JOINT NEIGHBOR SEMANTIC MODELING FOR MITIGATING OVER-SQUASHING IN HETEROGENEOUS SOCIAL EVENT DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Social Event Detection (SED) aims to identify real-world events from social media streams, where event semantics are distributed across messages connected by heterogeneous relations. Graph Neural Networks (GNNs) integrates this distributed semantic evidence through multi-hop neighborhood aggregation to learn event-discriminative message representations. However, compressing expanding neighborhoods into fixed-dimensional node representations can lead to over-squashing and the loss of critical event cues. Existing approaches largely address this problem through structural rewiring, leaving the organization of complementary neighbor semantics before propagation comparatively underexplored. In this work, we investigate a semantic bottleneck associated with over-squashing: aggregating individually encoded neighbors without explicitly modeling their complementary interactions can dilute or lose joint event semantics. To address this bottleneck, we propose **J**oint **N**eighbor **S**emantic **M**odeling (**JNSM**), a framework that organizes complementary neighbors into target-specific semantic units before propagation. JNSM first performs dual neighbor selection to construct candidate pairs from semantically close and distant neighbors. It then evaluates each pair using a target-centered semantic gain criterion that combines semantic contribution, redundancy suppression, and directional cancellation correction. High-gain pairs are selected as joint semantic units and encoded into learnable representations that capture complementary interactions and target-aware context. Through semantic-driven local rewiring, these representations are subsequently propagated as unified messages, enabling complementary event evidence to be jointly integrated into the target representation. Experiments on three heterogeneous SED benchmarks demonstrate competitive detection performance. Further analyses reveal consistent improvements in semantically diverse neighborhoods, supporting the effectiveness of semantic organization before propagation and its potential to alleviate semantic bottlenecks associated with over-squashing in heterogeneous SED.

1 INTRODUCTION

Social Event Detection (SED) aims to identify and organize social media messages associated with real-world events, supporting applications such as public opinion analysis Marozzo & Bessi (2018), fake news detection Shu et al. (2017), disaster response Liu et al. (2017), and trending event notification Macedo et al. (2015). With the rapid growth of social media, SED must process large-scale and continuously evolving streams, where messages are connected through users, entities, locations, timestamps, and other contextual signals Shang et al. (2016). Heterogeneous graphs therefore provide a natural representation of these interactions, modeling messages as nodes and their interactions through typed edges or meta-path-based relations.

Building on these representations, Graph Neural Networks (GNNs) have been widely adopted for SED to integrate textual features with relational dependencies through message passing. Existing heterogeneous SED methods, including ReDHAN Yu et al. (2023), HSED Qiu et al. (2024b), GraphHAM Qiu et al. (2024a), and PIGCN Yu et al. (2026), advance graph representation learning through heterogeneous attention, contrastive learning, hyperbolic geometry, and Helmholtz graph filtering. Other approaches enrich semantic evidence through data augmentation, as in SED-Aug Ma et al.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

(2025), or LLM-based text refinement and graph construction, as in LSED Qiu et al. (2025) and LHG Qiu et al. (2026). Within these graph-based approaches, neighborhood aggregation remains a central mechanism for incorporating information beyond the target message Veličković et al. (2017); Kipf & Welling (2017).

This mechanism is particularly valuable for SED because event semantics are often distributed across multi-hop interconnected messages rather than fully expressed in a single post Wu et al. (2021). For example, a target message such as "*so many people here, too crowded*" is ambiguous in isolation, whereas related messages containing "*Taylor Swift concert*", "*stadium queue*", and "*Sydney Olympic Park*" collectively reveal a coherent event context. By expanding their receptive fields, GNNs can integrate these complementary cues to recover event-level semantics that cannot be inferred from the target message alone Zhou et al. (2020).

However, this expansion also introduces a fundamental information bottleneck. As network depth increases, rapidly growing amounts of heterogeneous neighborhood evidence must be compressed into fixed-dimensional node representations, potentially leading to over-squashing and the loss of informative event cues Topping et al. (2022). Existing studies predominantly attribute this problem to structural bottlenecks caused by constrained information flow paths, motivating graph rewiring methods to improve long-range propagation. For example, FoSR Karhadkar et al. (2023) improves spectral connectivity, while LASER Barbero et al. (2024) introduces locality-preserving shortcut edges. Although these methods facilitate information propagation in terms of topology, improved connectivity alone does not ensure the preservation of joint neighbor semantics. Conventional message-passing GNNs typically propagate and aggregate neighbors independently, leaving complementary evidence to remain fragmented and potentially diluted or lost when compressed into the target representation. We refer to this information loss as a *semantic bottleneck*, providing a semantic perspective on over-squashing beyond structural information flow. This bottleneck is further complicated by redundant local evidence and individually weak but jointly informative distant cues. These limitations motivate explicitly identifying and jointly representing complementary neighbors before propagation to better preserve and integrate joint neighbor semantics into the target representation.

To address this bottleneck, we propose **J**oint **N**eighbor **S**emantic **M**odeling (**JNSM**), a semantic bottleneck-aware framework for heterogeneous SED. Rather than propagating neighboring messages independently, JNSM organizes complementary neighbors into target-specific joint semantic units before propagation. Specifically, JNSM first performs dual neighbor selection to retain semantically close and distant neighbors, from which candidate pairs are constructed. Each pair is then evaluated using a target-centered semantic gain criterion that combines semantic contribution, redundancy suppression, and directional cancellation correction. High-gain candidate pairs are selected as joint semantic units and encoded into learnable semantic unit representations that capture complementary interactions and target-aware context. Based on these representations, JNSM performs semantic-driven local rewiring and semantic unit propagation, enabling complementary evidence to be jointly represented and propagated as a unified message.

Our main contributions are summarized as follows:

- We introduce a semantic perspective on over-squashing in heterogeneous SED and define the semantic bottleneck, where independent neighbor aggregation can fragment, dilute, or lose the joint semantics of complementary neighbors during message passing.
- We propose Joint Neighbor Semantic Modeling (JNSM), a semantic bottleneck-aware framework that identifies complementary neighbors through target-centered semantic gain and represents them as joint semantic units before propagation. Through semantic-driven local rewiring and semantic unit propagation, JNSM preserves their joint semantics during message passing and reduces information loss under semantic compression.
- Experiments on three heterogeneous SED benchmarks demonstrate that JNSM achieves strong performance against state-of-the-art methods. Comprehensive ablation, sensitivity, and semantic bottleneck analyses further validate the effectiveness of joint neighbor semantic modeling for mitigating over-squashing.

2 PRELIMINARIES

Heterogeneous Social Graph. Following prior work Shang et al. (2016); Yu et al. (2023), we represent social media data as a heterogeneous information network $\mathcal{G} = (\mathcal{V}, \mathcal{R})$, where $\mathcal{V}$ contains message, user, entity, and word nodes, and $\mathcal{R}$ denotes the heterogeneous relations among them. To capture relation-specific semantics between messages, we construct message graphs using meta-paths.

Meta-path. A meta-path $p = \langle v_1 \xrightarrow{r_1} v_2 \cdots \xrightarrow{r_k} v_{k+1} \rangle$ defines a composite relation across heterogeneous node types. Using user-, entity-, and word-based meta-paths, we construct three relation-specific message graphs, denoted by $\mathcal{G}_P = \{\mathcal{G}_{p_u}, \mathcal{G}_{p_e}, \mathcal{G}_{p_w}\}$.

Joint Semantic Unit. For a target node v with neighborhood $\mathcal{N}(v)$, we define a *joint semantic unit* as a selected pair of target-specific complementary neighbors:

$$g^* = (u_i, u_j), \qquad u_i, u_j \in \mathcal{N}(v),\ u_i \neq u_j \tag{1}$$

The joint semantics of u_i and u_j are encoded into a unified representation conditioned on v, enabling JNSM to jointly model and propagate complementary neighbor evidence toward the target node.

3 METHODOLOGY

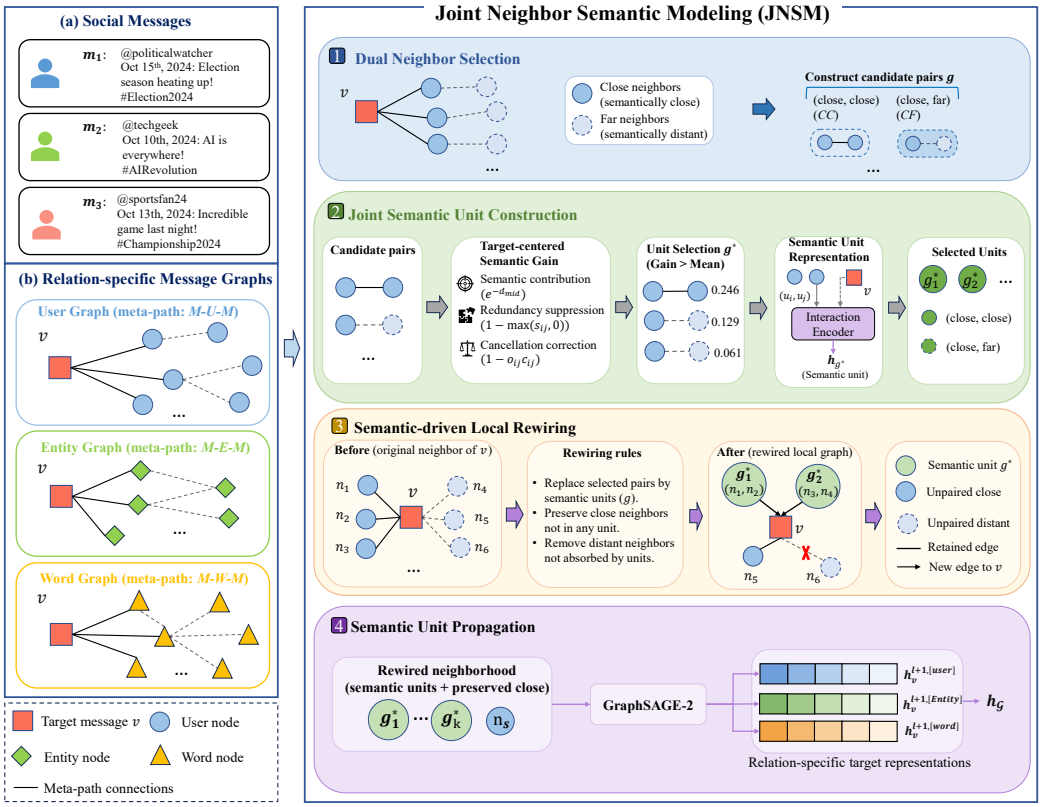

Figure 1: Overview of the proposed Joint Neighbor Semantic Modeling (JNSM) framework. For each relation-specific message graph, JNSM performs dual neighbor selection, constructs target-specific joint semantic units through semantic gain evaluation, conducts semantic-driven local rewiring, and propagates semantic unit messages to update the target representation. The resulting relation-specific representations are concatenated to form a global event-aware representation for event detection.

We propose **J**oint **N**eighbor **S**emantic **M**odeling (**JNSM**), a semantic bottleneck-aware framework for heterogeneous social event detection. Unlike conventional GNNs that propagate and aggregate neighbors independently, JNSM explicitly organizes complementary neighbors into target-specific joint semantic units before propagation, enabling their semantic interactions to be jointly represented and better preserved during message passing. This reduces information loss under semantic compression, thereby mitigating the semantic bottleneck associated with over-squashing.

As illustrated in Fig. 1, JNSM consists of four stages: (1) Dual Neighbor Selection, (2) Joint Semantic Unit Construction, (3) Semantic-Driven Local Rewiring, and (4) Semantic Unit Propagation.

3.1 Dual Neighbor Selection

Given a target node v at layer ℓ, JNSM partitions its sampled neighborhood into semantically close and distant sets, denoted by: $\mathcal{C}_v = \mathcal{C}_v^{\text{close}} \cup \mathcal{C}_v^{\text{distant}}$ (Validated in Exp. 4.4). The close-neighbor set captures locally consistent semantic evidence, while the distant-neighbor set provides informative complementary event cues beyond the close neighborhood.

Specifically, neighbors are ranked in ascending order of their semantic distance to the target:

$$d(v,u) = \left\| \mathbf{h}_v^{(\ell)} - \mathbf{h}_u^{(\ell)} \right\|_2 \tag{2}$$

The k nearest neighbors form the close-neighbor set $\mathcal{C}_v^{\text{close}} = \text{Top-}k(\mathcal{N}(v))$, while the remaining neighbors form the distant-neighbor set $\mathcal{C}_v^{\text{distant}} = \mathcal{N}(v) \setminus \mathcal{C}_v^{\text{close}}$.

3.2 Joint Semantic Unit Construction

Given the close and distant neighbor sets, JNSM constructs target-specific joint semantic units to explicitly model complementary neighbor semantics before propagation. This process consists of four steps: candidate pair construction, target-centered semantic gain evaluation, semantic unit selection, and semantic unit representation.

3.2.1 Candidate Pair Construction.

For each target node v, JNSM constructs two types of candidate pairs: close–close (CC) pairs that capture locally consistent semantics, and close–far (CF) pairs that incorporate complementary evidence from semantically distant neighbors (Validated in Exp. C.2.1). Formally, the candidate pairs is defined as

$$g_v = \left\{ (u_i, u_j) \;\middle|\; u_i \in \mathcal{C}_v^{\text{close}},\ u_j \in \mathcal{C}_v^{\text{close}} \cup \mathcal{C}_v^{\text{distant}},\ u_i \neq u_j \right\} \tag{3}$$

Each candidate pair is subsequently evaluated according to its semantic gain with respect to the target node.

3.2.2 Target-centered Semantic Gain.

JNSM evaluates each candidate pair $g_v = (u_i, u_j)$ using a target-centered semantic gain score that combines its joint semantic contribution, redundancy suppression, and directional cancellation correction. We first define the target-centered semantic offsets of the two neighbors:

$$\boldsymbol{\delta}_i = \mathbf{h}_{u_i}^{(\ell)} - \mathbf{h}_v^{(\ell)}, \qquad \boldsymbol{\delta}_j = \mathbf{h}_{u_j}^{(\ell)} - \mathbf{h}_v^{(\ell)} \tag{4}$$

Based on these offsets, the joint semantic contribution of the candidate pair relative to the target is characterized by the midpoint distance

$$d_{\text{mid}}(g,v) = \left\| \frac{\boldsymbol{\delta}_i + \boldsymbol{\delta}_j}{2} \right\|_2 \tag{5}$$

where a smaller $d_{\text{mid}}(g,v)$ indicates stronger alignment between the joint semantics and the target, corresponding to a higher semantic contribution score.

To assess semantic redundancy within each candidate pair, we compute cosine similarity between the two target-centered semantic offsets:

$$s_{ij} = \frac{\boldsymbol{\delta}_i^\top \boldsymbol{\delta}_j}{\|\boldsymbol{\delta}_i\|_2 \|\boldsymbol{\delta}_j\|_2} \tag{6}$$

The resulting redundancy suppression factor, $1 - \max(s_{ij}, 0)$, penalizes highly aligned neighbor pairs while favoring pairs with complementary semantic directions.

However, two strongly opposing neighbor offsets can produce an artificially small midpoint distance through vector cancellation, resulting in spuriously high semantic gain. To correct such cases, we assess cancellation

in terms of both magnitude and direction. We first define the average endpoint distance $d_{\text{end}}(g, v)$ and the corresponding magnitude cancellation score c_{ij} as

$$d_{\text{end}}(g, v) = \frac{\|\boldsymbol{\delta}_i\|_2 + \|\boldsymbol{\delta}_j\|_2}{2}, \qquad c_{ij} = \left[1 - \frac{d_{\text{mid}}(g, v)}{d_{\text{end}}(g, v) + \epsilon}\right]_0^1 \tag{7}$$

where $\epsilon > 0$ ensures numerical stability and $[\cdot]_0^1$ denotes clipping to $[0, 1$. The score c_{ij} measures the relative reduction in magnitude when the two offsets are averaged. Directional opposition o_{ij} is quantified as

$$o_{ij} = \left[\frac{-s_{ij} - \rho}{1 - \rho}\right]_0^1 \tag{8}$$

where $\rho = 0.95$ restricts the correction to severely opposing directions, i.e., $s_{ij} < -\rho$. The resulting correction factor, $1 - o_{ij}c_{ij}$, strongly penalizes a pair only when substantial magnitude cancellation coincides with severe directional opposition.

Combining semantic contribution, redundancy suppression, and cancellation correction, the target-centered semantic gain is formulated as

$$\text{Gain}(g, v) = \underbrace{\exp(-d_{\text{mid}}(g, v))}_{\text{semantic contribution}} \underbrace{(1 - \max(s_{ij}, 0))}_{\text{redundancy suppression}} \underbrace{(1 - o_{ij}c_{ij})}_{\text{cancellation correction}} \tag{9}$$

This criterion favors target-aligned, non-redundant complementary pairs while preventing severe directional cancellation from producing spuriously high semantic gains.

3.2.3 Semantic Unit Selection.

Given the semantic gains of candidate pairs, JNSM selects informative candidate pairs as **joint semantic units** using a target-adaptive mean threshold rather than a fixed global threshold. Specifically, it retains pairs whose semantic gains exceed both the target-specific mean and zero:

$$g_v^* = \left\{ g \in g_v \;\middle|\; \text{Gain}(g, v) > \max\left(\bar{\text{Gain}}_v, 0\right) \right\} \tag{10}$$

where $\bar{\text{Gain}}_v = \frac{1}{|g_v|} \sum_{g \in g_v} \text{Gain}(g, v)$ is the mean gain over the candidate pairs associated with target node v. From the retained candidates, pairs are greedily selected in descending gain order under a non-overlapping constraint, allowing each neighbor to participate in at most one unit. If none satisfies the criterion, the highest-gain pair is retained as a fallback.

3.2.4 Semantic Unit Representation.

For each selected joint semantic unit $g^* = (u_i, u_j)$, JNSM learns a unified representation that captures complementary neighbor semantics with respect to the target node v. We first construct a base representation by averaging the constituent neighbor embeddings:

$$\mathbf{b}_{g^*}^{(\ell)} = \frac{\mathbf{h}_{u_i}^{(\ell)} + \mathbf{h}_{u_j}^{(\ell)}}{2} \tag{11}$$

To enrich this base representation with within-unit semantic interactions, we employ a semantic unit encoder $\phi(\cdot)$ to learn a semantic residual $\Delta\mathbf{h}_{g^*,v}^{(\ell)}$. Its input $\mathbf{z}_{g^*,v}^{(\ell)}$ concatenates the base representation, semantic difference, multiplicative interaction, and target-aware context. The learned residual is used to refine the base representation to obtain the final semantic unit representation:

$$\Delta\mathbf{h}_{g^*,v}^{(\ell)} = \phi\left(\mathbf{z}_{g^*,v}^{(\ell)}\right), \qquad \mathbf{h}_{g^*}^{(\ell)} = \text{Norm}\left(\mathbf{b}_{g^*}^{(\ell)} + \alpha\Delta\mathbf{h}_{g^*,v}^{(\ell)}\right) \tag{12}$$

Here, α controls the residual strength. The resulting representation preserves the stable mean semantics while incorporating learnable semantic interactions for subsequent local rewiring and propagation.

3.3 Semantic-driven Local Rewiring

Given the selected joint semantic units, JNSM rewires each target node's local neighborhood for unit-level propagation. For each unit $g^* = (u_i, u_j)$, JNSM introduces a temporary semantic unit node with representation $\mathbf{h}_{g^*}^{(\ell)}$ and connects it to the target v, replacing the original edges from u_i and u_j. This allows their joint semantics to propagate as a unified message. Unpaired close

neighbors retain their direct connections to preserve locally consistent evidence, while connections from unpaired distant neighbors to the target are removed. The resulting neighborhood is

$$\widetilde{\mathcal{N}}(v) = g_v^* \cup \mathcal{C}_v^{\mathrm{close}^-} \tag{13}$$

where $\mathcal{C}_v^{\mathrm{close}^-}$ denotes the set of unpaired close neighbors. The rewired neighborhood thus combines joint semantic evidence from selected units with individual evidence from the remaining close neighbors.

3.4 Semantic Unit Propagation

After semantic-driven local rewiring, JNSM applies standard GraphSAGE aggregation over the rewired neighborhood $\widetilde{\mathcal{N}}(v)$ to integrate the representations of selected semantic units and unpaired close neighbors. The target representation is updated as

$$\mathbf{h}_v^{(\ell+1)} = \mathrm{GraphSAGE}\left(\mathbf{h}_v^{(\ell)}, \widetilde{\mathcal{N}}(v)\right) \tag{14}$$

3.5 Optimization Objective

To enhance intra-event cohesion and inter-event separability, JNSM is optimized using a supervised triplet loss:

$$\mathcal{L}_{\mathrm{tri}} = \sum_{(m_i, m_i^+, m_i^-)} \max\left\{\mathcal{D}(\mathbf{h}_{m_i}, \mathbf{h}_{m_i^+}) - \mathcal{D}(\mathbf{h}_{m_i}, \mathbf{h}_{m_i^-}) + C,\, 0\right\} \tag{15}$$

where m_i, m_i^+, and m_i^- denote the anchor, positive, and negative messages, respectively. $\mathcal{D}(\cdot,\cdot)$ denotes the Euclidean distance between message representations, and C is the margin, set to 8 by default.

4 Experiments

This section evaluates JNSM's overall detection performance and examines the role of joint neighbor semantic modeling in addressing semantic bottlenecks. We further assess the contributions of individual components, the effects of key design choices, and JNSM's scalability to larger heterogeneous social graphs. The experiments are designed to answer the following research questions:

- **RQ1** Does JNSM provide competitive and generalizable SED performance across heterogeneous datasets?
- **RQ2** Does JNSM effectively mitigate semantic bottlenecks in heterogeneous neighborhood aggregation?
- **RQ3** How does each major component contribute to the effectiveness of JNSM?
- **RQ4** How do different semantic unit construction and representation strategies affect JNSM?
- **RQ5** How well does JNSM scale to larger heterogeneous social graphs?

4.1 Experiment Setup

Datasets and Baselines. We evaluate JNSM in a supervised setting on three publicly available and widely used SED datasets: Twitter2012 McMinn et al. (2013), CrisisLexT Olteanu et al. (2015), and French Hasan et al. (2019). The baselines include conventional GNNs (DoubleGCN Kipf & Welling (2017), GraphSAGE Hamilton et al. (2017), GIN Xu et al. (2019), and GraphFormer Shi et al. (2021)), heterogeneous SED methods (ReDHAN Yu et al. (2023), HSED Qiu et al. (2024b), SED-Aug Ma et al. (2025), and PIGCN Yu et al. (2026)), and structural rewiring method LASER Barbero et al. (2024). Dataset statistics and baseline descriptions are provided in Appendix C.1.1.

Evaluation Metrics. Following prior SED studies, we evaluate event clustering performance using Normalized Mutual Information (NMI), Adjusted Mutual Information (AMI), and Adjusted Rand Index (ARI) Vinh et al. (2009). We also report runtime in minutes to evaluate computational efficiency.

4.2 Overall Offline Detection Performance (RQ1)

To answer RQ1, we evaluate all methods in the offline SED setting using SBERT embeddings as initial message features. Table 1 reports clustering performance across the three benchmark datasets. Runtime comparisons are provided in Appendix C.3.

JNSM achieves competitive performance across the three benchmarks. On Twitter2012, it obtains the highest NMI and AMI (94.42% and 91.81%), surpassing its GraphSAGE backbone by 0.36 and 0.54 percentage points,

Table 1: Offline SED Performance (Best in red, Second-best underlined)

Models	Twitter2012			CrisisLexT			French		
	NMI	AMI	ARI	NMI	AMI	ARI	NMI	AMI	ARI
Conventional GNNs									
DoubleGCN	92.95	89.57	83.95	87.69	87.67	93.32	75.69	70.51	49.50
GIN	91.11	87.02	81.06	87.18	87.17	92.94	77.45	72.78	67.28
GraphFormer	92.58	88.96	80.50	87.40	87.39	93.07	73.53	67.76	39.54
Heterogeneous SED Methods									
ReDHAN	92.41	88.79	81.07	86.76	86.74	92.31	72.61	66.69	55.44
HSED	85.95	79.13	67.93	76.03	76.00	84.93	65.73	58.09	24.51
SED-Aug	92.52	88.89	79.11	87.78	87.77	93.32	74.68	69.18	42.35
PIGCN	92.77	89.34	81.77	87.15	87.13	92.94	71.22	64.82	34.63
Structural Rewiring Method									
LASER	87.58	82.14	78.62	77.64	77.61	86.14	72.83	67.45	48.49
GraphSAGE	94.06	91.27	85.26	89.77	89.76	94.59	73.51	67.66	36.77
JNSM	94.42	91.81	84.77	90.07	90.05	94.84	74.52	68.91	39.27

respectively, with slightly lower ARI (84.77% vs. 85.26%). On CrisisLexT, JNSM ranks first across all three metrics, achieving 90.07% NMI, 90.05% AMI, and 94.84% ARI. On French, it improves upon GraphSAGE across all three metrics, although GIN remains the strongest method. These results support the effectiveness of organizing complementary neighbor semantics before propagation, with benefits that vary across datasets and metrics.

The comparison with LASER further highlights the value of semantic organization. JNSM outperforms this structural rewiring baseline across all three metrics on Twitter2012 and CrisisLexT with NMI gains of 6.84 and 12.43 percentage points, respectively. On French, it improves NMI and AMI by 1.69 and 1.46 percentage points, but achieves lower ARI (39.18% vs. 48.49%). Whereas LASER modifies graph connectivity to facilitate long-range information flow, JNSM explicitly models complementary neighbor semantics before propagation to alleviate the dilution or loss of semantic information. Its improvements over LASER in these comparisons suggest that organizing semantic evidence before aggregation offers a promising avenue for addressing semantic bottlenecks associated with over-squashing. Overall, these results demonstrate the effectiveness of joint neighbor semantic modeling for social event detection and support its potential to mitigate semantic information loss during message propagation.

4.3 Semantic Bottleneck Mitigation Analysis RQ2

To examine whether JNSM mitigates semantic bottlenecks, we focus on semantically diverse neighborhoods, where heterogeneous event evidence must be compressed into a fixed-dimensional target representation. For each target node v, we quantify neighborhood semantic diversity as the average deviation of neighbor representations from their semantic centroid:

$$D_v = \frac{1}{|\mathcal{N}(v)|} \sum_{u \in \mathcal{N}(v)} \left\| \mathbf{h}_u - \bar{\mathbf{h}}_{\mathcal{N}(v)} \right\|_2 \tag{16}$$

where $\bar{\mathbf{h}}_{\mathcal{N}(v)}$ denotes the mean representation of neighbors in $\mathcal{N}(\sqsubseteq)$. A larger D_v indicates greater semantic dispersion among neighboring messages and therefore more diverse evidence to be compressed during aggregation. Without explicitly organizing complementary neighbor semantics, this diversity may increase the risk of losing informative event cues under aggregation. We define the *high-diversity* regime as target nodes whose D_v values fall within the upper tertile of each dataset and compare GraphSAGE, LASER, and JNSM within this regime across the three benchmarks.

As shown in Table 2, JNSM achieves the strongest performance across all three metrics in high-diversity neighborhoods on all three benchmarks. Compared with GraphSAGE, JNSM improves NMI, AMI, and ARI by 1.76, 4.13, and 8.42 points on Twitter2012, with gains of 3.46–3.97 points on CrisisLexT and 1.94–2.62 points on French. JNSM also consistently outperforms LASER, particularly on CrisisLexT, suggesting that improving structural information flow alone may be insuffi-

Table 2: SED Performance (%) in High-Diversity Neighborhoods (mean±std)

Dataset	Method	NMI	AMI	ARI
Twitter2012	GraphSAGE	92.40±1.20	82.92±2.92	74.39±6.50
	LASER	84.19±1.26	64.55±2.81	57.19±4.63
	JNSM	**94.16**±1.17	**87.05**±2.60	**82.81**±5.35
CrisisLexT	GraphSAGE	66.42±3.28	66.16±3.31	80.47±2.78
	LASER	37.42±1.98	37.01±1.98	53.29±3.50
	JNSM	**69.88**±3.03	**69.64**±3.07	**84.44**±2.10
French	GraphSAGE	68.56±2.42	60.52±3.03	38.01±3.54
	LASER	66.78±1.58	58.42±2.03	36.13±4.61
	JNSM	**70.50**±3.27	**62.92**±4.15	**40.63**±7.01

Table 3: Component Ablation on Twitter2012 (mean±std,%)

Variant	NMI	AMI	ARI	Time (min)
Full JNSM	**94.42**±0.21	**91.81**±0.32	84.77±1.72	4.4
w/o Dual Selection	93.88±0.39	90.99±0.57	83.82±1.36	3.9
w/o Semantic Gain	93.88±0.65	90.98±0.99	82.33±2.02	3.5
w/o Unit Representation	93.38±1.39	90.25±2.12	82.69±5.03	3.2
w/o Local Rewiring	93.88±0.39	90.99±0.59	**85.56**±1.57	3.9

cient under high semantic diversity. By jointly representing complementary neighbors before propagation, JNSM better preserves their semantics during message passing, alleviating the semantic bottleneck associated with independent neighbor aggregation.

4.4 Component Ablations (RQ3)

To assess the contribution of each major component, we compare JNSM with four ablation variants, each disabling or replacing one component while remaining pipelines unchanged.

- **w/o Dual Selection**: Constructs candidate pairs from a unified sampled neighborhood without distinguishing semantically close and distant neighbors.
- **w/o Semantic Gain**: Replaces semantic gain-guided selection with random candidate pair selection under the same target-specific selection budget.
- **w/o Unit Representation**: Replaces the learnable semantic unit representation with parameter-free mean pooling of the two constituent neighbor embeddings.
- **w/o Local Rewiring**: Disables semantic-driven local rewiring and performs GraphSAGE propagation directly on the original neighborhood.

Table 3 shows that removing any JNSM component reduces mean NMI and AMI, supporting the contribution of all four stages for reliable joint neighbor semantic modeling. Replacing the learnable semantic unit representation with mean pooling produces the largest decreases in these metrics: NMI falls from $94.42\% \pm 0.21$ to $93.38\% \pm 1.39$ and AMI from $91.81\% \pm 0.32$ to $90.25\% \pm 2.12$, highlighting the importance of explicitly modeling within-unit semantic interactions beyond simple averaging. Removing dual neighbor selection or semantic gain also lowers all three clustering metrics, supporting the importance of constructing diverse candidates and selecting informative pairs. Disabling local rewiring reduces NMI and AMI but slightly increases ARI from 84.77% to 85.56%, reflecting a trade-off between preserving fine-grained pairwise distinctions through original neighbor connections, as captured by ARI, and promoting complementary semantic consistency through joint semantic units. Overall, the full model achieves the highest NMI and AMI with modest addi-

Table 4: Performance and Runtime on Larger Social Graphs (Best in red, Second-best underlined)

Models	Twitter2012 (11K nodes, 3M edges)				CrisisLexT (9K nodes, 33M edges)			
	NMI	AMI	ARI	Time	NMI	AMI	ARI	Time
DoubleGCN	85.74	80.14	52.95	25.6	91.53	91.51	95.00	26.0
GIN	84.49	78.64	62.29	28.7	90.96	90.94	94.54	30.5
GraphFormer	85.14	79.35	49.22	21.6	91.29	91.28	94.93	24.3
ReDHAN	83.57	77.36	72.52	51.5	92.31	92.29	95.62	28.5
HSED	79.09	70.57	31.87	35.9	80.62	80.58	87.00	14.3
SED-Aug	85.36	79.53	42.63	12.6	92.71	92.69	95.73	18.6
PIGCN	83.77	77.25	38.51	27.2	91.28	91.26	94.99	29.2
LASER	80.95	74.19	70.07	62.2	83.13	83.10	88.78	48.4
GraphSAGE	84.69	78.60	37.07	13.3	93.55	93.54	96.27	17.4
JNSM	84.12	77.74	35.32	11.0	94.14	94.13	96.58	14.9

tional runtime, supporting the combined effectiveness of identifying, representing, and propagating complementary neighbor semantics.

4.5 Scalability on Larger Social Graphs (RQ5)

To evaluate the scalability and effectiveness of JNSM on larger social graphs, we expand Twitter2012 from 4.7K nodes and 0.5M edges to 11K nodes and 3M edges, and CrisisLexT from 6K nodes and 17M edges to 9K nodes and 33M edges, using SBERT representations in both settings. These enlarged benchmarks examine whether JNSM maintains detection performance and computational efficiency as graph size and connectivity increase.

Table 4 shows that JNSM remains computationally efficient on larger heterogeneous graphs, requiring less runtime than GraphSAGE on both benchmarks. On CrisisLexT, it achieves the highest scores across all three clustering metrics, whereas on Twitter2012, it yields lower scores than its backbone. LASER achieves 70.07% ARI on Twitter2012, surpassing conventional GNN baselines and suggesting potential benefits from improving structural connectivity. However, JNSM outperforms LASER in NMI and AMI on Twitter2012 and across all three metrics on CrisisLexT, with substantially lower runtime. These results highlight the effectiveness and computational efficiency of joint neighbor semantic modeling on larger graphs.

Additional analysis of candidate pair construction, semantic gain design, adaptive thresholding, semantic unit representation, and close-neighbor sampling are provided in Appendix C.2.

5 Conclusion

In this paper, we propose **J**oint **N**eighbor **S**emantic **M**odeling (**JNSM**), a semantic bottleneck-aware framework for heterogeneous social event detection. JNSM addresses a semantic bottleneck associated with over-squashing, where independent neighbor aggregation can dilute or lose complementary event evidence when compressing diverse neighborhood information into fixed-dimensional target representations. To alleviate this bottleneck, JNSM explicitly models complementary neighbor interactions by organizing them into target-specific joint semantic units before aggregation and propagating their learned representations through semantic-driven local rewiring. Experiments on three benchmarks demonstrate competitive detection performance and computational efficiency, with consistent improvements in semantically diverse neighborhoods and over structural rewiring methods. These findings support semantic organization before propagation as a promising direction for alleviating semantic bottlenecks associated with over-squashing. Future work will extend JNSM to dynamic and streaming graphs through adaptive semantic unit construction.

AI USE STATEMENT

Generative AI tools were used to assist with research methodology refinement, experimental analysis, and interpretation of results. They were also used to support code refinement and to improve the clarity, structure, and readability of the manuscript. The authors critically reviewed, verified, and revised all AI-assisted suggestions, analyses, code modifications, and written content. The authors independently validated the experimental results and conclusions reported in this work. The authors take full responsibility for the methodology, implementation, experimental results, claims, and final content of this work.

REFERENCES

Albert Angel, Nick Koudas, Nikos Sarkas, and Divesh Srivastava. Dense subgraph maintenance under streaming edge weight updates for real-time story identification. *The VLDB Journal*, 23(2): 175–199, 2014.

Farzindar Atefeh and Wael Khreich. A survey of techniques for event detection in twitter. *Computational Intelligence*, 31(1):132–164, 2015.

Federico Barbero, Ameya Velingker, Amin Saberi, Michael Bronstein, and Francesco Di Giovanni. Locality-aware graph rewiring in gnns. In *International Conference on Learning Representations*, volume 2024, pp. 28716–28737, 2024.

Thomas M Cover, Joy A Thomas, and John Kieffer. *Elements of information theory*, volume 2. wiley New York, 1991.

Francesco Di Giovanni, Lorenzo Giusti, Federico Barbero, Giulia Luise, Pietro Lio, and Michael M Bronstein. On over-squashing in message passing neural networks: The impact of width, depth, and topology. In *International conference on machine learning*, pp. 7865–7885. PMLR, 2023.

Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

Mahmud Hasan, Mehmet A Orgun, and Rolf Schwitter. Real-time event detection from the twitter data stream using the twitternews+ framework. *Information Processing & Management*, 56(3): 1146–1165, 2019.

Kedar Karhadkar, Pradeep Kr Banerjee, and Guido Montúfar. Fosr: First-order spectral rewiring for addressing oversquashing in gnns. In *The Eleventh International Conference on Learning Representations (ICLR 2023)*, 2023.

Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2017.

Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the AAAI conference on artificial intelligence*, number 1, 2018.

Zhiwei Liu, Yang Yang, Zi Huang, Fumin Shen, Dongxiang Zhang, and Heng Tao Shen. Event early embedding: Predicting event volume dynamics at early stage. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 997–1000, 2017.

Congbo Ma, Yuxia Wang, Jia Wu, Jian Yang, Jing Du, Zitai Qiu, Qing Li, Hu Wang, and Preslav Nakov. Explicit and implicit data augmentation for social event detection. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 8402–8415, Vienna, Austria, 2025.

Augusto Q Macedo, Leandro B Marinho, and Rodrygo LT Santos. Context-aware event recommendation in event-based social networks. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pp. 123–130, 2015.

Fabrizio Marozzo and Alessandro Bessi. Analyzing polarization of social media users and news sites during political campaigns. *Social Network Analysis and Mining*, 8:1–13, 2018.

Andrew J McMinn, Yashar Moshfeghi, and Joemon M Jose. Building a large-scale corpus for evaluating event detection on twitter. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pp. 409–418, 2013.

Alexandra Olteanu, Sarah Vieweg, and Carlos Castillo. What to expect when the unexpected happens: Social media communications across crises. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pp. 994–1009, 2015.

Kenta Oono and Taiji Suzuki. Graph neural networks exponentially lose expressive power for node classification. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=S1ldO2EFPr.

Zitai Qiu, Congbo Ma, Jia Wu, and Jian Yang. An efficient automatic meta-path selection for social event detection via hyperbolic space. In *Proceedings of the ACM on Web Conference 2024*, WWW '24, pp. 2519–2529, New York, NY, USA, 2024a. Association for Computing Machinery. ISBN 9798400701719.

Zitai Qiu, Jia Wu, Jian Yang, Xing Su, and Charu C. Aggarwal. Heterogeneous social event detection via hyperbolic graph representations. *IEEE Transactions on Big Data*, pp. 1–15, 2024b.

Zitai Qiu, Congbo Ma, Jia Wu, and Jian Yang. Text is all you need: Llm-enhanced incremental social event detection. In *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 4666–4680, 2025.

Zitai Qiu, Rongwei Xu, Congbo Ma, Shan Xue, Jian Yang, Guanfeng Liu, Quan Z Sheng, Amin Beheshti, and Jia Wu. Lhg: Llm-enhanced and heterogeneous graph-induced for unsupervised social event detection. In *Proceedings of the ACM Web Conference 2026*, WWW '26, pp. 3612–3623, New York, NY, USA, 2026. Association for Computing Machinery.

Jingbo Shang, Meng Qu, Jialu Liu, Lance M Kaplan, Jiawei Han, and Jian Peng. Meta-path guided embedding for similarity search in large-scale heterogeneous information networks. *arXiv preprint arXiv:1610.09769*, 2016.

Yunsheng Shi, Zhengjie Huang, Shikun Feng, Hui Zhong, Wenjin Wang, and Yu Sun. Masked label prediction: Unified message passing model for semi-supervised classification, 2021.

Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, 19(1):22–36, 2017.

Jake Topping, Francesco Di Giovanni, Benjamin Paul Chamberlain, Xiaowen Dong, and Michael M Bronstein. Understanding over-squashing and bottlenecks on graphs via curvature. In *International Conference on Learning Representations (ICLR)*, 2022.

Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information theoretic measures for clusterings comparison: is a correction for chance necessary? In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML '09, pp. 1073–1080, New York, NY, USA, 2009. Association for Computing Machinery.

Li Wang, Wei Huang, Miao Zhang, Shirui Pan, Xiaojun Chang, and Steven Weidong Su. Pruning graph neural networks by evaluating edge properties. *Knowledge-Based Systems*, 256:109847, 2022.

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

Zhanghao Wu, Paras Jain, Matthew Wright, Azalia Mirhoseini, Joseph E Gonzalez, and Ion Stoica. Representing long-range context for graph neural networks with global attention. *Advances in neural information processing systems*, 34:13266–13279, 2021.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations (ICLR)*, 2019.

Yongsheng Yu, Jia Wu, and Jian Yang. Social event detection with reinforced deep heterogeneous graph attention network. In *2023 IEEE International Conference on Big Data (BigData)*, pp. 463–472, 2023.

Yongsheng Yu, Congbo Ma, Zitai Qiu, Shan Xue, Jian Yang, and Jia Wu. Pigcn: Physics-inspired graph convolution networks for heterogeneous social event detection. In *Proceedings of the ACM Web Conference 2026*, pp. 992–1003, 2026.

Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *AI open*, 1:57–81, 2020.

A Related Works

A.1 Social Event Detection

Early Social Event Detection (SED) methods primarily relied on topic modeling Atefeh & Khreich (2015) and text embedding or clustering techniques Angel et al. (2014). While effective at capturing broad topical patterns, text-centered approaches often struggle to model relational dependencies inherent in social streams, where event evidence spans messages, users, entities, timestamps, and hashtags. This limitation has motivated a shift from isolated text modeling toward graph-based SED, where social messages and their heterogeneous relations can be jointly modeled.

Recent SED methods increasingly leverage Graph Neural Networks (GNNs) to capture structural and semantic dependencies among social messages. ReDHAN Yu et al. (2023) incorporates dual contrastive learning to improve representation robustness against noisy neighborhood structures. HSED Qiu et al. (2024b) learns hyperbolic graph representations to capture hierarchical event structures, while GraphHAM Qiu et al. (2024a) combines automatic meta-path selection with semantic modeling in hyperbolic space. PIGCN Yu et al. (2026) introduces a physics-inspired approach, integrating Helmholtz graph filtering with Gaussian interaction-based edge weighting to model wave-like propagation and adapt neighbor contributions. These methods advance event representation learning through neighbor selection, learning objectives, geometric modeling, and propagation mechanisms.

Complementary approaches enrich the semantic evidence available for event detection. SED-Aug Ma et al. (2025) combines explicit LLM-based text augmentation with implicit feature-space perturbations to increase data diversity and improve model robustness. LSED Qiu et al. (2025) uses LLMs to normalize informal social media expressions before clustering, while LHG Qiu et al. (2026) leverages LLMs to implicitly construct heterogeneous graphs from semantic relations. These approaches improve event modeling through data augmentation, text refinement, and semantic relation discovery.

Despite these advances, explicitly organizing complementary neighbors into joint neighbor units before propagation remains comparatively underexplored. Conventional neighborhood aggregation does not explicitly model such joint semantics before compressing neighborhood evidence into a fixed-dimensional target representation, potentially diluting event cues that are informative in combination. In contrast, JNSM addresses this limitation by constructing target-specific joint semantic units that capture complementary neighbor interactions before propagation, aiming to preserve joint neighbor semantics and alleviate semantic bottlenecks during message passing.

A.2 Over-Squashing

Over-squashing is a fundamental limitation of message-passing GNNs, arising when information from rapidly expanding multi-hop neighborhoods is compressed into fixed-dimensional node representations. This limitation is particularly pronounced when relevant information originates from distant or poorly connected regions of the graph. Subsequent studies have further characterized over-squashing through structural bottlenecks, curvature, and long-range information propagation Topping et al. (2022); Di Giovanni et al. (2023).

A major line of research mitigates over-squashing through graph rewiring, which modifies graph topology to improve information flow. Curvature-based methods identify structural bottlenecks using geometric measures and adjust edges to facilitate communication Topping et al. (2022). Spectral approaches such as FoSR Karhadkar et al. (2023) enhance graph expansion and spectral connectivity to facilitate long-range propagation, while locality-aware methods such as LASER Barbero et al. (2024) balance improved connectivity with locality and sparsity to avoid excessive graph densification. These approaches primarily address the structural aspect of over-squashing by improving the paths through which information propagates.

However, improved connectivity does not by itself ensure that complementary neighborhood semantics are preserved during aggregation. In heterogeneous SED, event semantics is often distributed across neighbors whose cues become informative when considered jointly Zhou et al. (2020); Wu et al. (2021). For example, references to a location, person, and activity may collectively identify an event that no individual message fully describes. Without explicitly modeling these complementary relationships before aggregation, their joint semantics may be diluted or lost as neighborhood evidence is compressed into the target representation. This motivates a semantic perspective on over-squashing: beyond limitations on structural connections, information loss also arises from independently aggregating diverse neighborhood semantics.

Related approaches improve neighborhood quality through pruning or attention. Pruning-based graph learning Wang et al. (2022) removes less informative connections, while graph attention Veličković et al. (2017) adaptively weights neighbor contributions. Heterogeneous attention methods such as HAN Wang et al. (2019) further integrate node-level and meta-path-level information. Although these mechanisms select, weight, or combine neighborhood evidence, they do not explicitly construct target-specific units that jointly represent complementary neighbor semantics before propagation.

JNSM addresses this gap by organizing semantic evidence before neighborhood compression. It constructs candidate pairs from semantically close and distant neighbors, evaluates their target-centered semantic gain, and encodes selected complementary pairs as joint semantic units before aggregation. Propagating these learned units enables inter-neighbor interactions to be represented before target-level aggregation, helping preserve event evidence that may otherwise be diluted by independent neighbor aggregation. JNSM thus complements topology-oriented rewiring by addressing the semantic bottleneck associated with over-squashing.

B METHOD

B.1 THEORETICAL MOTIVATION FOR JOINT NEIGHBOR SEMANTIC MODELING

Semantic Bottleneck under Independent Aggregation. Event semantics in heterogeneous social graphs are often distributed across messages connected through users, entities, words, and temporal relations. A single neighbor may therefore provide only partial evidence, while complementary cues from multiple neighbors collectively reveal a more complete event context Zhou et al. (2020). Effective event representation thus benefits from modeling both individual neighbor contributions and their semantic interactions. A simplified neighborhood aggregation update for a target node v in GNNs can be expressed as:

$$\mathbf{h}_v^{(\ell+1)} = \sigma\Big(W^{(\ell)} \cdot \mathrm{AGG}\left(\{\mathbf{h}_u^{(\ell)} : u \in \mathcal{N}(v)\}\right)\Big) \tag{17}$$

where $\mathrm{AGG}(\cdot)$ denotes sum-, mean-, or attention-based aggregation Kipf & Welling (2017); Hamilton et al. (2017); Veličković et al. (2017).

However, these GNN operators typically propagate and aggregate neighbors independently, without explicitly organizing their complementary semantics before compression. Consider two neighbors u_i and u_j whose cues jointly characterize the event associated with a target node v. Because their representations are propagated separately, their complementary relationship may remain fragmented as they are combined with increasingly heterogeneous, redundant, noisy, or irrelevant neighborhood information in a fixed-dimensional target representation. Consequently, their joint event semantics may be diluted or lost during aggregation. We refer to this form of semantic information loss associated with over-squashing as the *semantic bottleneck*.

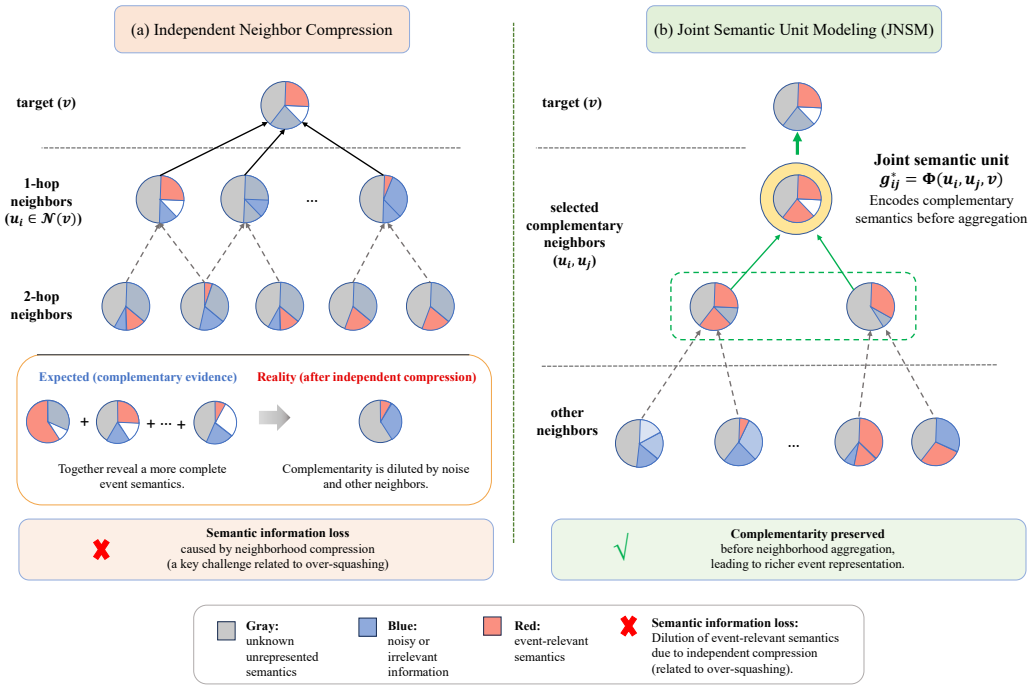

Figure 2: Illustration of the semantic bottleneck and joint neighbor semantic modeling. Independent aggregation compresses distributed neighbor evidence without explicitly modeling its complementarity, potentially diluting event-relevant semantics. JNSM selects complementary neighbors and jointly encodes their semantics before propagation. Gray, blue, and red denote unknown, noisy/irrelevant, and event-relevant semantics, respectively.

Increasing propagation depth does not resolve this semantic bottleneck and may instead intensify it. Deeper message passing expands the receptive field, introducing more heterogeneous neighborhood information that must be compressed into fixed-dimensional node representations Li et al. (2018); Oono & Suzuki (2020). The resulting compression pressure contributes to over-squashing Topping et al. (2022). Structural rewiring can facilitate long-range information flow by improving graph connectivity, but it does not explicitly organize complementary neighbor semantics into joint semantic units before propagation Karhadkar et al. (2023); Barbero et al. (2024). Consequently, improved connectivity can alleviate structural bottlenecks while leaving the semantic bottleneck unresolved.

JNSM addresses this limitation by explicitly modeling complementary neighbor interactions *before* neighborhood compression. Given a selected pair of complementary neighbors u_i and u_j, JNSM jointly encodes their representations with respect to the target node v:

$$\mathbf{h}_{g_{ij}^*} = \Phi\big(\mathbf{h}_{u_i}, \mathbf{h}_{u_j}, \mathbf{h}_v\big) \tag{18}$$

where $\Phi(\cdot)$ denotes the semantic unit encoder and $\mathbf{h}_{g_{ij}^*}$ is the resulting target-aware joint semantic representation. The semantic unit is then propagated as a unified entity, allowing the complementary event evidence of u_i and u_j to be preserved during aggregation. JNSM therefore changes the information-processing order from independently representing neighbors before aggregation to explicitly propagating complementary joint semantic units into the target representation:

$$\underbrace{\{u_i, u_j\} \to \mathrm{AGG} \to \mathbf{h}_v}_{\text{independent aggregation}} \quad \Longrightarrow \quad \underbrace{\{u_i, u_j\} \to \Phi(u_i, u_j, v) \to \mathrm{AGG} \to \mathbf{h}_v}_{\text{joint semantic modeling}} \tag{19}$$

Information-Theoretic Motivation for Semantic Complementarity. Joint semantic modeling is most useful when the selected neighbors contribute complementary rather than redundant event evidence. For two complementary neighbors $u_i, u_j \in \mathcal{N}(v)$, their joint information about the target can be decomposed using the chain rule of mutual information Cover et al. (1991):

$$I(v; u_i, u_j) = I(v; u_i) + I(v; u_j \mid u_i) \tag{20}$$

where $I(v; u_j \mid u_i)$ quantifies the additional target-related information contributed by u_j given u_i. A substantial conditional contribution therefore indicates that jointly considering u_i and u_j can capture complementary event evidence unavailable from u_i alone. This observation motivates selecting neighbor pairs whose joint information about the target exceeds that of either neighbor alone.

$$I(v; g) > \max(I(v; u_i), I(v; u_j)) \tag{21}$$

Here, $I(v; g)$ refers to the target-related information contained in the joint observation of u_i and u_j. The inequality indicates that each neighbor contributes information beyond that available from the other, providing an information-theoretic characterization of semantic complementarity. JNSM does not directly optimize this criterion; rather, it serves as the motivation for selecting complementary instead of redundant neighbor pairs.

Estimating these mutual-information quantities for every candidate pair is computationally expensive and statistically challenging in evolving heterogeneous graphs. JNSM therefore uses the target-centered semantic gain in Eq. 9 as a practical geometric criterion for pair selection. Its semantic contribution term favors pairs whose joint semantics are closely aligned with the target; redundancy suppression discourages pairs carrying highly overlapping semantic offsets; and cancellation correction prevents strongly opposing offsets from receiving spuriously high gains. Candidate pairs whose gains exceed the target-adaptive threshold are then selected as joint semantic units.

Overall, the semantic-bottleneck analysis explains *why* complementary neighbor semantics should be organized before neighborhood compression, while the information-theoretic analysis motivates *which* neighbor pairs should be jointly represented. Together, these perspectives provide the conceptual basis for semantic gain-guided joint neighbor semantic modeling as a means for alleviating semantic information loss associated with over-squashing.

B.2 Time Complexity

Let $C = |\mathcal{N}(v)|$ denote the sampled neighborhood size of target node v, d the embedding dimension, and M the number of selected semantic units. Constructing close–close and close–distant combinations yields at most $O(C^2)$ candidate pairs, and evaluating their semantic gains requires $O(C^2 d)$ time. Computing the target-adaptive threshold requires $O(C^2)$ time, while sorting the candidates for greedy non-overlapping selection requires $O(C^2 \log C)$ time in the worst case. Assuming a semantic unit encoder with hidden width proportional to d, encoding the M selected units requires $O(Md^2)$ time. After local rewiring, GraphSAGE aggregation over the selected semantic units and retained close neighbors requires $O((M+C)d)$ time, with an additional $O(d^2)$ cost for the feature transformation. Therefore, the per-target complexity of one JNSM layer is

$$O\big(C^2 d + C^2 \log C + Md^2 + (M + C)d + d^2\big)$$

Under non-overlapping selection, $M \leq \lfloor C/2 \rfloor$. Because C is bounded by the neighborhood sampling budget, the per-target computation remains bounded independently of the target node's full degree.

C Experiments

C.1 Experiment Setup

C.1.1 Datasets

We evaluate JNSM on three publicly available benchmarks for heterogeneous social event detection: Twitter2012, CrisisLexT, and French. Together, these datasets cover English- and French-language social streams and exhibit diverse event distributions and graph densities. Detailed dataset statistics are presented in Fig. 3.

Twitter2012 McMinn et al. (2013): It is a widely used SED benchmark comprising Twitter messages associated with real-world events. Its heterogeneous contextual information—including users, entities, and words—supports the construction of relation-specific message graphs and provides diverse neighborhood evidence for event detection.

CrisisLexT Olteanu et al. (2015): Following the experimental setting adopted in prior SED studies, we use messages from three natural disaster events: Hurricane Sandy, the Queensland Floods, and the West Texas Plant Explosion. The resulting densely connected message graphs provide a challenging setting for evaluating semantic modeling in highly connected neighborhoods.

French Hasan et al. (2019): It comprises French-language social media messages associated with real-world events. It provides a multilingual evaluation setting with graph structures and event distributions distinct from those of the English-language benchmarks.

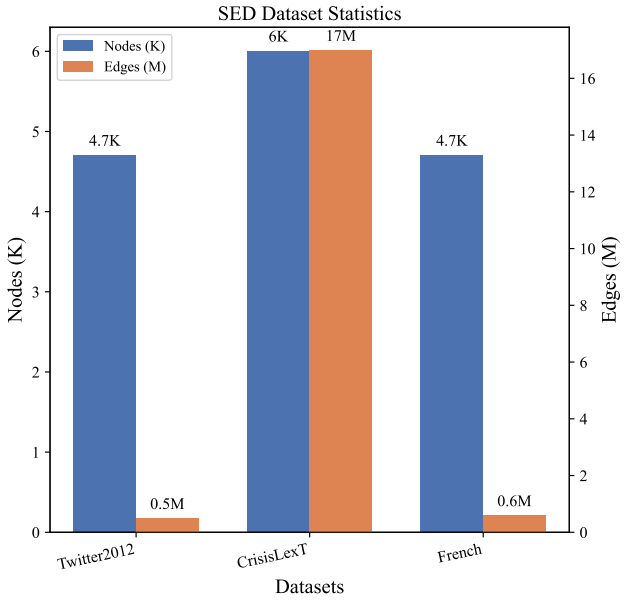

Figure 3: SED Dataset Statistics

C.1.2 BASELINES

- **GCN**[1] Kipf & Welling (2017): A classical message-passing GNN that learns node representations by applying graph convolution over local neighborhoods, serving as a standard graph-based baseline.
- **GraphSAGE**[2] Hamilton et al. (2017): An inductive GNN that learns node representations by sampling and aggregating neighborhood features. GraphSAGE also serves as the backbone of JNSM.
- **GIN**[3] Xu et al. (2019): An expressive message-passing GNN whose discriminative power matches that of the 1-WL test under appropriate conditions, providing a strong conventional GNN baseline.
- **GraphFormer**[4] Shi et al. (2021): A graph Transformer that incorporates graph structure into Transformer-based attention to capture dependencies among graph nodes.
- **ReDHAN**[5] Yu et al. (2023): A heterogeneous graph attention framework that combines reinforcement learning-based neighbor sampling with dual contrastive learning to reduce neighborhood noise and improve the robustness of event representations.
- **HSED** Qiu et al. (2024b): A supervised hyperbolic SED model that captures hierarchical event structures through hyperbolic representation learning and label-guided classification. We adapt HSED to the heterogeneous message graphs used under the common evaluation protocol.
- **SED-Aug**[6] Ma et al. (2025): An SED framework that combines explicit LLM-based text augmentation with implicit feature augmentation. For a controlled comparison, we retain its Gaussian perturbation of structure-fused representations while excluding the LLM-based augmentation component.
- **PIGCN**[7] Yu et al. (2026): A physics-inspired graph convolutional model that combines Helmholtz graph filtering with adaptive interaction-based edge weighting to model wave-like information propagation in heterogeneous social graphs.
- **LASER**[8] Barbero et al. (2024): A locality-aware graph rewiring method that introduces long-range connections while preserving graph locality and sparsity, serving as a structural rewiring baseline for comparison with JNSM.

[1] https://github.com/senadkurtisi/pytorch-GCN
[2] https://github.com/williamleif/GraphSAGE
[3] https://github.com/weihua916/powerful-gnns
[4] https://github.com/lucidrains/graph-transformer-pytorch
[5] https://github.com/yuyongsheng1990/Re-DHAN
[6] https://github.com/congboma/SED-Aug
[7] https://github.com/yuyongsheng1990/PIGCN
[8] https://github.com/Fedzbar/laser-release

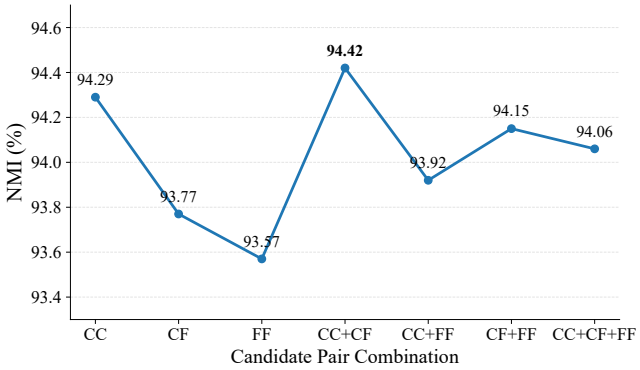

Figure 4: Impact of Different Candidate Pair Combinations on Twitter2012

C.1.3 Implementation Details.

All models are implemented in Python 3.11.7 using PyTorch Geometric and evaluated on a MacBook Air equipped with an 8-core Apple M3 CPU and 16GB of memory. Each dataset is partitioned into training, validation, and test sets in a 0.7/0.2/0.1 ratio using stratified random sampling to preserve the label distribution. The input, hidden, and relation-specific output dimensions are set to 302, 256, and 128, respectively. The three relation-specific representations are then concatenated to produce a 384-dimensional final embedding. All models are trained for up to 50 epochs using a learning rate of 10^{-3}, early stopping with a patience of 5. Unless otherwise specified, each experiment is repeated with five independent random seeds, and the mean performance is reported.

C.2 Experimental Analysis RQ4

C.2.1 Candidate Pair Combination Analysis

To investigate how neighbor pairing affects semantic unit construction, we consider three pair types according to their semantic proximity to the target: close–close (CC), close–far (CF), and far–far (FF). CC pairs model interactions between locally consistent semantic cues, whereas CF pairs capture potentially complementary semantics beyond the close neighborhood. FF pairs are included to determine whether interactions between two distant neighbors provide additional event evidence or primarily introduce semantic noise. We evaluate each pair type individually and in different combinations to identify the most effective candidate construction strategy.

As shown in Fig. 4, combining CC and CF pairs achieves the highest NMI of 94.42%, indicating that both locally consistent cues and potentially complementary evidence from distant neighbors contribute to candidate pair construction. In contrast, configurations involving FF pairs generally degrade performance, suggesting that interactions between two semantically distant neighbors are more likely to introduce noise than useful complementarity. Based on these findings, JNSM adopts the CC+CF combination for candidate pair construction, balancing local semantic consistency with complementary evidence beyond the close neighborhood.

C.2.2 Semantic Gain Design Analysis

To evaluate the design of the proposed semantic gain criterion, we individually ablate its three multiplicative factors in Eq. 9 while keeping all other components unchanged. Each factor captures a distinct aspect of candidate pair quality: (1) semantic contribution measures the target-centered contribution of the pair; (2) redundancy suppression penalizes semantically redundant neighbor interactions; and (3) cancellation correction prevents strongly opposing offsets from producing spuriously high contribution scores through vector cancellation.

As shown in Table 5, semantic contribution alone achieves competitive mean performance but exhibits relatively high variability across random seeds. Incorporating either redundancy suppression or cancellation correction slightly reduces the mean scores while improving stability, indicating that each factor regularizes candidate-pair evaluation rather than independently increasing average performance. Combining all three factors yields the highest NMI and AMI and the lowest standard deviations across all metrics. Although contribution alone achieves the highest mean ARI, its substantially greater variability suggests less reliable performance across

Table 5: Semantic Gain Design Analysis on Twitter2012 (mean±std, %)

Gain Factors	NMI	AMI	ARI
Contribution Only	94.38±0.66	91.75±0.99	**85.84**±2.79
Contribution + Cancellation	94.37±0.60	91.73±0.89	85.07±2.20
Contribution + Redundancy	94.25±0.43	91.56±0.66	84.95±2.78
Complete Semantic Gain	**94.42**±0.21	**91.81**±0.32	84.77±1.72

Table 6: Semantic Gain Threshold Analysis on Twitter2012 (mean±std, %)

Threshold	Rule	NMI	AMI	ARI
Fixed	$G_{g,v} > \tau$	94.14±0.31	91.42±0.45	83.14±0.52
Positive	$G_{g,v} > 0$	93.03±0.50	89.72±0.76	82.10±1.10
Adaptive Mean	$G_{g,v} > \max(\bar{G}_v, 0)$	**94.42**±0.21	**91.81**±0.32	**84.77**±1.72

Table 7: Semantic Unit Representation and Gain Scoring Analysis on Twitter2012 (NMI, %)

Unit Representation	Semantic Gain Strategy			
	Mean	Weighted	Learnable	Target
Mean Pooling	93.69	93.91	93.19	94.04
Learnable Encoder	93.92	93.71	93.72	94.75
Learnable Gate	93.91	92.95	93.83	94.60
Fixed Weighting	94.03	94.77	93.61	94.11
Average	93.89	93.84	93.59	**94.38**
Variance	**0.015**	0.420	0.059	0.093

runs. Overall, the complete semantic gain provides the most balanced and stable criterion for selecting informative complementary pairs.

C.2.3 Semantic Gain Threshold Analysis

To determine how candidate pairs should be filtered according to their semantic gains, we compare three thresholding criteria: a global fixed threshold, a zero threshold that retains all positive gain pairs, and a target-adaptive mean threshold. This comparison examines whether adapting the filtering criterion to target-specific gain distributions enables more effective semantic unit selection than global thresholding.

As shown in Table 6, the target-adaptive mean threshold achieves the best performance across all three metrics, reaching 94.42% NMI, 91.81% AMI, and 84.77% ARI, while exhibiting the lowest variance in NMI and AMI. In contrast, the positive-gain criterion is overly permissive by retaining candidate pairs with limited relative semantic benefit, whereas a global fixed threshold cannot adapt to target-specific gain distributions. These results support target-adaptive mean thresholding as a more effective criterion for selecting informative semantic units.

C.2.4 Semantic Unit Representation and Gain Scoring Analysis

To examine how semantic unit representation interacts with semantic gain evaluation, we compare four representation designs in combination with four semantic gain criteria on Twitter2012. In addition to reporting the performance of each configuration, we compute the mean and variance across representation designs to evaluate the overall effectiveness and sensitivity of each gain strategy.

As shown in Table 7, weighted semantic gain with fixed weighting achieves the highest individual NMI of 94.77%, exceeding target-centered gain with the learnable encoder by only 0.02 percentage points. However, weighted gain is highly sensitive to the choice of semantic unit representation, yielding the largest variance (0.420) and a lower average NMI of 93.84%. By comparison, target-centered gain achieves the highest average NMI (94.38%) with relatively lower variance across representation designs, indicating greater robustness to representation choice. Moreover, its combination with the learnable encoder reaches 94.75% NMI, closely matching the best individual result while explicitly modeling within-unit semantic interactions. Considering

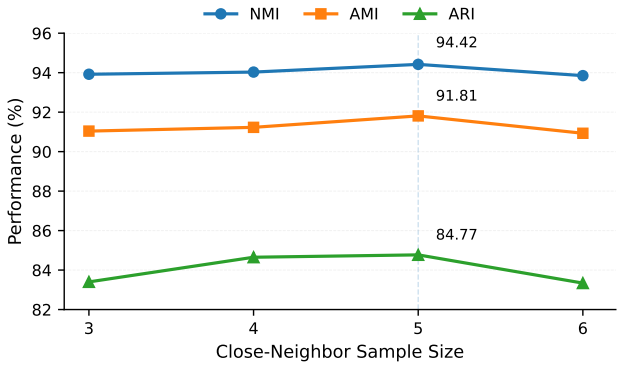

Figure 5: Sensitivity to Close Neighbor Sampling Size on Twitter2012

Table 8: Runtime comparison across three SED benchmarks (minutes; lower is better)

Method	Twitter2012	CrisisLexT	French
DoubleGCN	13.3	21.4	20.2
GraphFormer	11.5	21.9	19.5
GIN	14.4	21.1	19.4
ReDHAN	19.0	28.9	28.2
HSED	13.8	18.5	22.5
SED-Aug	10.3	19.3	18.3
PIGCN	13.6	26.4	22.1
LASER	18.4	30.1	23.9
GraphSAGE	12.7	23.4	19.6
JNSM	12.2	21.6	21.0

both effectiveness and robustness, JNSM therefore adopts target-centered gain with the learnable semantic unit encoder as its default configuration.

C.2.5 Close Neighbor Sampling Sensitivity

To assess the sensitivity of JNSM to the close neighbor sampling size, we vary the number of sampled close neighbors from 3 to 6 on Twitter2012. Fig. 5 reports the corresponding clustering performance and runtime, enabling evaluation of the trade-off between detection effectiveness and computational efficiency.

As the close neighbor sampling size increases from 3 to 5, clustering performance improves consistently, indicating that a broader close neighborhood provides richer local semantic evidence for constructing semantic units. The best performance is obtained with five close neighbors, reaching 94.42% NMI, 91.81% AMI, and 84.77% ARI. Increasing the sampling size to six reduces all three metrics, suggesting that the additional neighbors provide diminishing semantic benefit and may introduce less relevant evidence. Accordingly, we set the default close neighbor sampling size to five.

C.3 Runtime Analysis

We compare the computational efficiency of JNSM with all baseline methods across Twitter2012, CrisisLexT, and French. Table 8 reports the end-to-end runtime in minutes, with GraphSAGE serving as the backbone reference and LASER as the structural rewiring baseline.

JNSM exhibits runtime comparable to its GraphSAGE backbone, requiring 3.9% less time on Twitter2012 and 7.7% less on CrisisLexT, with a 7.1% increase on French. This efficiency is accompanied by higher NMI and AMI on Twitter2012 and consistent improvements across all three clustering metrics on CrisisLexT and French. Relative to LASER, JNSM reduces runtime by 33.7%, 28.2%, and 12.1% on Twitter2012, CrisisLexT, and French, respectively, while outperforming it across all clustering metrics. These results show that JNSM integrates joint neighbor semantics with end-to-end runtime comparable to its backbone while providing a more favorable performance-efficiency trade-off than the structural rewiring baseline.

Table 9: JNSM Placement Analysis on Twitter2012

Placement	NMI	AMI	ARI	Time (min)
Before Layer 1	93.76	90.89	84.66	14.5
Between Layers 1–2	**94.42**	**91.81**	**84.77**	**4.4**

C.4 Placement of Joint Neighbor Semantic Modeling

Within the two-layer architecture, JNSM can be applied either before the first GraphSAGE layer using the initial message representations or between the two layers using the intermediate representations. Because semantic unit construction relies on distinguishing informative and complementary neighbors, we compare these two placement strategies on Twitter2012 with SBERT features, with the results reported in Table 9.

Applying JNSM between the two GraphSAGE layers improves all three clustering metrics, increasing NMI from 93.76% to 94.42%, AMI from 90.89% to 91.81%, and ARI from 84.66% to 84.77%. This replacement also reduces runtime from 14.5 to 4.4 minutes. These results indicate that intermediate representations enriched with local neighborhood information provide a more effective and efficient basis for candidate evaluation and semantic unit construction than the initial message features. Accordingly, JNSM is applied between the first and second GraphSAGE layers in the default architecture.