Heterogeneous Graph Transformer with Poly-Tokenization

Zhiyuan Lu, Yuan Fang, Cheng Yang, Chuan Shi

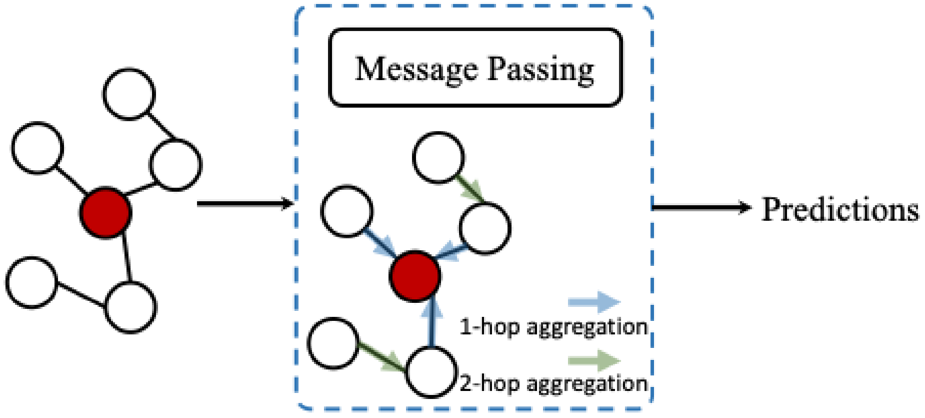
In the Proceedings of the 33rd International Joint Conference on Artificial Intelligence.

Background

Graph Neural Networks

$$h_v^l = M(h_v^{l-1}, \{h_i^{l-1} : i \in N_v\}; \theta^l)$$

Message passing function

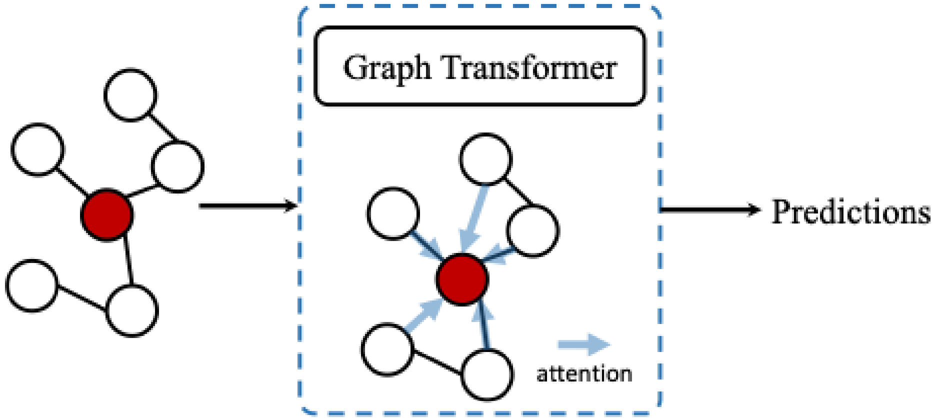


- Limited expressive power
- Over-smoothing problem

Graph Transformers

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Fully-connected self-attention

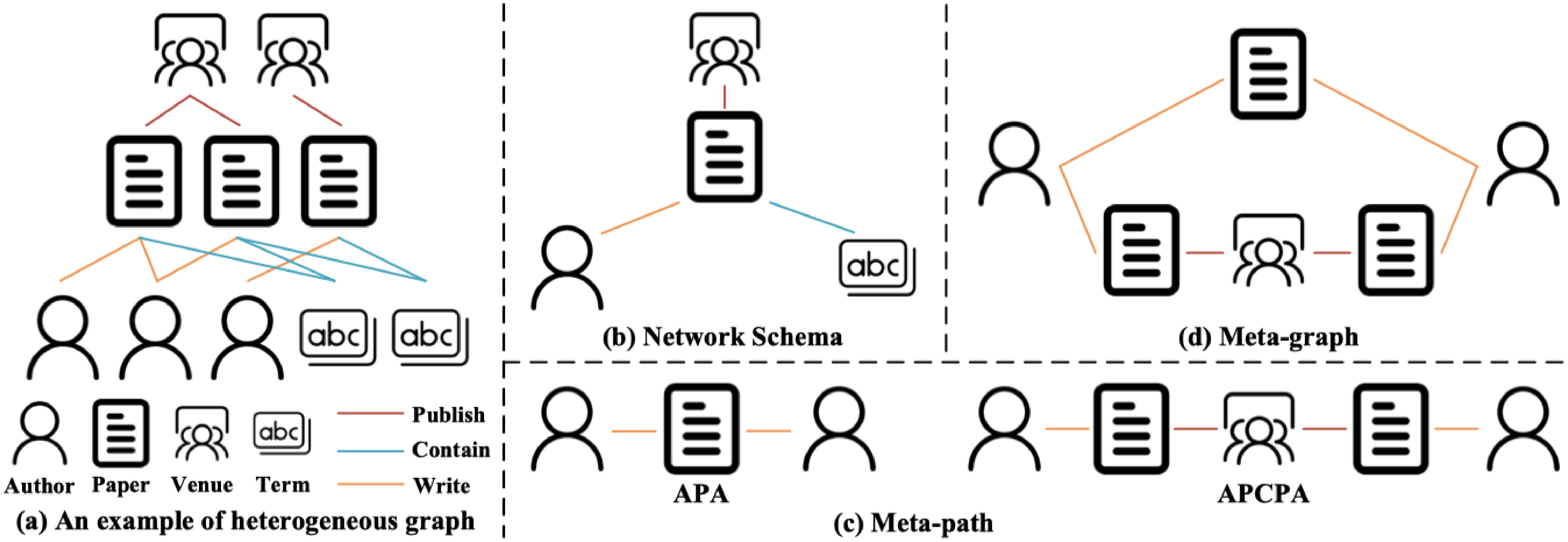


- Stronger expressive power
- Alleviate over-smoothing

Background

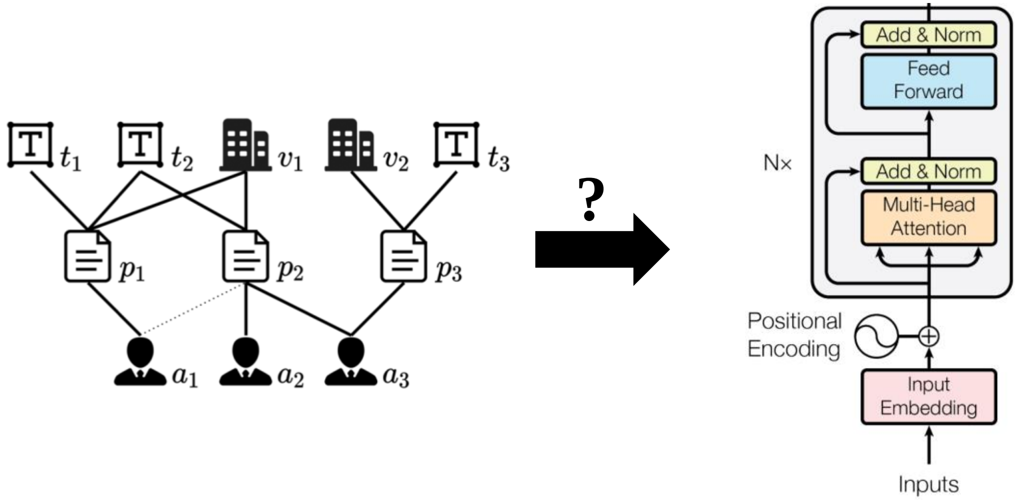
Heterogeneous Graph

HGs are powerful for representing complex real-world networks (e.g., academic networks, social networks).

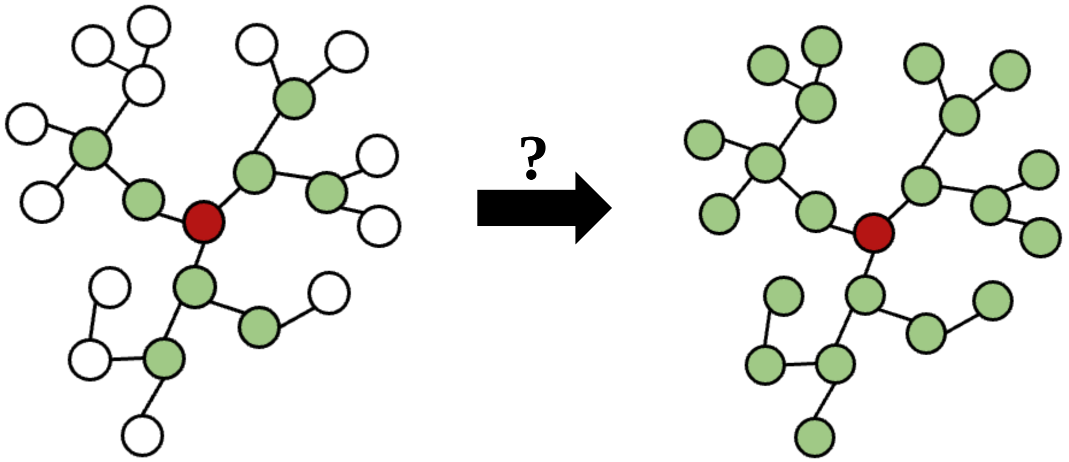


Motivation

1. Current graph transformer models struggle to integrate the rich and complex semantics inherent in heterogeneous graphs.



2. Existing methods often have limited receptive fields due to the quadratic complexity of transformer models.



The Overall Framework

The poly-tokenization mechanism

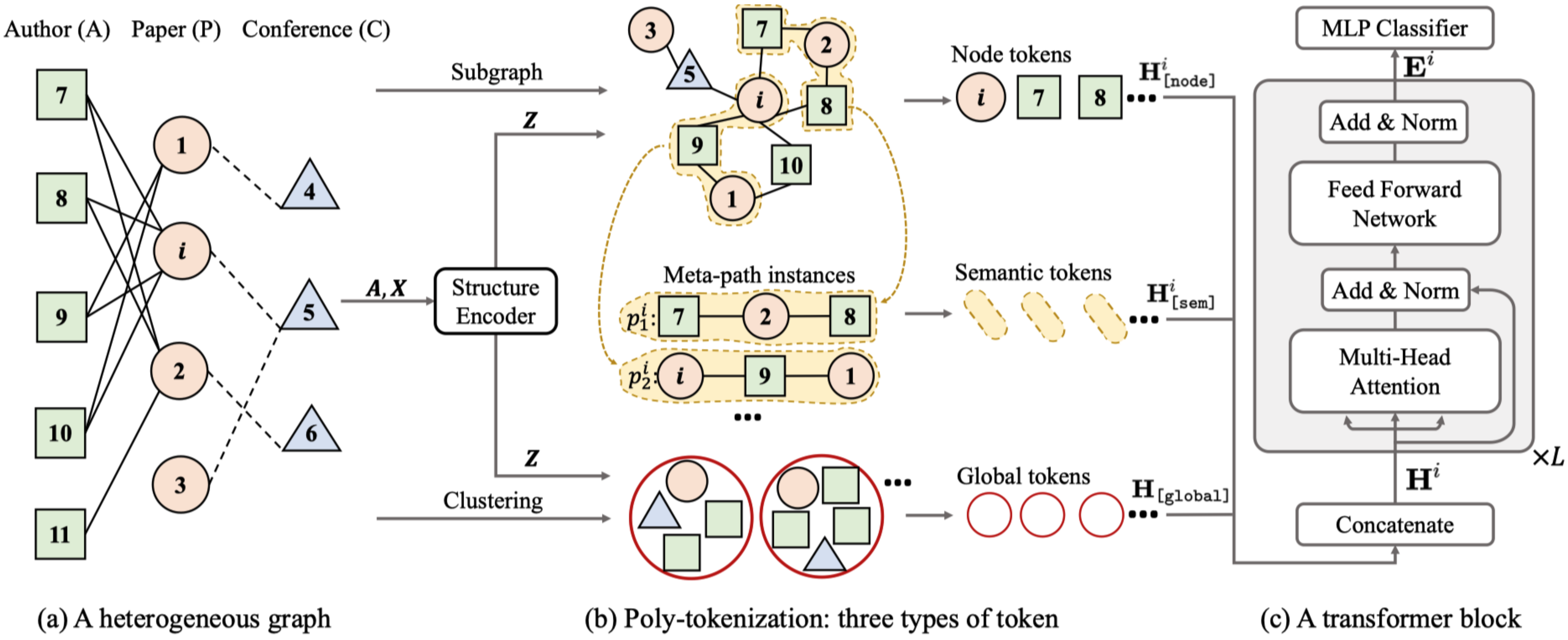
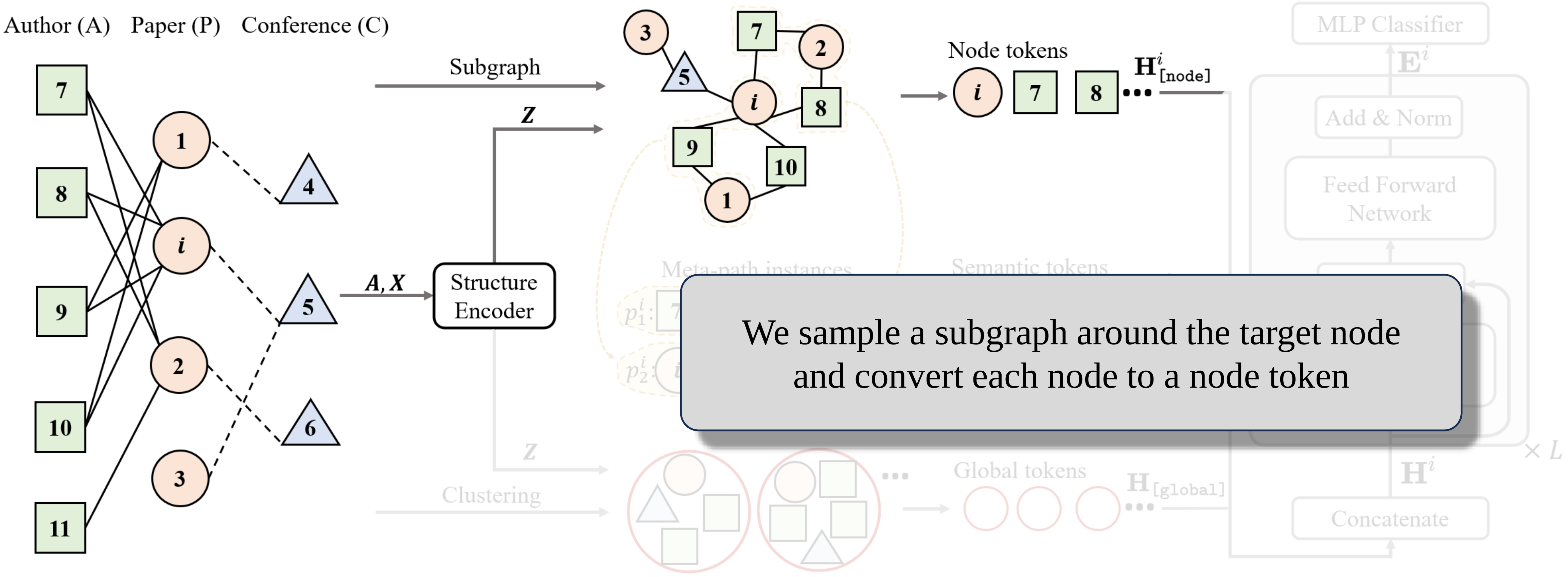


Figure 1: The overall framework of PHGT.

Node Token



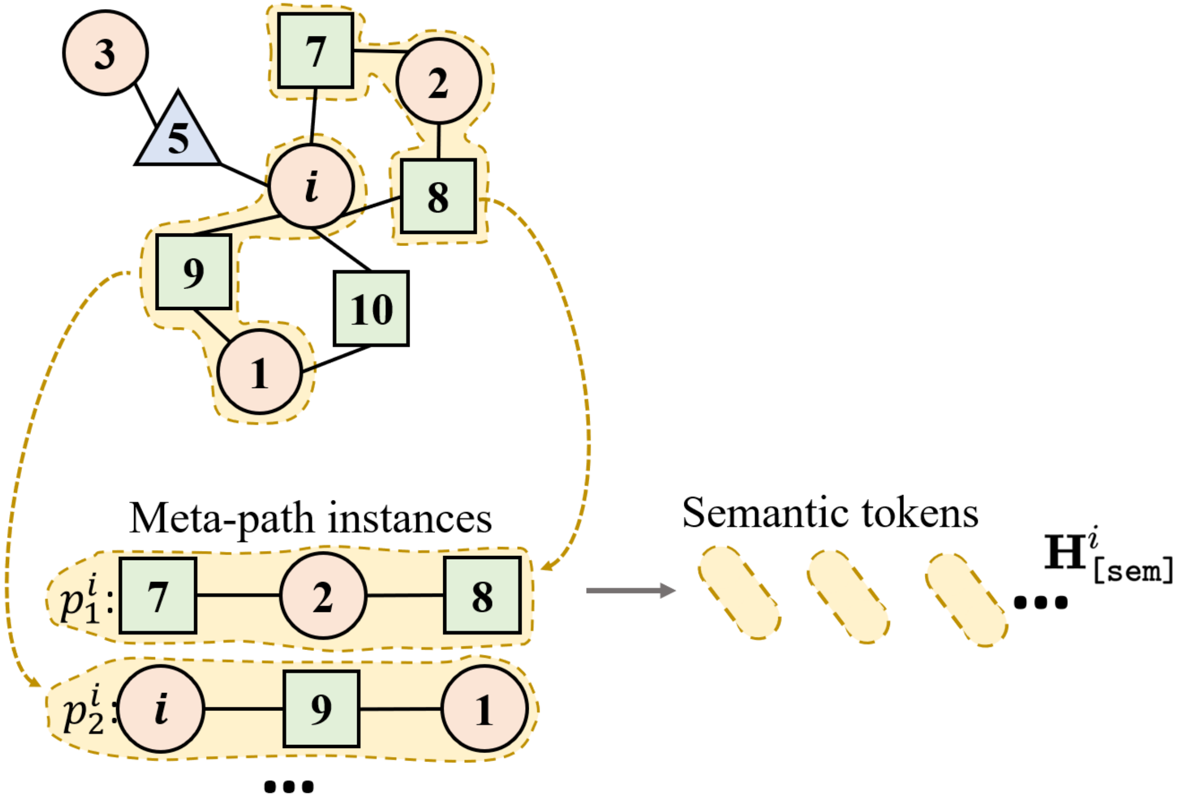
(a) A heterogeneous graph

(b) Poly-tokenization: three types of token

(c) A transformer block

We sample a subgraph around the target node and convert each node to a node token

Semantic Token



- Sample meta-path instances according to a pre-defined meta-path set.
- Each meta-path instance is converted to a semantic token

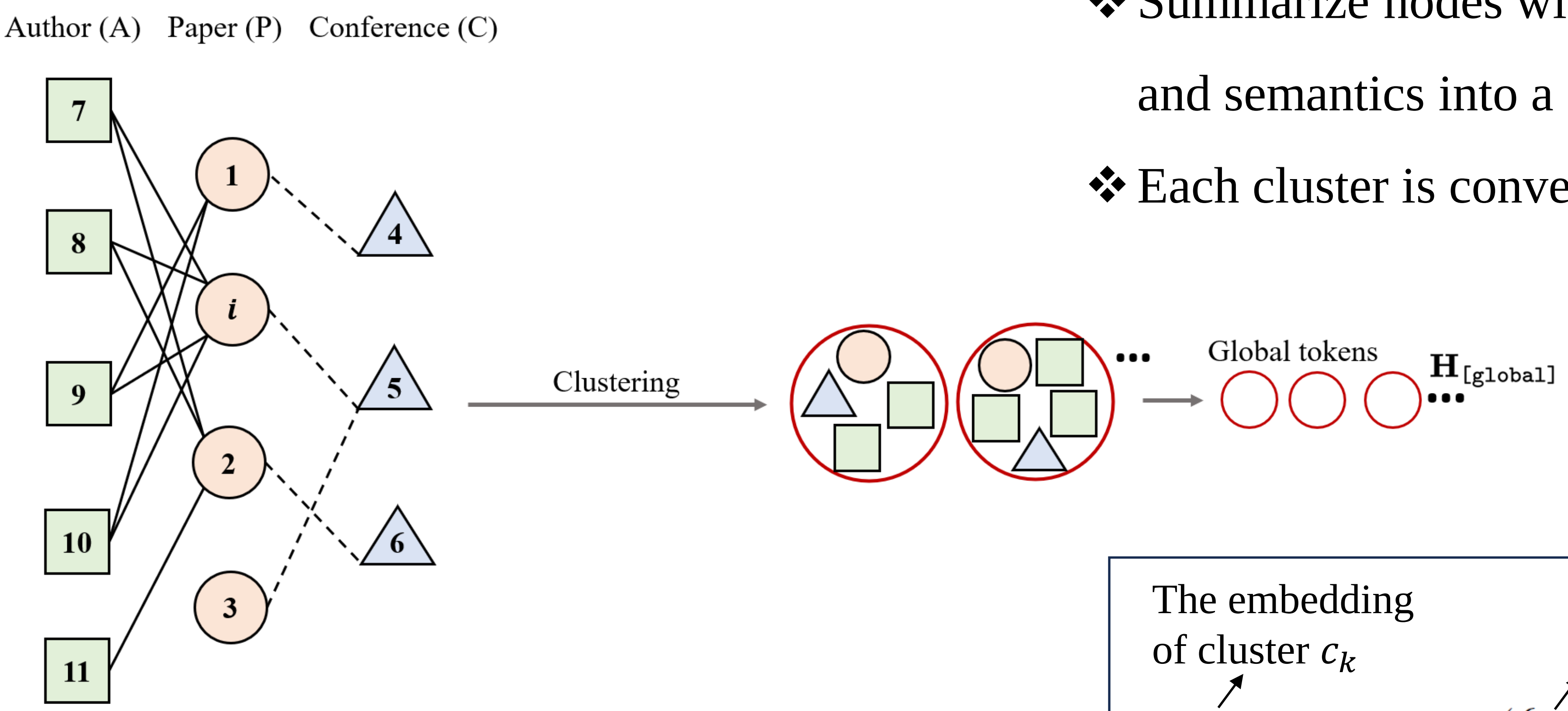
The embedding of path instance p_k^i

An instance of meta-path p_k

$$\mathbf{z}^i_{p_k} = \mathtt{ReadOut}(\{\mathbf{z}_u \mid u \in p_k^i\})$$

$$\mathbf{H}^i_{[\mathtt{sem}]} = [\mathbf{z}^i_{p_1}, \mathbf{z}^i_{p_2}, \ldots, \mathbf{z}^i_{p_N}]^\top$$

Global Token



- ❖ Summarize nodes with similar structure and semantics into a cluster.
- ❖ Each cluster is converted to a global token.

The embedding of cluster c_k

The embedding of node u in cluster c_k

$$\mathbf{z}_{c_k} = \mathtt{ReadOut}(\{\mathbf{z}_u \mid u \in c_k\})$$

$$\mathbf{H}_{[\mathtt{global}]} = [\mathbf{z}_{c_1}, \mathbf{z}_{c_2}, \ldots, \mathbf{z}_{c_M}]^\top$$

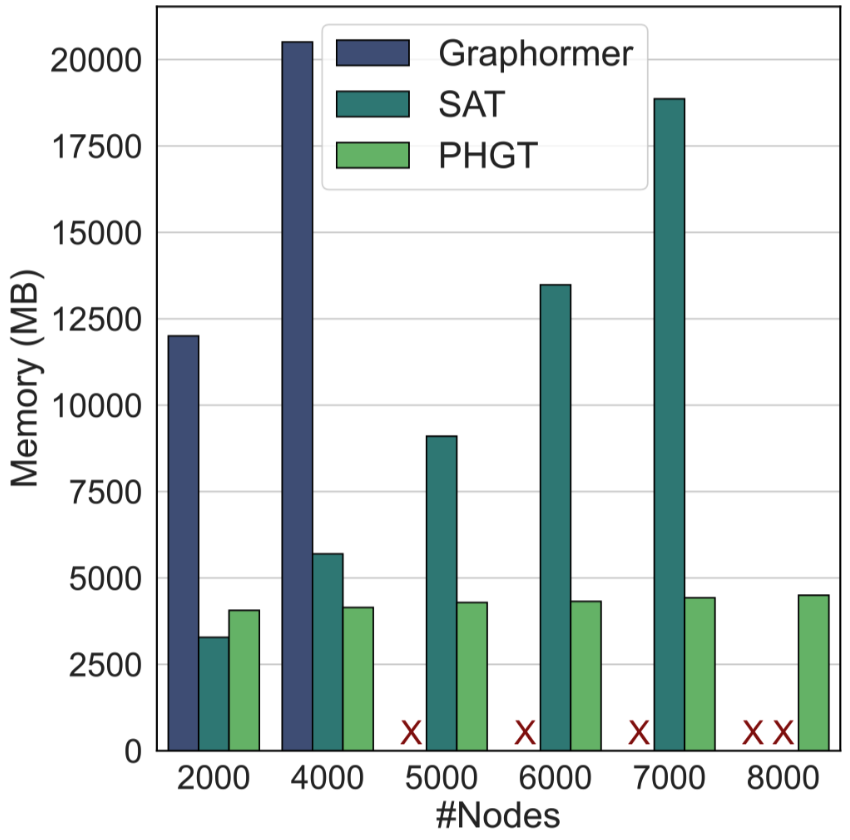
Heterogeneous Node Classification Experiments

Methods	DBLP		IMDB		ACM		Freebase	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1
RGCN	92.07±0.50	91.52±0.50	62.05±0.15	58.85±0.26	91.41±0.75	91.55±0.74	60.82±1.23	59.08±1.44
HAN	92.05±0.62	91.67±0.49	64.63±0.58	57.74±0.96	90.79±0.43	90.89±0.43	61.42±3.56	57.05±2.06
GTN	93.97±0.54	93.52±0.55	65.14±0.45	60.47±0.98	91.20±0.71	91.31±0.70	-	-
HetGNN	92.33±0.41	91.76±0.43	51.16±0.65	48.25±0.67	86.05±0.25	85.91±0.25	62.99±2.31	58.44±1.99
MAGNN	93.76±0.45	93.28±0.51	64.67±1.67	56.49±3.20	90.77±0.65	90.88±0.64	64.43±0.73	58.18±3.87
HGT	93.49±0.25	93.01±0.23	67.20±0.57	63.00±1.19	91.00±0.76	91.12±0.76	66.43±1.88	60.03±2.21
Simple-HGN	94.46±0.22	94.01±0.24	67.36±0.57	63.53±1.36	93.35±0.45	93.42±0.44	67.49±0.97	62.49±1.69
ANS-GT	93.15±0.51	92.75±0.43	66.65±0.35	62.52±0.61	92.55±0.54	93.67±0.62	67.33±0.61	61.24±0.57
NodeFormer	93.68±0.42	93.05±0.38	65.86±0.42	62.15±0.77	91.89±0.31	92.72±0.84	67.01±0.52	60.83±1.41
HINormer	94.94±0.21	94.57±0.23	67.83±0.34	64.65±0.53	93.15±0.36	93.28±0.43	67.78±0.39	**62.76±1.10**
PHGT	**95.33±0.18**	**94.96±0.17**	**68.81±0.08**	**65.91±0.30**	**93.72±0.40**	**93.79±0.39**	**68.74±1.42**	61.73±1.86

Ablation and Efficiency Studies

	DBLP	IMDB	ACM	Freebase
w/o both	94.80	68.35	93.34	67.58
w/o semantic token	94.94	68.58	93.41	67.73
w/o global token	94.91	68.54	93.55	68.06
PHGT	**95.33**	**68.81**	**93.72**	**68.89**

Ablation studies (micro-f1 score)



Memory usage as graph size increases

Conclusion

- PHGT addressed the two limitations of existing graph transformer models:
  - (1) the inability to capture heterogeneous semantics;
  - (2) the incapacity to model intricate long-range dependencies.

- Through comprehensive experiments on four benchmark datasets, we demonstrate the efficacy of our PHGT approach.

OpenHGNN shichuan.org

Thanks !

Heterogeneous Graph Transformer with Poly-Tokenization

Zhiyuan Lu, Yuan Fang, Cheng Yang, Chuan Shi