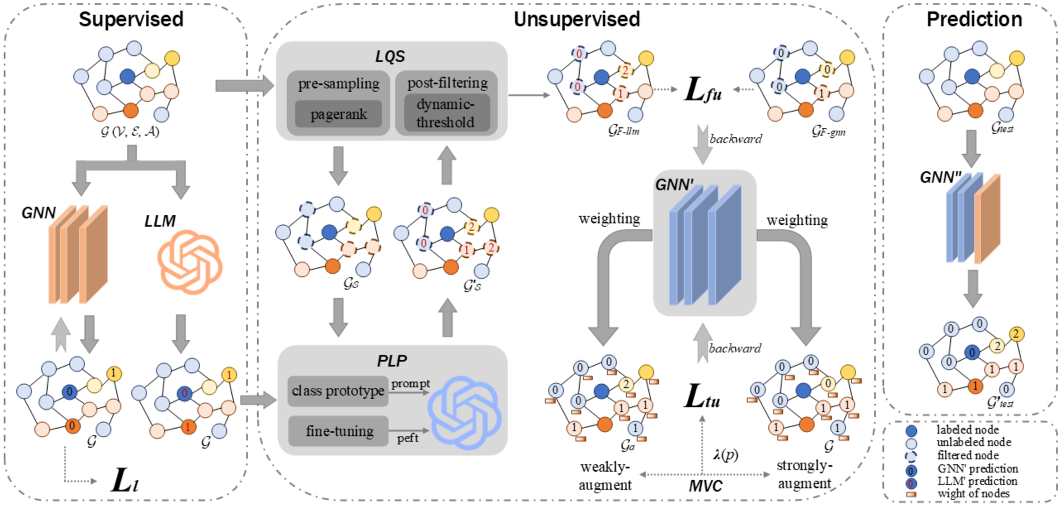
APL-LLM: Adaptive Pseudo-Labeling with Large Language Models for Few-shot Node Classification

Junyi LI, Chenweinan JIANG, Daixin WANG, Guo YE, Libang ZHANG, Huimei HE, Binbin HU, Zhiqiang ZHANG, Fuzhen ZHUANG

Frontiers of Computer Science, DOI: 10.1007/s11704-025-51174-6

Problems & Ideas

- Problems of conventional graph pseudo-labeling methods:
  - susceptibility to prediction errors from the base model.
  - issues caused by unforeseeable threshold settings
  - vulnerability to data distribution.
- Ideas: A label generation chain integrating LLM pseudo-labeling with prototypes, label quality screening and multi-view consistency to enhance downstream graph learning.



Overview of APL-LLM. It is composed of: (1) pseudo-labeling with prototypes makes summaries for each class, and generates pseudo-labels on the node set. (2) label quality screening filters high quality labels via pre-sampling and dynamic threshold mechanism. (3) training with multi-view enhanced consistency and finally prediction.

Main Contributions

- Contributions:
  - A general benchmark in utilizing LLMs predictions for graph-oriented tasks, with class prototypes to gain semantic guidance;
  - A novel approach for few-shot graph learning, named APL-LLM, where PLP and LQS interact to form a reliable label generation chain, and MVC mines the potential information of filtered data;
  - Extensive experiments show the superiority of APL-LLM over powerful baselines. It can effectively alleviate the problem of label sparsity.

Methods	Cora			Citeseer			PubMed			WikiCS		
	1-shot	3-shot	5-shot	1-shot	3-shot	5-shot	1-shot	3-shot	5-shot	1-shot	3-shot	5-shot
GCN	61.95 ± 2.03	63.50 ± 1.67	69.88 ± 1.24	53.82 ± 1.35	55.59 ± 1.07	59.05 ± 0.93	60.87 ± 1.92	64.03 ± 1.59	68.16 ± 1.21	51.59 ± 1.29	54.13 ± 1.43	56.90 ± 1.59
SAGE	62.84 ± 1.51	65.26 ± 1.45	69.09 ± 1.38	54.33 ± 1.16	57.02 ± 0.87	59.86 ± 0.98	62.74 ± 1.73	65.07 ± 1.38	67.29 ± 1.06	52.90 ± 1.19	55.08 ± 1.56	57.63 ± 1.21
Confidence	61.51 ± 1.68	64.91 ± 1.13	72.21 ± 1.36	54.98 ± 1.29	60.85 ± 1.03	65.59 ± 0.88	62.33 ± 1.65	64.96 ± 1.31	71.93 ± 0.87	50.83 ± 1.43	55.99 ± 1.37	63.05 ± 0.92
Li.et.	75.44 ± 0.13	79.15 ± 0.67	81.75 ± 0.82	64.59 ± 0.87	70.63 ± 0.64	72.93 ± 0.53	69.83 ± 0.65	76.60 ± 0.28	81.87 ± 0.11	65.54 ± 0.44	68.13 ± 0.66	70.97 ± 0.37
MoDis-MS	74.97 ± 0.54	78.76 ± 0.76	82.27 ± 0.15	67.15 ± 0.35	71.29 ± 0.55	72.24 ± 0.74	69.37 ± 0.74	72.12 ± 0.57	80.59 ± 0.56	66.95 ± 0.33	69.88 ± 1.08	71.43 ± 0.58
GLIM	75.21 ± 0.82	77.08 ± 0.71	81.26 ± 0.57	62.39 ± 0.83	65.03 ± 0.61	71.65 ± 0.35	67.94 ± 1.05	71.76 ± 0.83	76.58 ± 0.79	62.57 ± 0.74	66.12 ± 0.92	69.25 ± 1.26
LLM-GNN	76.11 ± 0.38	78.43 ± 0.55	80.62 ± 0.92	69.16 ± 0.39	71.01 ± 0.35	72.76 ± 0.77	79.84 ± 0.85	80.23 ± 0.49	80.45 ± 0.64	68.83 ± 0.93	70.06 ± 0.68	71.05 ± 0.42
APL-LLM	**77.35 ± 0.26**	**80.57 ± 0.35**	**83.16 ± 0.64**	**70.55 ± 0.21**	**72.42 ± 0.74**	**73.98 ± 0.50**	**80.58 ± 0.07**	**81.15 ± 0.17**	**82.67 ± 0.35**	**70.32 ± 0.45**	**71.59 ± 0.31**	**73.34 ± 0.18**

The node classification accuracy of algorithms (%). Bold indicates the best, underline indicates the second best.