

Bootstrapping Heterogeneous Graph Representation Learning via Large Language Models: A Generalized Approach

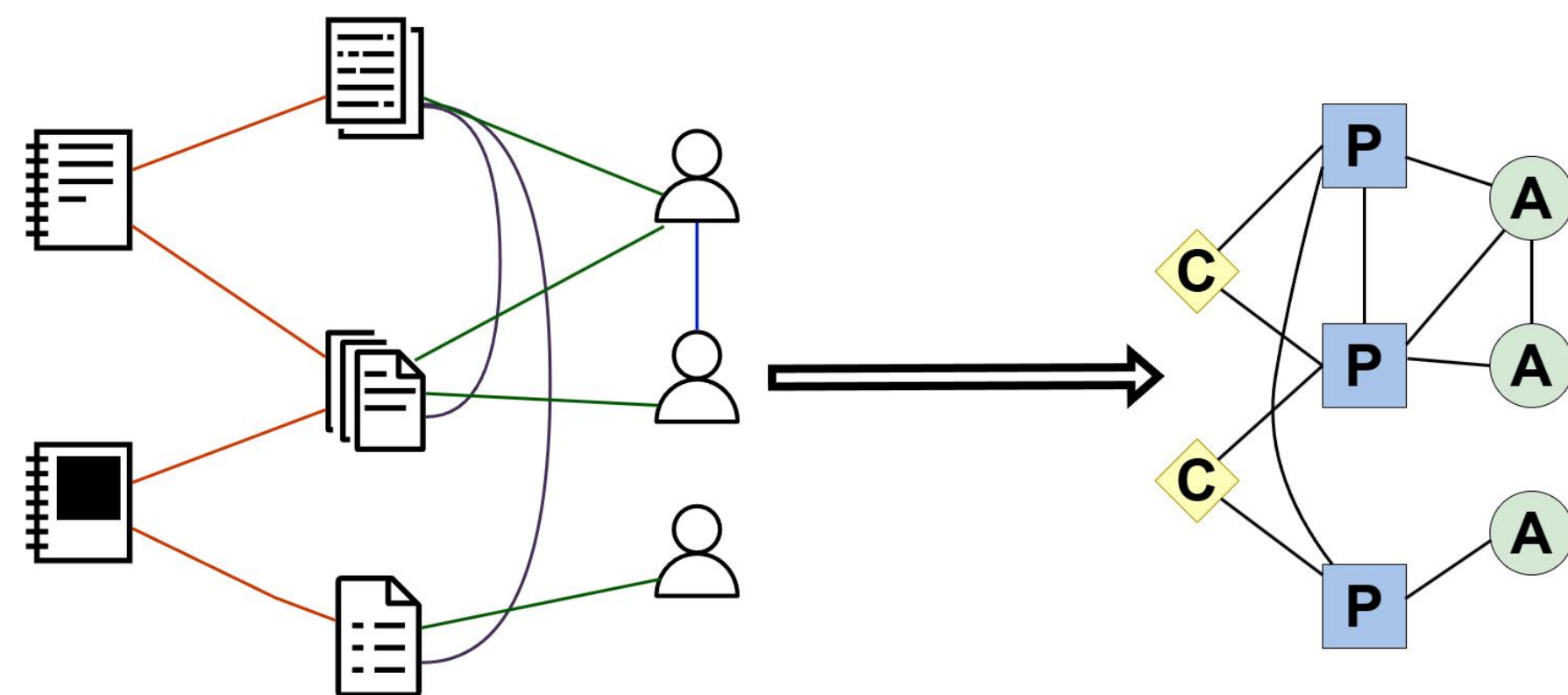
基于大型语言模型的广义异构图表示学习方法

章晨昊, 吴凤鸽, 赵军锁, 郑昌文, 刘华平

AAAI2025, 吴凤鸽, fengge@iscas.ac.cn

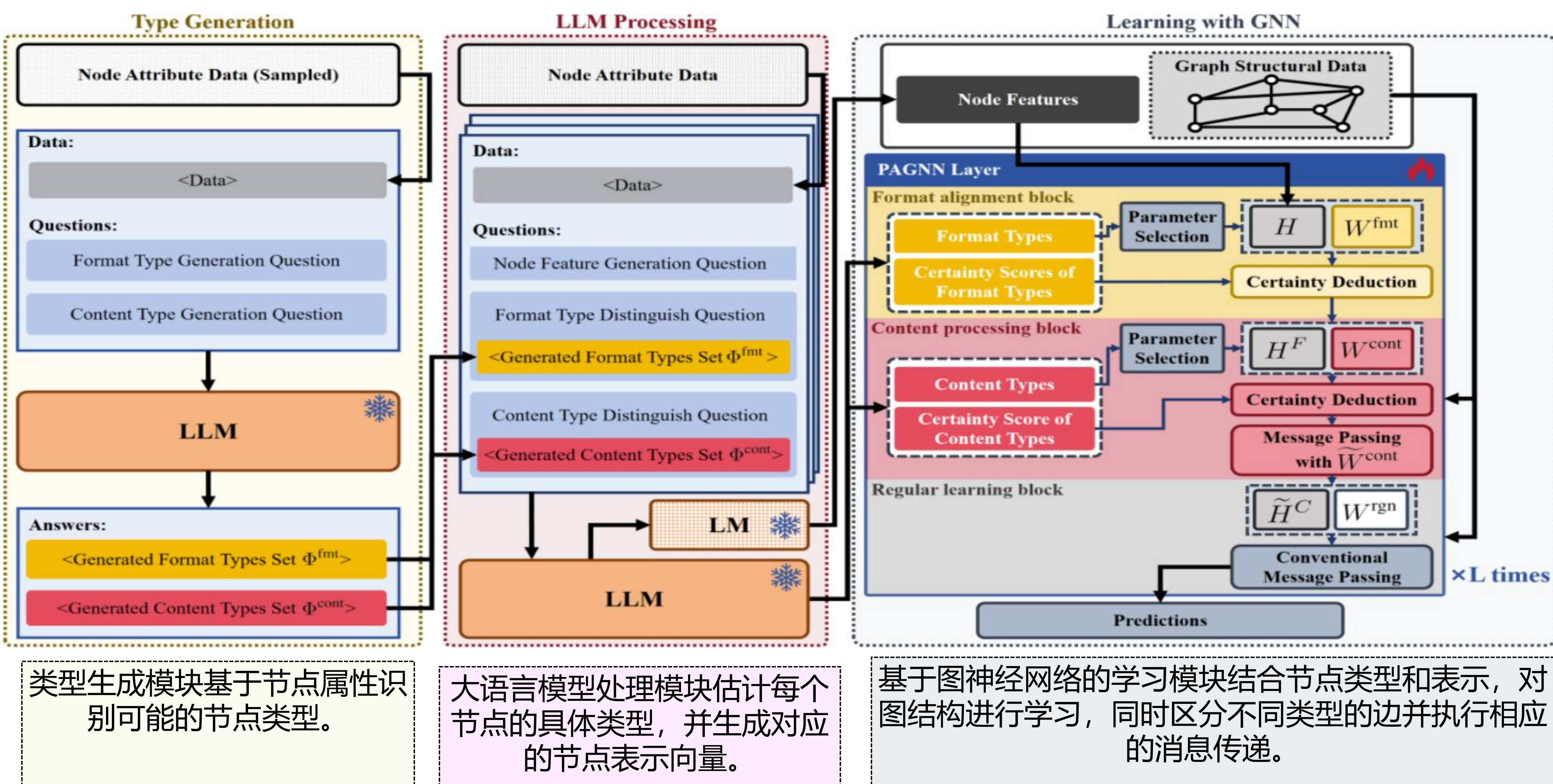
研究背景

- 复杂异构非结构化数据的处理通常需要大量的人力和资源。
- 现有的异构图神经网络因依赖先验知识和统一特征格式。
- 基于大语言模型的图表示学习方法能够对多种图表示进行统一对齐, 但通常忽略了异构图数据的复杂特性, 并且需要耗时的预处理步骤。



研究方法

- GHGRL方法将大语言模型与图神经网络的优势相结合, 能够在无需显式类型信息或特殊预处理的情况下处理任意格式和类型的节点与边的图数据。



效果评估

- 实验结果表明GHGRL在处理复杂异构数据时具有显著优势。与传统的异构图神经网络方法相比, GHGRL不仅在精度上具有明显提升, 还在处理灵活性和适应多样化数据方面表现出更高的鲁棒性。

Datasets	IMDB (10% Training)		IMDB (40% Training)		DBLP (10% Training)		DBLP (40% Training)		ACM (10% Training)		ACM (40% Training)	
Metrics	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1
GCN	57.47±0.72	58.43 ±1.15	60.13±0.76	60.38±1.19	89.09±0.32	89.8±0.34	88.94±0.38	89.61±0.40	89.47±0.23	90.23±0.24	89.19±0.28	89.95±0.27
GAT	60.12±0.79	60.79±1.26	62.85±1.28	63.1±0.83	89.66±0.26	90.93±0.23	91.40±0.19	91.79±0.21	92.23±0.96	92.27±0.95	92.26±0.86	92.38±0.81
HAN	61.28±0.12	61.26±0.15	62.78±0.38	62.15±0.26	91.23±0.51	92.10±0.62	91.92±0.48	92.52±0.59	90.58±0.40	90.56±0.39	92.70±0.45	92.75±0.42
MAGNN	57.78±2.85	57.97±1.82	59.92±1.24	60.07±0.89	92.24±0.49	92.70±0.51	93.21±0.42	93.68±0.43	89.46±0.64	89.71±0.53	91.25±0.24	91.33±0.35
SeHGNN	61.23±0.46	62.74±0.37	62.62±0.35	65.34±0.30	93.74±0.28	94.19±0.24	94.48±0.12	94.85±0.15	92.06±0.32	92.10±0.32	93.38±0.30	93.44±0.36
PSHGNN	61.35±0.79	62.25±0.42	67.21±0.66	67.55±0.56	92.89±0.09	93.46±0.07	93.98±0.12	94.29±0.10	91.07±0.26	91.00±0.24	93.78±0.23	93.77±0.19
HAN-w	58.31±0.32	58.26±0.31	59.83±0.33	59.02±0.35	87.54±0.78	87.93±0.58	88.16±0.68	88.64±0.69	90.08±0.34	90.02±0.32	91.98±0.38	91.86±0.37
MAGNN-w	57.02±1.23	57.36±0.86	59.44±1.06	59.76±0.93	90.24±0.49	90.65±0.63	90.32±0.77	91.32±0.82	88.95±0.18	89.26±0.20	91.23±0.16	91.33±0.25
SHEGNN-w	59.56±0.78	61.30±1.34	61.76±0.62	65.24±0.73	89.32±0.28	89.96±0.24	91.57±0.21	91.71±0.22	91.87±0.48	91.81±0.36	92.45±0.42	92.50±0.44
PSHGNN-w	59.68±0.55	61.04±0.36	65.52±0.52	66.03±0.44	89.78±0.23	90.46±0.25	91.58±0.12	91.93±0.10	91.01±0.26	90.97±0.24	92.97±0.23	93.02±0.19
TAPE	50.69±0.30	51.06±0.53	53.36±0.16	53.32±0.35	70.56±0.45	70.27±0.58	75.01±0.44	76.15±0.32	78.26±0.95	78.63±0.97	88.91±0.76	88.81±0.62
OFA	21.50±0.05	21.13 ±0.05	22.73±0.05	22.6±0.05	20.80±0.10	20.79±0.12	30.52±0.08	29.89±0.12	72.63±0.23	72.34±0.16	80.32±0.24	80.65±0.28
GOFA	32.12±0.20	32.29±0.16	33.75±0.51	33.82±0.26	35.90±0.36	35.43±0.38	44.03±0.50	44.52±0.41	78.91±0.56	78.92±0.73	84.28±0.33	84.21±0.79
GHGRL	69.73±0.53	70.11 ±0.57	72.13±0.64	72.46±0.62	89.85±0.23	90.48±0.18	91.67±0.34	92.17±0.32	92.71±0.36	92.69±0.30	94.21±0.44	94.63±0.42