西安電子科技大學
XIDIAN UNIVERSITY

NeurIPS 2022

Self-supervised Heterogeneous Graph Pre-training Based on Structural Clustering

Yaming Yang, Ziyu Guan, Zhe Wang, Wei Zhao, Cai Xu, Weigang Lu, Jianbin Huang

School of Computer Science and Technology, Xidian University

2022-10-13

Background

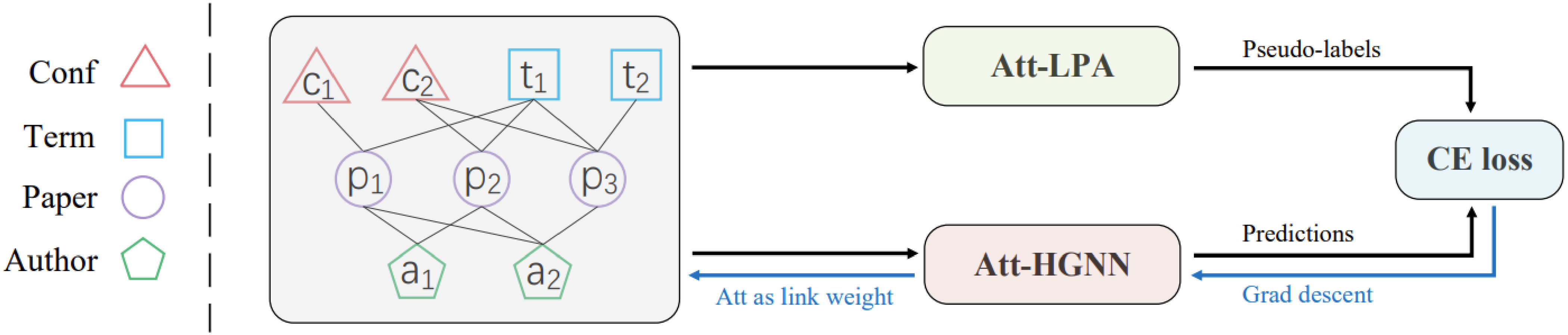
Problem

Pre-train heterogeneous graphs in a self-supervised manner.

Motivation

- Existing methods require high-quality positive and negative examples, limiting their flexibility and generalization ability.
- We propose a flexible framework SHGP, which does not need any positive examples or negative examples.

Method: Overall Architecture

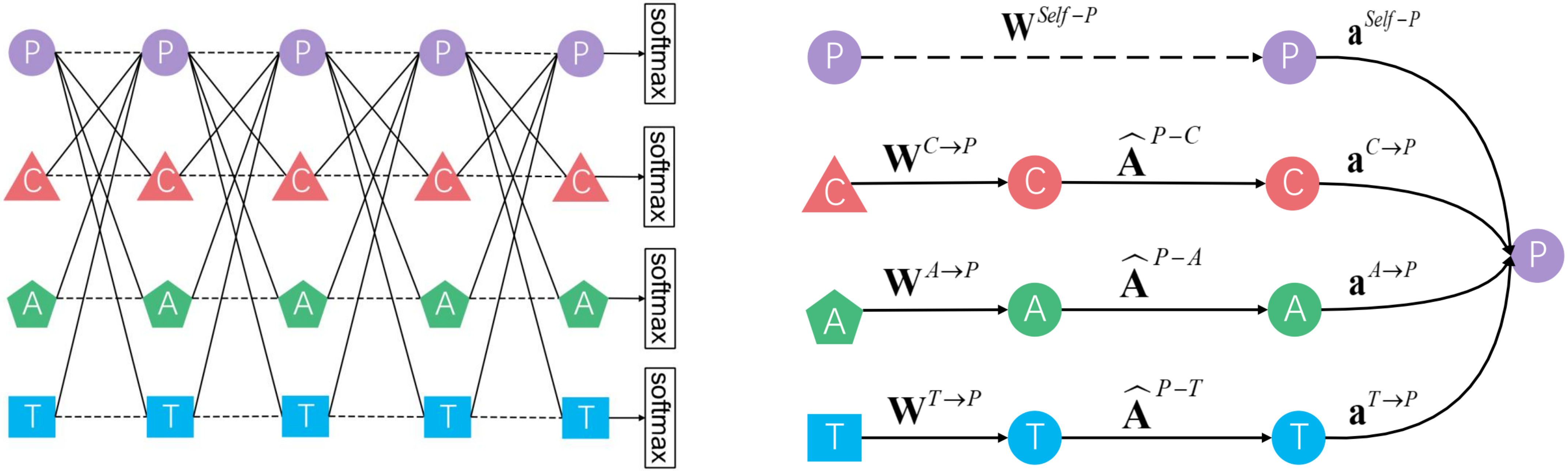


1. **Compute embeddings:** $\mathbf{H}^{[t]} = \textbf{Att-HGNN}(\mathcal{W}^{[t-1]}, \mathcal{G}, \mathcal{X})$ $\mathbf{P}^{[t]} = softmax(\mathbf{H}^{[t]} \cdot \mathbf{C}^{[t-1]})$
2. **Compute pseudo-labels:** $\mathbf{Y}^{[t]} = \textbf{Att-LPA}(\mathcal{W}^{[t-1]}, \mathcal{G}, \mathbf{Y}^{[t-1]})$
3. **Compute cross-entropy:** $\mathcal{L}^{[t]} = -\sum_{i \in \mathcal{V}} \sum_{c=1}^{K} \mathbf{Y}_{i,c}^{[t]} \ln \mathbf{P}_{i,c}^{[t]}$
4. **Gradient descent:** $\mathcal{W}^{[t]} = \mathcal{W}^{[t-1]} - \eta \cdot \nabla_{\mathcal{W}} \mathcal{L}^{[t]}$

Yaming Yang, et al., Self-supervised Heterogeneous Graph Pre-training based on Structural Clustering, NeurIPS 2022.

西安電子科技大學
XIDIAN UNIVERSITY

Method: Att-HGNN Encoder



1. **Feature Projection**: project different types of features into a common space.
2. **Object-level Aggregation**: aggregate one-type of neighbors by adjacency matrix.
3. **Type-level Aggregation**: aggregate different-types of neighbors by attention.

Yaming Yang, et al., Interpretable and Efficient Heterogeneous Graph Convolutional Network, IEEE TKDE 2021.

Experiments

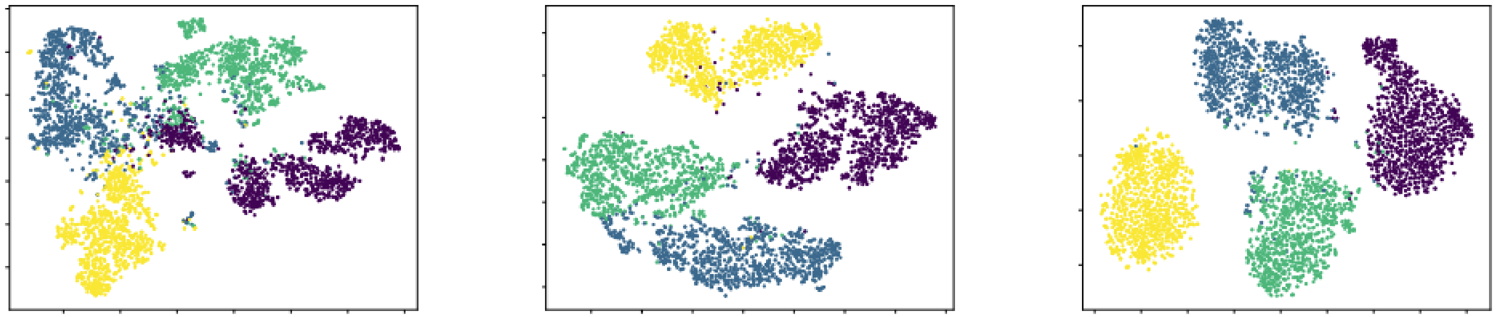
Datasets	Metrics	Train	HAN	HGCN	M2V	DMGI	HDGI	HeCo	H-DC	SHGP
MAG	Mic-F1	4%	90.07	93.16	88.97	94.43	94.10	95.75	85.03	**98.23**
		6%	91.83	95.18	89.94	93.80	93.68	95.93	85.16	**98.30**
		8%	92.17	97.13	90.15	94.36	94.27	96.08	86.03	**98.37**
	Mac-F1	4%	89.93	92.82	88.51	94.32	93.89	95.27	84.72	**98.24**
		6%	91.54	95.08	89.45	93.74	93.64	95.42	85.13	**98.33**
		8%	91.82	97.05	89.73	94.27	94.23	95.15	85.97	**98.41**
ACM	Mic-F1	4%	70.84	75.78	72.45	78.93	79.72	79.78	78.53	**80.31**
		6%	72.04	77.59	73.83	79.01	80.09	80.15	79.96	**80.78**
		8%	73.23	78.08	73.95	79.47	79.07	**80.94**	79.82	80.91
	Mac-F1	4%	61.50	64.61	53.01	59.37	60.57	65.91	64.89	**67.14**
		6%	60.23	64.04	51.86	59.15	61.09	65.63	64.37	**67.38**
		8%	62.37	65.73	53.72	59.42	59.99	67.15	65.11	**68.19**
DBLP	Mic-F1	4%	90.48	92.45	88.93	89.35	88.33	91.31	87.15	**93.70**
		6%	91.03	92.08	89.47	89.21	88.93	91.05	86.67	**93.92**
		8%	91.90	92.34	91.41	89.88	88.18	91.22	87.23	**94.13**
	Mac-F1	4%	90.01	92.13	88.49	88.21	87.69	90.53	87.03	**93.31**
		6%	90.51	91.71	88.97	88.03	88.75	90.26	86.53	**93.52**
		8%	91.35	92.04	89.83	88.57	87.38	90.42	87.11	**93.77**
IMDB	Mic-F1	4%	56.05	56.68	56.54	54.79	56.31	57.42	54.01	**58.51**
		6%	54.21	57.72	55.24	54.93	57.64	58.63	54.19	**59.76**
		8%	56.45	57.03	57.02	55.75	56.70	60.13	55.19	**61.60**
	Mac-F1	4%	39.04	36.66	27.03	37.95	30.84	38.66	34.72	**43.36**
		6%	36.63	39.38	26.51	38.67	36.35	39.43	36.61	**46.17**
		8%	38.20	40.54	27.86	39.89	34.64	40.00	38.03	**48.02**

Object Classification

Experiments

	MAG		ACM		DBLP		IMDB	
	NMI	ARI	NMI	ARI	NMI	ARI	NMI	ARI
M2V	39.67	43.75	32.53	28.49	49.50	56.73	1.43	1.03
DMGI	70.89	73.51	38.45	32.46	65.17	67.23	3.49	2.65
HDGI	73.96	77.15	39.13	32.34	59.98	62.33	4.15	2.96
HeCo	79.33	83.16	39.06	**32.69**	68.81	74.05	5.69	2.32
H-DC	42.75	49.01	18.60	19.75	47.15	53.15	1.57	1.12
SHGP	**90.65**	**93.00**	**39.42**	32.63	**73.30**	**77.31**	**6.33**	**3.10**

Object Clustering



(a) DMGI (b) HeCo (c) SHGP

Embedding Visualization

西安電子科技大學
XIDIAN UNIVERSITY

Summary

We propose SHGP, a novel heterogeneous graph pre-training framework.

- SHGP does not require any positive examples or negative examples.
- SHGP enjoys a high degree of flexibility.

Thank You!

SHGP paper

SHGP code

ie-HGCN paper

ie-HGCN code