HyperMixup: Hypergraph-Augmented with Higher-order Information Mixup

Kaixuan Yao[1], Zhuo Li[1], Jianqing Liang[1*], Jiye Liang[1], Ming Li[2], Feilong Cao[3]

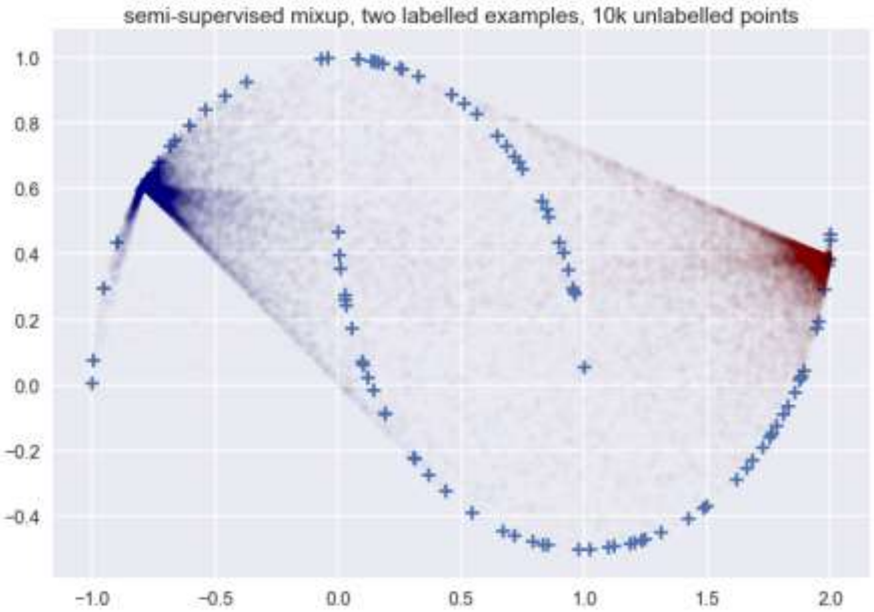
[1] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, School of Computer and Information Technology, Shanxi University, Taiyuan, China.

[2] Zhejiang Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University, Jinhua, China

[3] School of Mathematics, Institute of Mathematics and Cross-disciplinary Science, Zhejiang Normal University, China

The Problem: Hypergraph Learning is Data-Hungry

- **The Problem: Beyond Pairwise Mixup.**
  - **Hypergraphs:** Natural model for multi-way interactions.
  - **Challenge:** Standard Mixup violates group semantics.
  - **Need:** Augmentation that respects higher-order topology.



The HyperMixup Framework

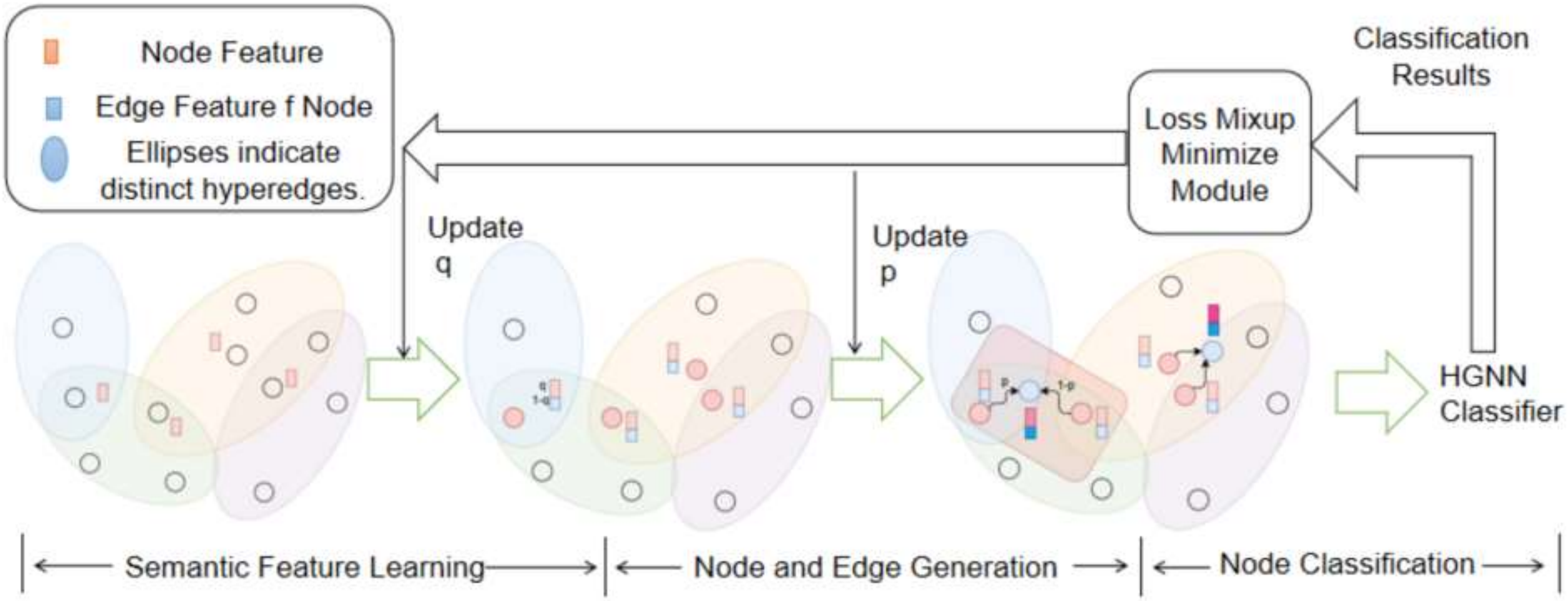


Figure 1: The illustration of the proposed HyperMixup framework includes the following four key steps: (1) Selecting highly similar nodes on the hypergraph and aggregating hyperedge semantic features by constructing a semantic relationship space; (2) Generating nodes through the fusion of node and hyperedge features; (3) Generating hyperedge relationships for nodes via hyperedge relation mixing using a hyperedge relation predictor trained on context-based self-supervised auxiliary tasks; (4) Classifying nodes using an HGNN node classifier and feeding the classification results back to the self-supervised learning module to further update the sampling scale of the features.

Method HyperMixup: Semantic-Aware Mixing

The CCA Pipeline: A Three-Step Process

Step 1: Structure-Aware Pairing

Select nodes using **dual similarity**:

$$s(v_i, v_j) = \underbrace{\cos(x_i, x_j)}_{\text{Feature}} + \mu \cdot \underbrace{\cos(x_{e_i}, x_{e_j})}_{\text{Hyperedge}}$$

Step 3: Feature Aggregation & Output

- Mix nodes: $\mathbf{x}' = \lambda \mathbf{x}_i + (1 - \lambda)\mathbf{x}_j$
- Enhance with hyperedge context.

Theoretical Foundation: Why It Works?

- **HyperMixup induces hypergraph-specific regularization.**
- **Theorem 1 (Regularization):**

$$\mathcal{L}_{mix} \approx \mathcal{L}_{std} + \underbrace{\mathcal{R}_1}_{\text{Gradient Alignment}} + \underbrace{\mathcal{R}_2}_{\text{Smoothness}} + \underbrace{\mathcal{R}_3}_{\text{Curvature}}$$

  - $\mathcal{R}_2 \propto \nabla f^T \Sigma_e \nabla f$: Penalizes sharp changes along hyperedge covariance Σ_e.
- **Provides robustness guarantees.**

Experiments

■Main Results

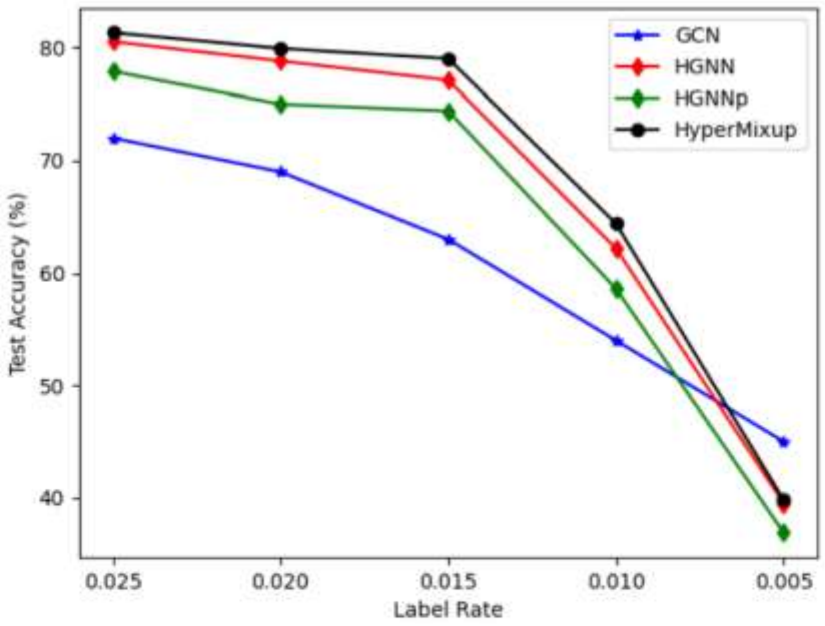
Table 2: Comparison of different methods: node classification Accuracy. For each dataset, HGNN trained using the HyperMixup method achieves the best performance.The best are highlighted in bold.

Method	Cora	Pubmed	CiteSeer	ModelNet40	NTU2012
GCN	81.50%	79.00%	70.30%	94.85%	80.43%
GAT	83.0%	79.00%	**72.5%**	95.75%	80.16%
GraphSAGE	83.2%	-%	-%	94.73%	80.7%
GraphConv	82.19%	-%	70.35%	95.66%	80.96%
HyperGCN	64.11%	73.09%	64.11%	95.46%	81.77%
Hyper-Atten	82.61%	79.00%	70.88%	96.11%	81.50%
HGNN	82.09%	78.60%	71.60%	96.80%	83.11%
HGNN+	76.71%	75.08%	66.43%	96.92%	84.18%
HyperMixup	**83.60%**	**79.50%**	72.20%	**97.04%**	**85.50%**

Table 3: Comparison with graph-based augmentations (Accuracy %)

Backbone	Method	Cora	PubMed	CiteSeer
GNN	Mixup	81.84±0.94	79.16±0.49	72.20±0.95
	GraphMixup	82.16±0.74	78.82±0.52	72.13±0.86
HGNN	Mixup	81.09±0.56	78.02±0.36	70.40±0.86
	GraphMixup	82.16±0.74	78.82±0.52	72.13±0.86
HGNN+	Mixup	76.70±0.86	74.90±0.14	66.20±0.84
HGNN	HyperMixup (Ours)	83.62±0.76	79.50±0.88	72.60±0.68
HGNN+	HyperMixup (Ours)	**84.02±0.52**	**80.04±0.32**	**73.02±0.82**

■Robustness Analysis



Conclusion

- **Core Contribution**: We introduced **HyperMixup**, the first hypergraph-aware data augmentation framework that **unifies feature mixing with higher-order topological constraints.**
- **Key Innovation**:
  - **Structure-Guided Mixing**: A principled pipeline from node pairing to topology reconstruction.
  - **Theoretical Depth**: Provides **regularization guarantees** and **robustness bounds**.
- **Proven Impact**: Delivers **consistent, state-of-the-art performance** across diverse tasks, especially powerful in **low-label and noisy regimes.**

Thanks