

# Self-Supervised Hypergraph Learning with Substructure Awareness for Hyperedge Prediction

Ming Li<sup>1</sup>, Huiting Wang<sup>2</sup>, Yuting Chen<sup>3,1\*</sup>, Lu Bai<sup>4\*</sup>, Lixin Cui<sup>5</sup>, Feilong Cao<sup>6</sup>, Ke Lv<sup>7,8</sup>

<sup>1</sup>Zhejiang Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University, Jinhua, China

<sup>2</sup>School of Computer Science and Technology, Zhejiang Normal University, Jinhua, China

<sup>3</sup>Centre for Learning Sciences and Technologies, The Chinese University of Hong Kong, Hong Kong, China

<sup>4</sup>School of Artificial Intelligence, Beijing Normal University, Beijing, China

<sup>5</sup>Central University of Finance and Economics, Beijing, China.

<sup>6</sup>School of Mathematical Sciences, Zhejiang Normal University, Jinhua, China

<sup>7</sup>School of Engineering Science, University of Chinese Academy of Sciences, Beijing, China

<sup>8</sup>Peng Cheng Laboratory, Shenzhen, China

mingli@zjnu.edu.cn, huiting\_wang@zjnu.edu.cn, yuting.chen@cuhk.edu.hk, bailu@bnu.edu.cn, cuilixin@cufe.edu.cn, caofeilong88@zjnu.edu.cn, luk@ucas.ac.cn

## Abstract

Hyperedge prediction plays a central role in hypergraph learning, enabling the inference of high-order relations among multiple entities. However, existing methods often rely on a simplistic *flat set assumption*, treating candidate hyperedges as unstructured collections of nodes and neglecting their potential internal compositionality. Furthermore, the severe scarcity of observed hyperedges poses a challenge for effective supervision. In this work, we propose **S<sup>3</sup>Hyper**, a **Substructure-contextualized Self-Supervised** framework for **Hyperedge** prediction, which jointly addresses these two challenges. Specifically, we design a substructure-contextualized hyperedge aggregator that models the internal hierarchy of candidate hyperedges by leveraging sub-hyperedge information. In parallel, we introduce an adaptive tri-directional contrastive learning module that incorporates node-level, hyperedge-level, and cross-level alignment objectives, supported by temperature-adaptive mechanisms. Experimental results on four public datasets demonstrate that S<sup>3</sup>Hyper consistently outperforms strong baselines, with ablation studies verifying the effectiveness of each component.

## 1 Introduction

Hypergraphs extend traditional graphs by allowing hyperedges to connect arbitrary-sized subsets of nodes, making them effective for modeling high-order interactions in complex systems such as co-authorship networks, group chats, and biochemical pathways (Wang and Kleinberg 2024; Millán et al. 2025; Li et al. 2025b). Among various hypergraph learning tasks, hyperedge prediction (Chen and Liu 2023), which aims to infer missing or future hyperedges, is fundamental for understanding group formation dynamics and enabling downstream applications like recommendation, community detection, and knowledge discovery (Kim et al. 2024; Zhang et al. 2025; Li et al. 2025a).

Despite initial advances, existing hyperedge prediction approaches face significant limitations that hinder their ability to fully exploit the expressive potential of hypergraphs (Zhang, Zou, and Ma 2020; Yadati et al. 2020; Hwang et al. 2022; Ko, Tong, and Kim 2025). One critical issue is the reliance on an implicit yet unexamined *flat set assumption*, wherein each candidate hyperedge is treated as an unordered collection of nodes, devoid of any internal structure or formation logic. However, in many real-world scenarios, hyperedges often emerge through a compositional process. For instance, in a co-authorship network, a new research group may arise by expanding upon existing sub-groups with established collaborations. Such substructures, which we refer to as *sub-hyperedges*, carry strong inductive signals about the legitimacy of a larger hyperedge. Unfortunately, most existing models overlook this intrinsic “part-to-whole” compositionality. Their aggregation modules, whether based on pooling functions or attention mechanisms, fail to contextualize the internal formation dynamics of a candidate hyperedge, thereby limiting predictive accuracy and generalization.

Another persistent challenge lies in the severe *positive sample scarcity* inherent in hyperedge prediction. The space of all possible hyperedges grows combinatorially with the number of nodes, while only a small fraction of valid hyperedges are typically observed. Some recent works attempt to mitigate this issue through contrastive self-supervised learning. However, they primarily focus on aligning similar entities (e.g., node-node or hyperedge-hyperedge pairs) across augmented views, ignoring the *affiliation logic* between different levels of granularity. In particular, they lack mechanisms to model how individual nodes collectively give rise to valid hyperedges, i.e., a gap we term the *affiliation gap*. This limits the model’s ability to learn representations that reflect not only similarity but also the underlying generative structure of hyperedges.

To address these limitations, we propose a novel framework, **Substructure-Contextualized Self-Supervised** learn-

\*Corresponding authors: Yuting Chen, Lu Bai  
Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

ing for **Hyperedge** prediction, termed **S<sup>3</sup>Hyper**. To overcome the flat set assumption, we design a **Substructure-Contextualized Hyperedge Aggregator**. This module takes into account the internal composition of candidate hyperedges by identifying and encoding meaningful sub-hyperedges from the known hypergraph. A substructure-aware node enhancement module extracts fine-grained local structure signals, which are then integrated via a context fusion layer and passed through a context-enhanced aggregation network to produce a representation that reflects the internal coherence of the candidate hyperedge. To mitigate the positive sample scarcity and bridge the affiliation gap, we propose an **Adaptive Tri-Directional Contrastive Learning** module. This component introduces three aligned contrastive objectives: (i) an adaptive node-level contrastive loss to enhance node representation learning with dynamic hardness-aware temperature adjustment; (ii) an adaptive hyperedge-level loss to improve hyperedge discrimination under limited supervision; and (iii) a novel node-hyperedge alignment loss that encourages consistent embedding spaces between nodes and their associated hyperedges. Together, these modules allow S<sup>3</sup>Hyper to reason about hyperedge plausibility not only from structural similarity but also from the generative logic of substructure composition and cross-level affiliation. We conduct extensive experiments on four public benchmark datasets to validate the effectiveness of S<sup>3</sup>Hyper. Compared with a range of strong baseline models, S<sup>3</sup>Hyper consistently achieves superior performance across all datasets. Furthermore, ablation studies confirm the effectiveness of each core component, substructure-contextualized aggregation and tri-directional contrastive learning, demonstrating their complementary roles in improving predictive accuracy.

In summary, our contribution is three-fold:

- We propose a new hyperedge predictor, **S<sup>3</sup>Hyper**, featuring a substructure-contextualized hyperedge aggregator. It explicitly models internal substructures of candidate hyperedges by combining a substructure-aware node enhancement module with a context-enhanced aggregation network, enabling plausibility estimation from compositional cues grounded in observed local sub-hyperedges.
- We develop an adaptive tri-directional contrastive learning scheme with three objectives—adaptive node loss, adaptive hyperedge loss, and node–hyperedge alignment. This scheme captures within-level and cross-level semantics, and adaptively tunes the temperature for hard negatives to alleviate the affiliation gap and sample sparsity.
- We conduct extensive experiments on four benchmark datasets, showing that **S<sup>3</sup>Hyper** consistently outperforms existing baselines across settings. In particular, it yields up to a 14.6% relative AUROC gain on NDC-class, demonstrating the benefit of substructure- and affiliation-aware modeling.

## 2 Preliminaries

**Basics on Hypergraphs.** A hypergraph is represented as  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , comprising a vertex set  $\mathcal{V}$  of size  $N = |\mathcal{V}|$ , a

hyperedge set  $\mathcal{E}$  of size  $M = |\mathcal{E}|$ . Suppose that vertices and hyperedges have feature dimensions  $d$  and  $o$ , respectively, we have the representation of vertex data as  $\mathbf{X} \in \mathbb{R}^{N \times d}$  and hyperedge data as  $\mathbf{Y} \in \mathbb{R}^{M \times o}$ .

The hypergraph structure, from a vertex perspective, is defined by an incidence matrix  $\mathbf{H} \in \{0, 1\}^{N \times M}$  where  $\mathbf{H}(v, e) = 1$  if vertex  $v$  is contained in hyperedge  $e$ , and 0 otherwise, as represented by:

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

The degrees of vertex  $v$  and hyperedge  $e$  are denoted by diagonal matrices  $\mathbf{D}_v \in \mathbb{R}^{N \times N}$  and  $\mathbf{D}_e \in \mathbb{R}^{M \times M}$ , calculated as  $\sum_{e \in \mathcal{E}} \mathbf{H}(v, e)$  and  $\sum_{v \in \mathcal{V}} \mathbf{H}(v, e)$ , respectively.

**Hyperedge Prediction Problem.** Given an observed hypergraph  $\mathcal{G}_{\text{obs}} = (\mathcal{V}, \mathcal{E}_{\text{obs}})$  over a vertex set  $\mathcal{V}$ , and a node feature matrix  $\mathbf{X} \in \mathbb{R}^{N \times d}$ , the objective of hyperedge prediction is to learn a scoring function  $f : \mathcal{P}(\mathcal{V}) \rightarrow \mathbb{R}$ , where  $\mathcal{P}(\mathcal{V})$  denotes the power set of  $\mathcal{V}$ . The function  $f(e')$  assigns a confidence score to any candidate hyperedge  $e' \notin \mathcal{E}_{\text{obs}}$ , reflecting its likelihood of being a valid or *ground-truth* hyperedge. A ground-truth hyperedge refers to a true relation that is either unobserved or expected to emerge in the future. This task is formulated as a binary classification problem: to determine whether a given candidate hyperedge  $e'$  should be classified as positive (i.e., part of the underlying true hyperedge set) or negative.

## 3 Proposed Method: S<sup>3</sup>Hyper

This section presents the proposed **S<sup>3</sup>Hyper** framework (see Figure 1). We begin with an overview of the hyperedge prediction pipeline in **Section 3.1**, followed by detailed discussions of its key components: the substructure-contextualized aggregator (**Section 3.2**) and tri-directional contrastive learning (**Section 3.3**).

### 3.1 Intuitive Overview of the Pipeline

**Hypergraph Encoding.** The framework first encodes the observed hypergraph  $\mathcal{G}_{\text{obs}} = (\mathcal{V}, \mathcal{E}_{\text{obs}})$  by learning representations for both nodes and hyperedges. We adopt a generic message-passing-based hypergraph neural network to capture high-order relational information through iterative updates between nodes and hyperedges. At each layer  $l$ , the embeddings are updated via:

$$\begin{aligned} \mathbf{U}^{(l)} &= \sigma \left( \mathbf{H}^\top \mathbf{Z}^{(l-1)} \mathbf{W}_V^{(l)} + b_V^{(l)} \right), \\ \mathbf{Z}^{(l)} &= \sigma \left( \mathbf{H} \mathbf{U}^{(l)} \mathbf{W}_E^{(l)} + b_E^{(l)} \right), \end{aligned} \quad (2)$$

where  $\mathbf{Z}^{(l)}$  and  $\mathbf{U}^{(l)}$  denote the node and hyperedge embeddings at layer  $l$ , respectively, with  $\mathbf{Z}^{(0)} = \mathbf{X}$ . The trainable parameters include weight matrices  $\mathbf{W}_V^{(l)}, \mathbf{W}_E^{(l)}$  and bias terms  $b_V^{(l)}, b_E^{(l)}$ . After  $L$  layers of propagation, we obtain the final node embeddings  $\mathbf{Z} = \mathbf{Z}^{(L)}$ , which are used in subsequent steps.

**Candidate Hyperedge Representation.** For each candidate hyperedge  $c \subseteq \mathcal{V}$ , the goal is to obtain a compact representation  $\mathbf{h}_c$  that reflects both its constituent node features and

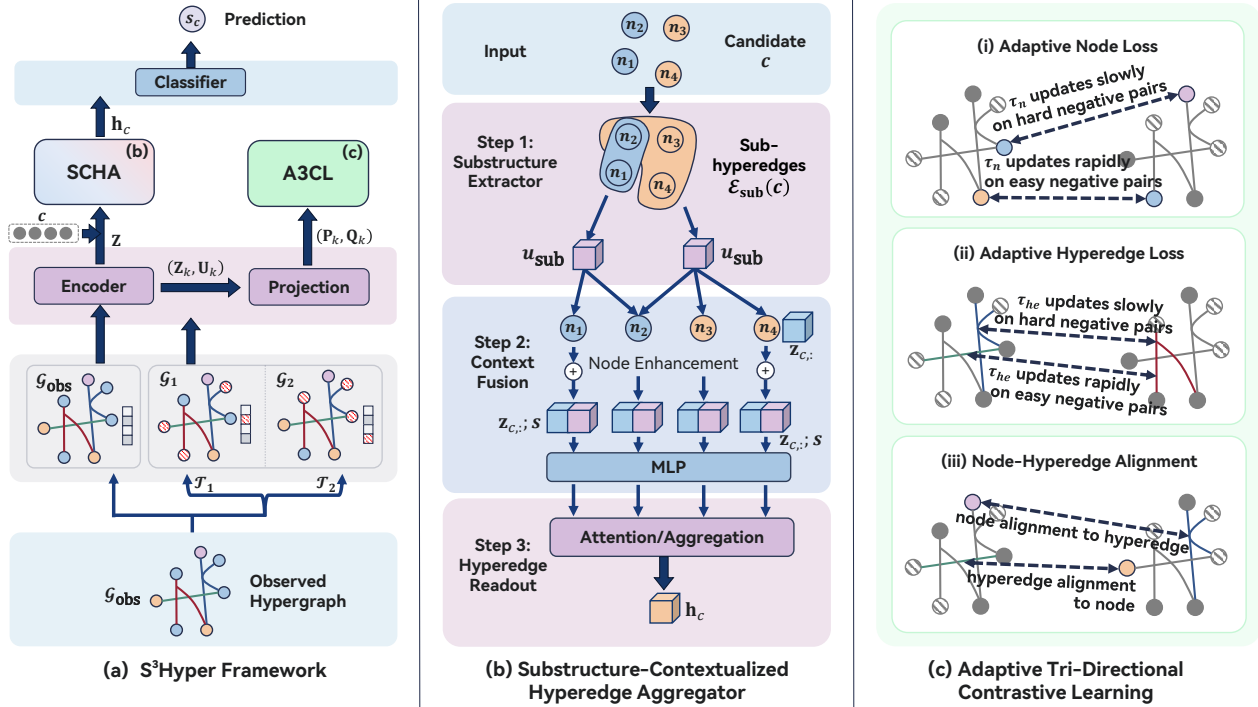


Figure 1: Overview of the  $S^3$ Hyper Framework: (a) presents the overall architecture, which integrates the Substructure-Contextualized Hyperedge Aggregator (SCHA) with the Adaptive Tri-Directional Contrastive Learning module (A3CL); (b) and (c) detail the internal mechanisms of SCHA and illustrate the three contrastive objectives (node-node, hyperedge-hyperedge, and node-hyperedge) employed in A3CL, respectively.

their structural context. Given the learned node embeddings  $\mathbf{Z}$ , a substructure-aware aggregation module (detailed in the next subsection) is applied to the subset  $\{z_v \mid v \in c\}$ , yielding the hyperedge-level representation  $\mathbf{h}_c \in \mathbb{R}^d$ .

**Hyperedge Prediction.** The aggregated representation  $\mathbf{h}_c$  is passed through a prediction head (e.g., a multi-layer perceptron) to produce a confidence score  $s_c \in [0, 1]$ , estimating the likelihood that  $c$  is a true hyperedge. The model is supervised using binary cross-entropy loss:

$$\mathcal{L}_{\text{pred}} = -\frac{1}{|C|} \sum_{c \in C} [y_c \log(s_c) + (1 - y_c) \log(1 - s_c)], \quad (3)$$

where  $C$  is the candidate set and  $y_c \in \{0, 1\}$  denotes the label indicating whether  $c$  is a ground-truth hyperedge.

### 3.2 Substructure-Contextualized Hyperedge Aggregator (SCHA)

To overcome the limitations of traditional aggregation methods that treat candidate hyperedges as flat, unordered sets, we introduce the substructure-contextualized hyperedge aggregator. The key intuition is that a hyperedge representation should reflect not only its constituent nodes but also the underlying substructure formed by known hyperedges. SCHA enhances node embeddings using this internal structure and then aggregates them into a unified hyperedge representation.

**Substructure Extractor.** Given a candidate hyperedge  $c = \{v'_1, v'_2, \dots, v'_k\}$ , we identify all its sub-hyperedges that

are already present in the observed hypergraph:

$$\mathcal{E}_{\text{sub}}(c) = \{e \subseteq c \mid e \in \mathcal{E}_{\text{obs}}\}.$$

For each sub-hyperedge  $e_j \in \mathcal{E}_{\text{sub}}(c)$ , we retrieve its embedding  $\mathbf{u}_{e_j}$ , either from the encoder or via mean pooling over constituent node embeddings. These form the substructure embedding matrix  $\mathbf{U}_{\text{sub}} \in \mathbb{R}^{m \times d}$ , where  $m = |\mathcal{E}_{\text{sub}}(c)|$ . A binary node-substructure incidence matrix  $\mathbf{M} \in \{0, 1\}^{k \times m}$  is constructed as:

$$\mathbf{M}_{ij} = \begin{cases} 1 & \text{if } v_i \in e_j, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

**Substructure-Aware Node Enhancement.** To incorporate structural context, each node aggregates the embeddings of the sub-hyperedges it participates in. Let  $\mathbf{1}_m$  be a length- $m$  vector of ones. The aggregated context matrix  $\mathbf{S} \in \mathbb{R}^{k \times d}$  is computed as:

$$\mathbf{S} = \text{diag}(\mathbf{M}\mathbf{1}_m)^{-1} \mathbf{M}\mathbf{U}_{\text{sub}}. \quad (5)$$

This operation can be replaced with max/sum pooling or attention-based alternatives to capture different structural statistics. The context-aware node embeddings  $\mathbf{Z}^{\text{aug}} \in \mathbb{R}^{k \times d}$  are then obtained by fusing the original node embeddings  $\mathbf{Z}_{c,:}$  with  $\mathbf{S}$  via a linear projection:

$$\mathbf{Z}^{\text{aug}} = \text{ReLU}([\mathbf{Z}_{c,:}; \mathbf{S}]\mathbf{W}_{\text{aug}}), \quad (6)$$

where  $\mathbf{W}_{\text{aug}} \in \mathbb{R}^{2d \times d}$  is a learnable matrix.

**Hyperedge Readout.** In the final stage, we aggregate the enriched node embeddings  $\{\mathbf{z}_1^{\text{aug}}, \dots, \mathbf{z}_k^{\text{aug}}\}$  to form a unified representation  $\mathbf{h}_c$  for candidate hyperedge  $c$ . To capture dependencies among the nodes, we apply a multi-head self-attention layer followed by layer normalization:

$$\mathbf{Z}^{\text{trans}} = \text{LN}(\mathbf{Z}^{\text{aug}} + \text{SelfAttention}(\mathbf{Z}^{\text{aug}})), \quad (7)$$

where  $\text{SelfAttention}(\cdot)$  is a standard multi-head attention module and  $\text{LN}(\cdot)$  denotes layer normalization. We then apply max pooling across the transformed embeddings to obtain the final hyperedge representation:

$$\mathbf{h}_c = \text{MaxPooling}(\mathbf{Z}^{\text{trans}}). \quad (8)$$

This representation, which encodes both internal substructure and inter-node interactions, is passed to a prediction head:

$$s_c = \text{Classifier}(\mathbf{h}_c), \quad (9)$$

where  $\text{Classifier}(\cdot)$  is a fully connected layer followed by a sigmoid activation.

### 3.3 Adaptive Tri-Directional Contrastive Learning for Node-Hyperedge Affiliation

To enhance embedding quality and bridge the affiliation gap between nodes and hyperedges, we introduce an auxiliary self-supervised module: **Adaptive Tri-Directional Contrastive Learning (A3CL)**. This module simultaneously optimizes three complementary contrastive objectives, node-level, hyperedge-level, and node-hyperedge alignment, to encourage structural consistency and affiliation awareness across views.

**Hypergraph Augmentation.** Following standard contrastive learning practices, we apply two random augmentations  $\mathcal{T}_1, \mathcal{T}_2 \sim \mathcal{T}$  to the input hypergraph  $\mathcal{G} = (\mathbf{X}, \mathbf{H})$ , generating two perturbed views  $\mathcal{G}_1 = (\mathbf{X}_1, \mathbf{H}_1)$  and  $\mathcal{G}_2 = (\mathbf{X}_2, \mathbf{H}_2)$ . We adopt node dropping and attribute masking as augmentation strategies, introducing perturbations at both structural and feature levels.

**Tri-Directional Contrastive Objectives.** Each view is encoded via a shared hypergraph encoder  $f_\theta(\cdot)$ , producing node and hyperedge embeddings  $(\mathbf{Z}_k, \mathbf{U}_k)$  for  $k \in \{1, 2\}$ . Two projection heads,  $g_\phi(\cdot)$  for nodes and  $g_\psi(\cdot)$  for hyperedges, are used to map encoder outputs into contrastive spaces:  $\mathbf{P}_k = g_\phi(\mathbf{Z}_k)$ ,  $\mathbf{Q}_k = g_\psi(\mathbf{U}_k)$ .

**(i) Adaptive Node-Level Loss.** This objective encourages consistency of node embeddings across views. We adopt a symmetric InfoNCE loss with adaptive temperature  $\tau_n^*$ :

$$\mathcal{L}_n = \frac{1}{2|\mathcal{V}|} \sum_{i=1}^{|\mathcal{V}|} [\ell_n(p_i^1, p_i^2) + \ell_n(p_i^2, p_i^1)], \quad (10)$$

$$\ell_n(p_i^1, p_i^2) = -\log \frac{\exp(\text{sim}(p_i^1, p_i^2)/\tau_n^*)}{\sum_{k=1}^{|\mathcal{V}|} \exp(\text{sim}(p_i^1, p_k^2)/\tau_n^*)}. \quad (11)$$

**(ii) Adaptive Hyperedge-Level Loss.** This loss aligns representations of the same hyperedge across views, using a symmetric InfoNCE loss with temperature  $\tau_{he}^*$ :

$$\mathcal{L}_{he} = \frac{1}{2|\mathcal{E}|} \sum_{j=1}^{|\mathcal{E}|} [\ell_{he}(q_j^1, q_j^2) + \ell_{he}(q_j^2, q_j^1)], \quad (12)$$

$$\ell_{he}(q_j^1, q_j^2) = -\log \frac{\exp(\text{sim}(q_j^1, q_j^2)/\tau_{he}^*)}{\sum_{k=1}^{|\mathcal{E}|} \exp(\text{sim}(q_j^1, q_k^2)/\tau_{he}^*)}. \quad (13)$$

**(iii) Node-Hyperedge Alignment Loss.** To explicitly model affiliation relationships, we introduce a node-hyperedge alignment loss. For each true membership pair  $(v_i, e_j)$ , we treat  $(p_i^1, q_j^2)$  as a positive pair and  $(p_i^1, q_k^2)$  as a negative, where  $e_k \neq v_i$ . A learnable discriminator  $D(\cdot, \cdot)$  scores each pair:

$$\ell_{n-he}(p_i^1, q_j^2) = -\left[ \log \frac{e^{D(p_i^1, q_j^2)}}{e^{D(p_i^1, q_j^2)} + e^{D(p_i^1, q_k^2)}} + \log \frac{e^{D(q_j^2, p_i^1)}}{e^{D(q_j^2, p_i^1)} + e^{D(q_j^2, p_l^1)}} \right],$$

where  $q_k^2, p_l^1$  are sampled negatives. The final loss is:

$$\mathcal{L}_{n-he} = \sum_{(v_i, e_j) \in \mathcal{M}} [\ell_{n-he}(p_i^1, q_j^2) + \ell_{n-he}(p_i^2, q_j^1)],$$

with  $\mathcal{M}$  denoting all observed node-hyperedge pairs.

**Adaptive Temperature Mechanism.** To dynamically adjust contrastive hardness, we introduce an adaptive temperature update rule:

$$\tau^{(t)} = \max \left( \tau^{(t-1)} - \eta \cdot \left[ \log \left( \frac{1}{N_p} \sum_{i \neq j} \exp(\rho \cdot d_{ij}) \right) \right]^{-1}, \tau_{\text{low}} \right)$$

where  $d_{ij} = \|\mathbf{e}_i - \mathbf{e}_j\|^2$  is the pairwise squared distance between embeddings, with  $\mathbf{e}_i$  and  $\mathbf{e}_j$  denoting either node or hyperedge embeddings depending on whether the update is applied to  $\tau_n$  or  $\tau_{he}$ .  $\eta$  is the update rate,  $\rho$  a scaling factor,  $N_p$  the number of negative pairs, and  $\tau_{\text{low}}$  the minimum temperature. This adaptive temperature mechanism is individually applied to the **(i) Adaptive Node-Level Loss** and **(ii) Adaptive Hyperedge-Level Loss**, allowing dynamic adjustment of  $\tau_n^*$  and  $\tau_{he}^*$  based on the evolving difficulty of distinguishing negative pairs during training.

### 3.4 Joint Training

The overall training objective of S<sup>3</sup>Hyper integrates the supervised prediction loss from the main channel,  $\mathcal{L}_{\text{pred}}$ , with the auxiliary contrastive loss,  $\mathcal{L}_{\text{A3CL}}$ , in a unified manner. The final loss function is defined as:

$$\mathcal{L} = \mathcal{L}_{\text{pred}} + \mu \cdot \mathcal{L}_{\text{A3CL}}, \quad (14)$$

where the auxiliary loss is composed of three components:

$$\mathcal{L}_{\text{A3CL}} = \lambda_n \mathcal{L}_n + \lambda_{he} \mathcal{L}_{he} + \lambda_{n-he} \mathcal{L}_{n-he}, \quad (15)$$

and  $\mu$ ,  $\lambda_n$ ,  $\lambda_{he}$ , and  $\lambda_{n-he}$  are hyperparameters controlling the relative contributions of each loss term. The sensitivity of these hyperparameters is empirically investigated in our subsequent experimental studies.

**Remark.** In summary, S<sup>3</sup>Hyper provides a unified and principled framework that simultaneously tackles two longstanding and underexplored challenges in hyperedge prediction. First, the **Substructure-Contextualized Aggregation** module explicitly incorporates internal substructure signals into the representation of candidate hyperedges, moving beyond the flat-set assumption commonly adopted in

Dataset	Test set	AUROC					Average Precision (AP)				
		SNS	MNS	CNS	MIX	Average	SNS	MNS	CNS	MIX	Average
Citeseer	HyperSAGNN	0.540	0.410	0.473	0.478	0.475	0.627	0.455	0.497	0.507	0.512
	NHP	<b>0.991</b>	0.701	0.510	0.817	0.751	<b>0.990</b>	0.731	0.520	0.768	0.751
	AHP	0.943	0.881	0.651	0.820	0.824	0.952	0.870	0.660	0.795	0.819
	CASH	0.925	0.921	0.720	0.857	0.856	0.928	0.919	0.701	0.831	0.845
	<b>S<sup>3</sup>Hyper</b>	0.940	<b>0.925</b>	<b>0.722</b>	<b>0.858</b>	<b>0.861</b>	0.940	<b>0.927</b>	<b>0.709</b>	<b>0.842</b>	<b>0.854</b>
Cora	HyperSAGNN	0.617	0.527	0.494	0.540	0.545	0.687	0.574	0.508	0.566	0.584
	NHP	0.943	0.641	0.472	0.774	0.703	0.949	0.678	0.509	0.744	0.718
	AHP	<b>0.964</b>	0.860	0.572	0.799	0.799	<b>0.961</b>	0.837	0.552	0.740	0.772
	CASH	0.923	0.867	0.671	0.824	0.822	0.915	0.854	0.644	0.789	0.801
	<b>S<sup>3</sup>Hyper</b>	0.919	<b>0.874</b>	<b>0.692</b>	<b>0.831</b>	<b>0.829</b>	0.910	<b>0.861</b>	<b>0.666</b>	<b>0.798</b>	<b>0.809</b>
NDC-class	HyperSAGNN	0.701	0.572	0.601	0.612	0.622	0.829	0.669	0.640	0.632	0.692
	NHP	0.839	0.786	0.714	0.721	0.765	0.577	0.375	0.272	0.219	0.361
	AHP	0.861	0.799	0.729	0.725	0.779	0.798	0.586	0.375	0.304	0.516
	CASH	0.881	0.719	0.653	0.756	0.752	0.852	0.727	0.675	0.750	0.751
	<b>S<sup>3</sup>Hyper</b>	<b>0.941</b>	<b>0.840</b>	<b>0.802</b>	<b>0.865</b>	<b>0.862</b>	<b>0.937</b>	<b>0.862</b>	<b>0.830</b>	<b>0.873</b>	<b>0.876</b>
DBLP	HyperSAGNN	0.448	0.574	0.572	0.530	0.531	0.562	0.602	0.586	0.577	0.582
	NHP	0.663	0.540	0.503	0.572	0.569	0.608	0.523	0.501	0.542	0.544
	AHP	<b>0.946</b>	0.820	0.568	0.778	0.778	<b>0.947</b>	0.815	0.561	0.735	0.764
	CASH	0.875	0.836	0.708	0.807	0.807	0.874	0.832	0.696	0.793	0.799
	<b>S<sup>3</sup>Hyper</b>	0.880	<b>0.855</b>	<b>0.738</b>	<b>0.826</b>	<b>0.825</b>	0.874	<b>0.844</b>	<b>0.716</b>	<b>0.806</b>	<b>0.810</b>

Table 1: Hyperedge prediction results on the four benchmark datasets. AUROC and Average Precision (AP) scores are reported across four distinct test sets: SNS, MNS, CNS, and MIX. The best-performing results in each column are indicated in **bold**.

prior work. Second, the **Adaptive Tri-Directional Contrastive Learning** module introduces a multi-perspective self-supervised signal that not only strengthens intra-view consistency but also enables meaningful alignment between nodes and hyperedges through affiliation-aware objectives. Together, these two strategies reinforce each other, allowing S<sup>3</sup>Hyper to produce structurally informed and affiliation-sensitive representations that are highly beneficial for hyperedge prediction.

## 4 Experiments

### 4.1 Experimental Setups

**Datasets.** We evaluate the proposed S<sup>3</sup>Hyper framework on four widely-used real-world hypergraph datasets from diverse domains, following the experimental setup in prior studies on hyperedge prediction (Dong, Sawin, and Bengio 2020; Ko, Tong, and Kim 2025): (1) **Cora** and **Citeseer** (Yadati et al. 2019), two co-citation networks where nodes represent academic papers and hyperedges denote sets of papers co-cited by the same reference; (2) **DBLP** (Yadati et al. 2019), a collaboration network in which nodes are researchers and hyperedges correspond to co-authorship groups; (3) **NDC-class** (Benson et al. 2018), a biomedical dataset where nodes are chemical components and hyperedges represent drug compounds formed by interacting ingredients. Node features for Cora, Citeseer, and DBLP are derived from bag-of-words encodings of paper abstracts, while NDC-class uses one-hot vectors indicating drug class labels.

**Baselines.** We compare S<sup>3</sup>Hyper with four representative hyperedge prediction models: **HyperSAGNN**(Zhang, Zou, and Ma 2020), a self-attention-based GNN tailored for variable-size hyperedges; **NHP**(Yadati et al. 2020), which

exploits hyperedge-aware message passing with max-min pooling; **AHP**(Hwang et al. 2022), which integrates adversarially sampled negatives; and **CASH**(Ko, Tong, and Kim 2025), a contrastive self-supervised model designed for contextual relation modeling within candidate groups. These baselines are selected for their strong performance and widespread adoption in the literature.

**Implementation Details.** We follow the evaluation protocol established in (Hwang et al. 2022) to ensure reproducibility and fairness. For each dataset, five independent random splits are performed. The set of existing hyperedges (i.e., positive samples) is divided into training (60%), validation (20%), and test (20%) subsets. Evaluation is conducted across four negative sampling strategies of increasing difficulty, which generate challenging negative examples for model discrimination: (i) **SNS**: Size-based negative sampling; (ii) **MNS**: Motif-based negative sampling; (iii) **CNS**: Clique-based negative sampling; (iv) **MIX**: A balanced mixture of all three types.

### 4.2 Overall Performance Comparison

Table 1 summarizes the predictive accuracy of S<sup>3</sup>Hyper across four benchmark hypergraph datasets under varying negative sampling strategies. Among the 32 evaluation cases (4 datasets × 2 metrics × 4 test settings), S<sup>3</sup>Hyper ranks first in 26, consistently outperforming all baselines. It achieves the best average AUROC and AP on every dataset, confirming its robustness and superior predictive capability. Notably, on the NDC-class dataset, S<sup>3</sup>Hyper yields an average AUROC of 0.862, outperforming HyperSAGNN, NHP, AHP, and CASH by 38.6%, 12.7%, 10.7%, and 14.6%, respectively. While NHP and AHP show competitive results on the SNS test set, their performance deteriorates signifi-

Dataset	Variants	AUROC					Average Precision (AP)				
		SNS	MNS	CNS	MIX	Average	SNS	MNS	CNS	MIX	Average
Citeseer	Baseline	0.842	0.839	0.623	0.771	0.769 ± 0.021	0.854	0.830	0.648	0.765	0.774 ± 0.016
	w/o SCHA	0.896	0.884	0.672	0.819	0.817 ± 0.023	0.903	0.868	0.680	0.805	0.814 ± 0.016
	w/o A3CL	0.909	0.911	0.682	0.832	0.833 ± 0.019	0.918	0.897	0.630	0.807	0.813 ± 0.018
	S <sup>3</sup> Hyper	<b>0.940</b>	<b>0.925</b>	<b>0.722</b>	<b>0.858</b>	<b>0.861 ± 0.016</b>	<b>0.940</b>	<b>0.927</b>	<b>0.709</b>	<b>0.842</b>	<b>0.854 ± 0.011</b>
	<i>Imp.</i>	+11.60%	+10.30%	+15.85%	+11.26%	+12.02%	+9.98%	11.75%	+9.39%	+10.08%	+10.36%
Cora	Baseline	0.817	0.738	0.536	0.699	0.697 ± 0.012	0.819	0.747	0.542	0.678	0.697 ± 0.015
	w/o SCHA	0.853	0.770	0.550	0.722	0.724 ± 0.022	0.861	0.779	0.541	0.698	0.720 ± 0.023
	w/o A3CL	0.912	0.870	0.653	0.817	0.813 ± 0.019	0.903	0.860	0.608	0.772	0.786 ± 0.020
	S <sup>3</sup> Hyper	<b>0.919</b>	<b>0.874</b>	<b>0.692</b>	<b>0.831</b>	<b>0.829 ± 0.013</b>	<b>0.910</b>	<b>0.861</b>	<b>0.666</b>	<b>0.798</b>	<b>0.809 ± 0.017</b>
	<i>Imp.</i>	+12.45%	+18.41%	+29.06%	+18.92%	+18.84%	+11.05%	+15.25%	+22.95%	+17.61%	+16.09%
NDC-class	Baseline	0.892	0.732	0.679	0.774	0.769 ± 0.017	0.865	0.740	0.702	0.767	0.769 ± 0.017
	w/o SCHA	0.893	0.743	0.693	0.781	0.778 ± 0.019	0.869	0.754	0.716	0.775	0.778 ± 0.021
	w/o A3CL	0.938	0.832	0.800	0.859	0.857 ± 0.014	0.935	0.856	0.828	0.867	0.872 ± 0.013
	S <sup>3</sup> Hyper	<b>0.941</b>	<b>0.840</b>	<b>0.802</b>	<b>0.865</b>	<b>0.862 ± 0.015</b>	<b>0.937</b>	<b>0.862</b>	<b>0.830</b>	<b>0.873</b>	<b>0.876 ± 0.013</b>
	<i>Imp.</i>	+5.53%	+14.74%	+17.99%	+11.72%	+12.03%	+8.35%	+16.45%	+18.29%	+13.79%	+13.92%
DBLP	Baseline	0.849	0.765	0.581	0.731	0.732 ± 0.006	0.860	0.761	0.583	0.713	0.729 ± 0.007
	w/o SCHA	0.879	0.812	0.614	0.770	0.769 ± 0.003	0.879	0.799	0.597	0.737	0.753 ± 0.010
	w/o A3CL	0.849	<b>0.860</b>	0.632	0.782	0.781 ± 0.007	0.853	<b>0.854</b>	0.591	0.744	0.760 ± 0.010
	S <sup>3</sup> Hyper	<b>0.880</b>	0.855	<b>0.738</b>	<b>0.826</b>	<b>0.825 ± 0.004</b>	<b>0.874</b>	0.844	<b>0.716</b>	<b>0.806</b>	<b>0.810 ± 0.004</b>
	<i>Imp.</i>	+3.59%	+11.74%	+27.01%	+12.91%	+12.68%	+1.60%	+10.95%	+22.84%	+12.97%	+11.07%

Table 2: Ablation results assessing the impact of key components in S<sup>3</sup>Hyper across four benchmark datasets.

cantly under the more challenging MNS and CNS settings. For example, both methods approach random performance ( $\approx 0.5$  AUROC) on the CNS set, suggesting potential overfitting to simpler negative samples. In contrast, S<sup>3</sup>Hyper maintains consistently high scores across all test configurations, highlighting its ability to handle varying levels of negative sample difficulty. On DBLP’s CNS set, it achieves the highest AUROC (0.738) and AP (0.716), demonstrating strong generalization on sparse and high-order structures.

These results underscore the effectiveness of S<sup>3</sup>Hyper in addressing two fundamental challenges in hyperedge prediction, namely, the flat set assumption and positive sample scarcity. By modeling substructure-contextualized signals and affiliation-aware group relations, S<sup>3</sup>Hyper effectively learns meaningful representations even in the presence of hard negatives. This consistent superiority across datasets and sampling strategies validates the design principles behind S<sup>3</sup>Hyper and confirms its broad applicability in real-world hypergraph scenarios.

### 4.3 Ablation Study

We conduct ablation studies to evaluate the individual contribution of each core component in S<sup>3</sup>Hyper. As shown in Table 2, we compare the full model against three variants:

- **Baseline:** A trivial model that removes both the Substructure-Contextualized Hyperedge Aggregator (SCHA) and the Adaptive Tri-Directional Contrastive Learning (A3CL) module, using max-min pooling as in prior work;
- **w/o SCHA:** A variant model that retains A3CL but discards SCHA;
- **w/o A3CL:** A variant model which applies SCHA but omits A3CL.

Across all datasets and test conditions, S<sup>3</sup>Hyper consistently achieves the highest AUROC and AP, confirming that both components contribute to performance gains. Compared to the baseline, S<sup>3</sup>Hyper improves average AUROC by 12.02%, 18.84%, 12.03%, and 12.68% on Citeseer, Cora, NDC-class, and DBLP, respectively. Notably, it demonstrates the largest gains on the challenging CNS setting, with up to +29.06% AUROC improvement (Cora), indicating that S<sup>3</sup>Hyper can effectively distinguish difficult negative samples that mimic plausible structures. The individual module contributions are also evident: w/o SCHA significantly underperforms the full model, confirming the importance of substructure-guided aggregation in overcoming the flat set assumption; w/o A3CL also lags behind, highlighting the role of the tri-directional contrastive objectives in alleviating the scarcity of positive samples and enhancing affiliation modeling. S<sup>3</sup>Hyper outperforms both variants, demonstrating the strong synergy between its two design pillars. The integration of SCHA and A3CL equips the model with complementary structural and semantic cues, leading to a deeper understanding of hyperedge formation and robust generalization across complex prediction scenarios.

### 4.4 Study on Aggregation Strategy

To assess the effectiveness of the proposed Substructure-Contextualized Hyperedge Aggregator (SCHA), we conduct a comparative study against three widely-used aggregation strategies in hyperedge prediction: mean pooling, max-min pooling, and attention. In this experiment, we substitute the SCHA module in S<sup>3</sup>Hyper with each baseline aggregator while keeping all other components (e.g., the encoder, A3CL module, and hyperparameters) unchanged to ensure a fair comparison.

As shown in Figure 2, SCHA attains the highest AU-

ROC on all four datasets. On Citeseer, SCHA reaches 0.86, slightly surpassing attention (0.85), while on NDC-class and DBLP it yields 0.86 and 0.82, exceeding the second-best methods by 8% and 3%, respectively. These results demonstrate the advantage of SCHA in modeling hyperedge structures. Unlike mean pooling, max-min pooling, and attention, which treat hyperedges as flat, unordered node sets and aggregate node features indiscriminately, SCHA explicitly discovers and integrates smaller collaborative substructures within each candidate hyperedge. By using a substructure discovery and part-to-whole aggregation mechanism, it evaluates the compositional validity of hyperedge formation, enabling more discriminative judgments about whether nodes form a coherent group and thus improving prediction performance. Overall, our findings empirically confirm the effectiveness of SCHA and reveal the performance bottlenecks of flat-set aggregation, underscoring the importance of structure-aware aggregation for accurate and generalizable hyperedge prediction.

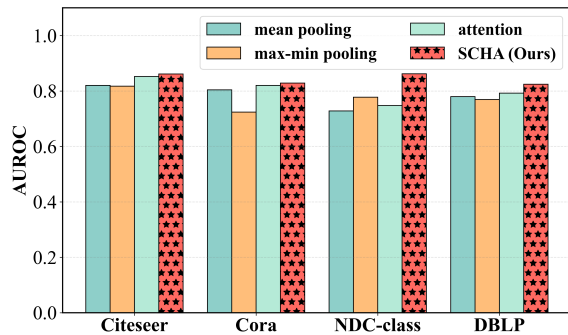


Figure 2: Performance comparison of different aggregation strategies. **SCHA** refers to our proposed Substructure-Contextualized Hyperedge Aggregator.

#### 4.5 Hyperparameter Sensitivity Analysis

We investigate the model’s sensitivity to the internal weighting of the three components within the A3CL loss: the node-level contrastive loss weight  $\lambda_n$ , the hyperedge-level loss weight  $\lambda_{he}$ , and the node-hyperedge alignment loss weight  $\lambda_{n-he}$ . In our study, each of these parameters is independently varied in the range  $[0.1, 1.0]$ , while the other two are held fixed. The AUROC results on the NDC-class and Cora datasets are shown in Figure 3. The curves remain nearly flat across all tested values, indicating that the model is highly stable and robust to these hyperparameter variations. This stability confirms that the three complementary contrastive objectives contribute meaningfully and harmoniously to representation learning, without requiring delicate tuning of their relative weights.

### 5 Related Work

Recent hyperedge prediction methods commonly adopt either pooling strategies (e.g., NHP (Yadati et al. 2020), AHP (Hwang et al. 2022)) or attention-based mechanisms (e.g., HyperSAGNN (Zhang, Zou, and Ma 2020),

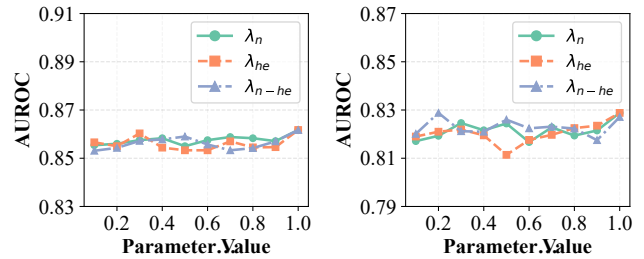


Figure 3: Impact of  $\lambda_n$ ,  $\lambda_{he}$ ,  $\lambda_{n-he}$ .

CASH (Ko, Tong, and Kim 2025)) to aggregate node features within a candidate hyperedge. A key limitation shared by these approaches is the assumption that nodes form a flat, unordered set, neglecting the internal substructure, such as frequently co-occurring node subsets, that may carry important signals about group formation. While incorporating substructures (e.g., motifs) is well-established in network science (Milo et al. 2002) and has gained traction in hypergraph representation learning (Benson, Gleich, and Leskovec 2016; Lee and Shin 2021), integrating such compositional cues into the hyperedge aggregation process remains largely unexplored in the context of hyperedge prediction. In parallel, contrastive self-supervised learning has been leveraged to improve representation quality under limited supervision, e.g., CHGNN (Song et al. 2024), HyGCL-AdT (Qian et al. 2024). These methods typically enhance node or hyperedge embeddings by enforcing consistency across augmented views of homogeneous entities (e.g., node-node or hyperedge-hyperedge pairs). However, they often overlook the cross-level semantics inherent in hypergraphs, particularly the “membership” relation that binds nodes to hyperedges, resulting in a missed opportunity to model the affiliation logic crucial for hyperedge inference.

## 6 Conclusion

This paper presents S<sup>3</sup>Hyper, a self-supervised hypergraph learning framework that integrates substructure-aware aggregation and cross-level contrastive learning for hyperedge prediction. S<sup>3</sup>Hyper provides an effective framework that touches on two key challenges in this task: the flat set assumption and the scarcity of positive samples. By modeling internal compositional signals and aligning node-hyperedge relationships through tri-directional supervision, our approach improves the quality of learned representations and enhances prediction accuracy. Extensive experiments validate the effectiveness of the proposed framework across multiple datasets. Future work may explore extending S<sup>3</sup>Hyper to dynamic or evolving hypergraphs, as well as investigating more effective strategies for sub-hyperedge selection in noisy settings. In addition, it is desirable to adapt the proposed framework to more challenging scenarios involving *heterogeneous* node types and/or *heterophilic* interaction patterns, posing new challenges for hyperedge prediction under structurally diverse conditions.

## Acknowledgements

Dr. Yuting Chen acknowledged a Project supported by Open Research Fund of Zhejiang Key Laboratory of Intelligent Education Technology and Application (No. 2025ZNJYKF002). This work was supported by the National Natural Science Foundation of China (No. 62536006, No. 62172370, No. 62576371, No. U23A20388, No. 62320106007).

## References

- Benson, A. R.; Abebe, R.; Schaub, M. T.; Jadbabaie, A.; and Kleinberg, J. 2018. Simplicial closure and higher-order link prediction. *Proceedings of the National Academy of Sciences*, E11221–E11230.
- Benson, A. R.; Gleich, D. F.; and Leskovec, J. 2016. Higher-order organization of complex networks. *Science*, 353(6295): 163–166.
- Chen, C.; and Liu, Y.-Y. 2023. A survey on hyperlink prediction. *IEEE Transactions on Neural Networks and Learning Systems*, 35(11): 15034–15050.
- Dong, Y.; Sawin, W.; and Bengio, Y. 2020. HNHN: Hypergraph networks with hyperedge neurons. *arXiv preprint arXiv:2006.12278*.
- Hwang, H.; Lee, S.; Park, C.; and Shin, K. 2022. AHP: Learning to negative sample for hyperedge prediction. In *SIGIR*, 2237–2242.
- Kim, S.; Lee, S. Y.; Gao, Y.; Antelmi, A.; Polato, M.; and Shin, K. 2024. A survey on hypergraph neural networks: An in-depth and step-by-step guide. In *KDD*, 6534–6544.
- Ko, Y.; Tong, H.; and Kim, S.-W. 2025. Enhancing hyperedge prediction with context-aware self-supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, 37(4): 1772–1784.
- Lee, G.; and Shin, K. 2021. THyMe+: Temporal hypergraph motifs and fast algorithms for exact counting. In *ICDM*, 310–319.
- Li, M.; Cheng, Y.; Bai, L.; Cao, F.; Lv, K.; Liang, J.; and Lio, P. 2025a. EduLLM: Leveraging large language models and framelet-based signed hypergraph neural networks for student performance prediction. In *ICML*, 34119–3414.
- Li, M.; Gu, Y.; Wang, Y.; Fang, Y.; Bai, L.; Zhuang, X.; and Lio, P. 2025b. When hypergraph meets heterophily: New benchmark datasets and baseline. In *AAAI*, 18377–18384.
- Millán, A. P.; Sun, H.; Giambagli, L.; Muolo, R.; Carletti, T.; Torres, J. J.; Radicchi, F.; Kurths, J.; and Bianconi, G. 2025. Topology shapes dynamics of higher-order networks. *Nature Physics*, 21: 353–361.
- Milo, R.; Shen-Orr, S.; Itzkovitz, S.; Kashtan, N.; Chklovskii, D.; and Alon, U. 2002. Network motifs: simple building blocks of complex networks. *Science*, 298(5594): 824–827.
- Qian, Y.; Ma, T.; Zhang, C.; and Ye, Y. 2024. Dual-level Hypergraph Contrastive Learning with Adaptive Temperature Enhancement. In *WWW*, 859–862.
- Song, Y.; Gu, Y.; Li, T.; Qi, J.; Liu, Z.; Jensen, C. S.; and Yu, G. 2024. CHGNN: A semi-supervised contrastive hypergraph learning network. *IEEE Transactions on Knowledge and Data Engineering*, 36(9): 4515–4530.
- Wang, Y.; and Kleinberg, J. 2024. From graphs to hypergraphs: Hypergraph projection and its reconstruction. In *ICLR*.
- Yadati, N.; Nimishakavi, M.; Yadav, P.; Nitin, V.; Louis, A.; and Talukdar, P. 2019. HyperGCN: A new method for training graph convolutional networks on hypergraphs. *NeurIPS*, 1511–1522.
- Yadati, N.; Nitin, V.; Nimishakavi, M.; Yadav, P.; Louis, A.; and Talukdar, P. 2020. NHP: Neural hypergraph link prediction. In *CIKM*, 1705–1714.
- Zhang, Q.; Yang, P.; Yu, J.; Wang, H.; He, X.; Yiu, S.-M.; and Yin, H. 2025. A survey on point-of-interest recommendation: Models, architectures, and security. *IEEE Transactions on Knowledge and Data Engineering*, 37(6): 3153–3172.
- Zhang, R.; Zou, Y.; and Ma, J. 2020. Hyper-SAGNN: a self-attention based graph neural network for hypergraphs. In *ICLR*.