Heterogeneous Graph Structure Learning for Graph Neural Networks

Jianan Zhao,[1,2*] Xiao Wang,[1*] Chuan Shi,[1†] Binbin Hu,[3] Guojie Song,[4] Yanfang Ye[2]

[1] School of CS, Beijing University of Posts and Telecommunications, Beijing, China
[2] Department of CDS, Case Western Reserve University, OH, USA
[3] Ant Group
[4] Key Laboratory of Machine Perception, Ministry of Education, Peking University

AAAI 2021

CONTENTS

AAAI 2021

1 Introduction

2 Proposed Method

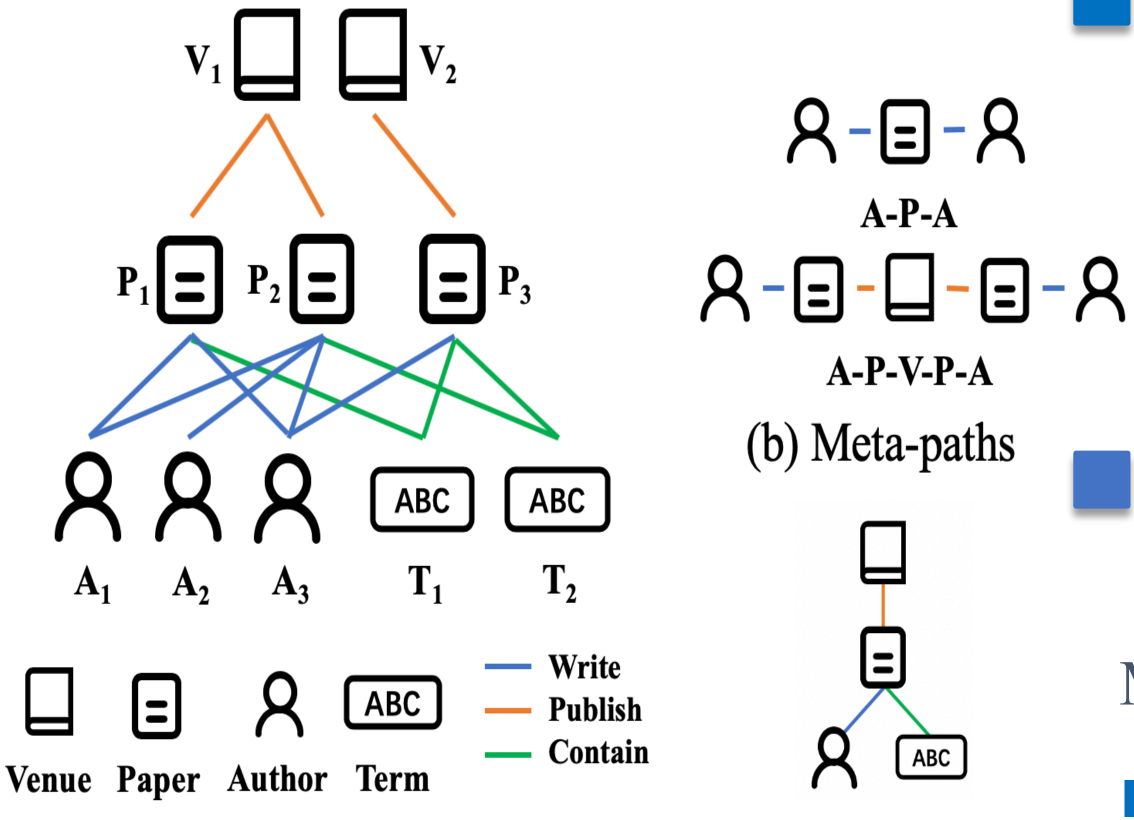
3 Experiments

4 Conclusions

AAAI 2021

CONTENTS

1. Introduction
2. Proposed Method
3. Experiments
4. Conclusions



(a) An example of HIN
(b) Meta-paths
(c) Network schema

Heterogeneous Graphs

- Include multiple types of nodes or links
- Contain rich semantics
- Flexible to characterize heterogeneous data

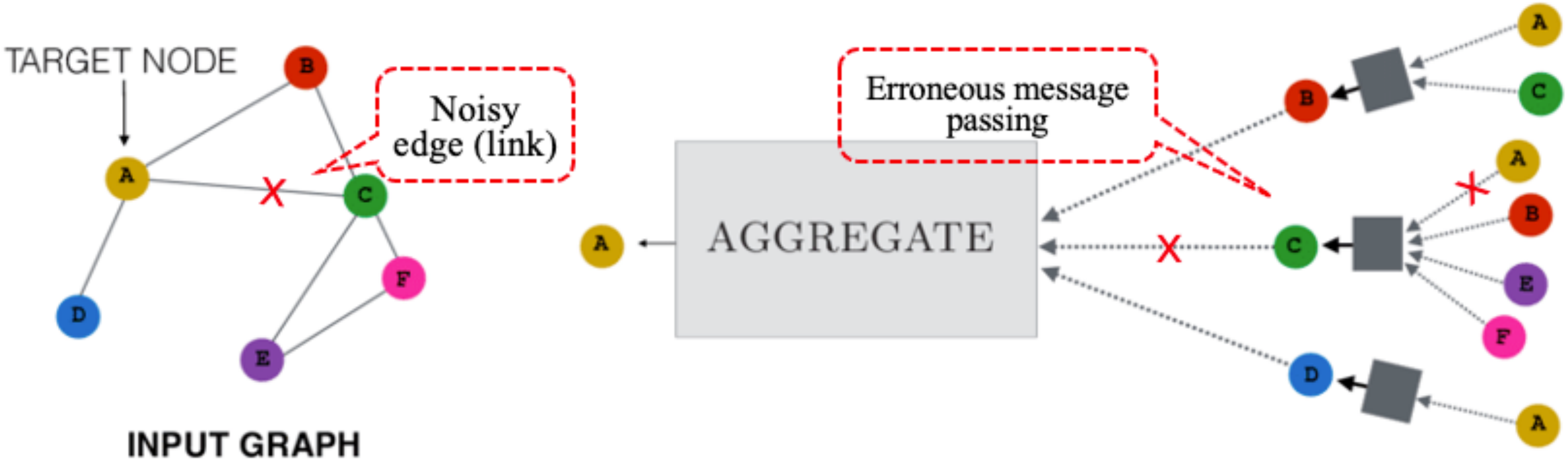
Heterogeneous Graph Neural Network (HGNNs)

Message passing on ...

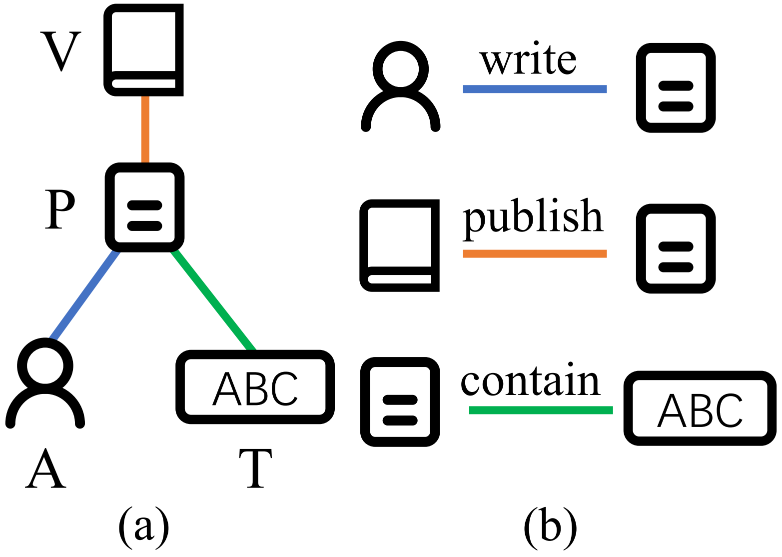
- Original graph structure (HetGNN, HGT, NSHE...)
- Meta-path based graph structure (HAN, MAGNN...)

Limitations of Heterogeneous GNNs (HGNNs)

- Most HGNNs perform message passing on the raw heterogeneous graph structure (or MP-based structure derived from it).
- Real-world heterogeneous graph structures are inevitably incomplete and noisy.



Solution: **Learn the heterogeneous graph structure** instead of rely on the raw graph structure!



Network schema (a) and relations (b) of a heterogeneous graph

Challenges

- C1: Heterogeneity in heterogeneous graphs
- C2: Complex interactions in heterogeneous graphs
  - Feature and topology interaction
  - High-order (meta-path) topology interactions

Solution: HGSL Framework

- Heterogeneity
  - Learn multiple heterogeneous subgraphs
- Complex Interactions
  - Feature similarity graphs and feature propagation graphs
  - Semantic graphs

CONTENTS

AAAI 2021

1 Introduction

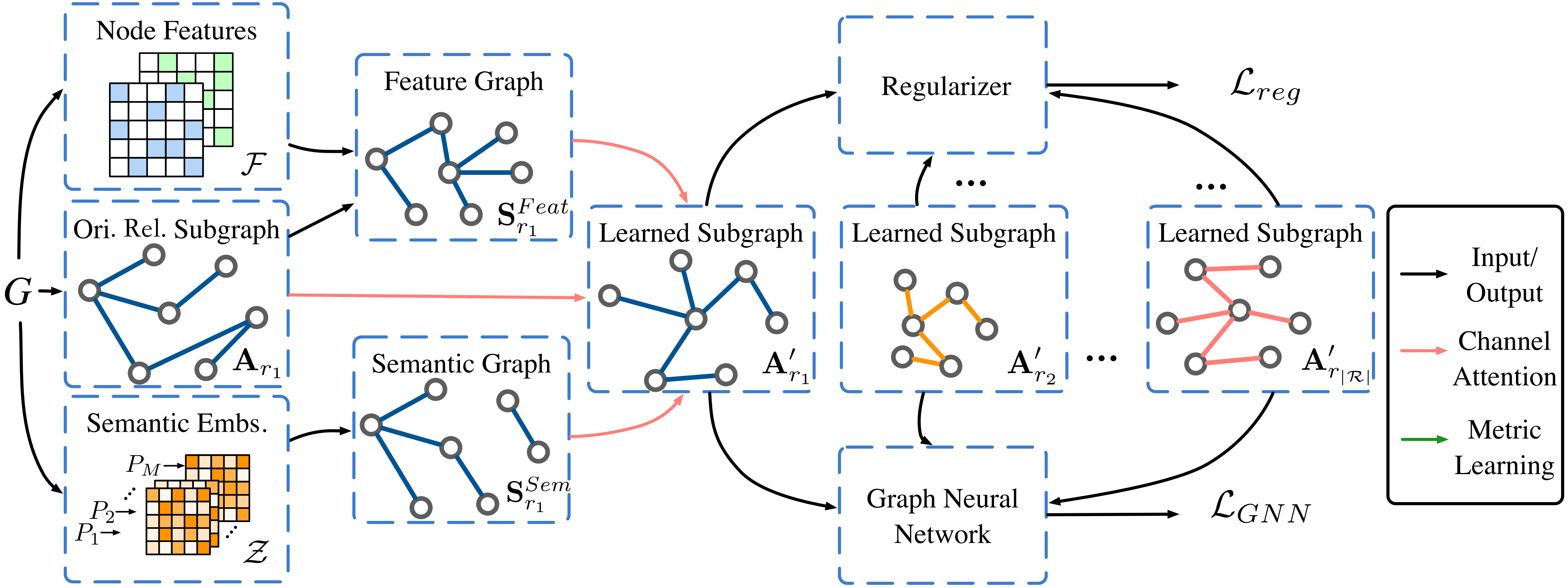
2 Proposed Method

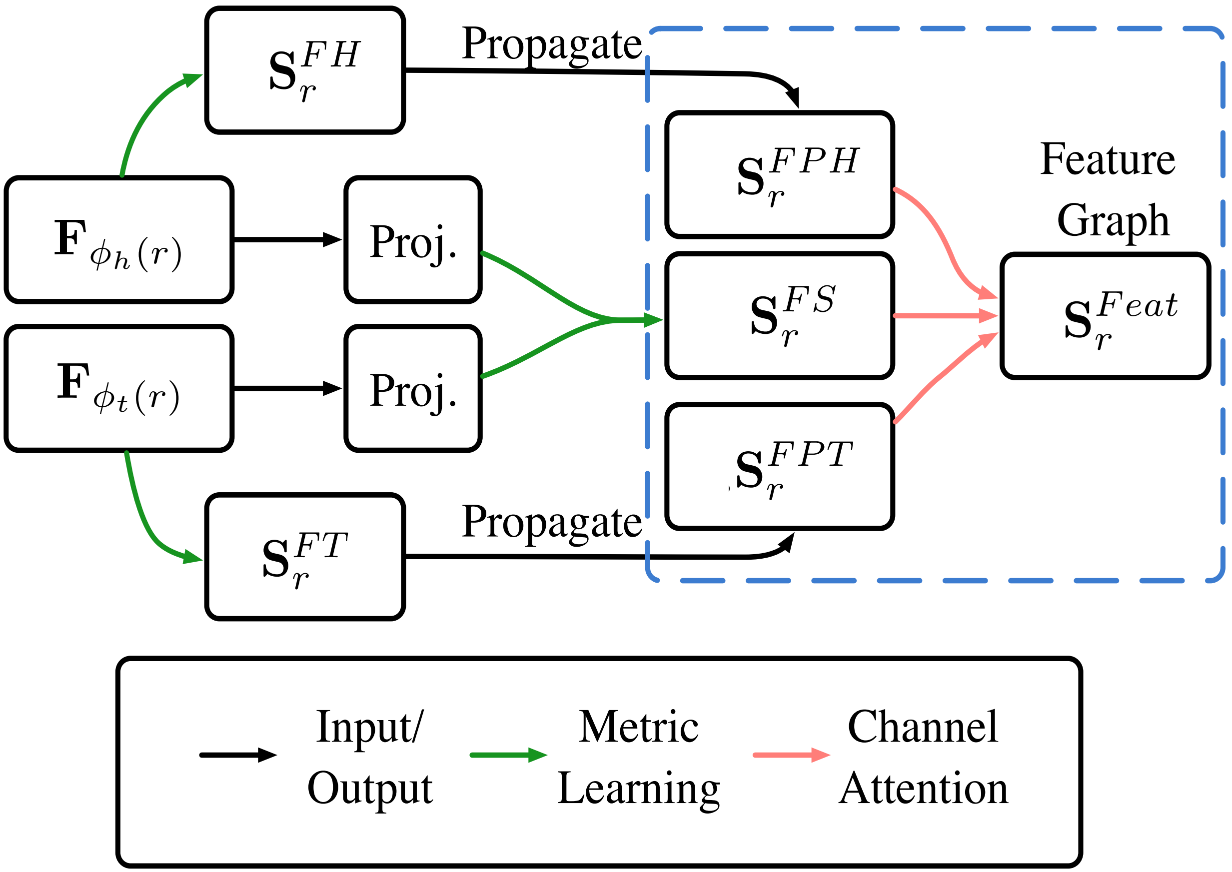
3 Experiments

4 Conclusions

Heterogeneous Graph Structure Learning for Graph Neural Networks (HGSL)

- Generate feature graph and semantic graph to fuse with the original graph.
- Jointly learn heterogeneous graph structure and GNN parameters for downstream tasks.



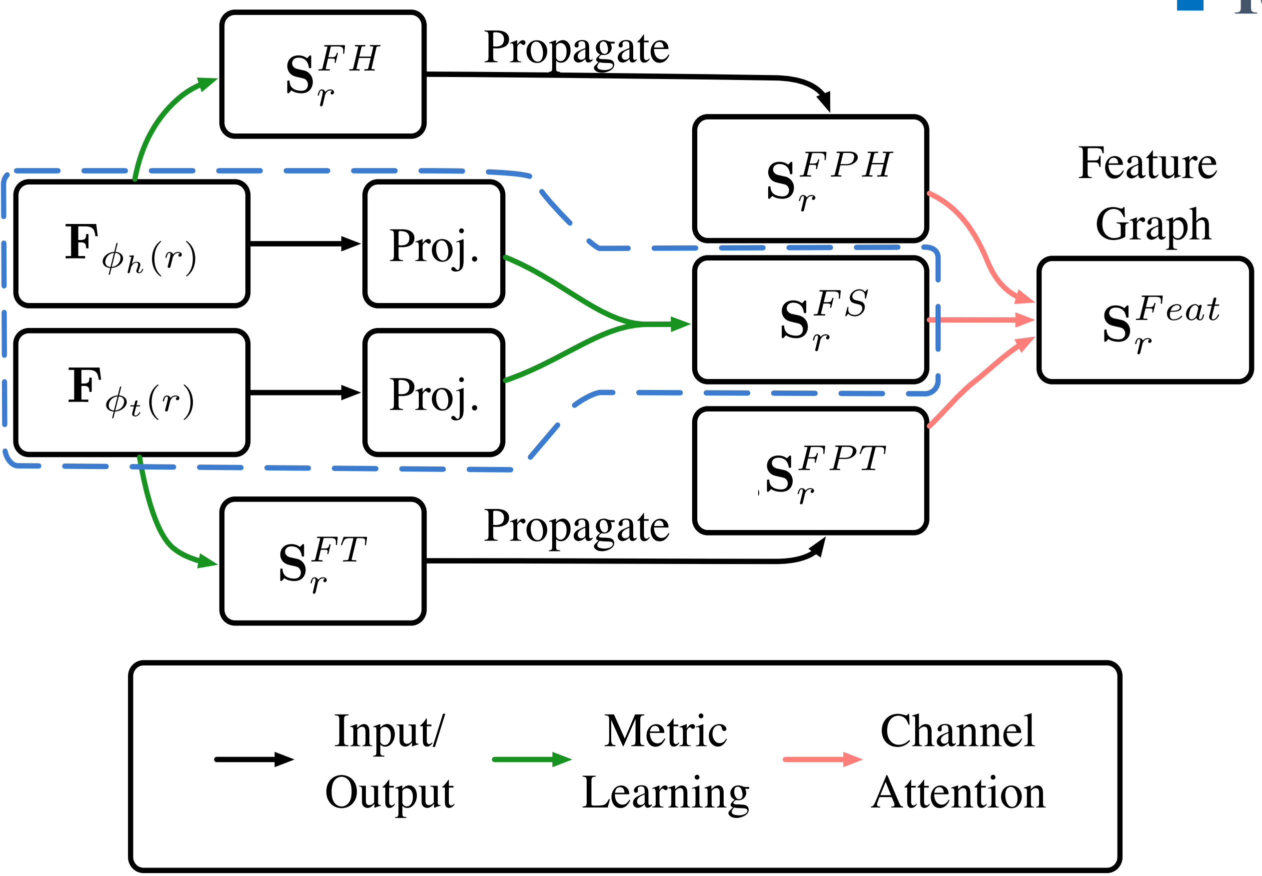


Feature graph generator

- **Feature Graph** $\mathbf{S}_r^{Feat}$
  - Graph structure generated by feature.
- **Feature similarity Graph** $\mathbf{S}_r^{FS}$
  - Idea: Node pairs with similar features may exist potential edges.
- **Feature Propagation Graphs** $\mathbf{S}_r^{FPH}$ *and* $\mathbf{S}_r^{FPT}$
  - Idea: Interactions between feature and topology may generate potential edges.
- **Channel Attention**

$$\mathbf{S}_r^{Feat} = \Psi_r^{Feat}([\mathbf{S}_r^{FS}, \mathbf{S}_r^{FPH}, \mathbf{S}_r^{FPT}]),$$

1×1 conv with parameters $\mathbf{W}_{\Psi,r}^{Feat} \in \mathbb{R}^{1\times1\times3}$



Feature graph generator

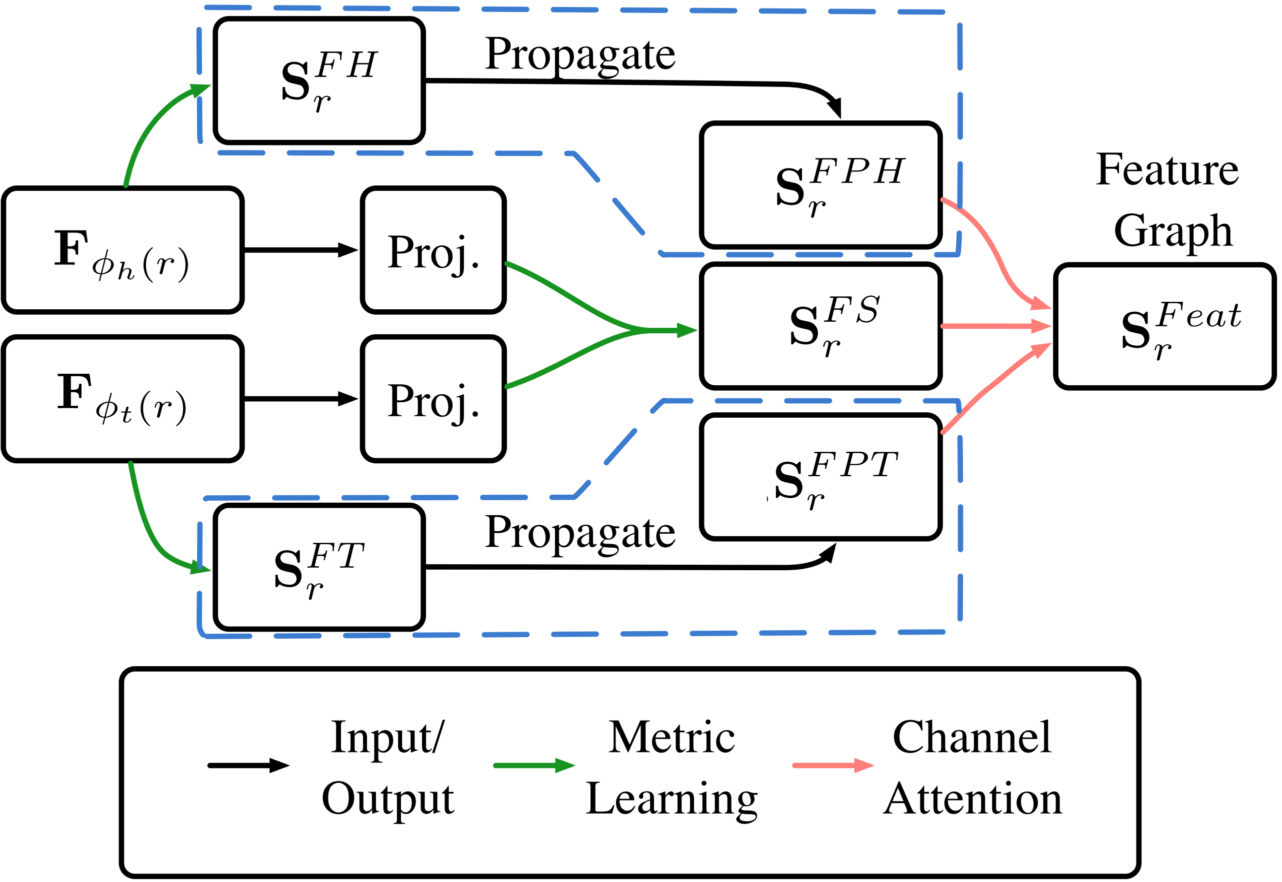
- **Feature Similarity Graph** $\mathbf{S}_r^{FS}$
  - Idea: Node pairs with similar features may exist potential edges.
  - Heterogeneous feature projection:

$$\mathbf{f}_i' = \sigma\left(\mathbf{f}_i \cdot \mathbf{W}_{\phi(v_i)} + \mathbf{b}_{\phi(v_i)}\right),$$

  - Graph structure learning by metric learning:

$$\mathbf{S}_r^{FS}[i,j] = \begin{cases} \Gamma_r^{FS}(\mathbf{f}_i', \mathbf{f}_j') & \Gamma_r^{FS}(\mathbf{f}_i', \mathbf{f}_j') > \epsilon^{FS} \\ 0 & \text{otherwise,} \end{cases}$$

$$\Gamma_r^{FS}(\mathbf{f}_i', \mathbf{f}_j') = \frac{1}{K}\sum_k^K \cos\left(\mathbf{w}_{k,r}^{FS} \odot \mathbf{f}_i', \mathbf{w}_{k,r}^{FS} \odot \mathbf{f}_j'\right)$$



- **Feature Propagation Graphs** $\mathbf{S}_r^{FPH}$ *and* $\mathbf{S}_r^{FPT}$
  - Idea: Interactions between feature and topology may generate potential edges.
  - Example: a user may be interested in movies watched by similar users (UU->UM) and movies similar to watched movies (UM->MM).
  - Homogeneous feature similarity

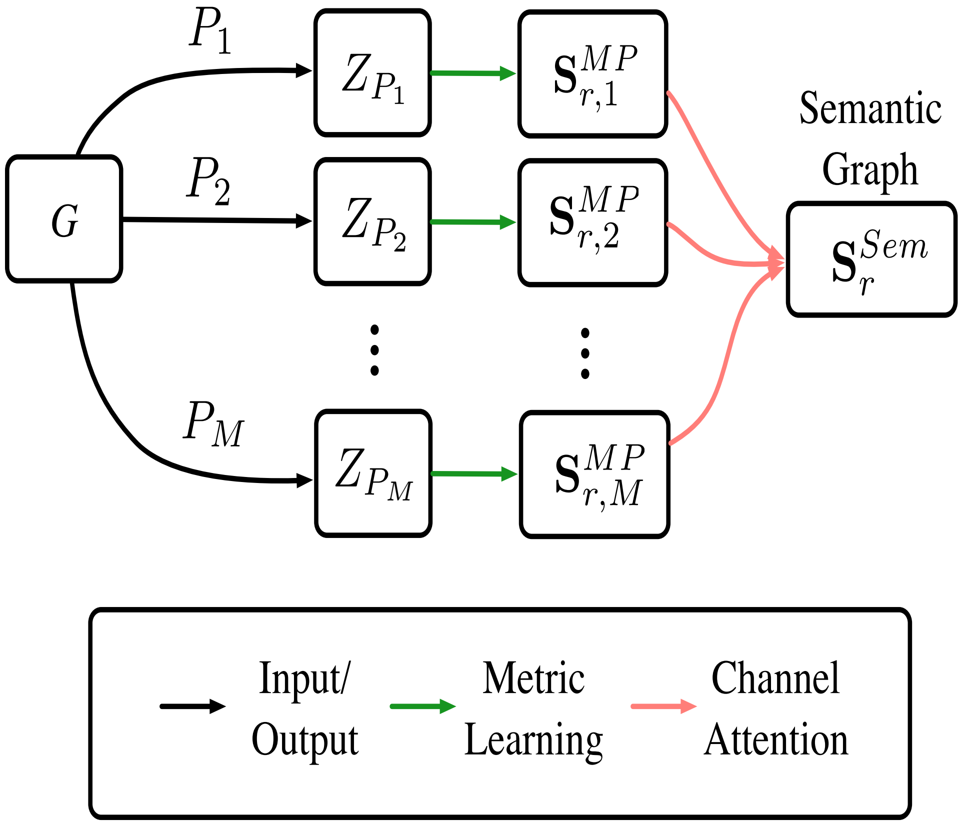
$$\mathbf{S}_r^{FH}[i,j] = \begin{cases} \Gamma_r^{FH}(\mathbf{f}_i, \mathbf{f}_j) & \Gamma_r^{FH}(\mathbf{f}_i, \mathbf{f}_j) > \epsilon^{FP} \\ 0 & \text{otherwise,} \end{cases}$$

e.g. user similarity

  - Feature propagation

$$\mathbf{S}_r^{FPH} = \mathbf{S}_r^{FH}\mathbf{A}_r \qquad \mathbf{S}_r^{FPT} = \mathbf{A}_r\mathbf{S}_r^{FT}$$

e.g. (UU -> UM) e.g. (UM -> MM)



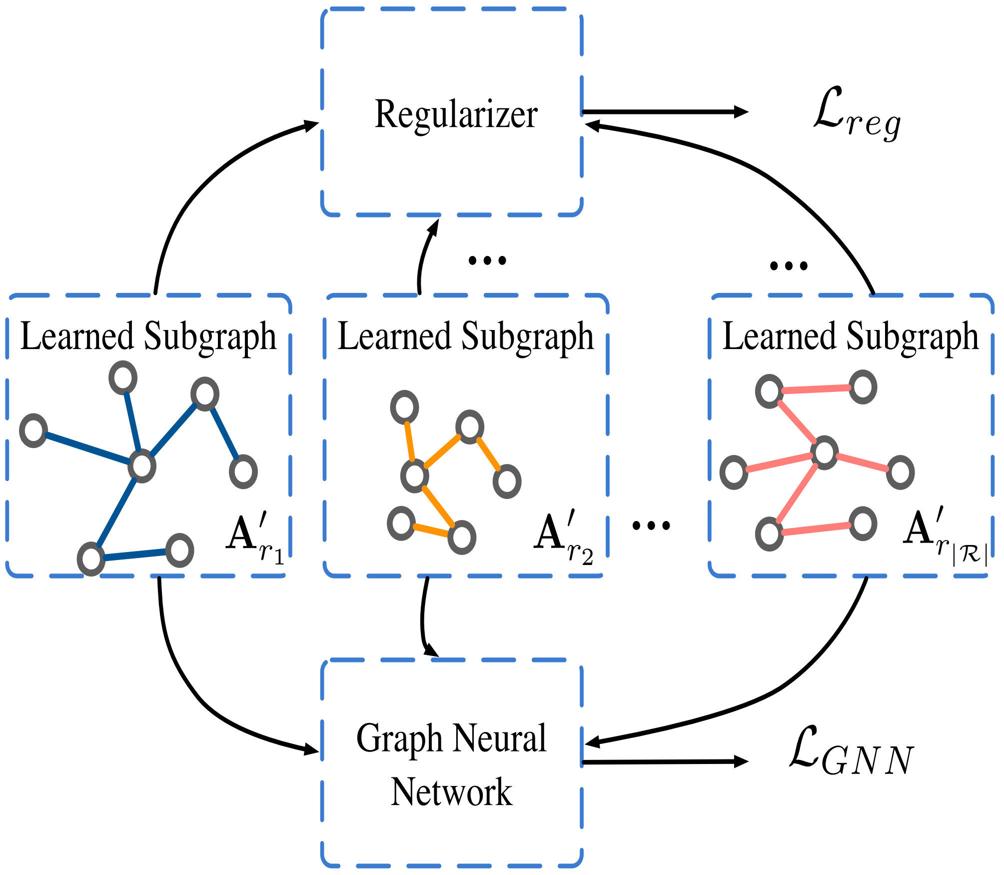
Semantic graph generator

- **Semantic Graph** $\mathbf{S}_r^{Sem}$
  - Idea: Use high-order topology to generate potential edges.
- **Semantic Candidate Graphs** $\mathbf{S}_{r,1}^{MP}, \mathbf{S}_{r,2}^{MP}, \ldots, \mathbf{S}_{r,M}^{MP}$
  - Metric learning on meta-path based node embeddings.

$$\mathbf{S}_{r,m}^{MP}[i,j] = \begin{cases} \Gamma_{r,m}^{MP}\left(\mathbf{z}_i^m, \mathbf{z}_j^m\right) & \Gamma_{r,m}^{MP}\left(\mathbf{z}_i^m, \mathbf{z}_j^m\right) > \epsilon^{MP} \\ 0 & \text{otherwise} \end{cases}$$

  - Adaptively fuse the semantic candidate graphs of different MP by channel attention.

$$\mathbf{S}_r^{Sem} = \Psi_r^{MP}\left(\left[\mathbf{S}_{r,1}^{MP}, \mathbf{S}_{r,2}^{MP}, \ldots, \mathbf{S}_{r,M}^{MP}\right]\right)$$



Optimization of HGSL

- **Relation Subgraph Generation**

$$\mathbf{A}'_r = \Psi_r([\mathbf{S}_r^{Feat}, \mathbf{S}_r^{Sem}, \mathbf{A}_r]),$$

- **Overall Heterogeneous Graph**

$$\mathcal{A}' = \{\mathbf{A}'_r, r \in \mathcal{R}\}.$$

- **GNN Loss**

$$\mathcal{L}_{GNN} = \sum_{v_i \in V_L} \ell\left(f_\theta(\mathbf{X}, \mathbf{A}')_i, y_i\right),$$

- **Regularization Loss**

$$\mathcal{L}_{reg} = \alpha \|\mathbf{A}'\|_1.$$

- **Overall Loss**

$$\mathcal{L} = \mathcal{L}_{GNN} + \mathcal{L}_{reg}.$$

CONTENTS

AAAI 2021

1 Introduction

2 Proposed Method

3 **Experiments**

4 Conclusions

Datasets

Dataset	# Nodes	# Edges	# Edge Type	# Features	# Training	# Validation	# Test
DBLP	7305	19816	4	334	600	300	2057
ACM	8994	25922	4	1902	600	300	2125
Yelp	3913	77176	6	82	300	300	2014

Experiments

- Node classification
- Ablation study
- Importance analysis of candidate graphs
- Parameter analysis

Baselines

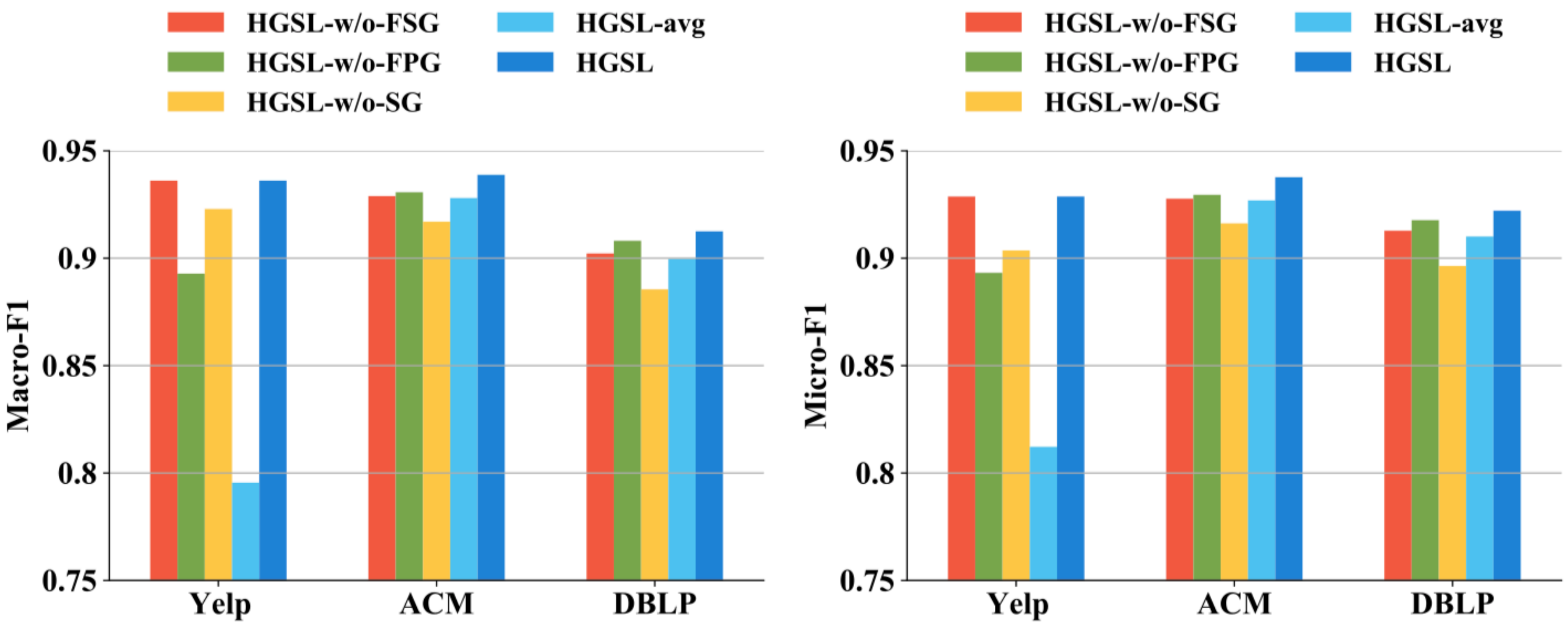
- Homogeneous
  - DeepWalk
  - GCN
  - GAT
  - GraphSage
- Heterogeneous
  - MP2Vec
  - HAN
  - HeGAN
  - GTN
- Graph structure learning related
  - LDS
  - Pro-GNN
  - Geom-GCN

Node Classification

	DBLP		ACM		Yelp	
	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
DeepWalk	88.00 ± 0.47	89.13 ± 0.41	80.65 ± 0.60	80.32 ± 0.61	68.68 ± 0.83	73.16 ± 0.96
GCN	83.38 ± 0.67	84.40 ± 0.64	91.32 ± 0.61	91.22 ± 0.64	82.95 ± 0.43	85.22 ± 0.55
GAT	77.59 ± 0.72	78.63 ± 0.72	92.96 ± 0.28	92.86 ± 0.29	84.35 ± 0.74	86.22 ± 0.56
GraphSage	78.37 ± 1.17	79.39 ± 1.17	91.19 ± 0.36	91.12 ± 0.36	93.06 ± 0.35	92.08 ± 0.31
MP2Vec	88.86 ± 0.19	89.98 ± 0.17	78.63 ± 1.11	78.27 ± 1.14	59.47 ± 0.57	65.11 ± 0.53
HAN	90.53 ± 0.24	91.47 ± 0.22	91.67 ± 0.39	91.57 ± 0.38	88.49 ± 1.73	88.78 ± 1.40
HeGAN	87.02 ± 0.37	88.34 ± 0.38	82.04 ± 0.77	81.80 ± 0.79	62.41 ± 0.76	68.17 ± 0.79
GTN	90.42 ± 1.29	91.41 ± 1.09	91.91 ± 0.58	91.78 ± 0.59	92.84 ± 0.28	92.19 ± 0.29
LDS	75.65 ± 0.20	76.63 ± 0.18	92.14 ± 0.16	92.07 ± 0.15	85.05 ± 0.16	86.05 ± 0.50
Pro-GNN	89.20 ± 0.15	90.28 ± 0.16	91.62 ± 1.28	91.55 ± 1.31	74.12 ± 2.03	77.45 ± 2.12
Geom-GCN	79.43 ± 1.01	80.94 ± 1.06	70.20 ± 1.23	70.00 ± 1.06	84.28 ± 0.70	85.36 ± 0.60
HGSL	**91.92 ± 0.11**	**92.77 ± 0.11**	**93.48 ± 0.59**	**93.37 ± 0.59**	**93.55 ± 0.52**	**92.76 ± 0.60**

Table 2: Performance evaluation of node classification (mean in percentage ± standard deviation).

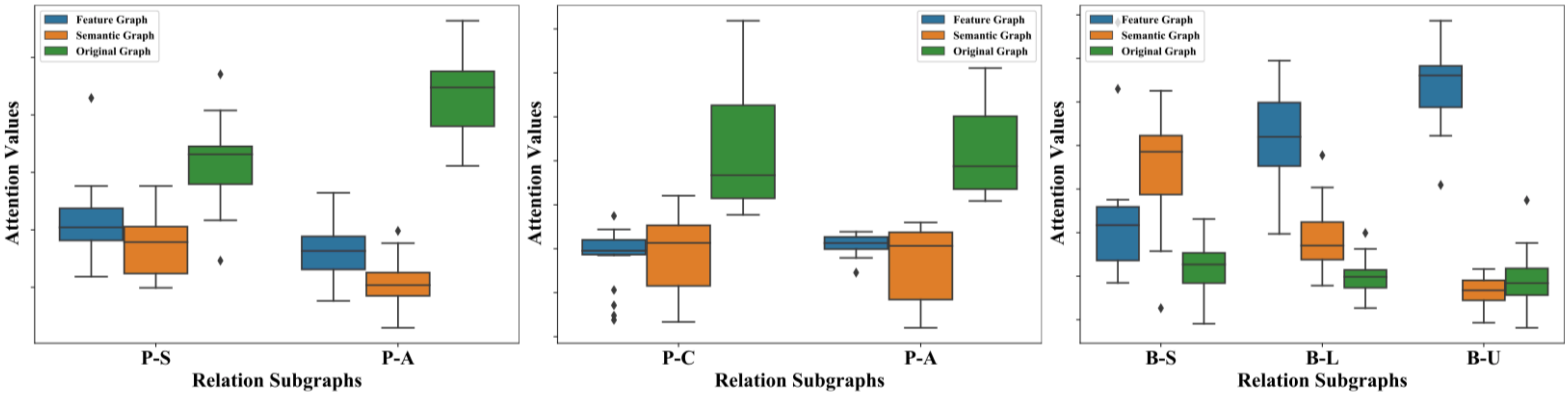
Ablation Study



(a) Macro-F1 (b) Micro-F1

Performance evaluation of variants of HGSL.

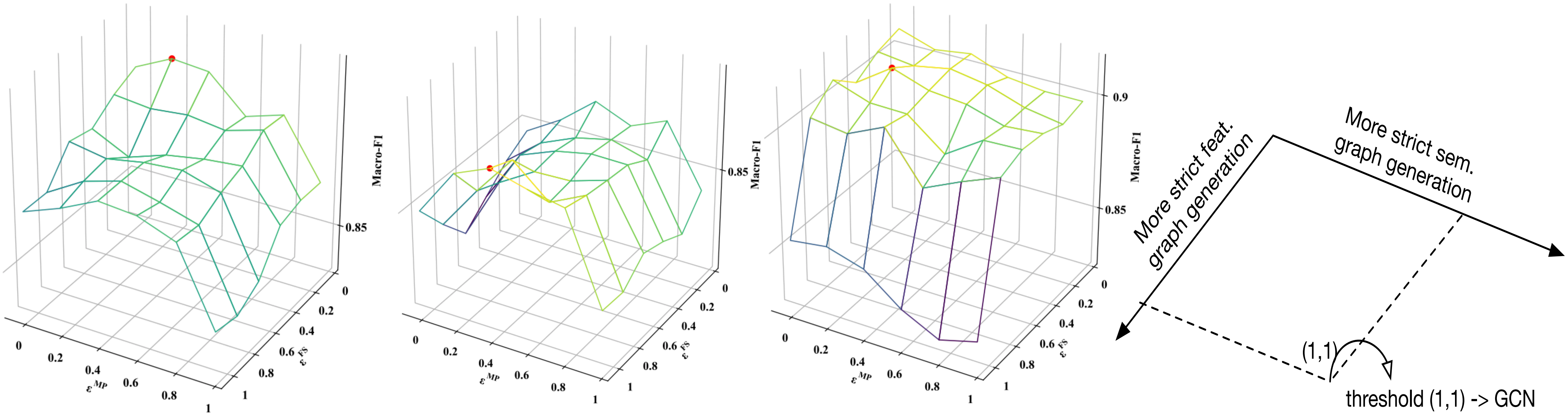
Channel Attention Analysis



(a) ACM (b) DBLP (c) Yelp

Channel attention distributions of relation subgraphs.

Parameter Sensitivity



(a) ACM (b) DBLP (c) Yelp

Parameter sensitivity of different thresholds.

CONTENTS

AAAI 2021

1 Introduction

2 Proposed Method

3 Experiments

4 Conclusions

Conclusion

- We make the first attempt to study the fundamental problem: How to learn an optimal heterogeneous graph structure for GNN towards downstream task.
- We propose a heterogeneous graph structure learning framework, where three kinds of graph structures (feature similarity graph, feature propagation graph, semantic graph) are generated, so as to comprehensively fuse an optimal heterogeneous graph for GNN.
- We conduct extensive experiments on three real-world datasets to validate the effectiveness of HGSL against the state-of-the-art methods.

- **Takeaway**
  - Graph structure learning is vital, since the raw graph is...
    - Noisy and incomplete
    - Hard (edges often represented by 0s or 1s)
  - Learn graphs from...
    - Feature similarity
    - High order topology (MP-based semantics)
    - Interactions of feature and topology

AAAI 2021

Thanks

Contact: jxz1244@case.edu