

# LLM-Empowered Few-Shot Node Classification on Incomplete Graphs with Real Node Degrees

## 基于真实度不完整图的大模型驱动少样本节点分类方法

李云, 杨翊, 朱嘉奇, 陈辉, 王宏安

CIKM 2024 (联系人: 杨翊 yangyi2012@iscas.ac.cn)

### 背景介绍

在现实场景中构建的图往往是不完整的, 这主要是由于隐私限制或资源有限等原因。在许多不完整图的场景中, 真实的节点度(比如社交网络中的粉丝数量、引文网络中出版物的参考文献数量等)是易于获取且包含丰富信息的, 它们能够反映出图的不完整程度。

### 主要贡献

(1) 提出了一个新的研究问题, 即基于带有真实节点度的不完整图节点分类; (2) 提出了一种迭代训练模型LLMDGCN, 整合LLMs和真实节点度以解决数据缺失问题; (3) 在四个基准数据集上进行的大量实验表明, 我们所提出的针对带有真实节点度的不完整图的节点分类方法具有有效性和鲁棒性。

### 核心方法

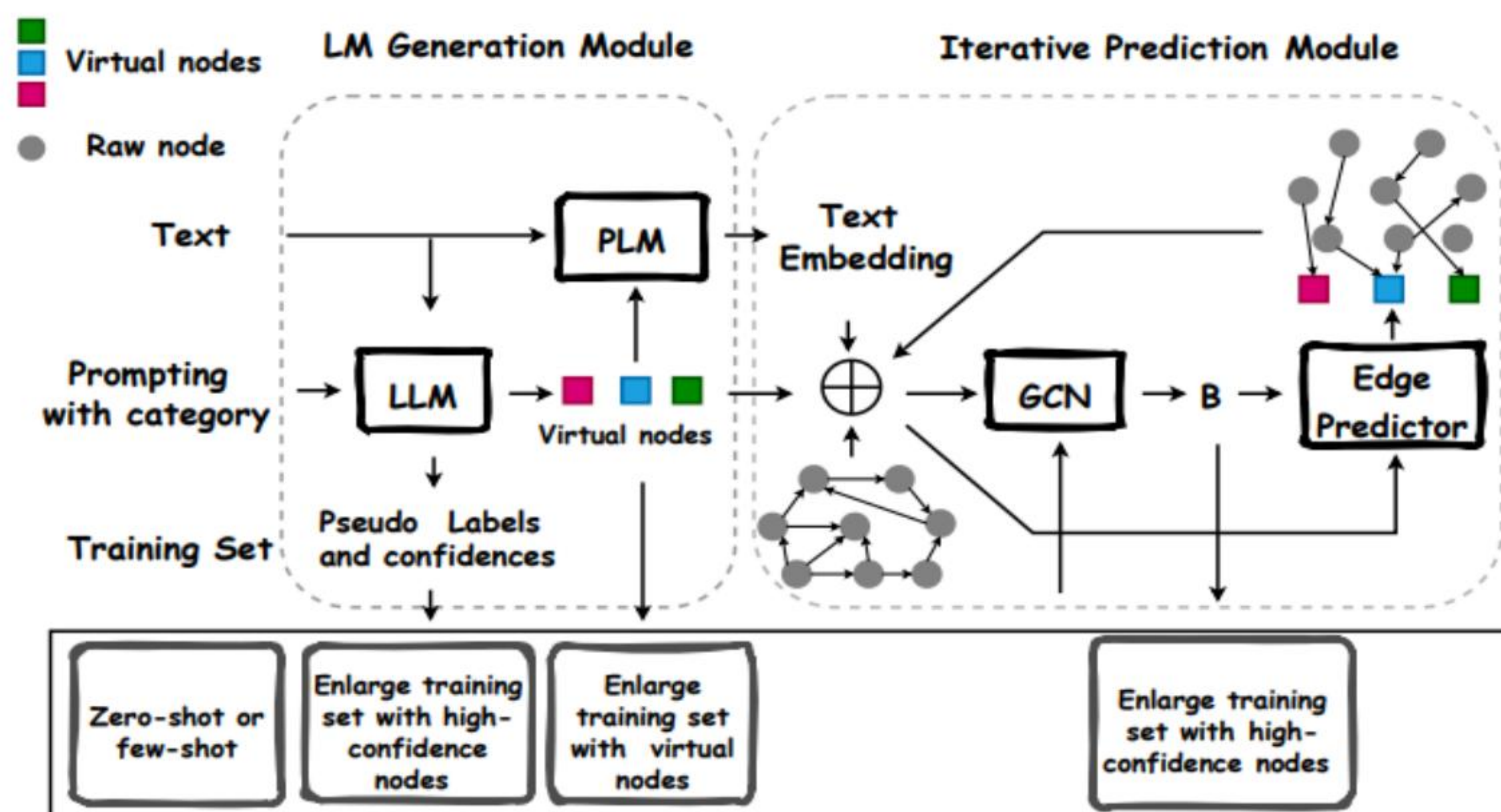


图1: LLMDGCN模型整体架构

- 语言模型生成模块: 利用LLMs补充不完整图中缺失的节点和初始伪标签, 再通过预训练语言模型获取节点初始嵌入。
- 迭代预测模块: 迭代训练图卷积网络(GCN)和边预测器: GCN生成伪标签, 边预测器结合真实节点度和类别间概率矩阵恢复缺失边, 更新后的图反哺GCN以提升伪标签准确性, 同时将高置信度节点加入训练集, 逐步优化分类效果

### 实验结果

Methods	0-shot	1-shot	2-shot	3-shot	4-shot	5-shot	10-shot
GCN	-	56.64(5.57)	59.20(5.90)	64.84(2.77)	66.24(2.02)	68.08(2.22)	71.40(1.51)
GAT	-	53.00(6.36)	58.36(4.71)	64.20(3.09)	64.96(2.43)	67.60(2.63)	71.20(1.36)
T2-GNN	-	53.82(5.54)	58.92(6.87)	63.92(4.43)	65.68(3.37)	67.32(2.13)	69.58(1.24)
D <sup>2</sup> PT	-	66.72(4.38)	68.26(4.26)	67.60(3.80)	67.60(2.91)	68.82(2.12)	68.22(1.80)
LLMGNN	69.48(2.52)	68.60(2.88)	70.00(2.43)	70.16(3.61)	70.74(2.83)	71.00(1.91)	71.88(2.21)
ENG	68.98(1.36)	68.52(1.24)	69.80(1.73)	69.92(1.79)	70.44(0.77)	70.10(1.28)	71.20(1.53)
<b>LLMDGCN</b>	<b>72.48(1.11)</b>	<b>73.04(1.01)</b>	<b>72.62(0.62)</b>	<b>72.66(1.25)</b>	<b>72.74(1.09)</b>	<b>73.18(1.62)</b>	<b>73.28(0.95)</b>

表1: Citeseer数据集的小样本节点分类准确率

Datasets	Cora		Citeseer		Pubmed		Religion	
	0-shot	1-shot	0-shot	1-shot	0-shot	1-shot	0-shot	1-shot
LLMDGCN(-V <sub>N</sub> , -V <sub>P</sub> )	-	59.50(7.83)	-	61.92(6.36)	-	63.32(3.36)	-	32.72(10.32)
LLMDGCN(-V <sub>N</sub> )	74.78(0.62)	76.14(1.85)	69.70(1.12)	70.46(1.15)	78.74(0.89)	77.72(2.22)	57.44(1.56)	59.86(3.04)
LLMDGCN(-V <sub>P</sub> )	73.42(0.38)	75.74(1.33)	72.38(0.87)	71.98(0.94)	73.20(1.67)	74.16(2.53)	52.14(2.02)	54.44(2.17)
LLMDGCN(-Edge, -Iter)	74.56(0.55)	74.56(1.87)	70.86(0.53)	70.94(1.15)	79.44(0.73)	79.16(1.77)	59.76(1.30)	60.98(1.63)
LLMDGCN(-Iter)	76.08(1.30)	74.76(2.04)	72.08(0.91)	72.86(1.00)	79.56(0.64)	79.16(1.53)	60.04(1.42)	61.08(1.55)
LLMDGCN(-Edge)	76.00(1.47)	75.16(1.52)	72.16(1.24)	72.24(0.93)	79.42(0.74)	79.08(2.21)	60.06(1.43)	60.94(1.35)
<b>LLMDGCN</b>	<b>76.20(1.20)</b>	<b>76.10(0.99)</b>	<b>72.48(1.11)</b>	<b>73.04(1.01)</b>	<b>79.52(0.68)</b>	<b>79.18(2.12)</b>	<b>60.22(1.01)</b>	<b>61.88(2.02)</b>

表2: 消融实验

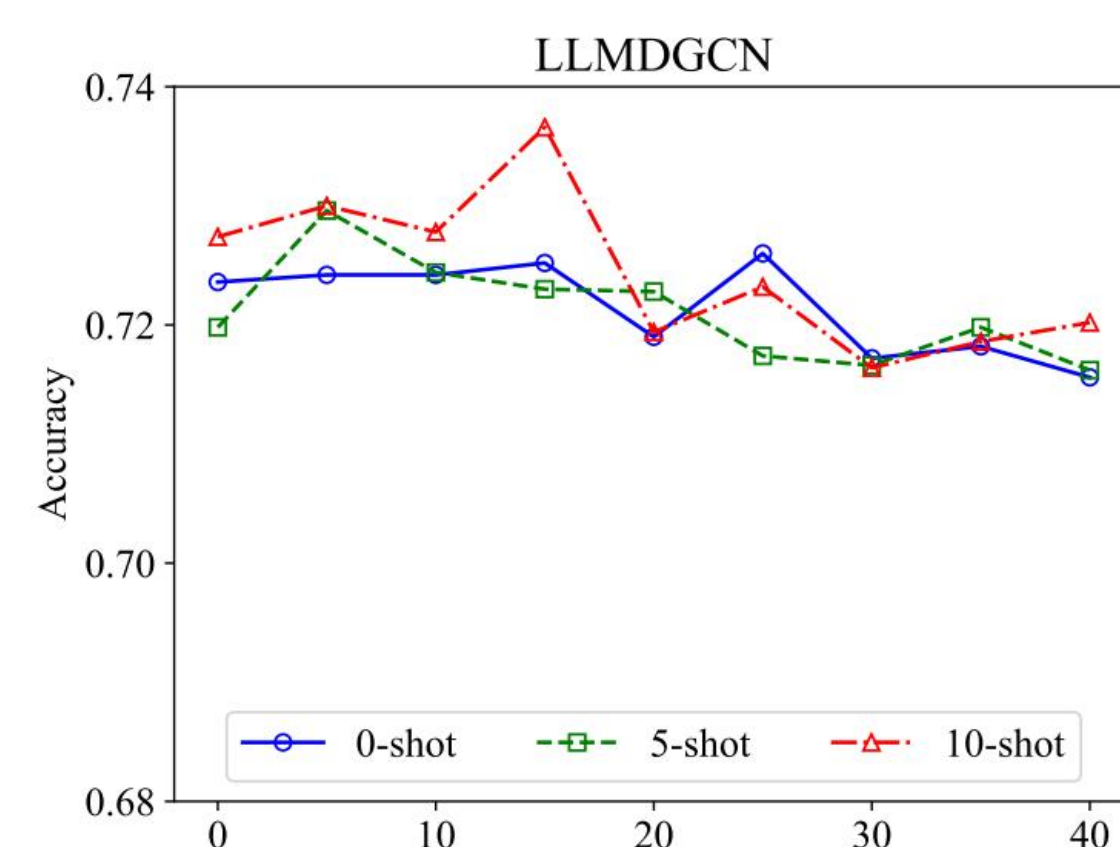


图2: 伪标签数量分析

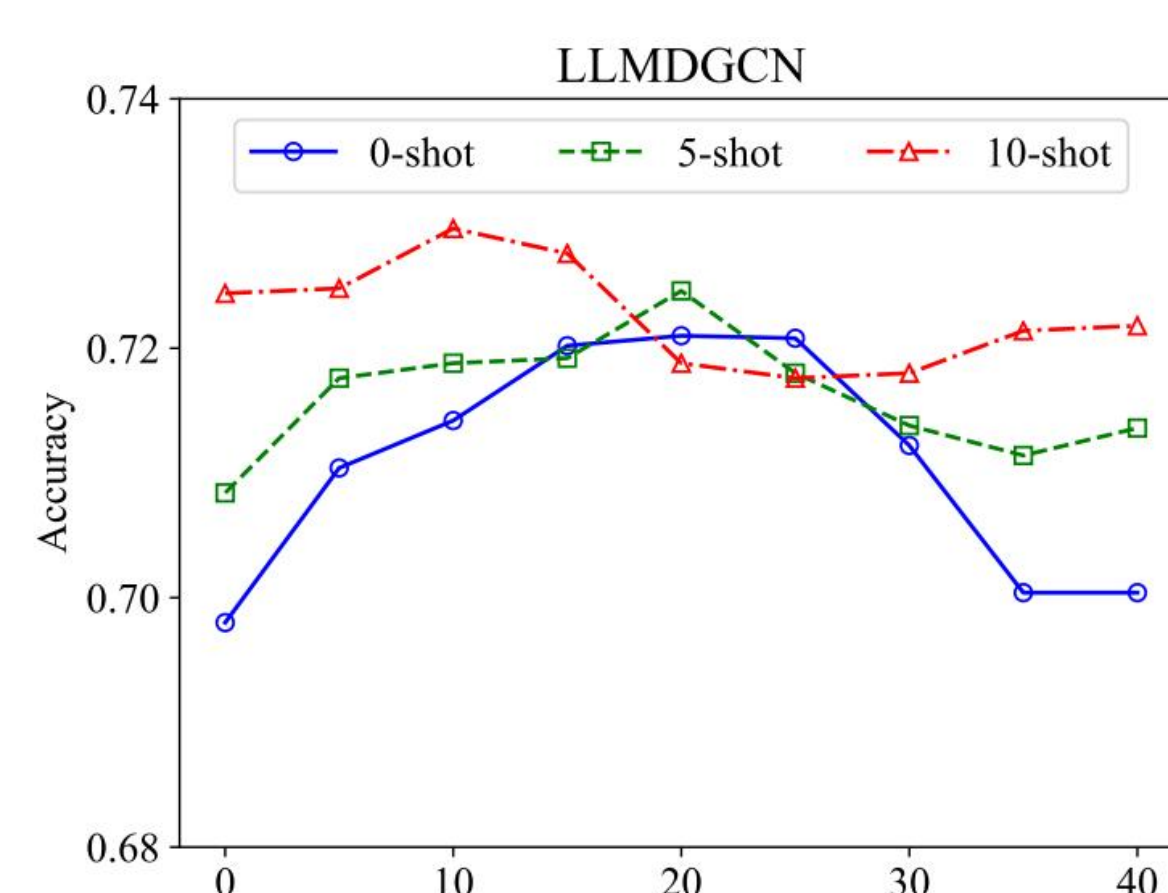


图3: 伪节点数量分析