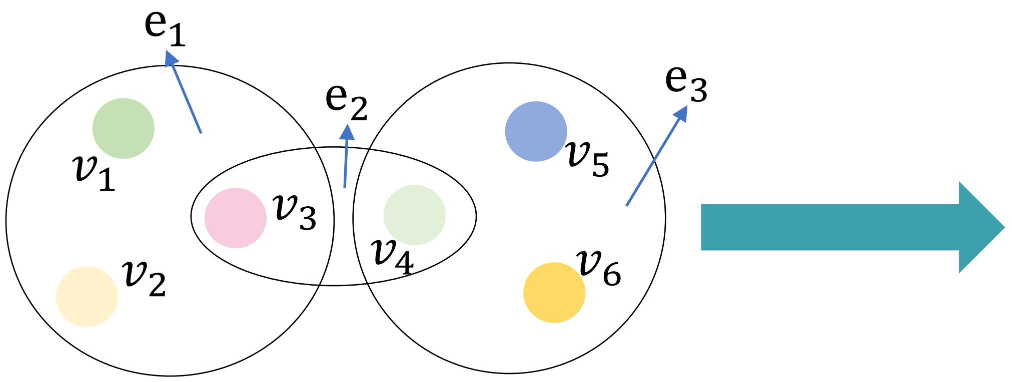
Hypergraph Transformer for Semi-Supervised Classification

Zexi Liu[1] Bohan Tang[2] Ziyuan Ye[3]
Xiaowen Dong[2] Siheng Chen[1,4] Yanfeng Wang[4,1]

1 Shanghai Jiao Tong University
2 University of Oxford
3 The Hong Kong Polytechnic University 香港理工大學
4 上海人工智能实验室 Shanghai Artificial Intelligence Laboratory

Hypergraph

Math formulation



$$\mathbf{H} \quad \begin{array}{c|c} & \begin{matrix} e_1 & e_2 & e_3 \end{matrix} \\ \begin{matrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{matrix} & \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

Hypergraph $\mathcal{H} = \{\mathcal{V}, \mathcal{E}, \mathbf{H}\}$

Node Set $\mathcal{V} = \{v_1, v_2, \ldots, v_n\}$

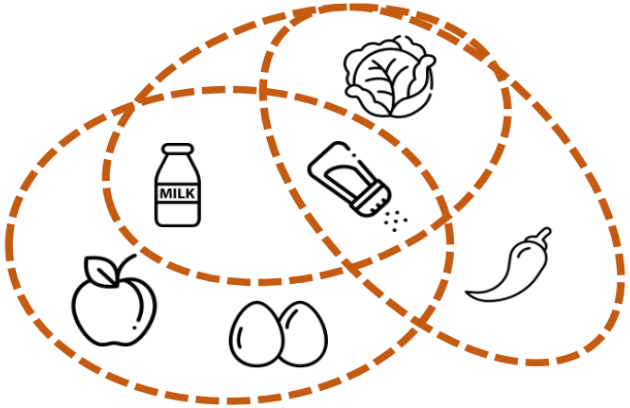
Edge Set $\mathcal{E} = \{e_1, e_2, \ldots, e_m\}$

Incidence Matrix $\mathbf{H} \in \{0,1\}^{n \times m}$

Hypergraph

Examples

Co-authorships



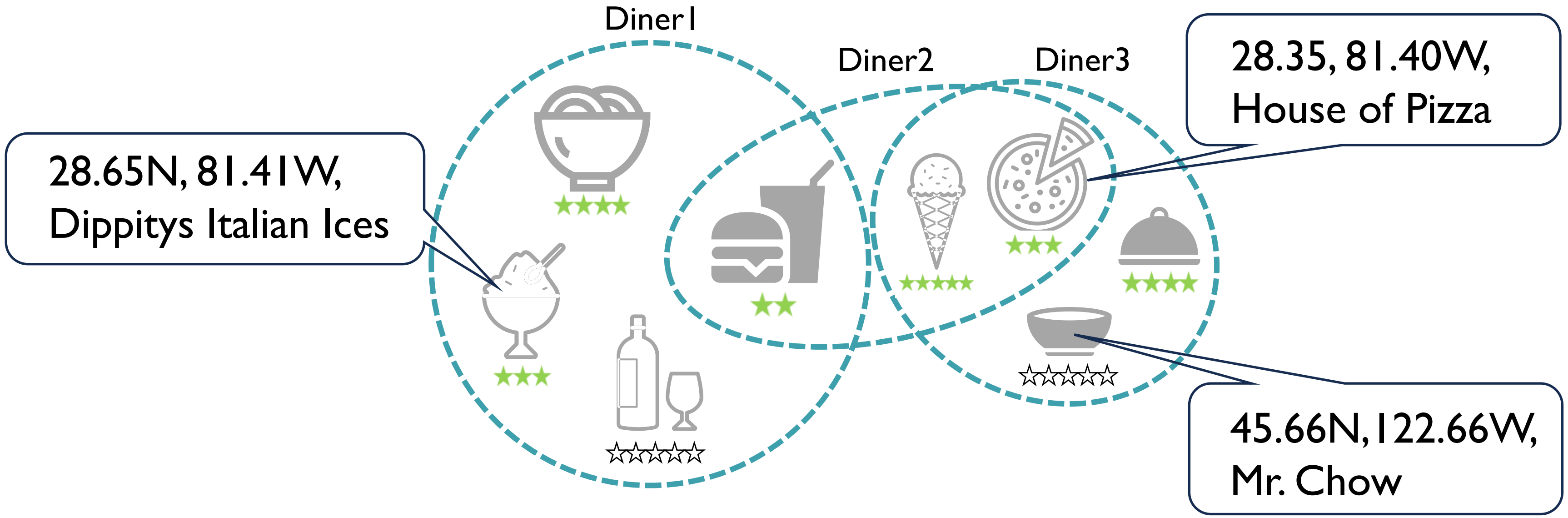
Cooking Recipe

A paper is co-authored by multiple authors

A recipe is composed of multiple ingredients

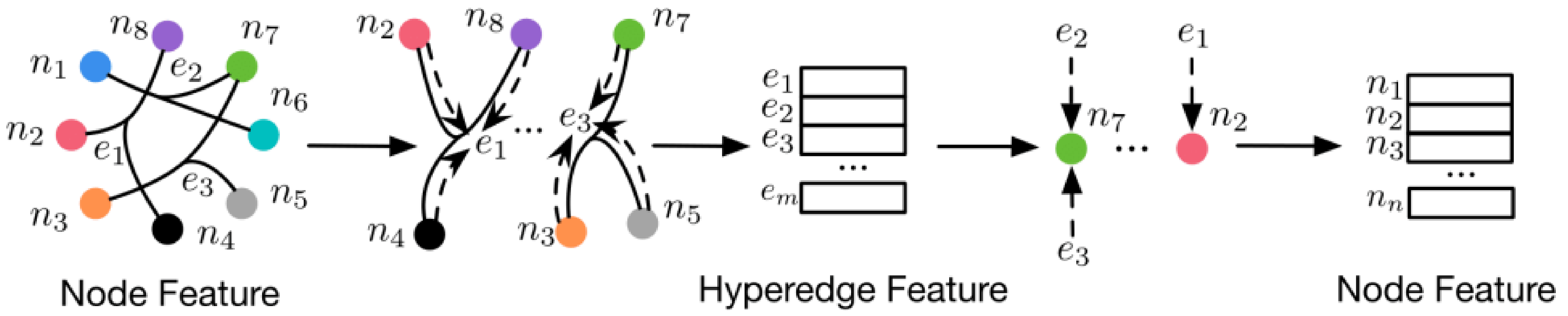
Semi-Supervised Node Classification on Hypergraphs

Given node features $\mathbf{X}_{\mathcal{V}} \in \mathbb{R}^{n \times d}$, labels $\mathbf{Y}_{lab} = \{\mathbf{y}_v\}_{v \in \mathcal{V}_{lab}}$ and hypergraph structure $\mathbf{H}$, we aim to classify nodes in $\mathcal{V} \setminus \mathcal{V}_{lab}$.



Hypergraph Neural Networks

Two-step message passing

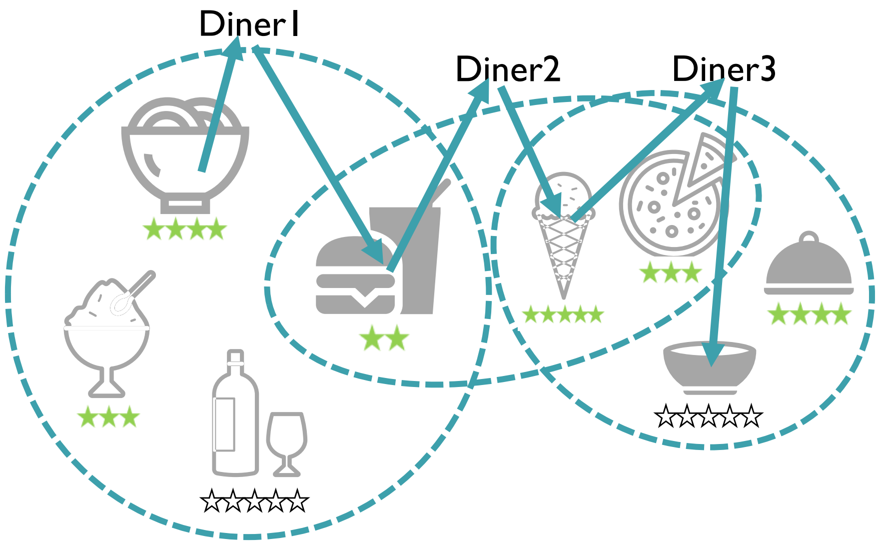


- Node features are sent to corresponding hyperedges to learn hyperedge embeddings
- Learned hyperedge embeddings are sent back to nodes to learn node embeddings

[1] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, and Yue Gao, "Hypergraph neural networks," AAAI 2019.

Hypergraph Neural Networks

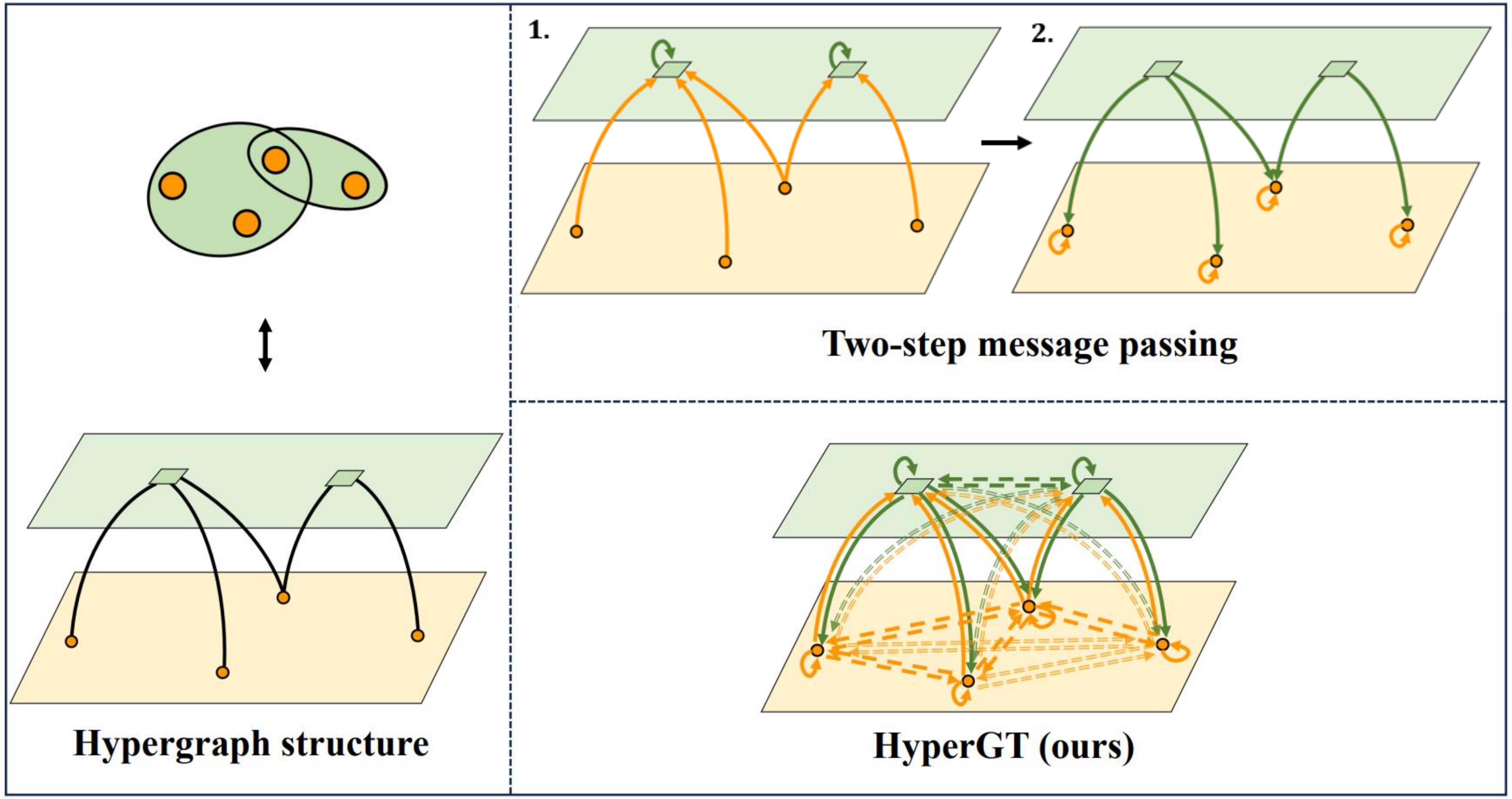
Two-step message passing



not efficiently exploiting the global information present in hypergraph-structured data!

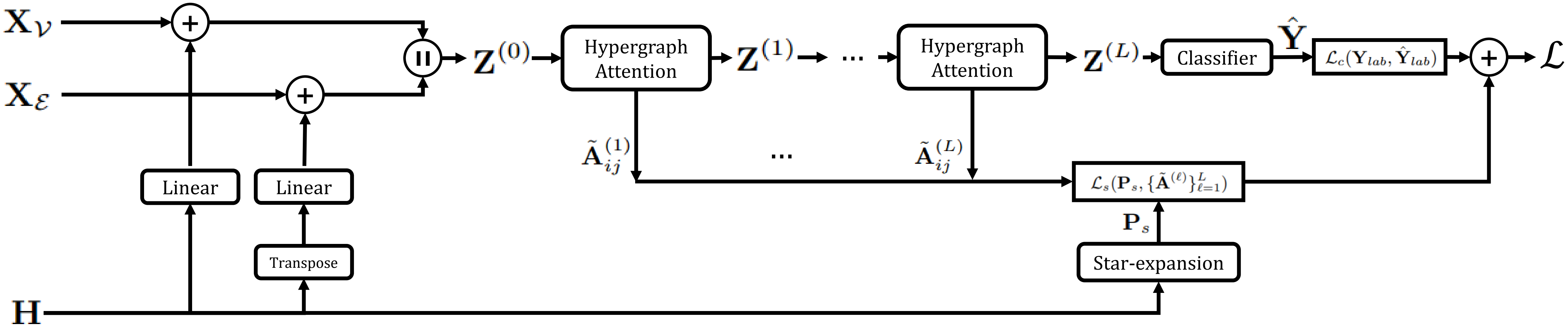
HyperGT: Interactions among all nodes and hyperedges

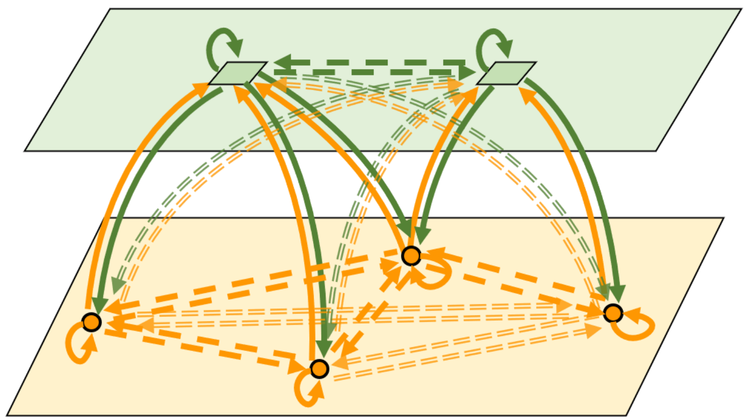
Capture both global and local interactions among all nodes and hyperedges in one single step



HyperGT's Overall Architecture

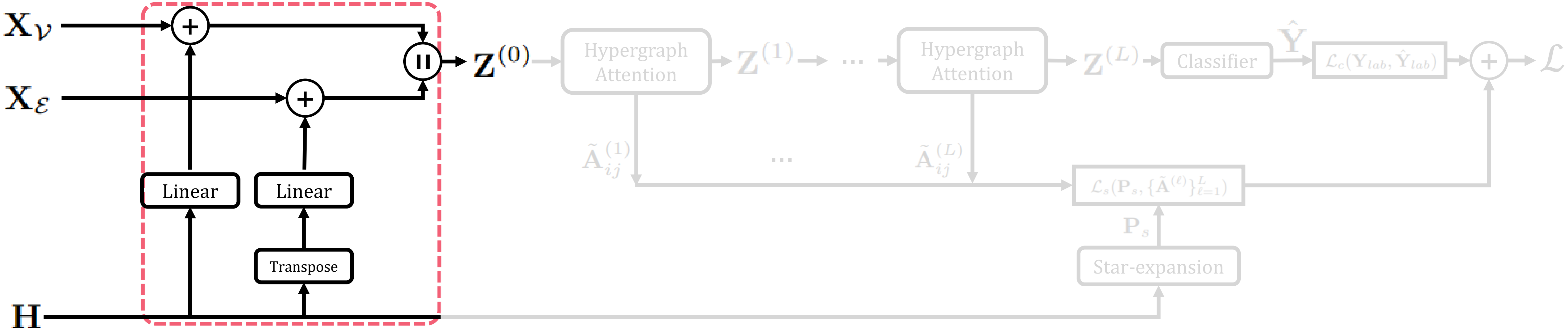
Transformer-based architecture to efficiently incorporates both global and local interactions in hypergraph

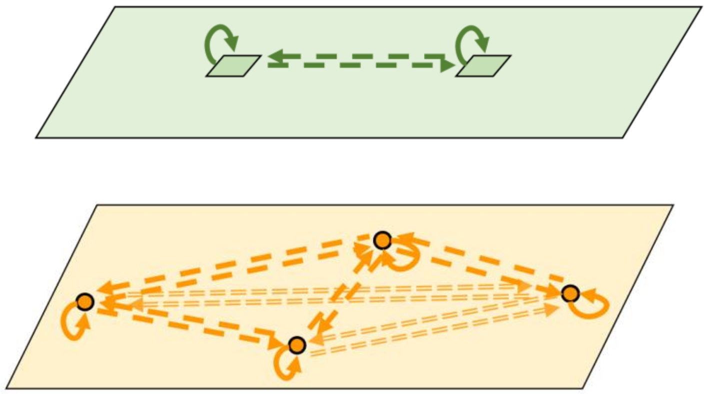




Hypergraph Incidence Matrix Based Positional Encoding

Valuable Structural insights:
offering local node-node and hyperedge-hyperedge interactions





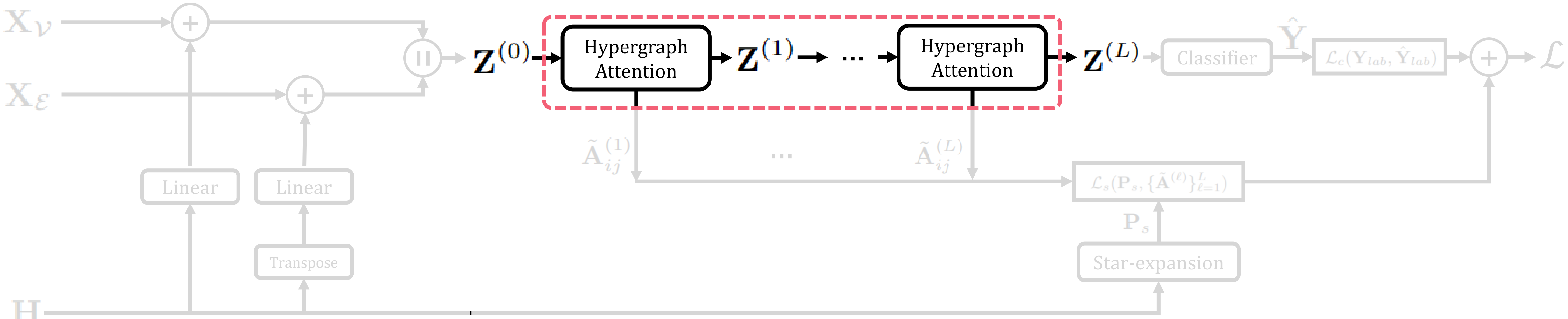
Input features

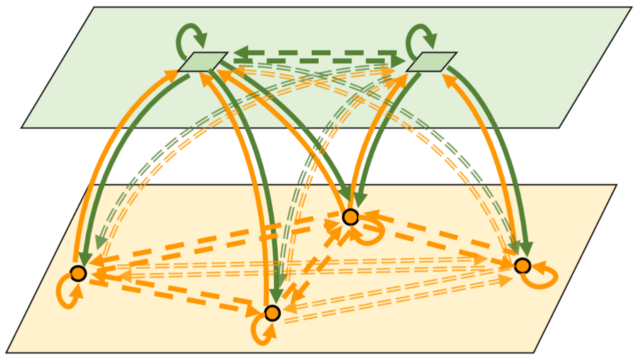
$$\mathbf{Z}^{(0)} = \mathbf{X} + \mathbf{P} = \begin{bmatrix} \mathbf{X}_{\mathcal{V}} + \mathbf{P}_{\mathcal{V}} \\ \mathbf{X}_{\mathcal{E}} + \mathbf{P}_{\mathcal{E}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{\mathcal{V}} + \mathbf{H}\mathbf{W}_{\mathcal{V}} \\ \mathbf{X}_{\mathcal{E}} + \mathbf{H}^{\top}\mathbf{W}_{\mathcal{E}} \end{bmatrix} \in \mathbb{R}^{(n+m)\times d}.$$

Theorem 1. Let $\mathbf{p}_u$, $\mathbf{p}_v$ be the positional encoding of nodes u and v, respectively. Then, $\|\mathbf{p}_u - \mathbf{p}_v\|_2 \leq C\sqrt{N_e}$, where C is a constant and N_e is the number of hyperedges only with either node u or v.

Hypergraph Attention

Efficient information propagation:
pairwise global interactions between all nodes and hyperedges in only one single step





Attention for each instance pair

$$\tilde{\mathbf{A}}_{ij}^{(\ell)} = \frac{\exp((\mathbf{z}_i^{(\ell)}\mathbf{W}_{\mathbf{Q}}^{(\ell)})(\mathbf{z}_j^{(\ell)}\mathbf{W}_{\mathbf{K}}^{(\ell)})^\top)}{\sum_{k=1}^{n+m}\exp((\mathbf{z}_i^{(\ell)}\mathbf{W}_{\mathbf{Q}}^{(\ell)})(\mathbf{z}_k^{(\ell)}\mathbf{W}_{\mathbf{K}}^{(\ell)})^\top)}$$

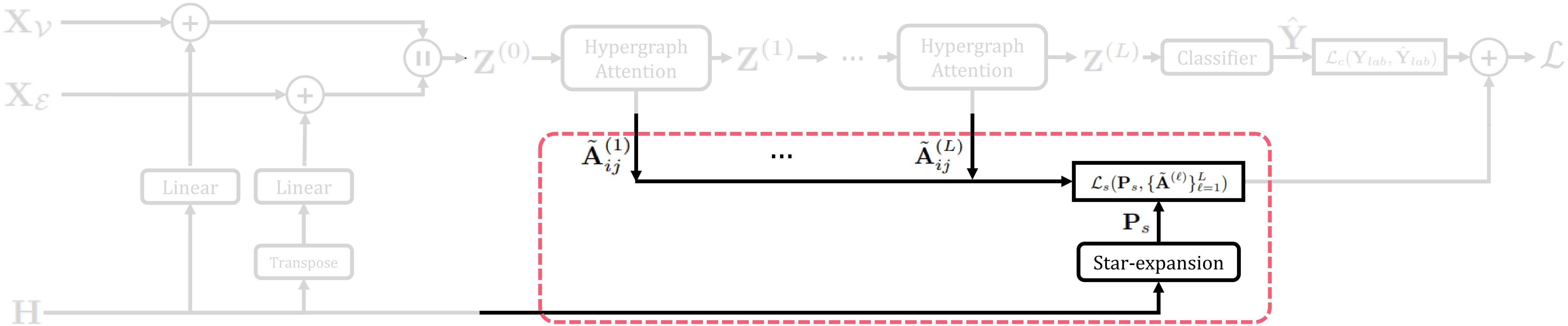
Updated representaion

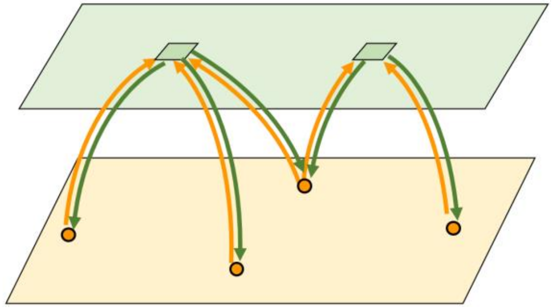
$$\mathbf{z}_i^{(\ell+1)} = \sum_{j=1}^{n+m} \tilde{\mathbf{A}}_{ij}^{(\ell)} \cdot (\mathbf{z}_j^{(\ell)}\mathbf{W}_{\mathbf{V}}^{(\ell)})$$

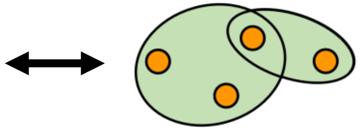
Hypergraph Structure Regularization

Connectivity Loseless Supervision:

utilize the node-hyperedge connection prior to guide the training of the attention matrix







Star expansion

$$\mathbf{A}_s = \begin{bmatrix} 0_{|\mathcal{V}|} & \mathbf{H} \\ \mathbf{H}^{\top} & 0_{|\mathcal{E}|} \end{bmatrix}$$

Probabilistic transition matrix

$$\mathbf{P}_s = \mathbf{D}_s^{-1}\mathbf{A}_s$$

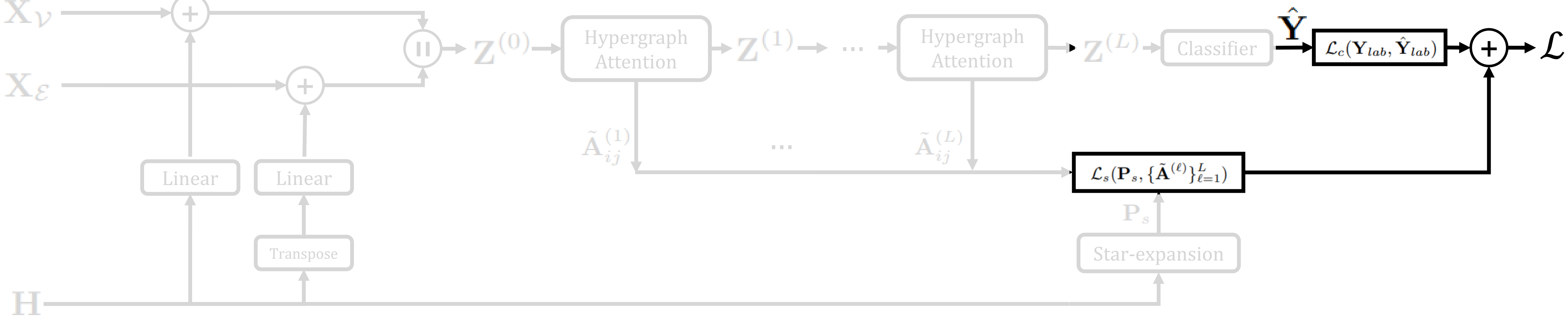
Structure regularization loss

$$\mathcal{L}_s(\mathbf{P}_s, \{\tilde{\mathbf{A}}^{(\ell)}\}_{\ell=1}^{L}) = -\frac{1}{(n+m)L}\sum_{\ell=1}^{L}\sum_{i,j=1}^{n+m}(\mathbf{P}_s)_{i,j}\log\tilde{\mathbf{A}}_{ij}^{(\ell)}$$

$$= -\frac{1}{(n+m)L}\sum_{l=1}^{L}\sum_{(i,j)\in\mathcal{E}_s}\frac{1}{d_i}\log\tilde{\mathbf{A}}_{ij}^{(\ell)},$$

HyperGT's Training Strategy

Balance node labels cross-entropy loss and hypergraph structure loss

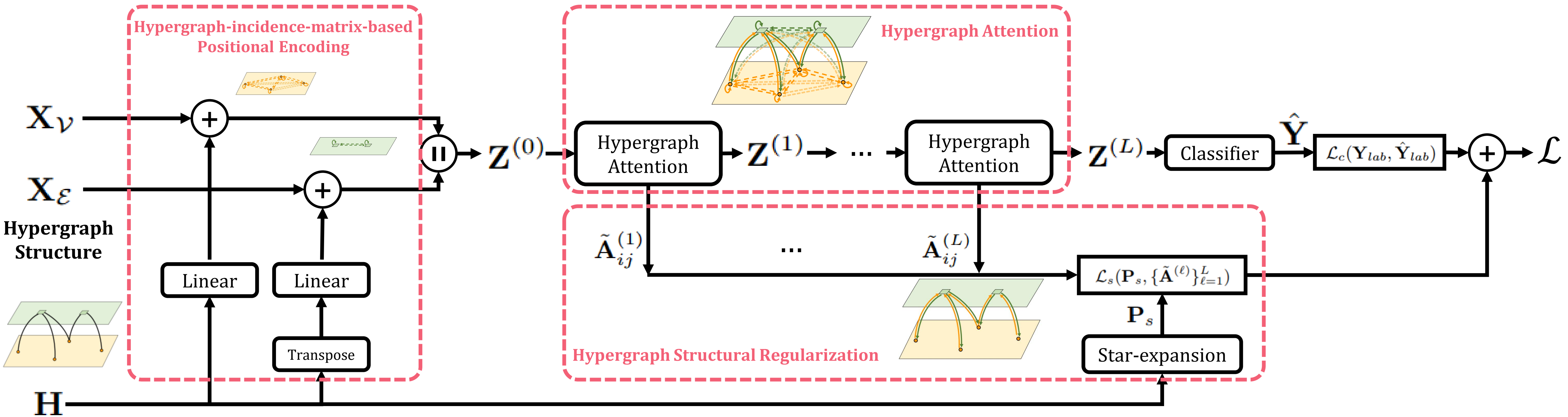


Supervised classification loss $$\mathcal{L}_c(\mathbf{Y}_{lab}, \hat{\mathbf{Y}}_{lab}) = -\frac{1}{|\mathcal{V}_{lab}|} \sum_{u \in \mathcal{V}_{lab}} \mathbf{y}_u \log(\hat{\mathbf{y}}_u^\top)$$

Structure regularization loss $$\mathcal{L}_s(\mathbf{P}_s, \{\tilde{\mathbf{A}}^{(\ell)}\}_{\ell=1}^{L}) = -\frac{1}{(n+m)L} \sum_{l=1}^{L} \sum_{(i,j) \in \mathcal{E}_s} \frac{1}{d_i} \log \tilde{\mathbf{A}}_{ij}^{(\ell)},$$

Final loss function $$\mathcal{L} = \mathcal{L}_c + \lambda \mathcal{L}_s$$

HyperGT's Overall Architecture



HyperGT effectively incorporates global interactions while preserving local connectivity patterns!

Experimental Results

Datasets

	Congress	Senate	Walmart	House
$\lvert\mathcal{V}\rvert$	1718	282	88860	1290
$\lvert\mathcal{E}\rvert$	83105	315	69906	341
# features	100	100	100	100
# class	2	2	11	2

Baselines

Nine representative HyperGNNs: HGNN (AAAI 2019), HyperGCN (NeurIPS 2019), HNHN(ICML 2020 workshop), HCHA (Pattern Recognition 2021),HyperND (2021), UniGNN (IJCAI 2021), AllDeepSets (ICLR 2022), AllSetTransformer (ICLR 2022), EDHNN (ICLR 2023).

Metric

$$Accuracy = \frac{N_{correct\ predictions}}{N_{total\ predictions}} * 100\%$$

Experimental Results

Test ACC

Superior classification accuracy across all datasets compared to previous hgnns

	Congress	Senate	Walmart	House
HGNN	91.26 ± 1.15	48.59 ± 4.52	62.00 ± 0.24	61.39 ± 2.96
HCHA	90.43 ± 1.20	48.62 ± 4.41	62.35 ± 0.26	61.36 ± 2.53
HNHN	53.35 ± 1.45	50.93 ± 6.33	47.18 ± 0.35	67.80 ± 2.59
HyperGCN	55.12 ± 1.96	42.45 ± 3.67	44.74 ± 2.81	48.32 ± 2.93
UniGCNII	94.81 ± 0.81	49.30 ± 4.25	54.45 ± 0.37	67.25 ± 2.57
HyperND	74.63 ± 3.62	52.82 ± 3.20	38.10 ± 3.86	51.70 ± 3.37
AllDeepSets	91.80 ± 1.53	48.17 ± 5.67	64.55 ± 0.33	67.82 ± 2.40
AllSetTransformer	92.16 ± 1.05	51.83 ± 5.22	65.46 ± 0.25	69.33 ± 2.20
ED-HNN	95.00 ± 0.99	64.79 ± 5.14	66.91 ± 0.41	72.45 ± 2.28
HyperGT(Ours)	**95.23 ± 0.73**	**65.49 ± 5.11**	**69.83 ± 0.39**	**74.55 ± 1.99**

Experimental Results

Ablation study

All of the components are helpful for modelling hypergraph data

node PE	hyperedge PE	structure regularization	ACC
-	-	-	45.67
✓	-	-	66.51
✓	✓	-	67.63
✓	✓	✓	69.83

Efficiency (ms/run)

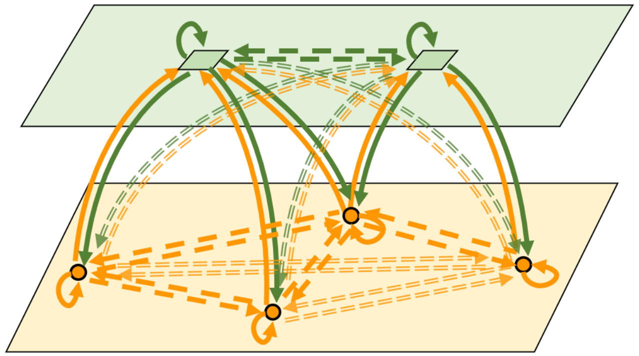
Maintains competitive inference speeds aligning with its theoretically low linear complexity [1]

HGNN	HCHA	UniGCNII	AllDeepset	AllSetTransformer	HyperGT
20.301	20.942	25.015	32.603	62.788	24.882

[1] Qitian Wu et al., "Nodeformer: A scalable graph structure learning transformer for node classification," NeurIPS 2022.

Summary

- **HyperGT efficiently incorporates both global and local interactions in hypergraph**
  - Hypergraph Attention: efficient propagate signals between all nodes and hyperedges
  - Hypergraph Incidence Matrix Based PE: offer local node-node & hyperedge-hyperedge correlations
  - Hypergraph Structure Regularization: capture connectivities between nodes and hyperedges



SHANGHAI JIAO TONG UNIVERSITY

Many Thanks!

Q&A

liuzexi@sjtu.edu.cn

https://github.com/zeroxleo/HyperGT