本文是对论文 ”**Classic GNNs are Strong Baselines: Reassessing GNNs for Node Classification”**

（作者：[Yuankai Luo）的非官方中文翻译。

仅供学习与交流使用，版权归原作者所有。

原文链接：<https://openreview.net/forum?id=xkljKdGe4E#discussion>

经典GNNs是强大的基准：重新评估用于节点分类的GNNs

Yuankai Luo1，Lei Shi2，Xiao-Ming Wu 3

（1. Beihang University, The Hong Kong Polytechnic University [luoyk@buaa.edu.cn](mailto:luoyk@buaa.edu.cn)；2. Beihang University [leishi@buaa.edu.cn](mailto:leishi@buaa.edu.cn); The Hong Kong Polytechnic University xiao-ming.wu@polyu.edu.hk）

摘要：图转换器（GTs）最近已成为传统消息传递图神经网络（GNNs）的流行替代品，因其理论上更优的表达能力和在标准节点分类基准测试中报告的卓越性能，通常显著优于GNNs。在本文中，我们进行了一项彻底的实证分析，以重新评估三种经典GNN模型（GCN、GAT和GraphSAGE）相对于GTs的性能。我们的研究结果表明，先前报告的GTs的优越性可能因GNNs中次优的超参数配置而被夸大了。值得注意的是，通过轻微的超参数调整，这些经典的GNN模型在所研究的18个不同数据集中的17个上实现了最先进的性能，达到甚至超过了最近的GTs的性能。此外，我们进行了详细的消融研究，以探究各种GNN配置——例如归一化、丢弃、残差连接和网络深度——对节点分类性能的影响。我们的研究旨在推动图机器学习领域更高标准的实证严谨性，鼓励对模型能力进行更准确的比较和评估。我们的实现可在 https://github.com/LUOyk1999/tunedGNN 获取。

关键词：

[第一章 引言 3](#_Toc210827547)

[第二章 经典的GNNs 5](#_Toc210827548)

[第三章 训练GNNs的关键超参数 7](#_Toc210827549)

[第四章 用于节点分类的实验设置 9](#_Toc210827550)

[第五章 实证研究发现 11](#_Toc210827551)

[5.1 经典GNNs在节点分类中的性能 11](#_Toc210827552)

[5.2 超参数对GNNs性能的影响 14](#_Toc210827553)

[第六章 结论 16](#_Toc210827554)

[参考文献 19](#_Toc210827555)

[A 数据集与实验详情 21](#_Toc210827556)

[A.1计算环境 21](#_Toc210827557)

[A.2 超参数与可复现性 21](#_Toc210827558)

[B 额外的基准测试结果 22](#_Toc210827559)

[B.1 在ogbn-proteins上使用边特征的GAT\* 22](#_Toc210827560)

[B.2 异质图上的更深网络 22](#_Toc210827561)

[B.3 Jumping Knowledge模式与早期结果 25](#_Toc210827562)

[C 可视化 25](#_Toc210827563)

[D 局限性与更广泛影响 26](#_Toc210827564)

1. 引言

节点分类是图机器学习中的一项基础任务，在社交网络分析、生物信息学和推荐系统等众多领域具有高影响力的应用。图神经网络（）已成为一类强大的模型来解决节点分类任务。通过迭代地聚合来自节点邻居的信息来运作，这是一个被称为消息传递的过程，它利用图结构和节点特征来学习用于分类的有用节点表示。虽然取得了显著成功，但研究也指出了其若干局限性，包括过度平滑（）、过度挤压（）、对异质性（）缺乏敏感度，以及在捕捉长距离依赖关系方面的挑战。

最近，图转换器（）作为的流行替代品而声名鹊起。与主要聚合局部邻域信息的不同，架构可以通过自注意力层捕捉任意一对节点之间的交互。在图级别的任务中取得了巨大成功，例如，涉及像分子图这样的小规模图的图分类任务。这一成功启发了许多工作去利用来解决节点分类任务，特别是在大规模图上，以解决前述的的局限性。尽管最先进的取得了许多进展并展现出有希望的结果，但值得注意的是，许多这类模型，无论是显式还是隐式地，仍然依赖来学习局部的节点表示，并将它们与全局注意力机制相结合以获得更全面的表示。

这促使我们重新思考：消息传递型在节点分类任务上的潜力是否在过去被低估了？尽管先前的研究在一定程度上探讨了这个问题，但这些研究在范围和全面性上存在局限性，涉及的节点和数据集数量有限，并且对超参数的理解不完整。在本研究中，我们全面地重新评估了三种经典模型——、和——在18个真实的基准数据集上的性能，这些数据集包括同质图、异质图和大规模图。我们检验了关键超参数对训练的影响，包括归一化、丢弃、残差连接和网络深度。我们将实证研究中的关键发现总结如下：

• 通过适当的超参数调整，经典的可以在包含数百万节点的同质图和异质图上实现极具竞争力的节点分类性能。值得注意的是，经典的超越了最先进的，在18个数据集中有17个取得了最高排名。这表明，先前声称的相对于的优越性可能被夸大了，原因可能是在对的评估中存在次优的超参数配置。

• 我们的消融研究为节点分类任务中的超参数提供了宝贵的见解。我们证明了(1)归一化（）对于大规模图至关重要；(2)丢弃（）始终是有益的；(3)残差连接（）可以显著提升性能，尤其是在异质图上；以及(4)在同质图上的倾向于在更深的网络层下表现更佳。

1. 经典的GNNs

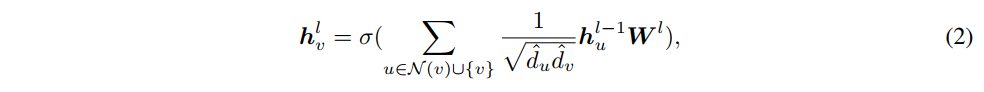
用于节点分类 我们将一个图定义为，其中表示节点集，表示边集。是节点特征矩阵，其中表示节点数量，表示节点特征的维度。是单热编码的标签矩阵，是类别数量。令为的邻接矩阵。

**消息传递图神经网络 ()** [19]在每一层计算节点表示：



其中表示与相邻的邻居节点，作为消息聚合函数，而是更新函数。初始时，每个节点以一个特征向量开始。函数从邻居节点聚合信息以更新其表示。最后一层的输出，即，是由生成的节点的表示。在这项工作中，我们专注于三种经典的模型：[28]、[20]和[68]，它们在学习节点表示的方法上有所不同。

**图卷积网络（）**[28]，标准的模型，被公式化为：



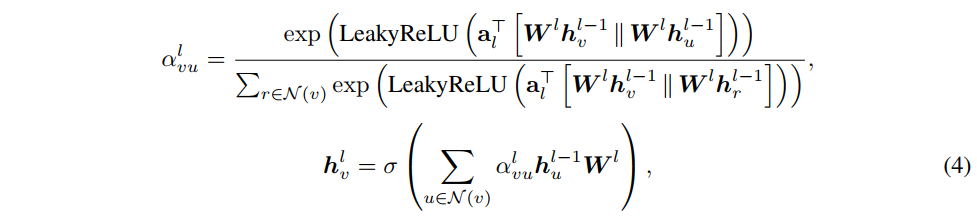
其中，表示节点的度，是第层可训练的权重矩阵，是激活函数，例如。

**GraphSAGE**[20]通过一种不同的方法学习节点表示：



其中和是可训练的权重矩阵，计算邻居节点的平均嵌入。

**图注意力网络（）**[68]采用掩码自注意力机制为邻居节点分配权重。对于一条边，的传播规则定义为：

其中是一个可训练的权重向量，是一个可训练的权重矩阵，表示拼接操作。

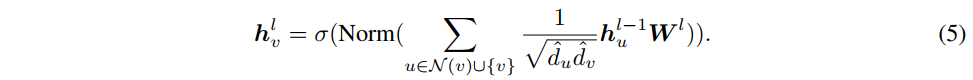
**节点分类**旨在预测未标记节点的标签。通常，对于任何节点，由最后一个层生成的节点表示会通过一个预测头，以获得预测标签。训练目标是最小化关于训练集中所有节点的总损失，其中表示真实标签，表示可训练的参数。

**同质图与异质图** 节点分类可以在同质图和异质图上进行。同质图的特点是边倾向于连接同类别的节点，而在异质图中，连接的节点可能属于非常不同的类别。模型隐式地假设同质性[48]，并且通常认为由于这种同质性假设，不能很好地泛化到异质图[90, 9]。然而，最近的研究[46, 40, 58, 42]通过实证表明，标准的在异质图上也表现良好。在本研究中，我们对经典的在同质图和异质图上的节点分类任务进行了全面的评估。

1. 训练GNNs的关键超参数

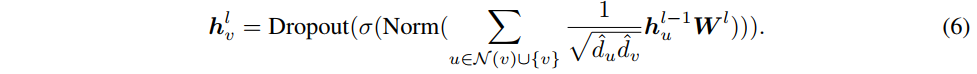
在本节中，我们概述了用于训练GNNs的关键超参数，包括归一化、丢弃、残差连接和网络深度。这些超参数被广泛应用于不同类型的神经网络中以提升模型性能。

**归一化**。具体来说，层归一化（Layer Normalization, LN）[2]或批归一化（Batch Normalization, BN）[26]可用于每个网络层的激活函数之前。以GCN为例：

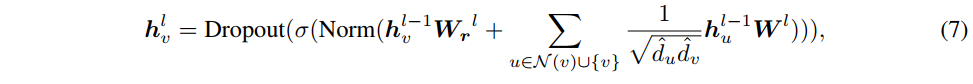


归一化技术通过减少训练过程中发生的协变量偏移（）来帮助稳定训练过程，这种偏移发生在每层节点嵌入的分布在训练期间发生变化时。对节点嵌入进行归一化有助于维持一个更一致的分布，从而允许使用更高的学习率并带来更快的收敛速度[5]。

**丢弃** (Dropout) 。丢弃[67]是一种最初在卷积神经网络（CNNs）中广泛用于通过减少隐藏神经元之间的协同适应（）来解决过拟合的技术，现已被发现在解决GNNs中的类似问题上也同样有效[58, 65]，在GNNs中，协同适应效应会通过消息传递在不同节点间传播和累积。通常，丢弃被应用于激活函数之后的特征嵌入上。



**残差连接**。残差连接[21]通过将层的输入直接连接到输出，显著增强了CNN的性能，从而有效缓解了梯度消失问题。它们首先被开创性的GCN论文[28]所采纳，并随后被整合到DeepGCNs[33]中以提升性能。形式上，线性残差连接可以如下整合到GNNs中：



其中是一个可训练的权重矩阵。这种配置可以减轻梯度不稳定性并增强GNN的表达能力[80]，解决了过度平滑[35]和过度挤压[1]的问题，因为线性分量()有助于保留可区分的节点表示[73]。

网络深度。更深的网络架构，例如深度CNNs[21, 25]，能够从数据中提取更复杂、更高层次的特征，有可能在各种预测任务上带来更好的性能。然而，GNNs在深度方面面临独特的挑战，例如过度平滑[35]，即随着网络深度的增加，节点表示变得无法区分。因此，在实践中，大多数GNNs采用浅层架构，通常由2到5层组成。尽管先前的研究，如DeepGCN[33]和DeeperGCN[34]，提倡使用深达56和112层的深度GNNs，但我们的研究结果表明，使用显著更浅的GNN架构，通常在2到10层范围内，就可以实现相当的性能。

1. 用于节点分类的实验设置

**数据集**。表1展示了这些数据集的统计数据和特性摘要。

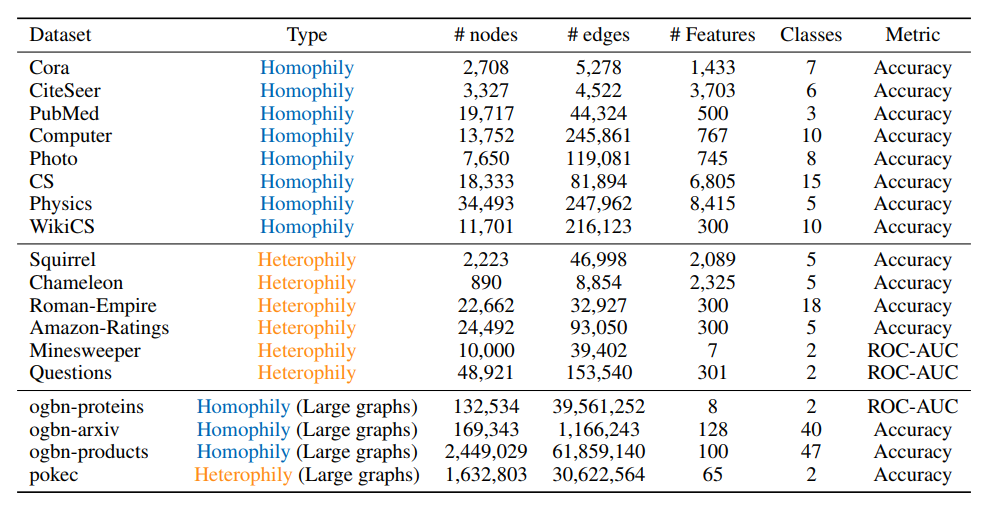


表 1：用于节点分类的数据集概览

\* **同质图**。Cora、CiteSeer和PubMed是三个广泛使用的引文网络[62]。我们遵循[28]中的半监督设置进行数据划分和度量。此外，Computer和Photo[63]是共同购买网络，其中节点代表商品，边表示连接的商品被频繁一起购买。CS和Physics[63]是合著网络，其中节点代表作者，边代表作者们至少合著过一篇论文。我们遵循广泛接受的训练/验证/测试集划分比例，即60%/20%/20%，以及[7, 64, 12]中准确率的度量标准。此外，我们利用WikiCS数据集，并使用[50]中提供的官方划分和度量标准。

**\* 异质图**。Squirrel和Chameleon[61]是两个著名的维基百科页面网络，专注于特定主题。根据[58]这篇异质图基准测试论文，这些数据集的原始划分在训练集和测试集之间引入了重叠节点，从而催生了一种过滤掉重叠节点的新数据划分方案以供评估。我们使用其提供的划分和度量标准。此外，我们利用了由同一来源[58]提出的另外四个异质数据集：Roman-Empire，其中节点对应于罗马帝国维基百科文章中的单词，边连接顺序或语法上相关的单词；Amazon-Ratings，其中节点代表产品，边连接频繁被共同购买的商品；Minesweeper，一个合成数据集，其中节点是100x100网格中的单元格，边连接相邻的单元格；以及Questions，其中节点代表来自Yandex问答网站的用户，边连接通过答案进行互动的用户。所有划分和评估指标均与源数据保持一致。

**\* 大规模图**。我们考虑了由开放图基准（Open Graph Benchmark, OGB）[24]最近发布的一系列大型图：ogbn-arxiv、ogbn-proteins和ogbn-products，其节点数量范围从0.16M到2.4M。我们维持OGB标准的评估设置。此外，我们还在社交网络pokec[32]上进行性能分析，该网络拥有1.6M节点，遵循[12]的评估设置。

**基准模型**。

我们的主要焦点是

经典的GNNs：[28]、[20]、[68]，

以及最先进的GNNs：[76]、[12]、[29]、[75]、[7]

和强大的GTs：[59]和[64]。此外，在综述[23, 53]中还存在各种其他的GTs，如[17, 15, 38, 86, 31, 3, 6, 82, 13]，它们在经验上显示出不如我们为节点分类任务所比较的GTs。

对于异质图，我们还考虑了五个专为异质性设计的模型，遵循[58]：[90]、[89]、[9]、[49]、[36]。请注意，我们采用了经验上最优的变体()，它展示了比先进的GNNs如[37]和[66]更优的性能。

我们报告的基线性能结果主要来自[12, 76, 58]，其余的则尽可能从它们各自的原始论文或官方排行榜中获取，因为这些结果是由经过精心调优的模型获得的。

**超参数配置**。我们对经典的GNNs进行超参数调优，与[12]中的超参数搜索空间保持一致。具体来说，我们使用学习率为{0.001, 0.005, 0.01}的优化器[27]和2500个周期的上限。我们将隐藏维度从{64, 256, 512}中进行调优。如第3节所讨论的，我们关注是否使用归一化(或)、残差连接，以及丢弃率{0.2, 0.3, 0.5, 0.7}和层数{1, 2, 3, 4, 5, 6, 7, 8, 9, 10}。此外，我们使用与经典GNNs相同的超参数搜索空间和训练环境来重新训练所有基线的GTs。对于每个GT特有的、在经典GNNs中不存在的超参数，我们根据原始GT论文中指定的搜索空间进行调优。我们报告5次使用不同初始化的独立运行的平均分数和标准差。Model\*表示我们自己的实现。 详细的实验设置和超参数在附录A中提供。

1. 实证研究发现
   1. 经典GNNs在节点分类中的性能

在本小节中，我们详细分析了三种经典模型与最先进的在节点分类任务上的性能对比。我们在同质图（表2）、异质图（表3）和大规模图（表4）上的实验结果表明，在18个数据集中，经典的通常优于或持平于先进的性能。值得注意的是，在评估的18个数据集中，经典的在其中的17个上取得了最高排名，展现了其强大的竞争力。我们在下文重点介绍我们的主要观察。

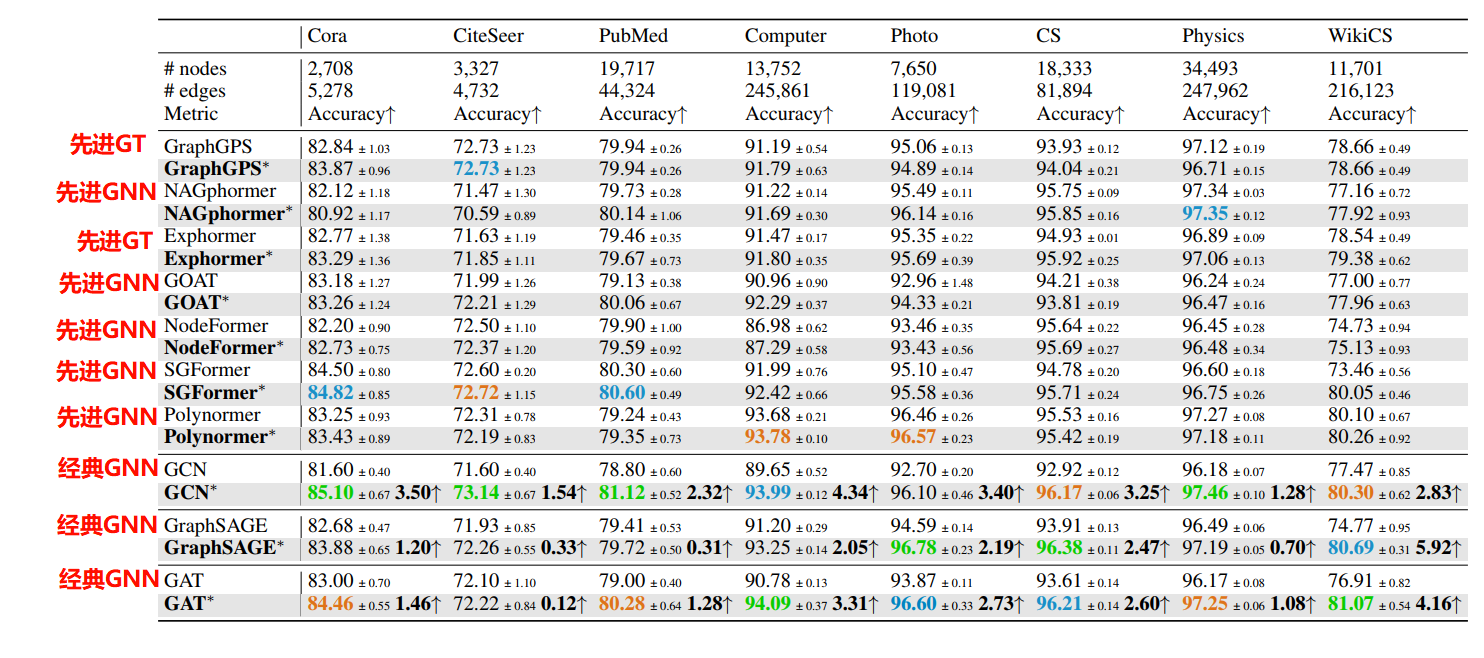


表2：同配图上的节点分类结果（%）。∗ 表示我们的实现，其余结果取自文献 [12, 76,]。最佳的第 1、2、3 名结果已高亮显示。

**关于同配图的观察 (表2) 经典的GNNs**。在对超参数进行轻微调整后，在同质图的节点分类任务上具有高度竞争力，在许多情况下甚至优于最先进的图转换器。

尽管先前报告的结果显示，大多数先进的在同质图上优于经典的，但我们对经典的实现可以在四个数据集中的两个上排名前二，其中和展现了非持续性的性能。具体来说，在CS和WikiCS上，经典的经历了约3%的准确率提升，达到了前三的性能。在WikiCS上，的准确率提升了4.16%，使其排名从第七位移至第一位，超过了。同样，在Photo和CS上，的性能优于和，成为顶级模型。在Cora、CiteSeer、PubMed和Physics上，调优为带来了显著的性能改进，准确率提升范围从1.54%到3.50%，使其在初始准确率低于先进的情况下，定位为性能最高的模型。

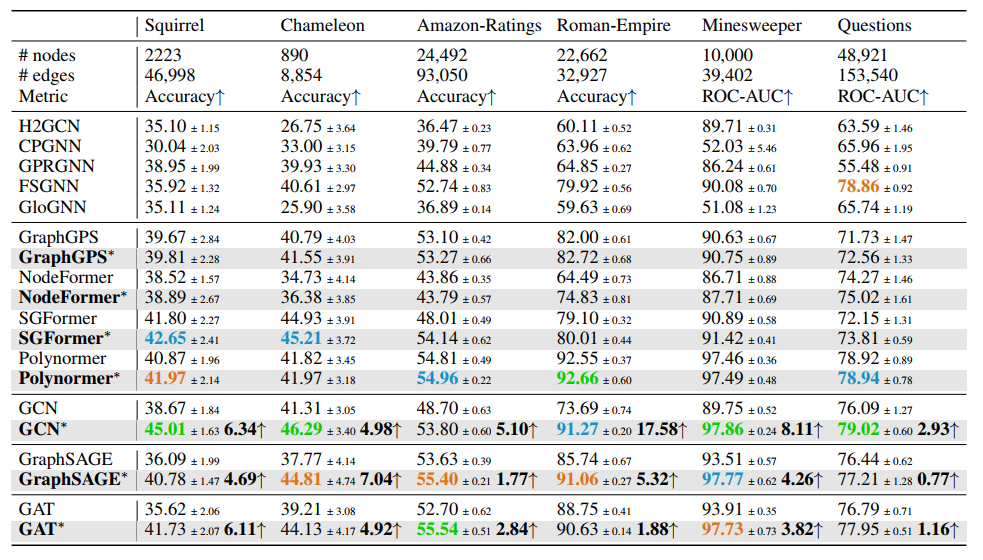


表 3：异配图上的节点分类结果（%）。∗ 表示我们的实现，其余结果取自文献 [12, 76, 58]。最佳的第1、第2、第3 名结果已高亮显示。

**关于异质图的观察 (表3)**。我们的实现显著增强了经典在异质图上的先前最佳结果，超越了为这类图量身定制的领先图转换器架构。这一进步不仅支持而且加强了[58]中的发现，即传统是异质图的有力竞争者，挑战了它们主要适用于同质图结构的普遍假设。

这三种经典的在六个异质图中的五个上占据了最高位置。具体来说，在像Chameleon和Squirrel这样著名的页面网络上，我们的实现分别将的准确率提升了4.98%和6.34%，使其在所有模型中名列第一。同样，在诸如Minesweeper和Questions等较大的异质图上，也展现出高性能，突显了其局部消息传递机制相对于的全局注意力的优越性。在Roman-Empire上，我们观察到的性能有17.58%的惊人提升。有趣的是，我们发现性能的提升主要源于残差连接，这将在消融研究中（见第5.2节）进一步分析。

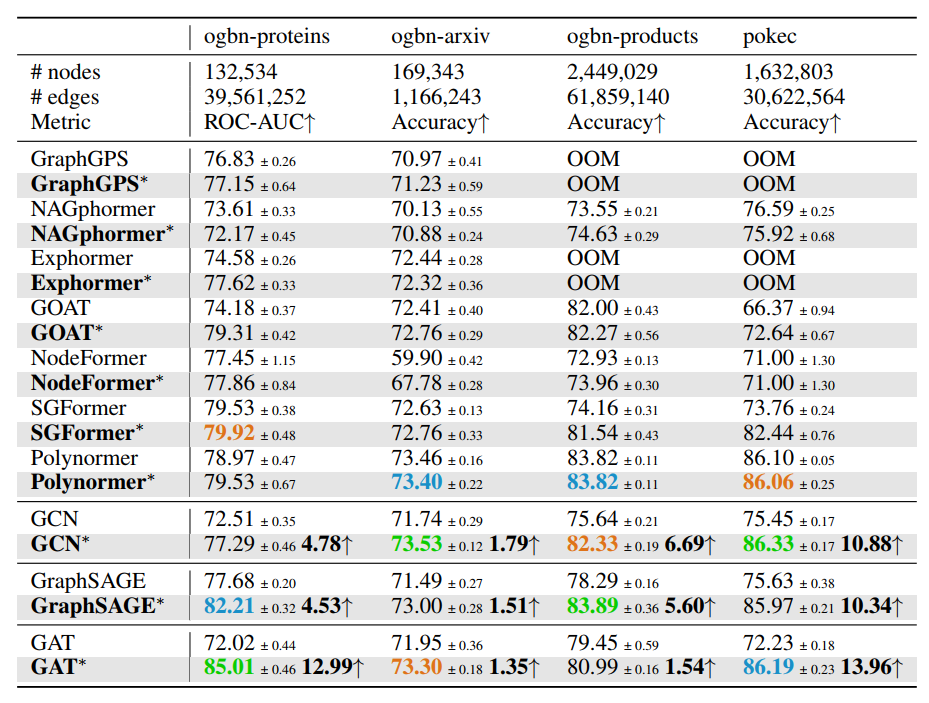


表4：大规模图上的节点分类结果(%)。带\*的表示我们的实现，而其他结果则取自[12, 76]。排名前第一、第二和第三的结果被高亮显示。OOM表示内存不足(Out of memory)。

**关于大规模图的观察 (表4)。**我们的实现显著提升了先前报告的经典的结果，在某些情况下准确率呈现了两位数的增长。它在这些大型图数据集上（无论是同质的还是异质的）均取得了最佳结果，并超越了最先进的图转换器。这表明消息传递机制在学习大规模图的节点表示方面仍然非常有效。

我们的经典GNNs实现展现出持续的卓越性能，在我们研究中包含的所有四个大规模数据集上均取得了最高排名。值得注意的是，GCN\*成为`ogbn-arxiv`和`pokec`上的领先模型，超越了所有被评估的先进GTs。此外，在`pokec`上，我们实现的所有三种经典GNNs都取得了超过10%的性能提升。在`ogbn-proteins`上，我们观察到GAT\*的性能有12.99%的绝对提升，显著超过SGFormer达5.09%（GAT\* - SGFromer）。同样，在`ogbn-products`上，GraphSAGE\*展现了显著的性能增长，在所有被评估的模型中获得了最佳性能。总而言之，一个基础的GNN可以在大规模图上取得目前已知的最佳结果，这表明当前的GTs尚未解决诸如过度平滑和长距离依赖等GNN问题。

* 1. 超参数对GNNs性能的影响

为了检验不同超参数在解释经典性能增强方面的独特贡献，我们进行了一系列消融分析，通过从、和中有选择地移除诸如归一化、丢弃、残差连接和网络深度等元素。这些消融操作的效果在同质图（见表5）、异质图（见表6）和大规模图（见表7）上进行了评估。我们的发现（下文将详细阐述）表明，移除单个组件会对模型准确率产生不同的影响。

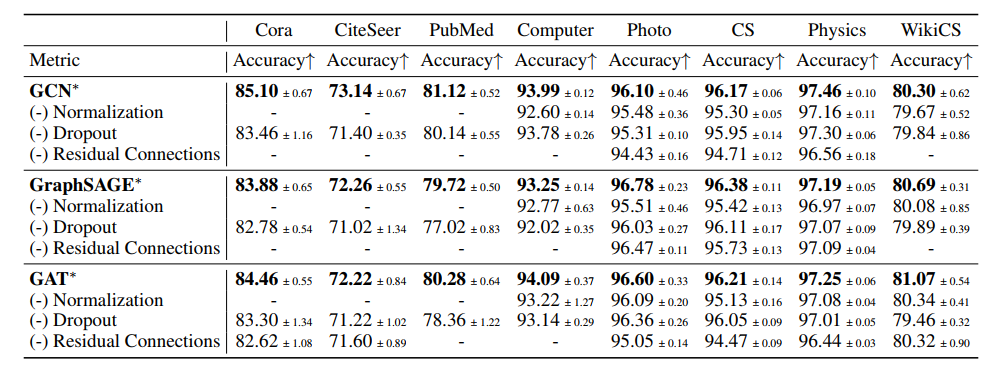


表5：同质图上的消融研究 (%)。- 表示对应的超参数未在GNN\*中使用，因为根据经验，它会导致较差的性能。(译者注：以GCN\*在Cora数据集上为例子，针对这个数据集，作者的原始模型就只有Dropout，消融实验也只有针对它的消融)

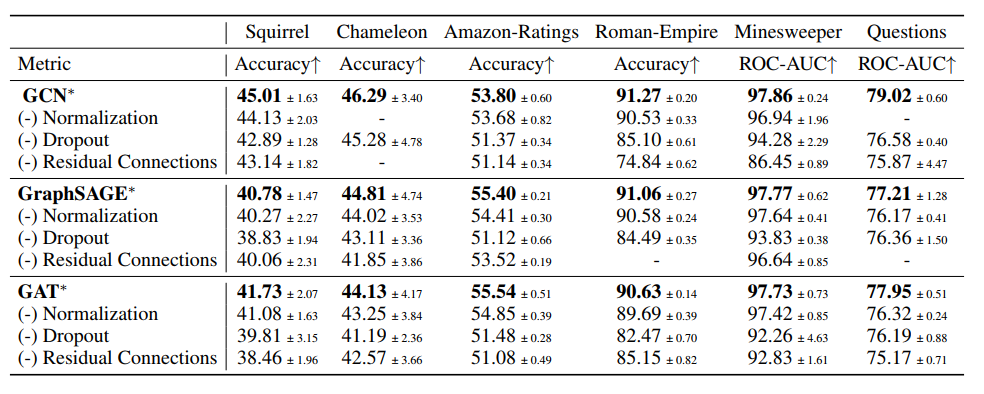


表6：异质图上的消融研究 (%)。

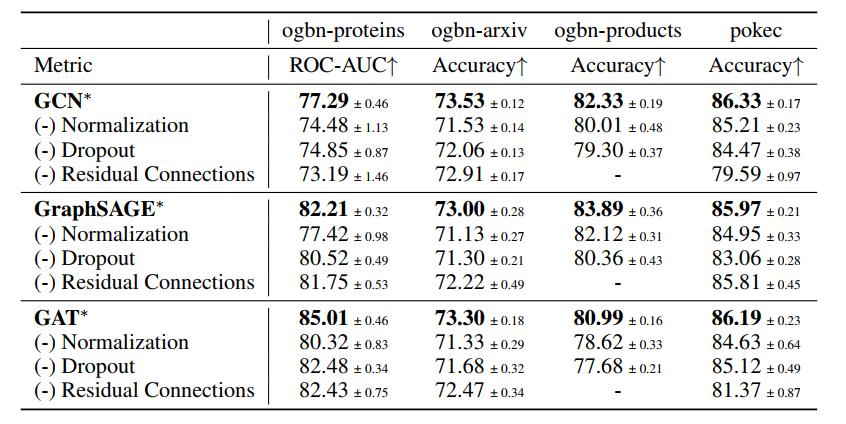


表7：大规模图上的消融研究 (%)。

**观察1：归一化（或）对于大规模图上的节点分类很重要，但对小规模图则不那么显著。**

我们观察到，移除归一化在小图上不会导致性能的实质性偏差。然而，归一化在大规模图上始终至关重要，移除它会导致和在ogbn-proteins上的准确率分别下降4.79%和4.69%。我们认为这是因为大图展现出更多样化的节点特征，导致整个图上存在不同的数据分布。归一化有助于标准化这些特征，从而确保一个更稳定的分布。

**观察2：丢弃（Dropout）被一致性地发现对于节点分类至关重要。**

我们的分析突显了丢弃在维持经典在同质图和异质图上的性能方面所起的关键作用，移除它会导致性能下降——例如，在PubMed上导致下降2.70%，在Roman-Empire上导致6.57%的下降。这一趋势在大型数据集上依然存在，其中移除丢弃会导致和在ogbn-proteins上分别出现2.44%和2.53%的性能下降。

**观察3：残差连接可以显著提升在特定数据集上的性能，在异质图上比在同质图上表现出更显著的效果。**

虽然在同质图上移除残差连接并不会一致性地导致显著的性能下降（在Cora、Photo和CS上观察到的差异约为2%），但其影响在诸如ogbn-proteins和pokec等大规模图上则更为可观。这种影响在异质图上甚至更为戏剧化，其中经典在移除残差连接后表现出最显著的准确率下降，例如在Roman-Empire上，下降了16.43%，下降了5.48%。同样，在Minesweeper上，也观察到了显著的性能下降，这强调了残差连接的关键重要性，尤其是在异质图上。这些图的复杂结构通常需要更深的网络层来有效捕捉节点间的多样化关系。在这种情况下，残差连接对于模型训练至关重要。

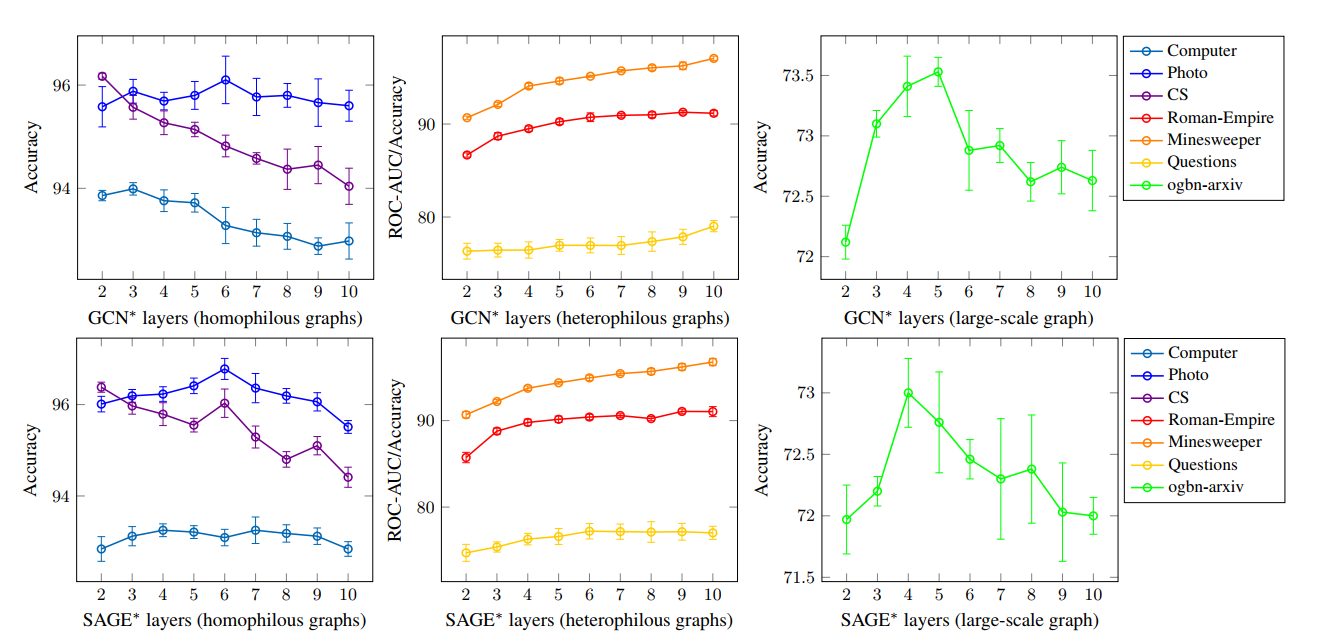


图1：关于网络层数的消融研究，从左到右分别展示了在同质图、异质图和大规模图上的结果。

**观察4：与同质图相比，更深的网络通常在异质图上带来更大的性能增益。**

如图1所示，和的性能趋势在不同图类型上是一致的。在同质图和ogbn-arxiv上，两个模型都在2到6层的范围内达到最佳性能。相比之下，在异质图上，它们的性能随着层数的增加而提高，这表明更深的网络对这些图更有益。我们在附录B中讨论了超过10层网络的情景。

1. 结论

我们的研究对基础GNN模型在节点分类任务中的功效进行了一次彻底的重新评估。通过广泛的实证分析，我们证明了这些经典的GNN模型在多种图数据集上能够达到甚至超越GTs的性能，从而挑战了GTs在节点分类任务中被普遍认为的优越性。此外，我们全面的消融研究为各种GNN配置如何影响性能提供了深刻的见解。我们希望我们的研究发现能推动图机器学习研究领域进行更严谨的实证评估。

参考文献

A 数据集与实验详情

A.1计算环境

我们的实现基于[16]和[72]。实验在一台配备了8块RTX 3090 GPU的单一工作站上进行。值得注意的是，仅有在pokec数据集上的实验是在一台配备了2块A100 GPU的独立工作站上执行的。

A.2 超参数与可复现性

对于经典的超参数设置，除了我们已涵盖的内容外，我们在表8、9、10中列出了其他设置。值得注意的是，对于异质图，我们将网络层数的搜索范围扩展以包括三个额外的设置：{12, 15, 20}（详见B.2节以进行进一步分析）。这一调整是基于我们的经验证据，即更深的网络倾向于在异质图上产生性能提升。函数作为非线性激活函数。关于超参数的更多细节可以在我们的代码中找到：<https://github.com/LUOyk1999/tunedGNN>。

对于每个模型特有的、在经典中不存在的超参数，我们根据原始论文中指定的搜索空间进行了调优：

\* : heads数量从{1, 2, 4}，GNNs从{, , }，位置编码方案从{, , }。

\* : heads数量从{1, 2, 4}，hops数量从{3, 10}。

\* : heads数量从{1, 2, 4}，位置编码方案从{, , }。

\* : heads数量从{1, 2, 4}，codebook大小从{1024, 2048, 4096}。

\* : heads数量从{1, 2, 4}，从{30, 50}，从{5, 10}，从{1, 2}，temperature从{0.25}。

\* : heads数量从{1, 2, 4, 8}。

\* : heads数量从{1, 2, 4}，GNN权重从{0.5, 0.8}。

由于ogbn-proteins、ogbn-products和pokec的数据规模巨大，阻止了在GPU内存上进行全批量训练，我们采用了不同的批量训练策略。对于ogbn-proteins，我们利用了优化的邻居采样方法[20]。对于pokec和ogbn-products，我们应用了先前由[12, 76, 75]使用的随机分区方法以实现小批量（）训练。对于其他数据集，我们采用全批量训练。

测试准确率由在验证集上报告最高结果的模型来实现。此外，我们报告的是5次使用不同初始化的独立运行的平均分数和标准差。

我们的代码在MIT许可下可用。

B 额外的基准测试结果

B.1 在ogbn-proteins上使用边特征的GAT\*

尽管[33]引入了训练深达56层的模型，且[34]进一步将其扩展到112层，但我们的实验表明这样的深度并非必要。具体来说，虽然在ogbn-proteins上取得了85.50%的准确率，但它利用了边特征作为输入，而这种配置在OGB数据集的标准基线[24]中并不常用。由于我们的实验在ogbn-proteins上没有整合边特征，为了维持公平的比较，我们在正文中排除了。

现在，我们采用与[73]中相同的方法，将边特征整合到中，结果如表11所示。一个6层的取得了87.47%的准确率，显著超过了的85.50%。这证明了无需像所提议的那样深；一个2到10层的范围通常是足够的。

B.2 异质图上的更深网络

在异质图上，经典的性能随着层数的增加而提高（上限为10层），这一点已由正文中的图1所证实。在本小节中，我们探索了超过10层的场景。具体来说，我们考虑了在Roman-Empire和Minesweeper数据集上，层数配置为12、15和20的和。结果如表12所示。最优层数()的变化可能源于不同图固有的独特结构。异质图可能具有更复杂的结构，因此需要一个更高的值。然而，在较大的值下观察到的轻微性能提升表明，非常深的网络可能不会产生显著更好的结果。总的来说，经典的最佳结果是在被限制在15层时取得的。

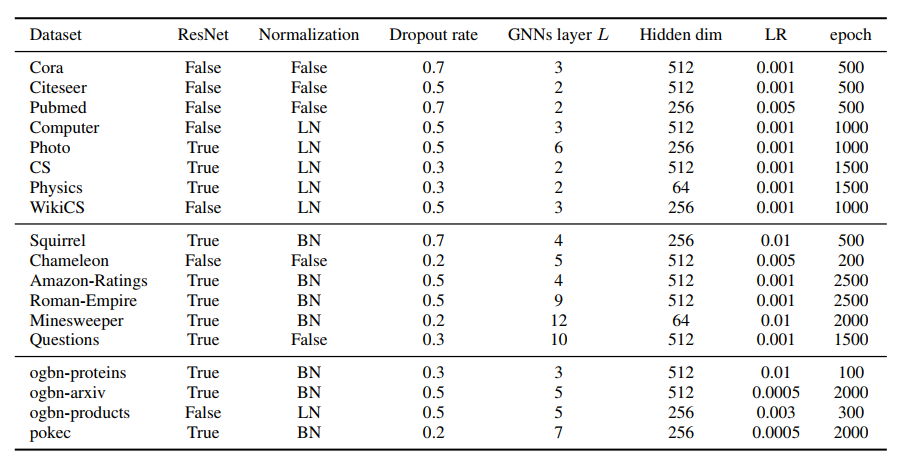


表 8：GCN∗ 的各数据集特定超参数设置

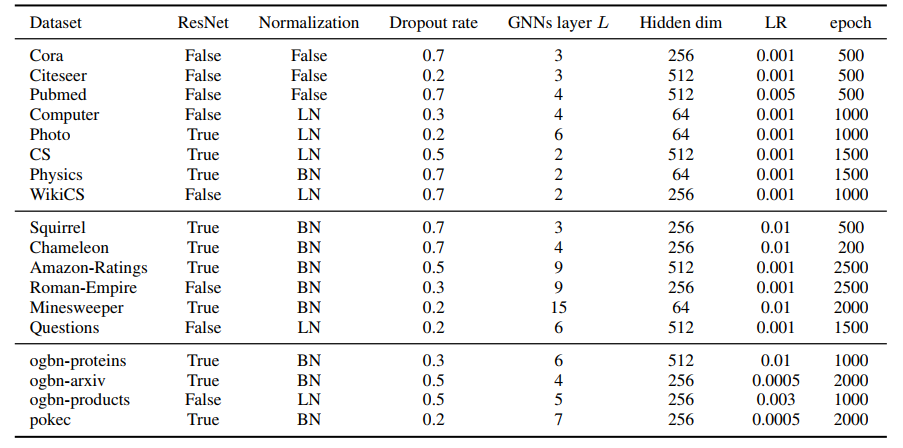


表 9：GraphSAGE∗ 的各数据集特定超参数设置

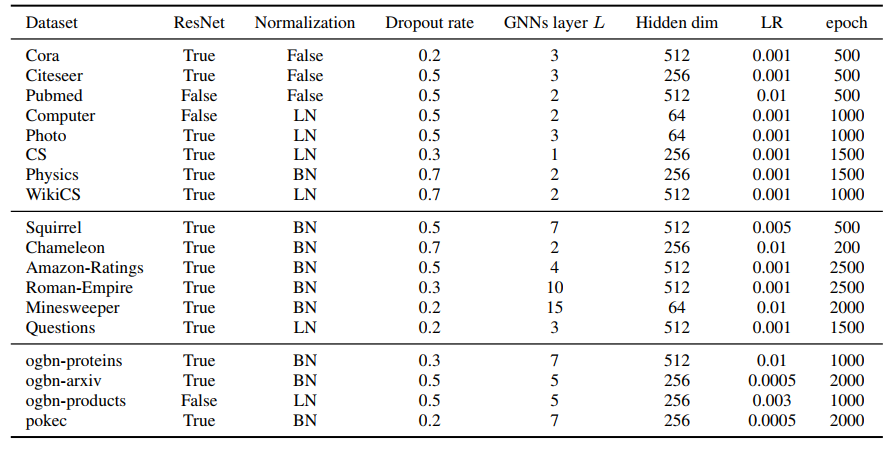


表 10：GAT∗ 的各数据集特定超参数设置

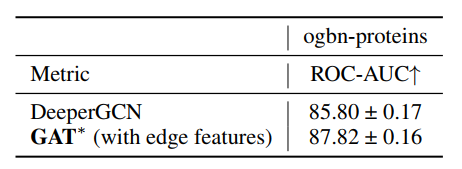


表 11：ogbn-proteins 数据集上的节点分类结果（%）

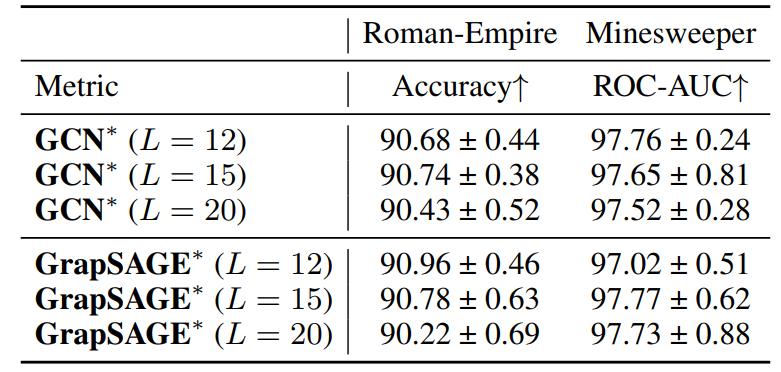


表 12：在异质图上关于层数 (L) 的消融研究结果（%）。

B.3 Jumping Knowledge模式与早期结果

Jumping Knowledge () 模式 [81] 聚合了来自不同层的表示，从而有效地捕捉图中不同邻域范围的信息。对于任意节点，模式的求和版本通过以下方式生成的表示：



其中是的层数。在我们之前的实验设置中，我们将视为的一种超参数配置。基于第3节中概述的超参数配置，我们扩展了调优空间，以包括是否使用的决策。在过去的实验中，我们没有执行详尽的搜索；相反，我们根据经验在该搜索空间内选择了子集，我们的早期结果报告在表13、14和15中（更多信息，请参考 <https://arxiv.org/abs/2406.08993v1>）。然而，在进行更详细的超参数调优后，我们发现可能并非必要。在大多数数据集中，不使用的结果与使用的结果相当，有时甚至更好。因此，在我们的论文中，我们从超参数调优的搜索空间中移除了。

C 可视化

在这里，我们展示了分类结果的t-SNE可视化。如图2所示，由GCN\*（我们的实现）生成的节点嵌入展现出比Polynormer\*生成的嵌入更大的类间距离。

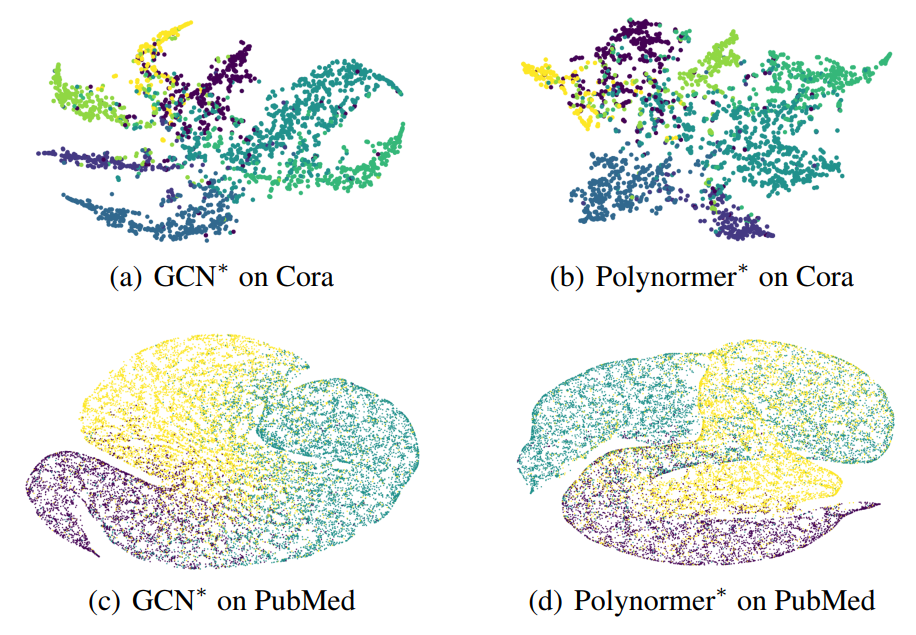


图 2：节点嵌入的 t-SNE 可视化结果。

D 局限性与更广泛影响

**更广泛影响**。本论文提出的工作其目标是推动机器学习领域的发展。我们的工作有许多潜在的社会影响，我们觉得在此处无需特别强调任何一项。

**局限性**。在本研究中，我们仅专注于节点分类任务，没有深入研究图分类[14, 44, 43]和链接预测[39, 87]任务。将我们的基准测试工作扩展到图级别和边级别的任务中，以涵盖经典的GNNs将是有益的。