

## 结构影响力及标签冲突感知的图课程学习方法

刘祖龙, 陈可佳

引用本文

刘祖龙, 陈可佳. 结构影响力及标签冲突感知的图课程学习方法[J]. 计算机科学, 2024, 51(10): 227-233.

LIU Zulong, CHEN Kejia. [Structural Influence and Label Conflict Aware Based Graph Curriculum Learning Approach](#) [J]. Computer Science, 2024, 51(10): 227-233.

---

## 相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

**Similar articles recommended (Please use Firefox or IE to view the article)**

[NLGAE:一种基于改进网络结构及损失函数的图自编码器节点分类模型](#)

NLGAE:A Graph Autoencoder Model Based on Improved Network Structure and Loss Functionfor Node Classification Task

计算机科学, 2024, 51(10): 234-246. <https://doi.org/10.11896/jsjcx.230700122>

[基于深度学习的Linux系统DKOM攻击检测](#)

Deep-learning Based DKOM Attack Detection for Linux System

计算机科学, 2024, 51(9): 383-392. <https://doi.org/10.11896/jsjcx.230700035>

[基于图神经网络的SSL/TLS加密恶意流量检测算法研究](#)

Study on SSL/TLS Encrypted Malicious Traffic Detection Algorithm Based on Graph Neural Networks

计算机科学, 2024, 51(9): 365-370. <https://doi.org/10.11896/jsjcx.230800079>

[基于知识图谱与邻域感知注意力机制的推荐算法研究](#)

Study on Recommendation Algorithms Based on Knowledge Graph and Neighbor Perception Attention Mechanism

计算机科学, 2024, 51(8): 313-323. <https://doi.org/10.11896/jsjcx.230500143>

[基于标签传播增强的多通道图卷积网络](#)

Multi-channel Graph Convolutional Networks Enhanced by Label Propagation Algorithm

计算机科学, 2024, 51(8): 304-312. <https://doi.org/10.11896/jsjcx.240100139>

# 结构影响力及标签冲突感知的图课程学习方法

刘祖龙<sup>1</sup> 陈可佳<sup>1,2,3</sup>

1 南京邮电大学计算机学院 南京 210023

2 江苏省大数据安全与智能处理重点实验室(南京邮电大学) 南京 210023

3 南京大学软件新技术重点实验室 南京 210023

(1021041210@njupt.edu.cn)

**摘要** 近年来,图神经网络(GNNs)已成为图学习领域的热点研究问题。受益于消息传递机制,GNNs在各类基于图的任务上均取得了优越的性能。现有的GNNs方法大多基于图中所有节点的训练难度相同的假设,然而,节点在结构影响力和邻域标签异配性等方面具有明显的差异。为此,提出了一种结构影响力及标签冲突感知的图课程学习方法(SILC-GCL),基于节点的训练难度对GNNs模型进行课程学习。首先,设计了一种综合考虑节点的PageRank影响力值以及邻域标签冲突程度的训练难度测量器;其次,采用了一个训练调度器,用于在每个训练阶段选择训练难度合适的节点并生成一个由易到难的训练节点序列;最后在GNNs骨架模型上进行训练。在6个现实网络数据集上进行的节点分类实验均验证了SILC-GCL方法的有效性。

**关键词:**图表示学习;图神经网络;课程学习;节点分类

**中图分类号** TP391

## Structural Influence and Label Conflict Aware Based Graph Curriculum Learning Approach

LIU Zulong<sup>1</sup> and CHEN Kejia<sup>1,2,3</sup>

1 School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

2 Jiangsu Key Laboratory of Big Data Security & Intelligent Processing (Nanjing University of Posts and Telecommunications), Nanjing 210023, China

3 State Key Laboratory for Novel Software Technology at Nanjing University, Nanjing 210023, China

**Abstract** In recent years, graph neural networks(GNNs) have emerged as a prominent research area in the field of graph learning. Leveraging the message passing mechanism, GNNs have showcased remarkable performance across diverse graph-based tasks. However, most existing GNNs methods assume uniform training difficulty across all nodes, disregarding the significant variability in the importance and contributions of different nodes. To address this problem, this paper proposes a structural influence and label conflict aware graph curriculum learning method(SILC-GCL), which takes the training difficulty of nodes into consideration. To begin with, a difficulty measure is designed through both the graph structure and node label semantics, calculating the PageRank value of nodes and the label conflict degree between nodes and their neighbors. Subsequently, a training scheduler is employed to select nodes with appropriate training difficulty at each training stage and then generate a sequence of training nodes from easy to difficult. Finally, SILC-GCL is trained based on backbone GNNs models. Experimental results of node classification on six benchmark datasets verify the effectiveness of SILC-GCL.

**Keywords** Graph representation learning, Graph neural networks, Curriculum learning, Node classification

## 1 引言

图是一种关系型的数据结构,被广泛用于对现实网络(如社会网络、蛋白质网络、语义网络等)进行建模。近年来,图神经

网络(Graph Neural Networks, GNNs)成为图机器学习领域的新兴研究方向,在节点分类、链接预测和图分类等基于图的任务上取得了比传统模型更加优越的性能<sup>[1-6]</sup>。GNNs主要通过聚合节点的邻域信息来学习节点的低维表示,其借助

到稿日期:2023-08-25 返修日期:2023-01-04

基金项目:国家自然科学基金(61876091);南京大学计算机软件新技术国家重点实验室开放课题(KFKT2022B01);南京邮电大学校级科研基金(NY221071)

This work was supported by the National Natural Science Foundation of China(61876091), Foundation of State Key Laboratory for Novel Software Technology at Nanjing University(KFKT2022B01) and Research Foundation of Nanjing University of Posts and Telecommunications(NY221071).

通信作者:陈可佳(chenkj@njupt.edu.cn)

消息传递和邻域聚合机制,能够保留节点的特征信息以及图的拓扑结构信息。

然而,现有的 GNNs 方法大多基于节点重要性相同的假设<sup>[7-8]</sup>。事实上,图中节点的重要性及其对学习的贡献往往存在差异。以节点的度(Degree)为例,通常图中只有少数节点拥有大量的连接,这使得节点度的分布呈现出无标度的幂律分布<sup>[9]</sup>。例如,图 1 展示了 3 个现实网络数据集中节点度的分布情况,其中纵坐标为节点度数,横坐标为各度数的节点数量的归一化结果。相比节点度,PageRank 值<sup>[10]</sup>更能有效表示节点重要性程度。因此,本文将 PageRank 值高的节点称为枢纽(Hub)节点,将 PageRank 值低的节点称为尾(Tail)节点。枢纽节点在图上拥有更大的重要性以及结构影响力,在 GNNs 的消息传递机制下能够聚合到更充分的局部邻域信息,学习到更高质量的节点表示;相反,尾节点则难以聚合到充分的邻域信息,节点表示效果较差。图 2 给出了 GCN 模型<sup>[7]</sup>在上述 3 个数据集上的节点分类结果,即对枢纽节点和尾节点分别统计其 Micro-F1 分数。由图 2 可知,GCN 在尾节点的分类性能显著低于枢纽节点的分类性能。

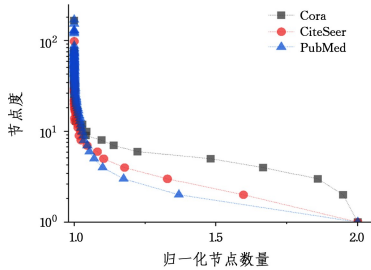


图 1 3 个现实网络数据集中节点度的分布情况  
Fig. 1 Node degree distribution of three datasets

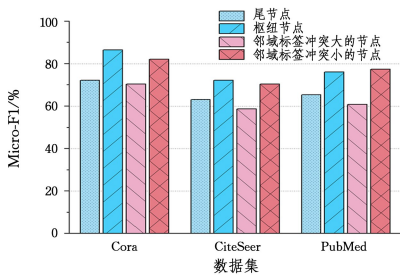


图 2 GCN 在不同难度节点上的 Micro-F1 分数  
Fig. 2 Micro-F1 scores of GCN trained on nodes with different difficulties

此外,大多 GNNs 模型还基于同质性或同配性(Homophily)的假设<sup>[11-12]</sup>,即节点倾向于连接属于同一类别的节点。然而,现实网络并不一定满足这一假设,图上节点与其邻居可能具有不同的标签,这一现象也被称为异配性(Heterophily)<sup>[13-14]</sup>。此时,节点聚合邻居信息时,会混入不同类别节点的特征,难以学习到类别相关的节点表示。如图 2 所示,GCN 在邻域标签冲突较大(即具有较多不同标签的邻居)时的节点分类结果显著劣于邻域标签冲突较小时的结果。

在机器学习领域,课程学习(Curriculum Learning)<sup>[15]</sup>是一种基于样本学习难度的训练策略。在模型训练初期,该策略首先选择训练难度较小的训练样本,然后逐步引入训练

难度更大的样本,直到模型收敛。课程学习在初始训练阶段排除了学习困难的样本,缓解了对噪声样本的过拟合问题,从而提高了模型的准确性以及鲁棒性。课程学习中最关键的部分是难度测量器,即如何衡量样本的训练难度。

针对现有 GNNs 存在的上述问题,本文提出了一种结构影响力及标签冲突感知的图课程学习方法(Structural Influence and Label Conflict aware Graph Curriculum Learning method, SILC-GCL)。该方法的主要目标是定义节点样本的训练难度并在训练过程中按从易到难的顺序逐步引入节点样本,从而增强 GNNs 的学习性能。具体来说,本文为图上的课程学习框架设计了一种同时考虑节点 PageRank 值以及邻域标签不一致性的样本难度测量器,即节点的 PageRank 值越大、与邻居之间的标签冲突越小,其训练难度就越小。此外,还采用了一个训练调度器,用于在每个训练阶段选择训练难度合适的节点,以减轻训练困难节点的负面影响。

## 2 相关工作

### 2.1 节点影响力感知的 GNNs

目前,已有一些工作关注到了图上的节点具有不同影响力的问题,并采用了不同的策略以缓解该问题<sup>[16-20]</sup>。Tail-GNN<sup>[16]</sup>根据节点度的高低来划分头、尾节点,通过学习到的邻域迁移向量,将头节点丰富的邻域信息迁移到尾节点,使得 GNNs 能够更有效地学习尾节点的表示。ColdBrew<sup>[17]</sup>利用教师-学生蒸馏模型解决 GNNs 的极端冷启动问题:教师模型学习图的整体结构信息,然后训练学生模型以模拟教师模型学习到的节点嵌入,从而为尾节点补充潜在的邻域信息。CenGCN<sup>[18]</sup>基于节点相似性与中心性重新分配枢纽节点、非枢纽节点与邻居之间的权重,从而缓解枢纽节点与非枢纽节点聚合邻居信息不均衡的问题。

### 2.2 图课程学习

课程学习是一种模仿人类学习认知过程的训练策略,旨在通过某种设计好的顺序(通常遵循由易到难的策略)对样本进行训练,以增强机器学习模型的泛化能力以及鲁棒性。目前,课程学习已被广泛应用于计算机视觉<sup>[21-22]</sup>、自然语言处理<sup>[23-24]</sup>等领域中。

在图机器学习领域,近期也出现了一系列采用课程学习框架的 GNNs 模型<sup>[25]</sup>。根据关注的下游任务不同,图课程学习可分为图级别<sup>[26-27]</sup>、链接级别<sup>[28]</sup>以及节点级别<sup>[29]</sup>等。本文主要关注节点级别的图课程学习方法。CLNode<sup>[29]</sup>首次将课程学习应用到了图学习中的节点分类任务,其定义了一个多视角的难度测量器,同时从节点标签以及节点特征两个角度来衡量节点的训练难度。GNN-CL<sup>[30]</sup>将课程学习与 SMOTE<sup>[31]</sup>生成器相结合,用于控制生成节点的数量,以解决节点类别不平衡的问题。HSAN<sup>[32]</sup>在图对比学习中引入课程学习框架,基于属性与结构相似度对负样本进行采样,逐渐引入训练难度更大的负样本。

## 3 预备知识

### 3.1 图神经网络

给定图  $G=(V,E,X)$ ,  $V$  为节点集,  $E$  为边集,  $X$  为节点

特征矩阵,节点 $v_i$ 的邻居节点集合表示为 $N(v_i)$ 。一般来说,GNNs学习节点表示包括两个关键步骤:1)聚合来自节点邻域的信息;2)从上一层学习的节点表示以及聚合的邻域信息中更新本层节点的表示。形式上,第 $l$ 层中节点 $v_i$ 表示为:

$$\mathbf{h}_{v_i}^l = f(\mathbf{h}_{v_i}^{l-1}, \{\mathbf{h}_u^{l-1} : u \in \mathcal{N}(v_i)\}; \Theta^l) \quad (1)$$

最终得到的节点表示 $\mathbf{h}_{v_i}^L$ 可用于各种下游任务。

### 3.2 节点分类

本文在节点分类任务上评估所提出的节点级图课程学习方法对GNNs性能的影响。

节点分类指利用一个分类器将节点表示映射到预测的标签上,映射的形式化表示为:

$$Y_{\text{pred}} = \text{classifier}(\sigma(\mathbf{H})) \quad (2)$$

其中,分类器一般选用多层感知机(MLP), $\mathbf{H}$ 为节点表示矩阵, $\sigma$ 为sigmoid函数。

### 3.3 课程学习

令 $\mathcal{C} = \langle Q_1, \dots, Q_t, \dots, Q_T \rangle$ 表示课程学习中的课程序列,

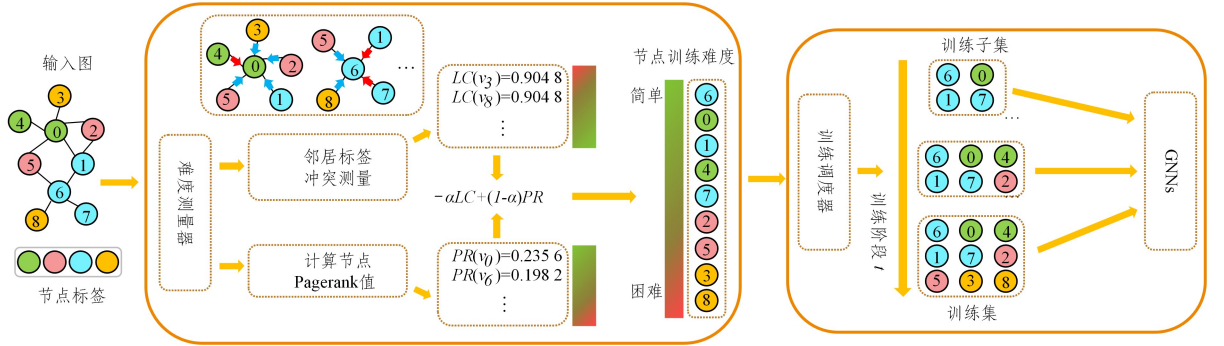


图3 SILC-GCL模型的框架

Fig. 3 Framework of SILC-GCL

## 4.1 训练难度测量器

### 4.1.1 计算PageRank值

如前文所述,节点的影响力越大,其局部信息的传播量也越大,学习也更加充分。常用的节点度(即节点的邻居数量)可以在一定程度上反映节点的局部重要性,但难以体现节点在全局结构中的影响。PageRank值的生成过程类似于GNNs的信息传递机制,能够更好地衡量节点的拓扑影响力。因此,本文提出将节点的PageRank值作为衡量节点训练难度的指标之一。

PageRank算法<sup>[10]</sup>的基本思想是在图上定义一个随机游走模型,即一阶马尔可夫链,描述随机游走者随机访问各个节点的行为。在一定条件下,访问各节点的概率会收敛达到平稳分布,此时节点的平稳访问概率值就是其PageRank值。

具体来说,节点 $v_i$ 的PageRank值可由式(3)得出:

$$PR(v_i) = d \left( \sum_{u \in \text{In}(v_i)} \frac{PR(u)}{|\mathcal{N}(u)|} \right) + \frac{1-d}{n} \quad (3)$$

其中, $\text{In}(v_i)$ 表示指向节点 $v_i$ 的节点集合; $|\mathcal{N}(u)|$ 表示节点 $u$ 的邻域集合; $d(0 \leq d \leq 1)$ 为阻尼因子,一般设定为0.85; $n$ 为图上节点的数量。计算无向图的PageRank值时,通常将其转换为有向图进行计算,即每条无向边对应两条有向边。经过若干次迭代后,得到所有训练节点的PageRank值的集合 $PR$ 。 $PR(v_i)$ 的值越大,节点 $v_i$ 具有更大的影响力,而GNNs

由 $T$ 个训练阶段上的训练数据子集组成。在第 $t$ 个训练阶段,数据子集 $Q_t$ 包含当前需要训练的节点样本。

$\mathcal{C}$ 中所有数据子集 $Q_t$ 的顺序由训练难度测量器和训练调度器两个部分决定。难度测量器首先估计每个样本 $x_i$ 的训练难度 $D(x_i)$ ,然后训练调度器根据 $D$ 在训练阶段 $t$ 生成 $Q_t$ 。

## 4 SILC-GCL方法

本文方法的整体框架如图3所示,主要设计了两个课程学习模块:

1)难度测量器:首先分别计算每个节点样本的PageRank值以及该节点与邻居节点之间的标签冲突程度。例如在图3中,节点0只有一个具有相同标签的邻居节点,节点6则有两个,因此节点0的邻域标签冲突程度要比节点6更高),然后综合这两个方面度量节点的训练难度。

2)训练调度器:基于节点的训练难度值设计训练度量器,分阶段地引入训练难度更大的节点。

也更易于学习 $v_i$ 的表示。

### 4.1.2 计算标签冲突程度

在现实图数据集中,节点和邻居之间的标签可能并不一致,这种情况被称为标签冲突或者异配性。邻域标签的不一致性会增加节点的学习难度,因此本文将节点标签的冲突程度作为衡量节点训练难度的另一个指标。

然而,大多数图数据集中的节点标签集合并不完整,在训练阶段仅能获取到少量节点的标签信息。因此,为了衡量节点邻域的标签冲突程度,本文首先基于节点分类任务预训练了一个GNN模型 $f_{\text{pre}}$ ,用于为图上的节点分配伪标签,具体可表示为:

$$\mathbf{H}_{\text{pre}} = f_{\text{pre}}(G) \quad (4)$$

$$\hat{Y} = \text{MLP}(\sigma(\mathbf{H}_{\text{pre}})) \quad (5)$$

其中, $\mathbf{H}_{\text{pre}} \in \mathbb{R}^{|\mathcal{V}| \times D}$ 表示由预训练的 $f_{\text{pre}}$ 学习得到的节点嵌入矩阵, $D$ 为节点嵌入的维度; $\hat{Y}$ 表示由多层感知机(MLP)为无标签节点预测的节点标签; $\tilde{Y}$ 为所有训练节点的标签集合,表示为:

$$\tilde{Y}_{v_i} = \begin{cases} Y_{v_i}, & v_i \in V_L \\ \hat{Y}_{v_i}, & v_i \in V_U \end{cases} \quad (6)$$

其中, $V_L$ 和 $V_U$ 分别表示有标签节点和无标签节点的集合,有标签节点保留其真实的标签。

随后,计算每个训练节点  $v_i$  与邻域节点之间的标签冲突程度。交叉熵(Cross Entropy)函数是分类任务中常用的一种损失函数,能够衡量两个不同分布之间的差异。因此,本文使用交叉熵函数来计算标签分布的冲突程度,具体表示为:

$$LC(v_i) = - \sum_{u \in \tilde{\mathcal{A}}(v_i)} \tilde{Y}_{v_i} \log(\tilde{Y}_u) \quad (7)$$

其中,邻域标签集合  $\tilde{\mathcal{A}}(v_i) = \mathcal{A}(v_i) \cup \{v_i\}$ ,即为该节点添加一个自环以考虑自身标签的影响。 $LC(v_i)$  的值越大,代表  $v_i$  与邻域之间的标签冲突程度越大,GNNs 也更难学习  $v_i$  的表示。

#### 4.1.3 综合训练难度

在得到节点重要程度  $PR$  以及标签冲突程度  $LC$  之后,本文综合评估节点的训练难度。由上文可知,更大的  $PR$  值代表节点更易于被训练,更大的  $LC$  值则代表节点更难以被训练,因此节点的最终训练难度可定义为:

$$D(v_i) = -\alpha LC(v_i) + (1-\alpha) PR(v_i) \quad (8)$$

其中,  $\alpha$  为超参数,用于控制两种训练难度值参与的权重。

#### 4.2 训练调度器

随后,本文设计了一种连续的训练调度器,以生成从易到难的训练节点序列,实现图课程学习。

连续训练调度器可视为一个步长函数  $s(t)$ , 将其将训练阶段  $t$  映射为一个标量  $\lambda \in (0, 1]$ , 即在训练阶段  $t$  时的训练样本的比例。该函数为单调非递减函数,从  $s(t_0) > 0$  开始,到  $s(T) = 1$  结束,  $T$  为训练阶段的总数。先前的工作已提出了各种步长函数<sup>[33-34]</sup>, 本文采用 3 种最常用的步长函数作为训练调度器,分别为:

1) 线性函数:

$$s(t) = \min\left(1, \lambda_0 + \frac{1-\lambda_0}{T_{\text{grow}}} \cdot t\right) \quad (9)$$

2) 根型函数:

$$s(t) = \min\left(1, \sqrt{\frac{1-\lambda_0^2}{T_{\text{grow}}} \cdot t + \lambda_0^2}\right) \quad (10)$$

3) 几何型函数:

$$s(t) = \min\left(1, 2^{(\log_2 \lambda_0 - \log_2 \lambda_0 * (t/T_{\text{grow}}))}\right) \quad (11)$$

其中,  $\lambda_0$  为简单训练样本比例的初始值,  $T_{\text{grow}}$  表示函数第一次取值为 1 的训练阶段。

线性函数以线性速度均匀增加训练样本的比例,根型函数可以更快引入困难的样本,而几何型函数则更关注简单样本的训练。本文实验部分对 3 种函数进行了更详细的分析。

## 5 实验

本文实验从以下 3 个方面对模型进行评价:

1) 基于节点分类任务,比较本文模型与相关基线方法的性能;

2) 对模型进行消融实验,以验证各个模块的有效性;

3) 分析模型对各项超参数的敏感性。

#### 5.1 数据集

本文在 7 个基准图数据集上进行了实验,其中 Cora<sup>[35]</sup>, CiteSeer<sup>[36]</sup>, PubMed<sup>[37]</sup> 以及 Arxiv<sup>[38]</sup> 为引文网络数据集,节点代表在计算机科学领域发表的论文,节点的特征为相应论文标题的词袋向量,边代表论文之间的引用关系。

Chameleon<sup>[39]</sup> 与 Squirrel<sup>[40]</sup> 是维基百科上具有特定主题的网络,节点代表网页,边代表网页之间的超链接。Actor<sup>[41]</sup> 为“电影-导演-演员-作者”网络的子图,仅包含演员节点,边表示这两位演员被收录在同一维基百科页面上。7 个数据集的详细统计信息如表 1 所列。同时,本文计算了各个数据集的平均标签冲突程度 Avg\_LC,用于衡量各个数据集的同配、异配性。

表 1 数据集统计信息

Table 1 Statistics of datasets

Dataset	Nodes	Edges	Features	Classes	Avg_LC
Cora	2708	5429	1433	7	0.2368
CiteSeer	3327	9270	3703	6	0.2808
PubMed	19717	88648	500	3	0.2172
Arxiv	169343	1166243	128	40	0.4569
Chameleon	2277	36101	2325	5	1.0954
Squirrel	5201	217073	2089	5	1.2467
Actor	7600	30019	932	5	1.1502

实验将数据集按照 7:2:1 的比例随机划分为训练集、验证集以及测试集,同时采用半监督学习设置,将训练集中每个类别的随机 20 个节点设置为有标签节点。

#### 5.2 比较方法

将本文方法与 3 个类别的对比方法进行了实验对比:

1) 3 种基准 GNNs, 包括 GCN<sup>[7]</sup>, GraphSAGE<sup>[8]</sup> 以及 GAT<sup>[42]</sup>, 这 3 种模型同时还用作本文的骨架(Backbone)模型;

2) 具有节点影响力感知的 GNNs, 包括 TailGNN<sup>[15]</sup>, ColdBrew<sup>[16]</sup> 以及 CenGCN<sup>[17]</sup>;

3) 图课程学习方法 CLNode<sup>[27]</sup>, 目前面向节点分类任务的方法较少,仅 CLNode 与本文工作最相关。

为了实验的公平性,所有比较方法均采用了其论文报告的最佳参数。

#### 5.3 节点分类结果

各方法在不同数据集上的节点分类 Micro-F1 和 Macro-F1 分数分别见表 2 和表 3, 包含各方法 10 次运行结果的平均结果以及标准差。考虑到本文方法可应用到不同的 GNN 模型中,实验将 3 个基线 GNN 方法(GCN, GraphSAGE, GAT) 作为骨架模型,以进一步验证本文方法的有效性和适用性。值得注意的是,由于 ColdBrew 以及 CenGCN 均以 GCN 为基础,难以拓展至其他 GNN 模型,因此本文仅在 GCN 作为骨架模型时与它们进行比较。

实验结果表明,本文方法在 7 个基准数据集上均取得了最优的分类结果,例如其 Micro-F1 值比骨架模型提升了 3%~4%。此外,在规模较大的数据集 Arxiv 上,本文方法仍能保持最优的性能,验证了该方法在现实复杂大规模图结构中的有效性。

ColdBrew 在节点分类任务上的性能提升不显著,这可能是由于其利用 MLP 作为学生模型,而 MLP 无法利用节点的邻域信息,因此难以以为尾节点找到潜在的邻居;TailGNN 的性能与原始模型相比有较为明显的提升,说明了区分头尾节点的必要性,但头节点邻域中的噪声信息也可能转移给尾节点,一定程度上会影响节点分类的性能;CenGCN 在 Cora, CiteSeer 以及 PubMed 上表现更佳,而在 Chameleon, Squirrel

以及 Actor 上表现欠佳,可能的原因是,该方法利用标签传播计算节点之间的权重,因此分类性能更依赖于数据集自身的标签质量。

CLNode 总体上取得了次优的结果,验证了课程学习框架的有效性,但其结果略低于本文的 SILC-GCL 方法,这说明基于节点特征相似度的难度测量器效果不如基于节点结构影响力的难度测量器。在 Squirrel 和 Actor 数据集上,CLNode+

GCN 和 CLNode+SAGE 的性能略低于 TailGNN,再一次验证了考虑节点的影响力有助于模型性能的提升。同时,CLNode 中两种指标的计算均基于强同配性假设,这使得模型性能可能受到数据集标签异配程度的影响。

总体来说,本文方法考虑了节点结构影响力的差异,并考虑了该差异对节点训练难度的影响,这是其分类性能优于 CLNode 的主要原因。

表 2 节点分类的 Micro-F1 分数

Table 2 Micro-F1 scores of node classification

Methods	Cora	CiteSeer	PubMed	Chameleon	Squirrel	Actor	Arxiv
GCN	83.72±0.67	64.74±1.32	73.34±0.57	34.37±2.16	26.08±1.01	22.53±1.63	70.26±1.94
ColdBrew	84.34±0.56	64.31±0.61	74.32±0.48	35.14±1.24	26.73±1.29	23.87±1.24	70.91±1.46
TailGCN	84.97±1.21	66.17±1.29	74.58±0.61	37.21±1.66	29.48±1.35	25.79±1.23	OOM
CenGCN	85.75±1.39	66.78±1.14	OOM	35.69±1.38	27.84±1.25	24.17±1.59	OOM
CLNode+GCN	86.91±0.75	67.59±1.24	75.64±1.31	37.74±1.47	28.91±1.52	25.34±1.67	71.69±1.62
SILC-GCL+GCN	<b>87.24±0.93</b>	<b>68.14±0.86</b>	<b>77.31±0.50</b>	<b>38.18±1.31</b>	<b>29.97±1.20</b>	<b>26.97±1.44</b>	<b>72.93±1.53</b>
GraphSAGE	80.11±1.41	65.09±1.47	73.33±0.42	35.96±2.18	29.90±1.69	29.67±1.91	69.73±1.81
TailSAGE	82.71±1.35	66.85±1.21	75.63±0.58	38.91±2.33	31.91±1.41	31.94±1.69	OOM
CLNode+SAGE	82.91±1.41	67.35±1.90	76.30±0.67	38.97±2.39	31.74±1.93	32.05±1.53	71.26±1.73
SILC-GCL+SAGE	<b>83.69±1.27</b>	<b>68.03±1.18</b>	<b>77.16±0.41</b>	<b>41.30±1.76</b>	<b>32.42±1.76</b>	<b>32.81±1.74</b>	<b>71.84±1.15</b>
GAT	77.90±1.30	58.80±1.41	72.90±1.06	35.70±2.86	25.57±1.82	21.55±1.34	70.03±1.67
TailGAT	78.85±1.56	59.97±1.85	74.83±1.26	37.04±2.30	26.41±2.60	22.97±2.01	OOM
CLNode+GAT	79.40±1.16	60.39±2.04	74.87±1.68	37.60±2.47	26.93±2.89	23.06±1.39	72.51±1.52
SILC-GCL+GAT	<b>80.19±1.49</b>	<b>61.55±1.70</b>	<b>75.41±0.78</b>	<b>38.86±1.43</b>	<b>27.28±0.97</b>	<b>23.84±0.81</b>	<b>73.28±1.45</b>

表 3 节点分类的 Macro-F1 分数

Table 3 Macro-F1 scores of node classification

Methods	Cora	CiteSeer	PubMed	Chameleon	Squirrel	Actor	Arxiv
GCN	82.56±0.60	57.18±0.97	70.96±0.92	28.77±1.63	20.89±1.40	21.01±1.33	64.38±1.55
ColdBrew	82.84±0.51	57.63±0.79	71.52±0.54	30.11±1.43	21.62±1.51	21.94±1.12	65.11±1.49
TailGCN	84.03±1.17	59.14±0.94	73.69±0.73	31.71±1.95	23.81±1.47	24.58±1.63	OOM
CenGCN	84.46±1.51	59.68±1.39	OOM	31.47±1.62	22.31±1.	23.09±1.74	OOM
CLNode+GCN	85.07±0.64	59.72±1.34	74.89±1.45	31.57±1.81	23.56±1.43	23.91±1.23	65.78±1.63
SILC-GCL+GCN	<b>85.51±0.80</b>	<b>60.34±0.77</b>	<b>76.06±0.54</b>	<b>32.10±1.99</b>	<b>24.12±1.22</b>	<b>24.96±1.17</b>	<b>66.34±1.46</b>
GraphSAGE	79.14±1.54	58.70±1.49	72.46±0.46	34.26±1.45	28.66±2.48	28.23±1.85	61.18±2.43
TailSAGE	81.69±1.63	59.74±1.44	74.57±0.61	36.84±2.21	30.13±1.79	30.61±1.58	OOM
CLNode+SAGE	81.83±1.74	60.24±1.97	75.19±0.50	33.15±2.11	29.94±1.82	30.75±1.49	62.76±1.64
SILC-GCL+SAGE	<b>82.16±1.05</b>	<b>61.07±1.77</b>	<b>76.69±0.38</b>	<b>37.70±1.99</b>	<b>30.94±1.22</b>	<b>31.41±1.99</b>	<b>63.26±1.72</b>
GAT	76.33±1.16	51.46±1.39	71.79±1.14	28.78±1.92	20.98±2.08	20.37±1.14	62.59±2.05
TailGAT	77.45±1.67	52.79±1.65	73.64±1.96	30.09±2.12	22.38±2.51	21.56±2.14	OOM
CLNode+GAT	77.83±1.26	53.22±2.07	74.07±1.81	30.87±1.79	22.74±2.49	21.92±1.87	63.74±1.81
SILC-GCL+GAT	<b>78.77±1.59</b>	<b>54.12±1.33</b>	<b>74.94±0.80</b>	<b>31.96±1.61</b>	<b>23.67±2.35</b>	<b>22.79±1.15</b>	<b>64.25±1.28</b>

## 5.4 消融研究

为了进一步研究本文模型中设计的两种难度测量器  $LC$  与  $PR$  的效果以及不同训练调度器对模型性能的影响,本文进行了一系列的消融实验,各实验均以 GCN 为骨架模型,并在 6 个数据集上进行了验证。

首先,为了观察  $LC$  与  $PR$  对本文模型性能的影响,本文分别将这两种指标单独作为模型的难度测量器。实验的结果如图 4 所示,其中 + $LC$  表示将  $LC$  作为难度测量器的度量, + $PR$  表示将  $PR$  作为难度测量器的度量, + $BOTH$  则表示本文提出的综合两者的难度测量器。由图 4 可知,综合考虑两种指标的难度测量器在 6 个数据集上均取得了最优的结果,这验证了 GNNs 中的样本存在不同的训练难度,而拓扑结构的难易程度也不应被忽略。与此同时,实验表明,在 Cora, CiteSeer, PubMed 以及 Arxiv 这 4 个具有更高同配性的图上, + $LC$  的性能要略低于 + $PR$ ; 而在 Chameleon, Squirrel 以及 Actor 这些具有更高异配性的图上则恰好相反。这说明,

当邻域的标签不一致性较高时,采用  $LC$  测量器更能鉴别训练困难的样本; 而当邻域的标签不一致性较低时,采用  $PR$  测量器更能鉴别困难样本。由于不同数据集可能存在不同的困难样本测量角度,因此本文综合考虑标签语义和拓扑结构两种度量策略是必要的。

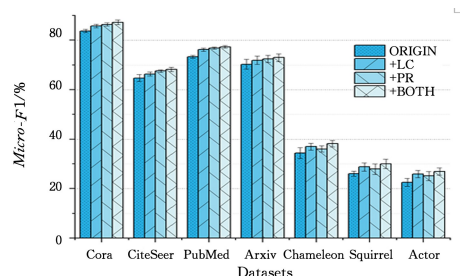


图 4 难度测量器的比较

Fig. 4 Comparison between difficulty measurers

其次,本文还考查了不同训练调度器(即线型、根型以及

几何型步长函数)对模型的影响,实验结果如图5所示。结果表明,几何型函数在所有数据集上均取得了最优性能。如前文所述,几何型函数在训练早期会更关注简单节点的训练,这使得在引入困难节点之前,GNNs能够充分学习容易节点的知识,从而减轻困难节点对模型训练的影响。

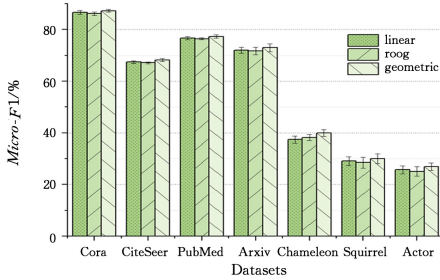


图5 训练调度器的比较

Fig. 5 Comparison between training schedulers

### 5.5 参数敏感性分析

本文研究了超参数 $\lambda_0$ 以及 $T_{grow}$ 对模型的影响。 $\lambda_0$ 控制训练节点的初始数量,而 $T_{grow}$ 控制引入困难节点的速度, $\lambda_0$ 的取值范围为 $[0.1, 0.2, \dots, 0.9]$ ,步长为0.1; $T_{grow}$ 的取值范围为 $[50, 100, \dots, 400]$ ,步长为50。 $\lambda_0$ 与 $T_{grow}$ 在各个数据集上均被设置为最优值。图6给出了基于GCN骨架模型的SILC-GCL在Cora数据集上的实验结果。

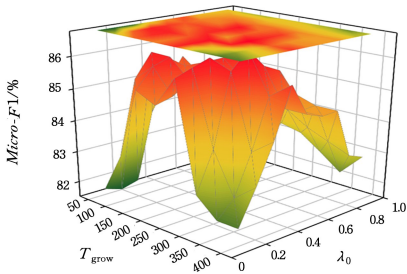


图6 超参数 $\lambda_0$ 、 $T_{grow}$ 的敏感性分析

Fig. 6 Sensitivity analysis of hyper-parameter  $\lambda_0$ ,  $T_{grow}$

从图6可知:1)随着 $\lambda_0$ 增大,模型的性能会先提升然后降低,在 $\lambda_0 \in [0.3, 0.6]$ 时,模型取得了较好的性能。这是因为, $\lambda_0$ 过小会导致初始训练节点数量过少,使得模型无法充分地学习;而 $\lambda_0$ 过大则可能会在训练初期引入困难节点,从而影响模型的训练性能。2) $T_{grow}$ 对模型也有类似的影响,在 $T_{grow} \in [200, 300]$ 时,模型取得了较好性能。这是因为, $T_{grow}$ 太小也会很快引入困难节点,从而降低模型性能; $T_{grow}$ 太大则会导致模型无法充分学习困难节点的信息,两种情况均会导致模型性能下降。

**结束语** 本文提出了一种结构影响力及标签冲突感知的图课程学习方法 SILC-GCL,该方法同时从图的结构信息以及节点标签两个角度来定义节点的训练难度并在训练过程中按从易到难的顺序逐步引入节点样本,从而增强 GNNs 的学习性能。在 6 个基准数据集上的实验充分验证了本文图课程学习方法的有效性。未来工作将继续探索在不同图学习任务上的图课程学习框架及其难度测量器的设计,以提高图学习模型的鲁棒性。

### 参考文献

- [1] WU Z, PAN S, CHEN F, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 32(1): 4-24.
- [2] GOYAL P, FERRARA E. Graph embedding techniques, applications, and performance: A survey [J]. Knowledge-Based Systems, 2018, 151: 78-94.
- [3] KUMAR A, SINGH S S, SINGH K, et al. Link prediction techniques, applications, and performance: A survey [J]. Physica A: Statistical Mechanics and Its Applications, 2020, 553: 124289.
- [4] ZHOU J, CUI G, HU S, et al. Graph neural networks: A review of methods and applications [J]. AI Open, 2020, 1: 57-81.
- [5] HE D, WANG T, ZHAI L, et al. Adversarial Representation Mechanism Learning for Network Embedding [J]. IEEE Transactions on Knowledge & Data Engineering, 2023, 35(2): 1200-1213.
- [6] JIN D, HUO C, LIANG C, et al. Heterogeneous graph neural network via attribute completion [C] // Proceedings of the ACM Web Conference. 2021: 391-400.
- [7] THOMAS N K, MAX W. Semi-Supervised Classification with Graph Convolutional Networks [C] // Proceedings of the International Conference on Learning Representations. 2017.
- [8] HAMILTON W, YING Z, LESKOVEC J. Inductive representation learning on large graphs [C] // Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017: 1025-1035.
- [9] BARABÁSI A L, BONABEAU E. Scale-free networks [J]. Scientific American, 2003, 288(5): 60-69.
- [10] PAGE L, BRIN S, MOTWANI R, et al. The pagerank citation ranking: Bring order to the web [R]. Stanford University: Technical Report, 1998.
- [11] CIOTTI V, BONAVENTURA M, NICOSIA V, et al. Homophily and missing links in citation networks [J]. EPJ Data Science, 2016, 5: 1-14.
- [12] LIU Y, ZHENG Y, ZHANG D, et al. Towards unsupervised deep graph structure learning [C] // Proceedings of the ACM Web Conference. 2022: 1392-1403.
- [13] ZHENG X, LIU Y, PAN S, et al. Graph neural networks for graphs with heterophily: A survey [J]. arXiv: 2202. 07082, 2022.
- [14] ZHU J, YAN Y, ZHAO L, et al. Beyond homophily in graph neural networks: Current limitations and effective designs [J]. Advances in Neural Information Processing Systems, 2020, 33: 7793-7804.
- [15] SOVIANY P, IONESCU R T, ROTA P, et al. Curriculum learning: A survey [J]. International Journal of Computer Vision, 2022, 130(6): 1526-1565.
- [16] LIU Z, NGUYEN T K, FANG Y. Tail-gnn: Tail-node graph neural networks [C] // Proceedings of the ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 1109-1119.
- [17] ZHENG W, HUANG E W, RAO N, et al. Cold Brew: Distilling Graph Node Representations with Incomplete or Missing Neighborhoods [C] // International Conference on Learning Representations.

- tations, 2021.
- [18] XIA F, WANG L, TANG T, et al. Cengen: Centralized convolutional networks with vertex imbalance for scale-free graphs[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 35(5):4555-4569.
- [19] SUN Q, LI J, YUAN H, et al. Position-aware structure learning for graph topology-imbalance by relieving under-reaching and over-squashing[C] // *Proceedings of the ACM International Conference on Information & Knowledge Management*. 2022: 1848-1857.
- [20] FU X, WEI Y, SUN Q, et al. Hyperbolic Geometric Graph Representation Learning for Hierarchy-imbalance Node Classification[C] // *Proceedings of the ACM Web Conference*. 2023:460-468.
- [21] GUO S, HUANG W, ZHANG H, et al. Curriculumnet: Weakly supervised learning from large-scale web images[C] // *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018:135-150.
- [22] JIANG L, MENG D, MITAMURA T, et al. Easy samples first: Self-paced reranking for zero-example multimedia search[C] // *Proceedings of the ACM International Conference on Multimedia*. 2014:547-556.
- [23] PLATANIOS E A, STRETCU O, NEUBIG G, et al. Competence-based Curriculum Learning for Neural Machine Translation[C] // *Proceedings of NAACL-HLT*. 2019:1162-1172.
- [24] TAY Y, WANG S, LUU A T, et al. Simple and Effective Curriculum Pointer-Generator Networks for Reading Comprehension over Long Narratives[C] // *Proceedings of the Annual Meeting of the Association for Computational Linguistics*. 2019:4922-4931.
- [25] LI H, WANG X, ZHU W. Curriculum Graph Machine Learning: A Survey[J]. *arXiv:2302.02926*, 2023.
- [26] WANG Y, WANG W, LIANG Y, et al. Curgraph: Curriculum learning for graph classification[C] // *Proceedings of the ACM Web Conference*. 2021:1238-1248.
- [27] CHU G, WANG X, SHI C, et al. CuCo: Graph Representation with Curriculum Contrastive Learning[C] // *Proceedings of the International Joint Conference on Artificial Intelligence*. 2021: 2300-2306.
- [28] WANG H, ZHOU K, ZHAO X, et al. Curriculum pre-training heterogeneous subgraph transformer for top-n recommendation [J]. *ACM Transactions on Information Systems*, 2023,41(1): 1-28.
- [29] WEI X, GONG X, ZHAN Y, et al. Cnode: Curriculum learning for node classification[C] // *Proceedings of the ACM International Conference on Web Search and Data Mining*. 2023: 670-678.
- [30] LI X, WEN L, DENG Y, et al. Graph neural network with curriculum learning for imbalanced node classification[J]. *arXiv: 2202.02529*, 2022.
- [31] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. *Journal of Artificial Intelligence Research*, 2002, 16:321-357.
- [32] LIU Y, YANG X, ZHOU S, et al. Hard sample aware network for contrastive deep graph clustering[C] // *Proceedings of the AAAI Conference on Artificial Intelligence*. 2023, 37(7): 8914-8922.
- [33] HACOEN G, WEINSHALL D. On the power of curriculum learning in training deep networks[C] // *Proceedings of International Conference on Machine Learning*. 2019:2535-2544.
- [34] PLATANIOS E A, STRETCU O, NEUBIG G, et al. Competence-based Curriculum Learning for Neural Machine Translation[C] // *Proceedings of NAACL-HLT*. 2019:1162-1172.
- [35] MCCALLUM A K, NIGAM K, RENNIE J, et al. Automating the construction of internet portals with machine learning[J]. *Information Retrieval*. 2000, 3:127-163.
- [36] GILES C L, BOLLACKER K D, LAWRENCE S. CiteSeer: An automatic citation indexing system[C] // *Proceedings of the ACM Conference on Digital Libraries*. 1998:89-98.
- [37] SEN P, NAMATA G, BILGIC M, et al. Collective classification in network data[J]. *AI Magazine*, 2008, 29(3):93.
- [38] HU W, FEY M, ZITNIK M, et al. Open graph benchmark: Datasets for machine learning on graphs[J]. *Advances in Neural Information Processing Systems*, 2020, 33:22118-22133.
- [39] PEI H, WEI B, CHANG K C C, et al. Geom-GCN: Geometric Graph Convolutional Networks[C] // *International Conference on Learning Representations*. 2019.
- [40] ROZEMBERCZKI B, ALLEN C, SARKAR R. Multi-Scale attributed node embedding [J]. *Journal of Complex Networks*, 2021, 9(1):1-22.
- [41] TANG J, SUN J, WANG C, et al. Social influence analysis in large-scale networks[C] // *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2009:807-816.
- [42] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph Attention Networks[C] // *International Conference on Learning Representations*. 2018.



**LIU Zulong**, born in 1997, postgraduate. His main research interests include graph representation learning and graph structure learning.



**CHEN Kejia**, born in 1980, Ph.D, associate professor. Her main research interests include complex network analysis and graph learning.