dedicated to the performance analysis of rumor detection models, with all experiments carried out under controlled and secure conditions. All applications of large models are also conducted within manageable limits.

References

Marina Azzimonti and Marcos Fernandes. 2023. Social media networks, fake news, and polarization. European journal of political economy, 76:102256.

Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on facebook. Science, 348(6239):1130–1132.

Rianne van den Berg, Thomas N Kipf, and Max Welling. 2017. Graph convolutional matrix completion. arXiv preprint arXiv:1706.02263.

Tian Bian, Xi Xiao, Tingyang Xu, Peilin Zhao, Wenbing Huang, Yu Rong, and Junzhou Huang. 2020. Rumor detection on social media with bi-directional graph convolutional networks. In Proceedings of the AAAI conference on artificial intelligence, volume 34, pages 549–556.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. Advances in neural information processing systems, 33:1877–1901.

Heng Chang, Yu Rong, Tingyang Xu, Wenbing Huang, Honglei Zhang, Peng Cui, Wenwu Zhu, and Junzhou Huang. 2020. A restricted black-box adversarial framework towards attacking graph embedding models. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 3389–3396.

C Chen, H Wang, M Shapiro, Y Xiao, F Wang, and K Shu. 2022a. Combating health misinformation in social media: Characterization, detection, intervention, and open issues.

Yongqiang Chen, Han Yang, Yonggang Zhang, Kaili Ma, Tongliang Liu, Bo Han, and James Cheng. 2022b. Understanding and improving graph injection attack by promoting unnoticeability. arXiv preprint arXiv:2202.08057.

Yuetian Chen and Mei Si. 2024. Reflections & resonance: Two-agent partnership for advancing llm-based story annotation. In Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), pages 13813–13818.

Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E Gonzalez, et al. 2023. Vicuna: An open-source chatbot impressing gpt-4 with 90%* chatgpt quality. See https://vicuna. lmsys. org (accessed 14 April 2023), 2(3):6.

Aakanksha Chowdhery, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, Hyung Won Chung, Charles Sutton, Sebastian Gehrmann, et al. 2023. Palm: Scaling language modeling with pathways. Journal of Machine Learning Research, 24(240):1–113.

Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable representation learning for heterogeneous networks. In Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining, pages 135–144.

Fangda Gu, Heng Chang, Wenwu Zhu, Somayeh Sojoudi, and Laurent El Ghaoui. 2020. Implicit graph neural networks. Advances in Neural Information Processing Systems, 33:11984–11995.

Jiawei Gu, Xuan Qian, Qian Zhang, Hongliang Zhang, and Fang Wu. 2023. Unsupervised domain adaptation for covid-19 classification based on balanced slice wasserstein distance. Computers in Biology and Medicine, page 107207.

Jian Guan, Jesse Dodge, David Wadden, Minlie Huang, and Hao Peng. 2023. Language models hallucinate, but may excel at fact verification. arXiv preprint arXiv:2310.14564.

Emil Julius Gumbel. 1954. Statistical theory of extreme values and some practical applications: a series of lectures, volume 33. US Government Printing Office.

Dongpeng Hou, Chao Gao, Xuelong Li, and Zhen Wang. 2024. Dag-aware variational autoencoder for social propagation graph generation. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 38, pages 8508–8516.

Zhenyu Hou, Xiao Liu, Yukuo Cen, Yuxiao Dong, Hongxia Yang, Chunjie Wang, and Jie Tang. 2022. Graphmae: Self-supervised masked graph autoencoders. In Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 594–604.

Beizhe Hu, Qiang Sheng, Juan Cao, Yuhui Shi, Yang Li, Danding Wang, and Peng Qi. 2024. Bad actor, good advisor: Exploring the role of large language models in fake news detection. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 38, pages 22105–22113.

Luca Iandoli, Simonetta Primario, and Giuseppe Zollo. 2021. The impact of group polarization on the quality of online debate in social media: A systematic literature review. Technological Forecasting and Social Change, 170:120924.

Eric Jang, Shixiang Gu, and Ben Poole. 2022. Categorical reparameterization with gumbel-softmax. In International Conference on Learning Representations.

Ling Min Serena Khoo, Hai Leong Chieu, Zhong Qian, and Jing Jiang. 2020. Interpretable rumor detection in microblogs by attending to user interactions. In Proceedings of the AAAI conference on artificial intelligence, volume 34, pages 8783–8790.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Thomas N Kipf and Max Welling. 2016. Variational graph auto-encoders. arXiv preprint arXiv:1611.07308.

Sumeet Kumar and Kathleen M Carley. 2019. Tree lstms with convolution units to predict stance and rumor veracity in social media conversations. In Proceedings of the 57th annual meeting of the association for computational linguistics, pages 5047–5058.

Brian Lester, Rami Al-Rfou, and Noah Constant. 2021. The power of scale for parameter-efficient prompt tuning. arXiv preprint arXiv:2104.08691.

Quanzhi Li, Qiong Zhang, and Luo Si. 2019a. Rumor detection by exploiting user credibility information, attention and multi-task learning. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 1173–1179.

Yichuan Li, Bohan Jiang, Kai Shu, and Huan Liu. 2020. Mm-covid: A multilingual and multimodal data repository for combating covid-19 disinformation. arXiv preprint arXiv:2011.04088.

Yuening Li, Xiao Huang, Jundong Li, Mengnan Du, and Na Zou. 2019b. Specae: Spectral autoencoder for anomaly detection in attributed networks. In Proceedings of the 28th ACM international conference on information and knowledge management, pages 2233–2236.

Hongzhan Lin, Jing Ma, Mingfei Cheng, Zhiwei Yang, Liangliang Chen, and Guang Chen. 2021. Rumor detection on twitter with claim-guided hierarchical graph attention networks. arXiv preprint arXiv:2110.04522.

Yang Liu and Yi-Fang Wu. 2018. Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In Proceedings of the AAAI conference on artificial intelligence, volume 32.

Jing Ma and Wei Gao. 2020. Debunking rumors on twitter with tree transformer. ACL.

Jing Ma, Wei Gao, Prasenjit Mitra, Sejeong Kwon, Bernard J Jansen, Kam-Fai Wong, and Meeyoung Cha. 2016. Detecting rumors from microblogs with recurrent neural networks.

Jing Ma, Wei Gao, Zhongyu Wei, Yueming Lu, and Kam-Fai Wong. 2015. Detect rumors using time series of social context information on microblogging websites. In Proceedings of the 24th ACM international on conference on information and knowledge management, pages 1751–1754.

Jing Ma, Wei Gao, and Kam-Fai Wong. 2018. Rumor detection on twitter with tree-structured recursive neural networks. Association for Computational Linguistics.

Erxue Min, Yu Rong, Yatao Bian, Tingyang Xu, Peilin Zhao, Junzhou Huang, and Sophia Ananiadou. 2022. Divide-and-conquer: Post-user interaction network for fake news detection on social media. In Proceedings of the ACM Web Conference 2022, pages 1148–1158.

Qiong Nan, Danding Wang, Yongchun Zhu, Qiang Sheng, Yuhui Shi, Juan Cao, and Jintao Li. 2022. Improving fake news detection of influential domain via domain-and instance-level transfer. In Proceedings of the 29th International Conference on Computational Linguistics, pages 2834–2848.

TB OpenAI. 2022. Chatgpt: Optimizing language models for dialogue. openai.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. Advances in neural information processing systems, 35:27730–27744.

Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2017. A stylometric inquiry into hyperpartisan and fake news. arXiv preprint arXiv:1702.05638.

Mohiuddin Md Abdul Qudar and Vijay Mago. 2020. Tweetbert: a pretrained language representation model for twitter text analysis. arXiv preprint arXiv:2010.11091.

Lanyu Shang, Yang Zhang, Bozhang Chen, Ruohan Zong, Zhenrui Yue, Huimin Zeng, Na Wei, and Dong Wang. 2024. Mmadapt: A knowledge-guided multi-source multi-class domain adaptive framework for early health misinformation detection. In Proceedings of the ACM on Web Conference 2024, pages 4653–4663.

Chengcheng Shao, Giovanni Luca Ciampaglia, Onur Varol, Kai-Cheng Yang, Alessandro Flammini, and Filippo Menczer. 2018. The spread of low-credibility content by social bots. Nature communications, 9(1):1–9.

Kai Shu, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big data, 8(3):171–188.

Ling Sun, Yuan Rao, Yuqian Lan, Bingcan Xia, and Yangyang Li. 2023. Hg-sl: Jointly learning of global and local user spreading behavior for fake news early detection. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 37, pages 5248–5256.

Tiening Sun, Zhong Qian, Sujun Dong, Peifeng Li, and Qiaoming Zhu. 2022. Rumor detection on social media with graph adversarial contrastive learning. In Proceedings of the ACM Web Conference 2022, pages 2789–2797.

Xiang Tao, Liang Wang, Qiang Liu, Shu Wu, and Liang Wang. 2024. Semantic evolvement enhanced graph autoencoder for rumor detection. In Proceedings of the ACM on Web Conference 2024, pages 4150–4159.

Yijun Tian, Kaiwen Dong, Chunhui Zhang, Chuxu Zhang, and Nitesh V Chawla. 2023. Heterogeneous graph masked autoencoders. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 37, pages 9997–10005.

H Touvron, T Lavril, G Izacard, X Martinet, MA Lachaux, T Lacroix, B Rozière, N Goyal, E Hambro, F Azhar, et al. 2023a. Open and efficient foundation language models. Preprint at arXiv. https://doi.org/10.48550/arXiv, 2302.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023b. Llama: Open and efficient foundation language models. arXiv preprint arXiv:2302.13971.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. Advances in neural information processing systems, 30.

Dandan Wang, Yadong Zhou, Yuxing Qian, and Yunmei Liu. 2022. The echo chamber effect of rumor rebuttal behavior of users in the early stage of covid-19 epidemic in china. Computers in Human Behavior, 128:107088.

Haoran Wang, Yingtong Dou, Canyu Chen, Lichao Sun, Philip S Yu, and Kai Shu. 2023. Attacking fake news detectors via manipulating news social engagement. In Proceedings of the ACM Web Conference 2023, pages 3978–3986.

Xiao Wang, Houye Ji, Chuan Shi, Bai Wang, Yanfang Ye, Peng Cui, and Philip S Yu. 2019. Heterogeneous graph attention network. In The world wide web conference, pages 2022–2032.

Lingwei Wei, Dou Hu, Wei Zhou, and Songlin Hu. 2022. Uncertainty-aware propagation structure reconstruction for fake news detection. In Proceedings of the 29th International Conference on Computational Linguistics, pages 2759–2768.

Jiaying Wu, Shen Li, Ailin Deng, Miao Xiong, and Bryan Hooi. 2023. Prompt-and-align: prompt-based social alignment for few-shot fake news detection. In Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, pages 2726–2736.

Chang Yang, Peng Zhang, Wenbo Qiao, Hui Gao, and Jiaming Zhao. 2023. Rumor detection on social media with crowd intelligence and chatgpt-assisted networks. In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing, pages 5705–5717.

Xiaoyu Yang, Yuefei Lyu, Tian Tian, Yifei Liu, Yudong Liu, and Xi Zhang. 2021. Rumor detection on social media with graph structured adversarial learning. In Proceedings of the twenty-ninth international conference on international joint conferences on artificial intelligence, pages 1417–1423.

Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. 2016. Hierarchical attention networks for document classification. In Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies, pages 1480–1489.

Shu Yin, Peican Zhu, Lianwei Wu, Chao Gao, and Zhen Wang. 2024. Gamc: An unsupervised method for fake news detection using graph autoencoder with masking. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 38, pages 347–355.

Zhenrui Yue, Huimin Zeng, Yimeng Lu, Lanyu Shang, Yang Zhang, and Dong Wang. 2024. Evidence-driven retrieval augmented response generation for online misinformation. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pages 5628–5643.

Kaiwei Zhang, Junchi Yu, Haichao Shi, Jian Liang, and Xiao-Yu Zhang. 2023. Rumor detection with diverse counterfactual evidence. In Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pages 3321–3331.

Zoie Zhao, Sophie Song, Bridget Duah, Jamie Macbeth, Scott Carter, Monica P Van, Nayeli Suseth Bravo, Matthew Klenk, Kate Sick, and Alexandre LS Filipowicz. 2023. More human than human: Llm-generated narratives outperform human-llm interleaved narratives. In Proceedings of the 15th Conference on Creativity and Cognition, pages 368–370.

Xinyi Zhou, Kai Shu, Vir V Phoha, Huan Liu, and Reza Zafarani. 2022. "this is fake! shared it by mistake": Assessing the intent of fake news spreaders. In Proceedings of the ACM Web Conference 2022, pages 3685–3694.

Peican Zhu, Zechen Pan, Yang Liu, Tian Jiwei, Keke Tang, and Zhen Wang. 2024. A general black-box adversarial attack on graph-based fake news detectors. pages 568–576.

Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tolmie. 2016. Analysing how people orient to and spread rumours in social media by looking at conversational threads. PloS one, 11(3):e0150989.

LogicSortGPT is an intelligent assistant designed to rank posts and extract viewpoints based on thematic relevance and narrative coherence. Given a core topic, multiple posts, and a prioritized sequence of perspectives, your tasks are as follows:

Task 1: Rank the paragraphs by their relevance to the core topic, adherence to the importance sequence of perspectives, logical progression, and narrative quality.

Task 2: Identify and extract any new viewpoints from the comments that differ from the existing perspectives. Add these to the original set of perspectives. If no new viewpoints are identified, retain the original set.

Requirements:

1. Ensure alignment with the core topic.
2. New viewpoints must be relevant to the core topic.
3. Maintain narrative logic to enhance the flow and coherence of the story.
4. The order of paragraphs should reflect a smooth logical transition and the evolution of perspectives.
5. The output should provide the optimal paragraph sequence by number, ensuring alignment with the topic without altering the content.

Based on the topic: **{{source post}},** the following is the set of perspectives:

[1] **{{perspective_1}}**

[2] **{{perspective_2}}**

(...additional perspectives)

There are **{{num}}** paragraphs, each identified by a number. Sort them and output the sequence of paragraph numbers:

[1] **{{passage_1}}**

[2] **{{passage_2}}**

(...additional paragraphs)

Output the new set of perspectives and the ordered sequence in JSON format.

- [**New Perspective Set**]: If there are new perspectives, respond with "Yes" and the content of the New Perspective Set. Otherwise, output "No”.
- [**Ordered Sequence**]: Provide the sorted paragraph numbers, separated by commas.

(a) Prompt A

LogicSortGPT is an intelligent assistant designed to revise prompts based on feedback and rank perspectives for thematic relevance and clarity. Given the feedback and a prioritized sequence of perspectives, your tasks are as follows:

Task 1: Revise **Prompt_A** based on the **{{New Perspective Set}}** . Use the newly added perspectives (positioned at the end) to enhance existing perspectives without altering their original meaning, but with appropriate supplementation where necessary.

Task 2: Rank the perspectives by importance, where importance is defined as their contribution to determining whether the **{{source post}}** contains false information. Retain the most important **{{top-k}}** perspectives while maintaining the original format. All other sections remain unchanged.

Based on the news content: **{{source post}}**, **{{New Perspective Set}}**, **{{Prompt_A}}**, output the revised **[Prompt_A]** and **[Perspective Set]** in JSON format.

(b) Prompt B

Figure 7: Examples of prompt A and prompt B.

Table 3: The statistics of three datasets.

Dataset	MC-Fake	Weibo	MM-COVID
# Event Count	8,347	4,664	3,820
# True Rumors	3,381	2,351	1,832
# False Rumors	4,966	2,313	1,988
# Connections Count	153,606,2	3,805,656	71,885
# Users Count	201,481,5	2,746,818	51,794

.1 The Details of Experimental Setup

Datasets. To evaluate our model, we used three real-world rumor datasets: MC-Fake (Shu et al., 2020), Weibo (Ma et al., 2016), and MM-COVID (Li et al., 2020). These datasets include social contexts (tweets, retweets, replies, retweet and reply relationships) and cover diverse topics (politics, entertainment, health, COVID-19, Syrian war). MM-COVID focuses on COVID-19 misinformation. All datasets have binary labels: false rumor (F) and true rumor (T). MC-Fake includes many disconnected graphs, with Weibo and MM-COVID also containing some disconnected data. The labels for Weibo are based on reports from the Sina Community Management Center (Ma et al., 2016). MC-Fake labels come from debunking websites [1] [2], while MM-COVID labels are provided by Snopes [3] and Poynter [4], where domain experts fact-check and classify information. Snopes is an independent English-language publication owned by Snopes Media Group. All statistics are shown in Table 3.

Baselines. We compare our proposed model with the following baseline and state-of-the-art models on all three datasets: Traditional Text-only Approach: **TweetBert** (Qudar and Mago, 2020): A pre-trained language model on 850M tweets, fine-tuned for rumor verification. LLM-assisted Approaches: **LLaMA2**$_{text}$ (13b version) (Touvron et al., 2023b): Generates authenticity predictions directly from news statements and comment texts. **LLaMA2**$_{text}$ **+ SPT**: Fine-tunes LLaMA2$_{text}$ using soft prompt tuning (Lester et al., 2021). **ChatGPT**$_{text}$ (gpt-3.5-turbo-0613) (Ouyang et al., 2022): Similar to LLaMA2$_{text}$, generating predictions from text information. **ChatGPT**$_{full}$: Incorporates news statements, comment texts and user information for prediction. **CICAN** (Yang et al., 2023): A model that integrates collective intelligence with a ChatGPT-assisted network, using a semantic structure mining module to enhance semantic content capture. Graph-based Approaches: **GACL** (Sun et al., 2022): A GCN-based model using contrastive learning to improve robustness and generalization. **GARD** (Tao et al., 2024): A graph autoencoder model that captures semantic evolution in propagation and uses self-supervised feature reconstruction to enhance rumor detection. Heterogeneous Graph-based Approaches: **PSIN** (Min et al., 2022): A divide-and-conquer model analyzing post-user interactions, improving accuracy by focusing on social context interactions. **HAN** (Yang et al., 2016): Uses BiGRUs to explore hierarchical structures in news articles, learning embeddings from sentence contributions to differentiate fake from genuine news. **HG-SL** (Sun et al., 2023): Detects fake news by learning user spreading behaviors, combining global interaction and local context learning.

Experimental Setup. All models were implemented using the PyTorch framework and executed on an NVIDIA Tesla V100 GPU. In the heterogeneous graph propagation structure enhancement module of NMGIA, HAN (Wang et al., 2019) was adopted as the default encoder and decoder. The

[1]https://www.politifact.com/

[2]https://www.gossipcop.com/

[3]www.snopes.com

[4]www.poynter.org/coronavirusfactsalliance/

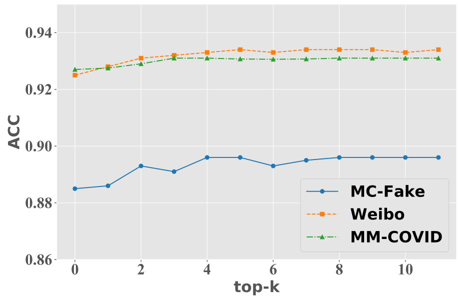


Figure 8: Evaluation of the value of top-k on NMGIA.

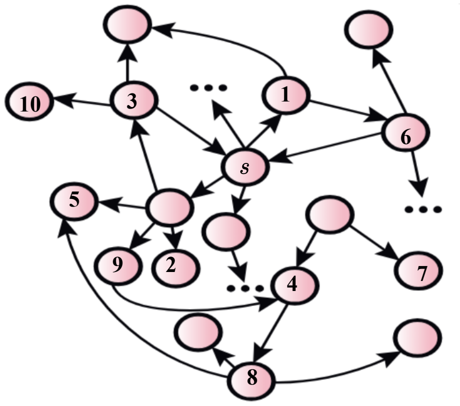


Figure 9: Propagation graph of the example in the case study

threshold for the adaptive sampling strategy was set to 0.5. In the text content enhancement module, the sliding window size for post ranking was set to 20 with a stride of 10, and the top-k values were set to $\{5, 8, 4\}$ for MC-Fake, Weibo, and MM-COVID, respectively. We utilized the cost-effective and efficient ChatGPT (gpt-3.5-turbo) as our LLM. All methods were evaluated with 5-fold cross-validation. For the English datasets MC-Fake and MM-COVID, we employed pre-trained Tweet-BERT for the embedding layer, while BERT was used for the Chinese Weibo dataset, both with 768 hidden units. During training, cross-entropy loss was used to assess the difference between predictions and true labels. Parameter updates were performed using the Adam optimizer (Kingma and Ba, 2014). We explored learning rates from 1e-4 to 5e-3, early stopping patience values from 5 to 20, and tested dropout and replacement rates from 0 to 0.5, with a step size of 0.1. All models were implemented using PyTorch and executed on an NVIDIA Tesla V100 GPU. Baseline models adhered to configurations from their original publications.

Table 4: The impact of sliding window hyperparameters on accuracy.

Window size	Step size	MC-Fake	Weibo	MM-COVID
20	10	**0.896**	**0.934**	**0.931**
40	20	0.890	0.930	0.928
60	30	0.888	0.927	0.928

.2 The Details of Parameter Analysis

We evaluated the impact of different top-k values on the model, focusing on the extraction of the most critical viewpoints for text enhancement. As shown in Figure 8, $k = \{5, 8, 4\}$ yielded the best and most consistent performance across the three datasets. This suggests that even a small number of viewpoints, extracted by the LLM, can enhance the model's understanding of the text by providing deeper insight.

Additionally, we analyzed the effect of the sliding window strategy by varying the window size while setting the stride to half the window size, as shown in Table 4. In this configuration, most comments were ranked twice per iteration. Results indicate that a window size of 20 performed best in capturing the viewpoint and narrative semantic features of rumor comments, while a window size of 60 performed the worst. This may be due to the challenges posed by large window sizes in handling long contexts and numerous posts. In contrast, MM-COVID was less affected by window size, likely due to the smaller number of posts per news event. Results from MC-Fake and Weibo indicate that our model and sliding window approach are better suited for more complex real-world environments.

Table 5: Examples of narrative ordering (partial)

Claim:Lindsay Lohan wears a hijab, sparking mixed feelings on social media. Content: Lindsay Lohan was spotted wearing a hijab at a conservative fashion show. The actress attended the second London Modest Fashion Week, dressed appropriately. Before the show, she posted a video on Instagram where she learned how to tie a hijab, expressing her desire to wear it traditionally.
1. Heard Lohan's new film role might involve cultural elements. Maybe this hijab is part of her character prep? **Explanation**: Early in the discussion, it is suggested that Lohan's wearing of the hijab might be related to a film role, setting a possible context of a non-personal or cultural decision.
2. It's for a movie, people! Let's not jump to conclusions. **Explanation**: Directly supporting the previous comment, explicitly stating the hijab is for a film role, reinforcing the credibility of this preparation.
3. Lohan's been spending a lot of time in Middle Eastern countries. This could be more than just a fashion statement. **Explanation**: Provides background information, explaining that Lohan might have a deeper understanding and respect for the culture due to her experiences in Middle Eastern countries.
4. Lohan in a hijab? Isn't this the actress known for breaking norms? Curious about the shift. **Explanation**: Raises a question as to why Lohan, previously known for breaking norms, would choose to wear a hijab, introducing a discussion of a possible personal transformation.
5. Emmmmm.... exploring or exploiting? Lohan's history suggests she's no stranger to controversy for attention. **Explanation**: Discusses whether Lohan's actions might be an exploration or exploitation of cultural differences for attention, adding to the skepticism about her motives.
6. While Lohan embraces a hijab, let's not forget women in some regions are forced to wear it without choice. **Explanation**: Reminds people not to overlook the sensitivity and complexity of the hijab globally, especially in the context of rights and freedom.
7. Lohan's hijab at a fashion show could be a powerful statement, or just another celebrity stunt. Thoughts? **Explanation**: Questions whether this is merely another PR stunt, suggesting that Lohan's actions might not stem from a genuine respect for the culture.
8. If it's genuine, it's a brave move in today's divisive climate. If not, it's highly problematic. **Explanation**: Discusses the authenticity of Lohan's actions and their potential social impact, questioning their sincerity.
9. Seeing Lohan in a hijab brings modest fashion into the limelight, but let's keep the focus on choice. **Explanation**: Emphasizes the importance of choice, even in discussions of fashion and cultural expression.
10. A celeb can wear a hijab and be applauded, while others are persecuted for the same. Let's discuss this disparity. **Explanation**: Raises the issue of double standards in wearing a hijab under different circumstances, highlighting societal reactions and potential injustices.
Viewpoint set: v_1: Celebrity Influence and Authenticity: Comments debate if Lindsay Lohan's hijab is a true cultural gesture or merely a publicity stunt, questioning the authenticity of her motives. v_2: Cultural Sensitivity and Appropriation: Discussions focus on the differences between a celebrity choosing to wear a hijab and women who are forced to wear it, addressing concerns about cultural sensitivity and exploitation. v_3: Personal Growth or Role Preparation: Speculations suggest Lohan's hijab might be linked to personal developments from her experiences in the Middle East or as preparation for a film role, implying a deeper narrative. v_4: Public and Media Perception: Observations note the different reactions celebrities and ordinary people receive for similar cultural or fashion choices, pointing out societal double standards.

Table 6: Viewpoint Set Processed by Agent B

Viewpoint set:
v_1': Cultural Sensitivity and Appropriation: Discussions emphasize the contrast between a celebrity's voluntary adoption of a hijab and the compulsory wearing by many women, highlighting issues of cultural sensitivity and potential exploitation.
v_2': Celebrity Influence and Authenticity: There is debate about whether Lohan's wearing of a hijab is a genuine cultural gesture or a calculated move to attract attention, questioning the authenticity of her actions.
v_3': Personal Growth or Role Preparation: Speculation suggests Lohan's choice may be influenced by personal growth from her time in Middle Eastern countries or preparation for a role involving cultural elements, providing context for her actions.
v_4': Public and Media Perception: Commentary points out the double standards in public and media responses to celebrities versus ordinary individuals adopting similar cultural attire, impacting how the claim is perceived.