HyperDet: Source Detection in Hypergraphs via Interactive Relationship Construction and Feature-rich Attention Fusion

Le Cheng[1,2] , **Peican Zhu**[1*] , **Yangming Guo**[3] , **Keke Tang**[4*] , **Chao Gao**[1] and **Zhen Wang**[3*]

[1]School of Artificial Intelligence, Optics and Electronics, Northwestern Polytechnical University
[2]School of Computer Science, Northwestern Polytechnical University
[3]School of Cybersecurity, Northwestern Polytechnical University
[4]Cyberspace Institute of Advanced Technology, Guangzhou University
{ericcan, w-zhen}@nwpu.edu.cn, tangbohutbh@gmail.com

Abstract

Hypergraphs offer superior modeling capabilities for social networks, particularly in capturing group phenomena that extend beyond pairwise interactions in rumor propagation. Existing approaches in rumor source detection predominantly focus on dyadic interactions, which inadequately address the complexity of more intricate relational structures. In this study, we present a novel approach for Source Detection in Hypergraphs (HyperDet) via Interactive Relationship Construction and Feature-rich Attention Fusion. Specifically, our methodology employs an Interactive Relationship Construction module to accurately model both the static topology and dynamic interactions among users, followed by the Feature-rich Attention Fusion module, which autonomously learns node features and discriminates between nodes using a self-attention mechanism, thereby effectively learning node representations under the framework of accurately modeled higher-order relationships. Extensive experimental validation confirms the efficacy of our HyperDet approach, showcasing its superiority relative to current state-of-the-art methods.

1 Introduction

Source detection in graphs offers a viable mathematical approach to identifying the sources of propagation information such as rumors [Shah and Zaman, 2011; Zhu *et al.*, 2022]. Over recent years, various methodologies have been proposed to address this issue [Wang *et al.*, 2025; Dong *et al.*, 2024]. Initially, techniques based on source centrality theory [Shah and Zaman, 2011] and maximum likelihood estimation [Yang *et al.*, 2020], including methods such as LPSI [Wang *et al.*, 2017], OJC [Zhu *et al.*, 2017], and MLE [Pinto *et al.*, 2012; Cheng *et al.*, 2025]. More recently, with the rapid advancement of graph neural networks [Kipf and Welling, 2017; Jin *et al.*, 2023], an increasing number of researchers have employed strategies to learn node representations by embedding node features and modeling propagation information
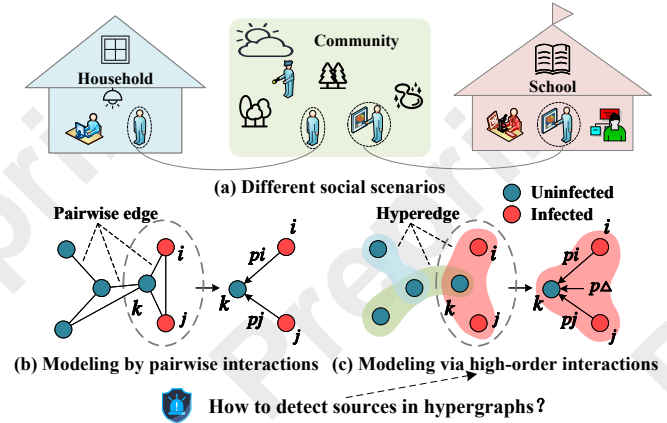
---

*Corresponding authors.

Figure 1: Social scenario modeling. (a) Various social scenarios. (b) Pairwise interactions in exiting methods. (c) Higher-order interactions in our proposed approach. This accurate yet complex hypergraph modeling introduces new challenges to source detection.

[Ling *et al.*, 2022; Wang *et al.*, 2022]. These approaches have set benchmarks in identifying sources [Cheng *et al.*, 2024a].

However, existing methods predominantly model graphs through dyadic interactions, assuming that user interactions are pairwise. In practice, interactions often extend beyond pairwise engagements, occurring among triads or larger groups, thus constituting higher-order interactions [Battiston *et al.*, 2020; Gao *et al.*, 2022]. To our knowledge, no existing work has addressed source detection considering interactions that extend beyond the pairwise level. When networks exhibit higher-order interactions, the influence/pressure among group peers accelerates the information propagation process, in stark contrast to binary interactions, as illustrated in Fig. 1.

Hypergraphs offer a superior framework for modeling the aforementioned higher-order interactions. However, source detection in hypergraphs faces two significant challenges. First, propagation on hypergraphs exhibit both static low/high-order topological structures and dynamic user interactions, posing a primary challenge in accurate modeling. Second, node features are often manually engineered and of low dimensionality, which can hinder model performance. Furthermore, the varying significance of different nodes in information propagation, if treated uniformly, af-

fects the model's learning and convergence. To intuitively address these issues, we should: 1) Establishing relationships between nodes through both static topology and dynamic interactions to enhance subsequent message propagation. 2) Enabling the model to autonomously learn and enrich node features, while distinguishing between nodes based on their information transmission capabilities.

In this paper, we propose a novel Source Detection approach in Hypergraphs (HyperDet) via Interactive Relationship Construction and Feature-rich Attention Fusion. Initially, to accurately depict user relationships and node features, an Interactive Relationship Construction module (IRC) is developed to leverage the static topology of the hypergraph as well as the dynamic interactions between infected and uninfected subgraphs to form the nodes' interactive relationships; combining node states, propagation information, and positional encoding to constitute raw node features. Subsequently, to automatically learn and enrich node features, and to focus on nodes with stronger information propagation capabilities, a Feature-rich Attention Fusion module (FAF) is introduced. This module includes a designed hypergraph autoencoder to extract latent node features, and a multi-head attention mechanism to automatically assign different attention coefficients to various nodes. Lastly, to counteract model prediction biases due to significant sample size differences between sources and non-sources, we introduce a class balancing mechanism. The effectiveness of our proposed method is validated across eight public datasets, and extensive experiments demonstrate that HyperDet outperforms state-of-the-art methods in hypergraph-based source detection.

Overall, our contributions can be summarized as follows:

- We are the first to formalize the problem of source detection in hypergraphs and to propose a methodology to address this challenge.

- We introduce a novel approach for source detection in hypergraphs through Interactive Relationship Construction and Feature-rich Attention Fusion.

- We demonstrate the effectiveness of our proposed method over baseline approaches through extensive experimental validation.

2 Related Work

General Source Detection. Various methods have been proposed for source detection [Bao *et al.*, 2024]. Based on the source centrality theory [Prakash *et al.*, 2012; Shah and Zaman, 2011], LPSI selects local outliers by iterating node labels [Wang *et al.*, 2017], EPA computes the infection time for each node and selects the node with the longest duration of infection [Ali *et al.*, 2019], and OJC identifies nodes that cover the infected area with the smallest radius [Zhu *et al.*, 2017]. However, these methods do not account for user diversity and have limited application scenarios.

Graph Learning-based Source Detection. Utilizing graph neural networks (GNNs), GCNSI and SIGN employ user states as inputs to classify nodes and identify sources [Dong *et al.*, 2019; Li *et al.*, 2021], while GCSSI considers the last infected node, referred to as the wavefront [Dong *et al.*,

2022]. From the perspective of network structure, ResGCN enhances the influence of initial features through a residual structure [Shah *et al.*, 2020]. Nevertheless, these methods are primarily challenged by class imbalance and lack integration with the information propagation process.

Addressing graph diffusion explicitly, IVGD initially learns the information propagation process and subsequently reverses this knowledge for source detection [Wang *et al.*, 2022]. SL-VAE deciphers various propagation patterns on graphs [Ling *et al.*, 2022], and GIN-SD accommodates nodes with partial information [Cheng *et al.*, 2024b]. Despite these advances, the above methods predominantly model social network propagation through pairwise interactions. This simplification overlooks higher-order interactions that are prevalent in group dynamics, a crucial aspect that remains unaddressed by these models [Battiston *et al.*, 2020].

Hypergraph Neural Networks. To model higher-order social phenomena, hypergraphs have been proposed with hyperedges containing multiple nodes beyond mere pairwise connections [Zhou *et al.*, 2006; Battiston *et al.*, 2020]. Building on the higher-order structure of hypergraphs, HGNN introduces a two-stage message passing approach to learn node representations, consisting of hyperedge information aggregation and node information aggregation [Feng *et al.*, 2019; Gao *et al.*, 2022]. Inspired by the Graph Attention Network [Veličković *et al.*, 2017], Hyper-Atten enhances the model's learning capability by computing attention coefficients for nodes and their associated hyperedges [Bai *et al.*, 2021].

Based on GNNs, UniGNN directly applies GNNs to hypergraphs using hyperedges as intermediaries [Huang and Yang, 2021], while HyperGCN first transforms hypergraphs into weighted graphs before applying GCNs for node representation learning [Yadati *et al.*, 2019]. Integrating existing approaches, AllSet proposes an effective model that unifies current hypergraph learning methods [Chien *et al.*, 2022]. As an extension of simple pairwise graphs, hypergraphs demonstrate a stronger and more flexible data modeling capability, widely applied and exhibiting high efficiency across multiple domains [Jiao *et al.*, 2024].

3 Problem Formulation

Preliminary on Hypergraphs. The low/high-order interactions in social networks can be modeled using a hypergraph $G = (V, E, \mathbf{\Omega})$, where $V = \{v_1, v_2, ..., v_n\}$ represents the set of nodes; $E = \{e_1, e_2, ..., e_m\}$ denotes the set of hyperedges, with each hyperedge $e_i = \{v_{i1}, v_{i2}, ..., v_{ij}\}$, where $i \in \{1, 2, ..., m\}$ and $j \in \{1, 2, ..., n\}$; and $\mathbf{\Omega} \in \mathbb{R}^{m \times m}$ is a diagonal matrix, each element of which represents the weight of the corresponding hyperedge. Based on the relationship between node v and hyperedge e, the hypergraph can be represented by an incidence matrix $\mathbf{H} \in \mathbb{R}^{n \times m}$, i.e.,

$$\mathbf{H}_{ve} = \begin{cases} 1, & v \in e \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Consequently, the degree of node v in the hypergraph can be defined as $\mathbf{D}_{vv} = \sum_{e=1}^{m} \mathbf{\Omega}_{ee} \mathbf{H}_{ve}$, and the degree of hyperedge e can be defined as $\mathbf{D}_{ee} = \sum_{v=1}^{n} \mathbf{H}_{ve}$, where the diagonal matrices $\mathbf{D}_V \in \mathbb{R}^{n \times n}$ and $\mathbf{D}_E \in \mathbb{R}^{m \times m}$ respectively represent the degree matrices for nodes and hyperedges.
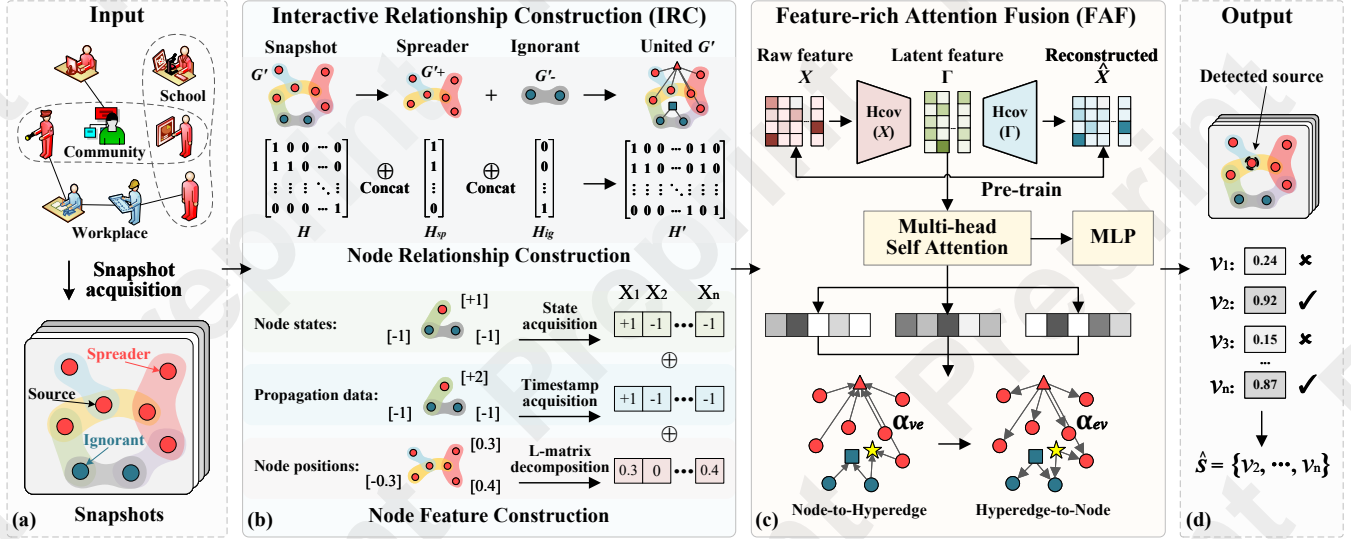
Figure 2: Framework of HyperDet. (a) The snapshot acquisition process. (b) The Interactive Relationship Construction (IRC) module, which constructs node relationships through static topology and dynamic interactions, and initially builds raw node features. (c) The Feature-rich Attention Fusion (FAF) module, which first learns node representations via autoencoding and then differentiates nodes using a self-attention mechanism to enhance model learning capabilities. (d) Model output, where nodes are classified based on predicted probabilities to identify the source set $\hat{s}$.

Propagation Process on Hypergraphs. The information propagation on hypergraphs unfolds over the evolutionary time t, where at $t = 0$, the set of source nodes s in the network transitions from the ignorant (susceptible) state to the spreader (infected) state. As time progresses, each spreader v_i disseminates information to its neighbors with a probability p_i. Unlike the propagation processes on bipartite graphs, the propagation on hypergraphs not only involves pairwise interactions but also includes peer influence/pressure within groups represented by hyperedges. Consequently, an ignorant node transitions to the spreader state with an additional probability $p_\triangle$, as depicted in Fig. 1. Several classical models, such as the IC, SI, SIR, and SIS models [Battiston *et al.*, 2020; de Arruda *et al.*, 2020], have been proposed to characterize the propagation process on hypergraphs.

In summary, the propagation process on hypergraphs can be represented by the time series $\{G'(t), t \geq 0\}$, where $G'(t)$ denotes the network snapshot at time t, with nodes being categorized into two types based on the states, i.e., G_+ (spreader) and G_- (ignorant).

Source Detection in Hypergraphs. As the propagation proceeds, when $\delta\%$ of the nodes in the network reach the spreader state, a network snapshot G' is captured, which includes the network's topological structure T, the states of the nodes N, and the propagation information P. Consequently, source detection in a hypergraph can be formalized as:

$$\hat{s} = f(G'(T, N, P)), \tag{2}$$

where $f(\cdot)$ represents the corresponding source detection method, and $\hat{s}$ denotes the detected source set.

Discussion. For source detection in hypergraphs, the network snapshot G' not only encapsulates static low/high-order topological structures but also contains dynamic user interaction information, underscoring the necessity to accurately model the relationships between nodes. Furthermore, the node features are manually engineered with low dimensionality and the attributes of nodes are diverse; thus, enabling the model to automatically learn, enrich these node features, and differentiate between nodes becomes crucial for enhancing model performance.

4 Method

In this section, we introduce the proposed HyperDet, which primarily consists of two modules: the Interactive Relationship Construction module (IRC) and the Feature-rich Attention Fusion module (FAF), as illustrated in Fig. 2.

4.1 Interactive Relationship Construction (IRC)

The IRC module comprises two submodules, i.e., Node Relationship Construction and Node Feature Construction.

Node Relationship Construction. The hypergraph snapshot G' incorporates not only the static topological structure but also the dynamic interaction information among nodes, we model these two aspects using hyperedges. For the static topological structure, it is represented by the incidence matrix $\boldsymbol{H}$, where each hyperedge e encompasses nodes that can interact with each other. Regarding the dynamic interaction information, nodes in G' that are in ignorant and spreader states are encompassed within specific hyperedges, denoted as $e_{ig} = \{v_i | v_i \in G_-\}$ and $e_{sp} = \{v_j | v_j \in G_+\}$, respectively, which can further be transformed into the incidence matrices $\boldsymbol{H}_{ig} \in \{0,1\}^{n \times 1}$ and $\boldsymbol{H}_{sp} \in \{0,1\}^{n \times 1}$. Therefore, the overall incidence matrix can be concatenated as:

$$\boldsymbol{H'} \in \{0,1\}^{n \times (m+2)} = \boldsymbol{H} \parallel \boldsymbol{H}_{ig} \parallel \boldsymbol{H}_{sp}. \tag{3}$$

Node Feature Construction. Various types of information, including node states, propagation data, and node positions, are embedded as node features.

The hypergraph snapshot $G'(t)$ captures the state of node v_i at time t within the network: spreader ($v_i \in G_+$), indicating that v_i is influenced by and propagating the rumor, and ignorant ($v_i \in G_-$), meaning v_i has not received the rumor or deems the information unreliable. Consequently, the state feature X_i^1 can be represented as:

$$X_i^1 = \begin{cases} +1, & v_i \in G_+ \\ -1, & \text{otherwise.} \end{cases} \quad (4)$$

Social platforms such as Facebook and Twitter record timestamps when users forward information, which are critical for understanding the propagation dynamics. Therefore, for node v_i, we incorporate the information timestamp as its propagation features:

$$X_i^2 = \begin{cases} t_i, & v_i \in G_+ \\ -1, & \text{otherwise.} \end{cases} \quad (5)$$

In spatial domain convolution, unlike the message passing aggregated from local neighbors, the positional relationships between nodes play a crucial role in facilitating global message propagation. Additionally, encoding the different positions of infected nodes assists the model in learning the locational characteristics of the source. Given the generative nature of Laplacian positional encoding, we decompose the Laplacian matrix of the infected subgraph for encoding the positional features of nodes.

We first extract the infected subgraph G'_+ from G', and its symmetric normalized Laplacian matrix is computed as:

$$L_+^{sym} = I - D_{V+}^{-\frac{1}{2}} H_+ D_{E+}^{-1} H_+^T D_{V+}^{-\frac{1}{2}}, \quad (6)$$

where H_+ is the incidence matrix of G'_+, D_{V+} and D_{E+} respectively represent the corresponding degree matrices of the nodes and edges. Subsequently, the matrix L_+^{sym} is subjected to factorization:

$$\triangle_{L_+^{sym}} = \Psi \Lambda \Psi^T, \quad (7)$$

here, Λ is a diagonal matrix whose elements are eigenvalues, while matrix Ψ consists of the corresponding eigenvectors. Given that the first component of the eigenvectors weakly differentiates between node structure and organizational information, we select k smallest non-trivial eigenvectors starting from the second component as Ψ_i for node v_i's positional encoding ($k \ll n$). Therefore, the positional feature of node v_i is represented as:

$$X_i^3 = \begin{cases} \Psi_i, & v_i \in G'_+ \\ -1, & \text{otherwise.} \end{cases} \quad (8)$$

In conclusion, the raw feature vector of node v_i can be synthesized by concatenating several component features to achieve a comprehensive representation:

$$X_i = \left[\|_{x=1}^3 X_i^x \right]. \quad (9)$$

4.2 Feature-rich Attention Fusion (FAF)

The FAF module first automates the learning and enrichment of raw node features, and then differentiates between nodes using a self-attention mechanism.

Feature Augmentation via Autoencoding. In contemporary graph-based learning paradigms, the manually engineered raw node features may not adequately support model learning and stability due to their simplistic representation and limited discriminative power. To address this, we design an autoencoder with encoder-decoder architecture. Specifically, the feature vector X undergoes transformation by the encoder $\text{En}(\cdot)$, resulting in an embedding Γ, which is subsequently reconstructed back into $\hat{X}$ by the decoder $\text{De}(\cdot)$. This process is encapsulated as:

$$\hat{X} = \text{De}\left(\text{En}\left(X, H'\right), H'\right). \quad (10)$$

Each layer of this architecture involves successive transformations through hypergraph convolution and attention mechanism $\text{HConv}_{\text{att}}(\cdot)$. The process can be formalized as:

$$X^{(l+1)} = \text{LReLU}\left(\text{HConv}_{\text{att}}\left(X^{(l)}, \Theta^{(l)}, H', \Omega\right)\right), \quad (11)$$

where $X^{(l)} \in \mathbb{R}^{l_w \times n}$ represents the features of nodes at layer l and $X^{(0)} = X$, $\Theta^{(l)} \in \mathbb{R}^{l_w \times m}$ denotes the hyperedge attributes at the lth layer, l_w is the dimensionality of both node features and hyperedge attributes. H' and Ω are the incidence matrix and hyperedge weights matrix, respectively. $\text{LReLU}(\cdot)$ serves as the activation function. The attention coefficients α are computed based on the relative importance of nodes in relation to their associated hyperedges:

$$\alpha_{ij} = \frac{\exp\left(\vec{a}^T \text{LReLU}\left(W\left[X_i \| \theta_j\right]\right)\right)}{\sum_{e_k \in \mathcal{N}(v_i)} \exp\left(\vec{a}^T \text{LReLU}\left(W\left[X_i \| \theta_k\right]\right)\right)}, \quad (12)$$

where $\vec{a} \in \mathbb{R}^{2l'_w}$ represents a learnable parameter of the attention mechanism, $W \in \mathbb{R}^{l'_w \times l_w}$ is the weight matrix, and $\mathcal{N}(v_i)$ denotes the set of hyperedges that include node v_i within the hypergraph. X_i and θ_j represent the features of node v_i and the attributes of hyperedge e_j, respectively.

Building on the aforementioned architectures, each layer of hypergraph attention convolution is delineated into two distinct operations, i.e.,

(1) Node-to-Hyperedge Convolution In this process, each node transmits its features to the connected hyperedges:

$$\theta'_e = \sum_{v \in \mathcal{N}(e)} \alpha_{ve} W X_v, \quad (13)$$

where $\mathcal{N}(e)$ represents the set of nodes being connected to the hyperedge e.

(2) Hyperedge-to-Node Convolution Conversely, in the hyperedge-to-node convolution, the features of hyperedges are aggregated back to the nodes:

$$X'_v = \sum_{e \in \mathcal{N}(v)} \alpha_{ev} W' \theta'_e. \quad (14)$$

To facilitate effective learning, the autoencoder is initially pre-trained by minimizing the discrepancy between the original node features and the reconstructed features, fostering an accurate representation of the latent space conducive to downstream tasks. The loss function utilized during this pre-training phase is defined as:

$$\mathcal{L}_{ae} = \|X - \hat{X}\|^2. \quad (15)$$

Differential Attention Fusion. The latent node representations $\boldsymbol{\Gamma}$ learned by the autoencoder often require nuanced processing due to the diversity of nodes. To address this, we employ hypergraph attention convolution to allocate distinct attention coefficients to different nodes. The transformation for $\boldsymbol{\Gamma}$ is formalized by:

$$\boldsymbol{\Gamma}' = \text{LReLU}\left(\text{HConv}_{\text{att}}\left(\boldsymbol{\Gamma}, \boldsymbol{\Theta}, \boldsymbol{H}', \boldsymbol{\Omega}\right)\right), \qquad (16)$$

where $\boldsymbol{H}'$ is the incidence matrix constructed in Eq. (3) within the IRC module.

Furthermore, to enhance the discriminative power of our model, a multi-head attention mechanism is employed. This approach involves the fusion of K independent attention computations to refine the node representations, which are then concatenated to serve as the input for the subsequent layer. The process is formalized as:

$$\boldsymbol{\Gamma}'' = \big\|_{k=1}^{K} \text{LReLU}\left(\text{HConv}_{\text{att}}\left(\boldsymbol{\Gamma}'^{k}\right)\right). \qquad (17)$$

In the final layer of attention module, to align dimensions and synthesize the information across different heads, the representations from K attention heads are averaged:

$$\boldsymbol{\Gamma}''' = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{\Gamma}''^{k}. \qquad (18)$$

This operation ensures the final node representation encapsulates a comprehensive view, mitigating the bias that might arise from a single convolution channel. Ultimately, these node representations are projected into a 2-dimensional space and transformed via a softmax function to calculate the probabilities of each node belonging to the source/non-source set:

$$P(v_x) = \frac{e^{z_i}}{\sum_j e^{z_j}}, \ \vec{z} = \boldsymbol{\Gamma}'''_x. \qquad (19)$$

This sophisticated architecture leverages both the detailed local interactions and the global structure of the hypergraphs, thereby ensuring that our model effectively discerns the nuanced roles of nodes in information propagation.

4.3 Loss Function and Training

To address the significant disparity in sample sizes between the source node set and the non-source set, thereby mitigating prediction bias, we design a balancing coefficient:

$$\rho = \frac{|s|}{n - |s|}, \qquad (20)$$

where n and $|s|$ represent the number of nodes and sources, respectively. This approach ensures that the contribution of each node is weighted equally. Integrating this class balance mechanism, we configure the loss function for the attention fusion module as:

$$\mathcal{L}_{af} = \sum_{v_i \in s} \mathcal{L}_i + \rho \sum_{v_j \in (V-s)} \mathcal{L}_j, \qquad (21)$$

here, $\mathcal{L}$ denotes the cross-entropy loss, for sample x and its label y, $\mathcal{L}(x,y) = -y\log(\hat{y})$, $\hat{y}$ is the predicted probability.

During the training process, we fine-tune the autoencoder, and the overall loss function is formulated as:

$$\mathcal{L}_{HyperDet} = \mathcal{L}_{ae} + \mathcal{L}_{af} + \lambda\|w\|_2, \qquad (22)$$

where λ is the regularization parameter, and $\|\cdot\|^2$ represents the L2-norm, adding a penalty for model complexity to prevent overfitting.

For the proposed HyperDet system, the IRC module establishes node relationships and features. Subsequently, the FAF module undertakes the automatic learning and enrichment of node features, followed by the application of multi-head attention to differentiate nodes based on their significance and roles within the network. This structured approach ensures effective identification of source nodes within diverse and intricate network structures.

5 Experiments

5.1 Experimental Settings

Implementation. Due to the independent attributes of users in social networks, the short-term nature and the group effects of rumor propagation, we employ a heterogeneous independent cascade model on hypergraphs to simulate the spread of rumors. Initially, 5% of the nodes are selected as sources, with each node's propagation probability p following a uniform distribution $U(0, 0.5)$. The group propagation probability $p_{\triangle}$ is proportional to the ratio of infected nodes within a hyperedge, represented as $p_{\triangle} = 0.3(|e \cap G_+|/|e|)$. A network snapshot is captured when 30% of the nodes become spreaders, and the final ratio of the training set to the test set is 8:2. For the learning process, the autoencoder and the HyperDet system are set with learning rates of 0.01 and 0.005, respectively, with the latent node feature dimensions at 64. For small networks (G_1-G_4), there are three attention heads and 500 hidden neurons; medium networks (G_5-G_7) use two heads and 500 neurons; and the large network (G_8) has one head and 400 neurons due to space constraints. The hypergraphs undergo clique expansion before being applied in the baseline methods. All experiments are conducted on a workstation equipped with four NVIDIA RTX 3090 GPUs.

Datasets. We evaluate the methods using eight diverse and widely utilized datasets across various scales and domains, including Zoo [Asuncion *et al.*, 2007], House [Chodrow *et al.*, 2021], NTU2012 [Chen *et al.*, 2003], Mushroom [Asuncion *et al.*, 2007], ModelNet40 [Wu *et al.*, 2015], 20News [Asuncion *et al.*, 2007], PubMed [Yadati *et al.*, 2019], and Walmart [Amburg *et al.*, 2020].

Evaluation Metrics. Accuracy (ACC), F1-Score, and Area Under the Curve (AUC) are used to assess method performance. ACC measures the correct classification rate of samples. The F1-Score comprises Precision, which quantifies the ratio of true sources in the predicted set $\hat{s}$, calculated as $|\hat{s} \cap s| / |\hat{s}|$, and Recall, which assesses the proportion of true sources correctly detected, computed as $|\hat{s} \cap s| / |s|$. AUC evaluates the model's capacity to differentiate between source and non-source classes at all threshold levels. These metrics provide a comprehensive framework for assessing the effectiveness and reliability of source detection methods.

Baselines. Several representative methods that published recent years are considered as baselines, including those based on source centrality such as LPSI [Wang *et al.*, 2017] and EPA [Ali *et al.*, 2019]; methods that account for user

Methods	Zoo			House			NTU2012			Mushroom		
	ACC	F-Score	AUC	ACC	F-Score	AUC	ACC	F-Score	AUC	ACC	F-Score	AUC
LPSI	0.798	0.348	0.785	0.806	0.341	0.804	0.795	0.316	0.823	0.810	0.332	0.780
EPA	0.793	0.356	0.790	0.812	0.336	0.812	0.806	0.302	0.815	0.816	0.329	0.764
GCNSI	0.819	0.378	0.805	0.826	0.322	0.809	0.814	0.297	0.829	0.809	0.358	0.798
SIGN	0.824	0.457	0.816	0.829	0.417	0.810	0.819	0.408	0.826	0.811	0.436	0.820
GCSSI	0.817	0.418	0.801	0.815	0.397	0.799	0.813	0.384	0.825	0.810	0.395	0.810
ResGCN	0.829	0.497	0.810	0.824	0.459	0.814	0.805	0.439	0.829	0.821	0.447	0.826
IVGD	0.835	0.578	0.809	0.815	0.513	0.829	0.852	0.526	0.841	0.842	0.510	0.858
SL-VAE	0.834	0.598	0.815	0.841	0.528	0.815	0.840	0.517	0.832	0.852	0.526	0.884
GIN-SD	0.856	0.682	0.849	0.853	0.617	0.847	0.866	0.629	0.850	0.861	0.631	0.878
HyperDet (ours)	**0.901**	**0.792**	**0.892**	**0.884**	**0.726**	**0.875**	**0.877**	**0.715**	**0.869**	**0.912**	**0.801**	**0.923**

Methods	ModelNet40			20News			PubMed			Walmart		
	ACC	F-Score	AUC	ACC	F-Score	AUC	ACC	F-Score	AUC	ACC	F-Score	AUC
LPSI	0.764	0.235	0.786	0.732	0.255	0.758	0.749	0.181	0.762	0.705	0.125	0.712
EPA	0.785	0.249	0.791	0.756	0.267	0.764	0.762	0.192	0.778	0.710	0.114	0.725
GCNSI	0.810	0.254	0.801	0.795	0.218	0.785	0.805	0.188	0.789	0.806	0.180	0.799
SIGN	0.820	0.429	0.818	0.809	0.359	0.796	0.804	0.229	0.784	0.819	0.236	0.805
GCSSI	0.808	0.378	0.786	0.813	0.357	0.782	0.803	0.189	0.786	0.796	0.195	0.795
ResGCN	0.814	0.482	0.813	0.820	0.426	0.811	0.806	0.235	0.817	0.819	0.224	0.818
IVGD	0.826	0.610	0.823	0.821	0.534	0.827	0.824	0.531	0.819	0.815	0.523	0.824
SL-VAE	0.836	0.579	0.824	0.835	0.561	0.819	0.818	0.530	0.819	0.829	0.514	0.836
GIN-SD	0.826	0.607	0.816	0.839	0.580	0.821	0.834	0.574	0.826	0.829	0.527	0.836
HyperDet (ours)	**0.887**	**0.786**	**0.885**	**0.869**	**0.748**	**0.882**	**0.870**	**0.723**	**0.871**	**0.868**	**0.627**	**0.859**

Table 1: The performance of source detection across all datasets for each method, with the best results highlighted in bold.

states like GCNSI [Dong *et al.*, 2019], SIGN [Li *et al.*, 2021], GCSSI [Dong *et al.*, 2022] and ResGCN [Shah *et al.*, 2020]; and approaches integrating propagation information such as IVGD [Wang *et al.*, 2022], SL-VAE [Ling *et al.*, 2022], and GIN-SD [Cheng *et al.*, 2024b].

5.2 Performance Analyses

Comparison with State-of-the-art Methods. The comparative results are compiled in Table 1. From the outcomes, we can discern several insights: 1) All methods exhibit relatively high ACC, while F-Scores are generally lower. This discrepancy primarily arises because ACC accounts for all samples, including non-sources, thus a greater divergence between ACC and F-Score indicates a more severe impact of class imbalance issues on the model, as observed with the LPSI, EPA, and GCNSI methods. 2) Methods that integrate propagation information, such as IVGD, SL-VAE, and GIN-SD, demonstrate superiority over those based solely on source centrality theory or user states. This advantage is largely due to the inherent randomness of the propagation process, which cannot be effectively countered by merely considering user states. Overall, in the context of all evaluated methods, the presence of higher-order interactions means that the propagation process is not confined to pairwise interactions, making it challenging for baseline methods based on pairwise interactions to capture propagation dynamics effectively. Consequently, these methods underperform compared to their optimal outcomes. However, the proposed HyperDet achieves the best results across all datasets. Specifically, HyperDet improves performance by 23%-30% over methods that based on source centrality theory and user states, and by 8%-15% over those consider propagation information. This improvement is attributable primarily to HyperDet's strategies: 1) Designing higher-order interaction relationships among nodes through both static topology and dynamic interactions. 2) Enriching node features and discriminatively learning node representations through autoencoding and multi-head attention mechanisms. 3) Eliminating predictive bias in the model through a class balancing mechanism.

Visualization. To provide an intuitive representation of source detection results, we visualize the correctly detected sources of representative methods on the House network, as shown in Fig. 3. This visualization facilitates an immediate understanding of the comparative performance and effectiveness of each method in a real-world network context.
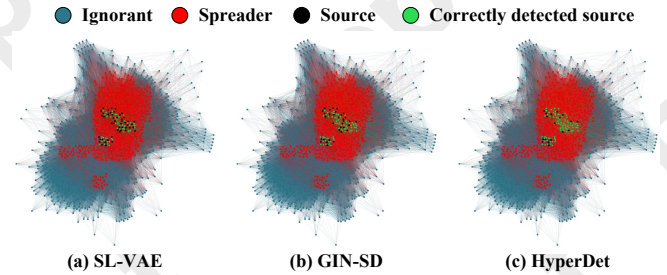

(a) SL-VAE   (b) GIN-SD   (c) HyperDet

Figure 3: Visualization of source detection results on House.

Performance on Early Detection. Early detection of rumor sources is crucial to prevent further escalation of harm. We analyze the performance of various methods as the scale of rumor propagation increases from 10% to 30%, in an increment of 5%. The results are summarized in Fig. 4. As the scale of propagation expands, the accuracy of all methods gradually declines. This trend is primarily due to the increasing number of infected nodes that need to be differentiated, underscoring the importance of early intervention in rumor source detection. Moreover, HyperDet outperforms baseline methods across all datasets, demonstrating its efficiency.
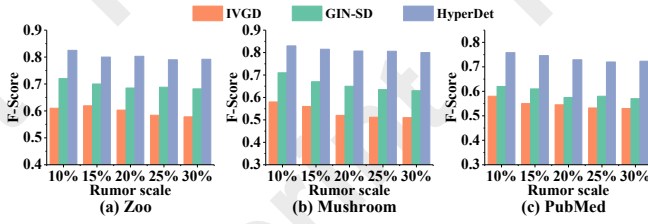
Figure 4: The performance in early rumor sources detection.

Impact of Data Incompleteness. The data incomplete rates are increased from 0 to 0.25 in steps of 0.05, indicating that a corresponding proportion of nodes lack features. The results are shown in Fig. 5. As the proportion of incomplete data increases, there is a noticeable decline in the performance of all methods. This decrease is primarily due to the incomplete features of nodes, which hinder the model's ability to effectively learn and converge. However, the performance of HyperDet decreases the least, which is mainly because in addition to node states and propagation features, we also consider structural features, specifically positional encodings. This integration diminishes the impact of feature loss on the model's performance, demonstrating the robustness and adaptability of the proposed HyperDet.
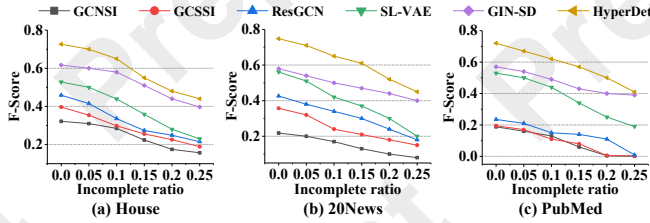

Figure 5: Impact of data incompleteness on source detection.

Computational Efficiency. The runtime of HyperDet compared to methods that integrate propagation information is detailed in Table 2. Due to the requirement for additional modules to learn propagation dynamics, methods such as IVGD and SL-VAE exhibit longer runtimes. In contrast, HyperDet incorporates propagation information into feature embeddings, thereby reducing model complexity. This demonstrates its enhanced efficiency.

Datasets	IVGD	SL-VAE	HyperDet
NTU2012	179.71	198.53	**126.11**
Mushroom	264.30	295.05	**212.86**

Table 2: The runtime (s) comparison across different methods on NTU2012 and Mushroom datasets.

5.3 Ablation Study and Other Analyses

Effect of Interactive Relationship Construction. We initially remove the high-order structure from the hypergraph; results from w/o H indicate a significant performance decline compared to configurations that retain this structure, as shown in Table 3. Additionally, by preserving the high-order topology while eliminating the dynamic interaction construction, as in w/o D, we observe an improvement over w/o H, yet performance remains inferior to the full HyperDet model. These

Module	Methods	ModelNet40		PubMed	
		ACC	F-Score	ACC	F-Score
IRC	w/o H	0.835	0.629	0.840	0.611
	w/o D	0.854	0.753	0.847	0.705
FAF	w/o E	0.874	0.771	0.852	0.705
	w/o A	0.837	0.739	0.846	0.697
	w/ AL	0.825	0.731	0.839	0.689
	w/ AS	0.774	0.358	0.781	0.260
-	HyperDet	**0.887**	**0.786**	**0.870**	**0.723**

Table 3: Performance of different HyperDet variants.

findings underscore the importance of both static topological structures and dynamic interactions in capturing the high-order relationships between nodes.

Effect of Feature-rich Attention Fusion. Removing the autoencoder, as indicted in Table 3, w/o E demonstrates that manually engineered features with sparse dimensions degrade model performance. Regarding attention, we design three variants: treating nodes and hyperedges uniformly, w/o A achieves the best results. w/ AL is next, primarily due to the presence of small-degree yet crucial bridge nodes. w/ AS performs the worst, mainly because it overly focuses on nodes with lesser information transmission capabilities. These findings emphasize the importance of automatic learning and enrichment of features, as well as adaptive attention in enhancing model efficacy.

Effects of Different Information Diffusion Models. The performance of HyperDet under various propagation models is depicted in Table 4. Experimental results affirm that HyperDet consistently exhibits stable performance not only in the influence-based IC model but also in the infection-based SI, SIS, and SIR models. Notably, in the SIS and SIR models, the sources may randomly recover and thus become obscured, leading to a decline in performance. These findings substantiate the versatility and broad applicability of HyperDet across diverse modeling scenarios.

Models	House			NTU2012		
	ACC	F-Score	AUC	ACC	F-Score	AUC
SI	0.878	0.728	0.881	0.879	0.711	0.873
SIS	0.863	0.692	0.851	0.859	0.685	0.862
SIR	0.854	0.687	0.850	0.862	0.680	0.865
IC	0.884	0.726	0.875	0.877	0.715	0.869

Table 4: The performance of HyperDet under different information propagation models.

6 Conclusion

This paper explores the phenomenon of group interactions within social networks and introduces a novel source detection approach, HyperDet, applied to hypergraphs to address this challenge. The key idea lies in integrating static topology with dynamic interactions to construct node relationships, followed by enhancing model learning capabilities through feature enrichment and adaptive attention mechanisms. Extensive comparative and ablation experiments on hypergraphs demonstrate HyperDet's efficacy and robustness against state-of-the-art methods. We hope this work can inspire source detection researchers and advances the development of effective modeling in this field.

Acknowledgements

References

[Ali *et al.*, 2019] Syed Shafat Ali, Tarique Anwar, Ajay Rastogi, and Syed Afzal Murtaza Rizvi. EPA: Exoneration and prominence based age for infection source identification. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 891–900, 2019.

[Amburg *et al.*, 2020] Ilya Amburg, Nate Veldt, and Austin Benson. Clustering in graphs and hypergraphs with categorical edge labels. In *Proceedings of The Web Conference 2020*, pages 706–717, 2020.

[Asuncion *et al.*, 2007] Arthur Asuncion, David Newman, et al. Uci machine learning repository, 2007.

[Bai *et al.*, 2021] Song Bai, Feihu Zhang, and Philip HS Torr. Hypergraph convolution and hypergraph attention. *Pattern Recognition*, 110:107637, 2021.

[Bao *et al.*, 2024] Qing Bao, Ying Jiang, Wang Zhang, Pengfei Jiao, and Jing Su. Graph contrastive learning for source localization in social networks. *Information Sciences*, page 121090, 2024.

[Battiston *et al.*, 2020] Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, and Giovanni Petri. Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 874:1–92, 2020.

[Chen *et al.*, 2003] Ding-Yun Chen, Xiao-Pei Tian, Yu-Te Shen, and Ming Ouhyoung. On visual similarity based 3d model retrieval. In *Computer Graphics Forum*, volume 22, pages 223–232. Wiley Online Library, 2003.

[Cheng *et al.*, 2024a] Le Cheng, Peican Zhu, Chao Gao, Zhen Wang, and Xuelong Li. A heuristic framework for sources detection in social networks via graph convolutional networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2024.

[Cheng *et al.*, 2024b] Le Cheng, Peican Zhu, Keke Tang, Chao Gao, and Zhen Wang. GIN-SD: source detection in graphs with incomplete nodes via positional encoding and attentive fusion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 55–63, 2024.

[Cheng *et al.*, 2025] Le Cheng, Peican Zhu, Keke Tang, Chao Gao, and Zhen Wang. Efficient source detection in incomplete networks via sensor deployment and source approaching. *IEEE Transactions on Information Forensics and Security*, 2025.

[Chien *et al.*, 2022] Eli Chien, Chao Pan, Jianhao Peng, and Olgica Milenkovic. You are allset: A multiset function framework for hypergraph neural networks. In *ICLR*, 2022.

[Chodrow *et al.*, 2021] Philip S Chodrow, Nate Veldt, and Austin R Benson. Generative hypergraph clustering: From blockmodels to modularity. *Science Advances*, 7(28):eabh1303, 2021.

[de Arruda *et al.*, 2020] Guilherme Ferraz de Arruda, Giovanni Petri, and Yamir Moreno. Social contagion models on hypergraphs. *Physical Review Research*, 2(2):023032, 2020.

[Dong *et al.*, 2019] Ming Dong, Bolong Zheng, Nguyen Quoc Viet Hung, Han Su, and Guohui Li. Multiple rumor source detection with graph convolutional networks. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pages 569–578, 2019.

[Dong *et al.*, 2022] Ming Dong, Bolong Zheng, Guohui Li, Chenliang Li, Kai Zheng, and Xiaofang Zhou. Wavefront-based multiple rumor sources identification by multi-task learning. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 6(5):1068–1078, 2022.

[Dong *et al.*, 2024] Yiqi Dong, Dongxiao He, Xiaobao Wang, Youzhu Jin, Meng Ge, Carl Yang, and Di Jin. Unveiling implicit deceptive patterns in multi-modal fake news via neuro-symbolic reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 8354–8362, 2024.

[Feng *et al.*, 2019] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao. Hypergraph neural networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3558–3565, 2019.

[Gao *et al.*, 2022] Yue Gao, Yifan Feng, Shuyi Ji, and Rongrong Ji. HGNN+: General hypergraph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3181–3199, 2022.

[Huang and Yang, 2021] Jing Huang and Jie Yang. Unignn: a unified framework for graph and hypergraph neural networks. In *IJCAI*, pages 2563–2569, 2021.

[Jiao *et al.*, 2024] Pengfei Jiao, Hongqian Chen, Qing Bao, Wang Zhang, and Huaming Wu. Enhancing multi-scale diffusion prediction via sequential hypergraphs and adversarial learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 8571–8581, 2024.

[Jin *et al.*, 2023] Di Jin, Bingdao Feng, Siqi Guo, Xiaobao Wang, Jianguo Wei, and Zhen Wang. Local-global defense against unsupervised adversarial attacks on graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 8105–8113, 2023.

[Kipf and Welling, 2017] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR*, 2017.

[Li *et al.*, 2021] Liang Li, Jianye Zhou, Yuewen Jiang, and Biqing Huang. Propagation source identification of infectious diseases with graph convolutional networks. *Journal of Biomedical Informatics*, 116:103720, 2021.

[Ling *et al.*, 2022] Chen Ling, Junji Jiang, Junxiang Wang, and Zhao Liang. Source localization of graph diffusion via variational autoencoders for graph inverse problems. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1010–1020, 2022.

[Pinto *et al.*, 2012] Pedro C Pinto, Patrick Thiran, and Martin Vetterli. Locating the source of diffusion in large-scale networks. *Physical Review Letters*, 109(6):068702, 2012.

[Prakash *et al.*, 2012] B Aditya Prakash, Jilles Vreeken, and Christos Faloutsos. Spotting culprits in epidemics: How many and which ones? In *2012 IEEE 12th International Conference on Data Mining*, pages 11–20. IEEE, 2012.

[Shah and Zaman, 2011] Devavrat Shah and Tauhid Zaman. Rumors in a network: Who's the culprit? *IEEE Transactions on Information Theory*, 57(8):5163–5181, 2011.

[Shah *et al.*, 2020] Chintan Shah, Nima Dehmamy, Nicola Perra, Matteo Chinazzi, Albert-László Barabási, Alessandro Vespignani, and Rose Yu. Finding patient zero: Learning contagion source with graph neural networks. *arXiv preprint arXiv:2006.11913*, 2020.

[Veličković *et al.*, 2017] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2017.

[Wang *et al.*, 2017] Zheng Wang, Chaokun Wang, Jisheng Pei, and Xiaojun Ye. Multiple source detection without knowing the underlying propagation model. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[Wang *et al.*, 2022] Junxiang Wang, Junji Jiang, and Liang Zhao. An invertible graph diffusion neural network for source localization. In *Proceedings of the ACM Web Conference 2022*, pages 1058–1069, 2022.

[Wang *et al.*, 2025] Xiaobao Wang, Yujing Wang, Dongxiao He, Zhe Yu, Yawen Li, Longbiao Wang, Jianwu Dang, and Di Jin. Elevating knowledge-enhanced entity and relationship understanding for sarcasm detection. *IEEE Transactions on Knowledge and Data Engineering*, 2025.

[Wu *et al.*, 2015] Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1912–1920, 2015.

[Yadati *et al.*, 2019] Naganand Yadati, Madhav Nimishakavi, Prateek Yadav, Vikram Nitin, Anand Louis, and Partha Talukdar. Hypergcn: A new method for training graph convolutional networks on hypergraphs. *Advances in Neural Information Processing Systems*, 32, 2019.

[Yang *et al.*, 2020] Fan Yang, Shuhong Yang, Yong Peng, Yabing Yao, Zhiwen Wang, Houjun Li, Jingxian Liu, Ruisheng Zhang, and Chungui Li. Locating the propagation source in complex networks with a direction-induced search based gaussian estimator. *Knowledge-Based Systems*, 195:105674, 2020.

[Zhou *et al.*, 2006] Dengyong Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. *Advances in Neural Information Processing Systems*, 19, 2006.

[Zhu *et al.*, 2017] Kai Zhu, Zhen Chen, and Lei Ying. Catch'em all: Locating multiple diffusion sources in networks with partial observations. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[Zhu *et al.*, 2022] Peican Zhu, Le Cheng, Chao Gao, Zhen Wang, and Xuelong Li. Locating multi-sources in social networks with a low infection rate. *IEEE Transactions on Network Science and Engineering*, 9(3):1853–1865, 2022.