
从对比性链接中提炼自我知识，在不传递信息的情况下 对图节点进行分类

罗毅

中国电子科技大学计算机科学与工程学院
成都 611731, 中国
cf020031308@163.com

陈**爱国** 

中国电子科技大学计算机科学与工程学院
成都 611731, 中国
agchen@uestc.edu.cn

柯岩

中国电子科技大学计算机科学与工程学院
成都 611731, 中国
kyan@uestc.edu.cn

凌天

中国电子科技大学计算机科学与工程学院
成都 611731, 中国
lingtian@uestc.edu.cn

摘要

如今，遵循信息传递范式的图谱神经网络（GNNs）已成为图形数据学习的主流方式。这种模式下的模型必须花费额外的空间来用邻接矩阵查找相邻的节点，并花费额外的时间来汇总相邻节点的多个信息。为了解决这个问题，我们开发了一种叫做LinkDist的方法，将连接节点对的自我知识提炼成多层感知器（MLP），而不需要聚合信息。用8个真实世界的数据集进行的实验表明，从LinkDist得到的MLP可以在不知道节点的邻接关系的情况下预测节点的标签，但在半监督 and 全监督节点分类的情况下，可以达到与GNN相当的准确性。此外，LinkDist得益于其非消息传递范式，我们还可以从任意抽样的节点对中提取自我知识，以对比的方式进一步提高LinkDist的性能。

1 简介

图是许多包含成对关系的现实世界数据的通用表示。例如，学术论文和他们的引用关系可以构建一个图，其中的论文是

*通讯作者。电子邮件：agchen@uestc.edu.cn 预印本。正

在审查中。

表示为节点，其编码的摘要为节点特征。如果一篇论文引用了另一篇论文，它们的节点在该图中是相连的。像这样的图通常被称为引文网络[15]。同样，我们也有共同购买网络[14]、共同作者网络[16]、蛋白质-蛋白质相互作用网络（PPI）、事后共同评论网络[8]等等[17]。

由于图的广泛应用，研究人员提出了越来越多的方法来挖掘图中的信息。图形神经网络（GNNs）[18]是一个被充分研究的方法系列。他们结合节点特征和图的结构信息来研究下游任务的节点表示，如节点分类，其目标是通过给节点分配适当的标签来进行分类。

许多图式神经网络遵循消息传递范式[5]，其中包含三个步骤：

1. 每个节点将其表示传递给其相邻的节点
2. 每个节点将其收到的表征进行汇总
3. 每个节点对聚合的信息进行转换，以获得其最新的表示方法

例如，图卷积网络（GCN）[12]中的一个层通过计算每个节点收到的信息的平均值来汇总这些信息，并通过线性变换来转换这些平均信息。

这种范式将神经网络应用于图，使图上的学习变得灵活而强大。然而，它仍然有一些缺点。

- 预测单个节点的标签必须汇总多个相邻节点的信息，这使得计算量很大，并且需要花费额外的空间来查找相邻节点（通常是用邻接矩阵）。
- 传递消息在实践中并不容易。出于批处理的目的，我们必须按节点的度数（作为其相邻节点的数量）进行分组，如果图太大，无法操作其邻接矩阵，则按组调用消息传递。
- 依靠相邻节点的信息，当插入新的节点或新的边时，需要花费图的时间和精力将信息传播到所有的节点。

纠正上述一些问题的一个方法是知识提炼[6]。知识蒸馏是将知识从训练有素的教师网络转移到较小的学生网络，以达到部署的目的。例如，当设备上的空间和计算资源太有限，无法存储邻接矩阵或及时汇总多个消息时，我们可以将训练有素的GCN的知识提炼成多层感知器（MLP）[2]并将MLP单独部署到这些设备上。部署的MLP可以预测，而不需要聚合相邻节点的信息，因此它可以是轻量级和快速的。然而，当图形发生变化时，MLP仍然要再次研究来自GCN的每一个节点的最新知识。这使得知识提炼法难以应用于动态图。

然而，我们观察到，正如我们在第4节中的实验所显示的那样，从训练好的GCN中提炼出来的MLP有时可以胜过原来的GCN。这表明，如果我们能找到一种方法将结构知识提炼到MLP中，那么消息传递可能就不是必须的。受此启发，我们建议从节点对中提炼知识。具体来说，我们通过MLP预测一个节点的特征，并通过分叉层预测其邻近节点的标签。同时，我们通过对它们的知识进行提炼，使这两个预测的一致性最大化。之后，我们省略分叉层，得到一个包含蒸馏结构知识的MLP。这种被命名为*LinkDist*的方法是轻量级的，快速的，并且容易实现。当图发生变化时，*LinkDist*会调整其参数，而不是将消息传播到所有节点。在解决消息传播中的所有三个挑战的同时，第4节中的节点分类实验表明，从*LinkDist*中得到的MLPs可以与GNNs的预测精度相当。

2 相关作品

自我知识蒸馏 除了知识蒸馏的师生方案，还有一种类似的技术，即自我知识蒸馏[21]，通过蒸馏模型，使其正规化。

自己的知识而不需要教师模型。实现这一目标的有效方法是诱导出与相关数据一致的预测，如同标签节点的通用标签分布[20]。我们的方法LinkDist也有类似的工作。它通过预测一个节点的标签与该节点本身及其相邻节点一致来提炼自己的知识。

对比学习 对比学习[13]是一种技术，它通过把相似的点拉到一起，把不相似的点推开，来学习区分没有标签的数据点。同样，我们提出了一种对比性的方法来训练LinkDist。当我们通过最小化两个连接节点的预测差异来提炼自知之明时，我们也最大化了两个任意采样节点的预测的差异。此外，与无标签学习的对比性学习不同，我们把用连接节点做出的预测拉到地面真实标签上，把用任意抽样的节点对做出的预测推离它们。

3 我们的方法：链接点

在这一节中，我们描述了LinkDist，它利用一个分叉的多层感知器，从节点的特征和位置上对节点进行预测。它的训练方式是在边缘集上迭代，并从链接中提炼出自我知识，以引起一致的预测。然后，我们提出一种对比训练的方法，从不存在的链接中提炼知识。训练之后，我们以两种模式评估LinkDist。一种是更快，一种是更准确。

3.1 叉式多层感知器

我们的目标是学习一个映射 $M: x \rightarrow (z, s)$ ，该映射接收一个节点 v 的特征 x ，并预测该节点的标签 y 和其相邻节点的标签。在图形数据中，聚合相邻节点的信息可以帮助一个节点确定其标签。反之亦然，用一个节点的特征部分恢复相邻节点的标签是合理的。

映射 $M: x \rightarrow (z, s)$ 可以实现为一个分叉的多层感知器，有隐藏层、输出层和推理层。我们将节点 v 的特征 x 送入隐藏层，将其编码为隐藏表征 h 。然后，输出层和推理层使用隐藏表征 h 分别产生标签分布的对数 z, s ，其中 z 是 v 的标签 y 的近似值， s 是 v 的相邻节点的可能标签分布。

3.2 LinkDist的培训

我们在图1中说明了LinkDist的训练情况。在训练阶段，我们对边缘集 $E = \{(v_i, v_j), \dots\}$ 进行迭代，将每个节点对 (v_i, v_j) 的特征 x_i, x_j 输入分叉MLP，得到它们的标签 y_i, y_j 的近似值 z_i, z_j 和它们相邻节点的标签的近似值 s_i, s_j 。

地面真理标签监督 如果节点 v_i 被分配了一个地面真理标签 y_i ，我们通过最小化其交叉熵损失来监督 z_i 与 y_i 。由于 v_i 和 v_j 是彼此的邻接节点，标签 y_i 也是 s_j 的可能分布，所以我们也用 y_i 来监督 s_j 。同样，如果节点 v_j 被分配了一个地面真理标签 y_j ，我们用 y_j 来监督 z_j 和 s_i 。

此外，我们迭代的边缘集的标签分布与节点集的标签分布不同。为了预测节点集的标签，我们在计算交叉熵损失 L_{CE} 时给标签分配权重 y/y_{ne} ，其中 y_n 是训练集的标签分布， y_e 是边缘集的标签分布，未标记的节点被屏蔽。

链接提炼 对于每个节点，我们有两个来自不同来源的标签预测。一个是来自它的特征。另一个是来自它的相邻节点。通过最小化它们之间的差异 L_{MSE} ，我们可以将它们中的知识相互提炼出来，使通过其特征信息和来自其相邻节点的结构知识来预测一个节点的标签的一

致性最大化。

总的来说，我们对 $L_{CE} + \alpha \cdot L_{MSE}$ 进行优化，以投合地面真实的标签监督和链接蒸馏。默认情况下，我们设定超参数 $\alpha = 1 - n_e / e$ ，其中 n_e 是标记节点出现的次数。

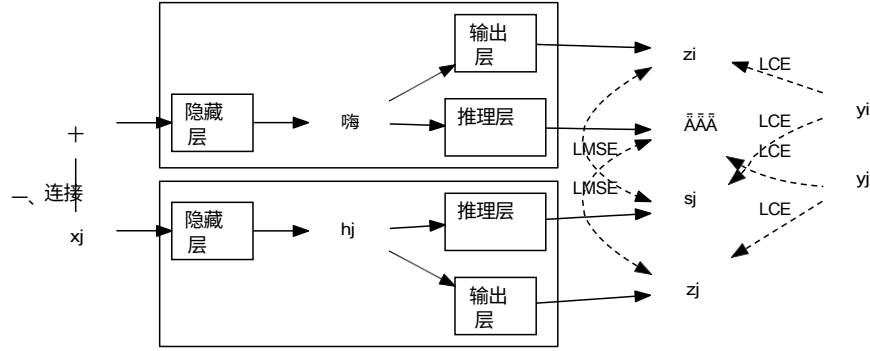


图1： LinkDist的训练。两个连接节点 (v_i, v_j) 的特征 x_i, x_j 被送入具有隐藏层、输出层和推理层的分叉MLP。分叉MLP产生 z_i, s_i, z_j, s_j 作为 v_i 的标签、 v_i 的相邻节点的标签、 v_j 的标签和 v_j 的相邻节点的标签的近似值。它们由 v_i, v_j, v_j (如果有的话) 的真实标签 y_i, y_j 监督。 s_i 和 s_j 包含标记的节点的暗知识和额外的知识。未标记的节点。因此，我们将知识从 s_i 提炼到 z_i ，从 s_j 提炼到 z_j 。

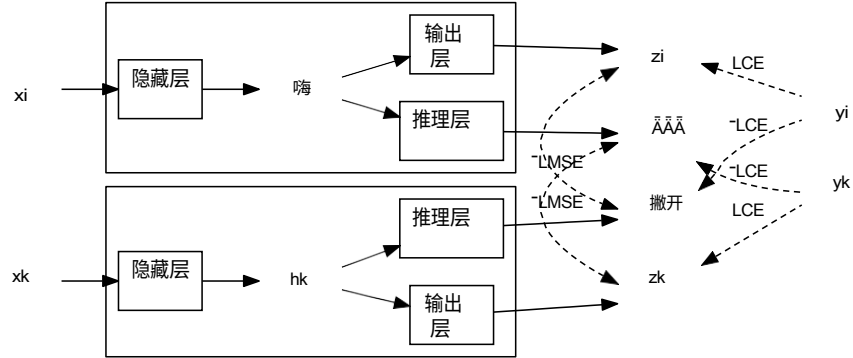


图2：负链接的对比性训练。两个任意抽样的节点 (v_i, v_k) 的特征 x_i, x_k 被送入分叉MLP。分叉MLP产生 z_i, s_i, z_k, s_k 作为 v_i 的标签、 v_i 的相邻节点的标签、 v_k 的标签和 v_k 的相邻节点的标签的近似值。 z_i, z_k 由 v_i, v_j 的地面真实标签 y_i, y_k 监督。由于节点 v_k 不可能连接 v_i ，我们将 s_i 推离 z_k 和 y_k ，分别用均方误差和交叉熵来衡量它们的差异最大化。同样地，我们将 s_k 推离 z_i 和 y_i 。

在边缘集中。换句话说，我们拥有的包含监督信息的标记节点越多，从相邻节点推断出的知识就越不重要。

3.3 负面链接的对比性训练

为了利用没有联系的节点对，我们提出了一种对比性训练的方法，如图2所示。

表1： 8个真实世界数据集的摘要

数据集	#节点	#特点	#Classes	# 边缘	平均学位
Cora	2708	1433	7	5278	3.90
景点介绍	3327	3703	6	4552	2.77
公共医学杂志	19717	500	3	44324	4.50
科拉满	19793	8710	70	63421	6.41
亚马逊照片	7650	745	8	119081	31.13
亚马逊电脑	13752	767	10	245861	35.76
共同作者CS	18333	6805	15	81894	8.93
合著者物理学	34493	8415	5	247962	14.38

我们对任意的节点对进行抽样，构建负链接[7]，并将每个节点对 (v_i, v_k) 的特征 x_i, x_k 送入分叉的MLP，得到它们的标签 y_i, z_k 的近似值 y_i, y_k 和它们相邻节点的标签的近似值 s_i, s_k 。

虽然我们用 y_i 来监督 z_i ，用 y_k 来监督 z_k ，就像我们在上一节3.2中所做的那样，我们最大限度地提高 s_i, z_k 的交叉熵和 s_k, z_i 的交叉熵。这是因为节点 v_k 是任意采样的，它不可能连接 v_i 。因此，节点 v_i 的标签的近似值 z_i 可能与分配给节点 v_k 的相邻节点的标签的近似值 s_k 不同，并且 z_k 可能与 s_i 不同。同样，我们最大化 z_i, s_k 的MSE和 z_k, s_i 的MSE来从负链接中提取自我知识。

这种对比性训练方法必须与前面3.2节中描述的训练过程一起采用。在这项工作中，我们将对比性训练的模型命名为 *CoLinkDist*。

3.4 LinkDist的两种评估模式

LinkDist中的分叉MLP用其输出层输出节点的标签分布 z ，用其推理层输出节点的相邻节点的标签分布 s 。当省略推理层及其输出 s 时，分叉的MLP成为一个MLP，但有从链接中提炼的结构知识。从 LinkDist 和 CoLinkDist 派生出来的MLP，被命名为 *LinkDistMLP* 和 *CoLinkDistMLP*，可以在不聚合其相邻节点的信息的情况下快速对图节点进行分类，因此它们可以在没有邻接矩阵或其他节点特征的情况下轻量级部署。此外，第4节中的实验表明，LinkDistMLP的准确度比普通MLP的准确度高出很多，接近于GCN的准确度。

当相邻节点的邻接矩阵和特征可用时，LinkDist也可以利用消息传递来提高其准确性。它将来自节点 v_i 本身的预测和从其相邻节点的预测来估计其可能的标签 \hat{y}_i ：

$$\hat{y}_i = z_i + \alpha \sum_{v_j \in N(v_i)} s_j$$

其中 $N(v_i)$ 是与节点 v_i 相邻的节点的集合。

4 实验：节点分类

在本节中，我们对8个真实世界的数据集进行了节点分类实验。重现这项工作的代码发表在 <https://github.com/cf020031308/LinkDist>。

4.1 数据集

为了全面评估我们的方法，我们用8个真实世界的数据集进行实验，包括4个引文网络、2个

共同购买网络和2个共同著作网络。它们是Cora、Citeseer、Pubmed[15]、Cora Full[3]、Amazon Photo、Amazon Computer[14]、Coauthor CS和Coauthor Physics[16]。这些数据集在很多方面都非常不同。我们在表1中对它们进行了总结。

半监督 and 全监督 在8个数据集中，Cora、Citeseer和Pubmed以其预定的数据集分割方式分布。详细来说，在这3个数据集中，训练集都包含了

每类有20个节点，验证集包含500个节点，测试集包含1000个节点。在Cora、Citeseer和Pubmed上，我们相应地使用这些拆分来做半监督的节点分类实验。在其他数据集上，我们把节点分成训练集、验证集和测试集，与Cora、Citeseer和Pubmed的情况完全一样。特别是，我们从每个类别中最多随机选择20个节点来构建训练集，500个节点来构建验证集，1000个节点来构建测试集。

除了这种半监督的设置，我们还用完全监督的设置进一步分割每个数据集。具体来说，我们在一个数据集中随机选择60%的节点来构建其训练集。然后，我们将其余的节点平均分为验证集和测试集。

归纳式和归纳式 在半监督式节点分类中，我们对归纳式和归纳式进行了实验。在过渡性设置下，方法在训练阶段使用所有节点的特征和连接。在归纳设置下，两个评估集（验证集和测试集）中的节点在训练阶段被屏蔽。在评估集中有端点的连接也是如此。这种设置符合实际情况，在这种情况下，图会不断地扩大新的节点[4]。

在全监督的节点分类中，大多数节点要么在评价集中，要么与评价集中的节点相连。因此，我们只实验了过渡性的设置。

4.2 基线

我们实现了3种方法作为基线。

图卷积网络（GCN） LinkDist在消息传递范式下工作，可以利用大多数GNN的所有信息。因此，我们将LinkDist与GCN进行比较，以考察其在常见情况下的适用性。由于归纳设置，我们实现GCN的归一化邻接矩阵为 $(D + I)^{-1}(A + I)$ 而不是 $(D + I)^{-1}(A + I)(D + I)^{-1}$ 。I是身份矩阵，A是邻接矩阵，D是一个对角线矩阵，其中每个对角线条目都是 d_i 是节点v的学位 i 。

多层感知器（MLP） 在非信息传递模式下，LinkDistMLP的工作方式就像MLP一样。然而，由于从链接中提炼出的知识，LinkDistMLP应该超越它。所以我们在这里放置一个MLP，作为LinkDistMLP的下限。

从GCN中提炼出的MLP（GCN2MLP） 我们向MLP传授由训练好的GCN预测的评价集外的所有节点的软标签逻辑。提炼出来的MLP与MLP和LinkDistMLP的结构完全相同，但继承了GCN的结构知识。我们认为这是LinkDistMLP的一个适当的基准，以衡量它离 "完美 "还有多远。

所有MLP和GCN的实现都有3个全连接的层。LinkDist有2个隐藏层，1个输出层和1个推理层，都实现为全连接层。所有层内的隐藏表示都有相同的大小256。每两层之间有1个批量归一化[10]，1个层归一化[1]，1个Dropout[9]，有0.5的可能性擦除一个位置和1个LeakyReLU[19]激活器。

所有的方法都可以根据在评估阶段是否利用相邻节点的信息而分为两组。MLP、GCN2MLP、LinkDistMLP和CoLinkDistMLP不涉及信息传递，而GCN、LinkDist和CoLinkDist涉及信息传递。

4.3 结果

我们使用Adam[11]，学习率为0.01来优化参数。GCN的训练是全批进行的。MLP和LinkDist是以迷你批次训练的，批次大小设置为1024。由于MLP和GCN在节点集上迭代，而LinkDist在边缘集上迭代，我们对前两个模型进行了200个epochs的训练，但对LinkDist进行了大约 $200/\bar{d}$ epochs的训练，其中 \bar{d} 是图的平均程度。

在每一个训练纪元之后，我们在两个评估集上评估该方法，产生一对准确率分数。运行后，我们得到测试集的准确度分数与最高的

表2：半监督式和归纳式节点分类的准确率分数（%）。方法按其在评估阶段是否聚集相邻节点的信息分为两组。我们对每组中的每个数据集的最高准确率进行了加粗。

方法	Cora	景点介绍	公共医学杂志	科拉满	亚马逊照片	亚马逊电脑	共同作者CS	合著者物理学
MLP	56.28	54.74	70.33	41.91	75.89	66.14	90.11	89.53
GCN2MLP	67.28	64.19	76.72	53.59	87.83	79.36	93.13	93.20
链接DistMLP	66.38	67.85	72.78	47.15	86.16	76.44	89.20	90.68
协同工作	68.93	69.85	74.32	49.88	86.34	76.48	90.16	91.38
GCN	77.12	62.98	73.00	56.88	87.63	79.61	88.21	92.67
链接地址	75.72	71.19	74.84	56.09	89.67	82.36	91.64	92.84
协同工作	77.14	71.74	74.77	57.33	90.56	81.93	92.20	93.04

表3：半监督和过渡性节点分类的准确度分数（%）。方法按其在评估阶段是否聚集相邻节点的信息分为两组。我们对每组中的每个数据集的最高准确率进行了加粗。

方法	Cora	景点介绍	公共医学杂志	科拉满	亚马逊照片	亚马逊电脑	共同作者CS	合著者物理学
MLP	56.28	54.74	70.33	41.91	75.89	66.14	90.11	89.53
GCN2MLP	67.61	63.29	77.87	54.15	87.11	78.21	92.49	93.42
链接DistMLP	80.79	70.26	72.41	51.78	88.19	78.23	89.83	90.72
协同工作	81.19	70.96	75.41	53.43	88.90	78.77	90.66	91.63
GCN	76.47	63.11	73.99	57.58	87.00	77.95	87.93	92.79
链接地址	81.05	70.27	74.06	55.87	90.14	82.65	90.94	92.06
协同工作	81.39	70.79	75.64	57.05	90.54	82.53	91.88	92.74

表4：全监督和过渡性节点分类的准确度分数（%）。方法按其在评估阶段是否聚集相邻节点的信息分为两组。我们对每组中的每个数据集的最高准确率进行了加粗。

方法	Cora	景点介绍	公共医学杂志	科拉满	亚马逊照片	亚马逊电脑	共同作者CS	合著者物理学
MLP	74.71	71.95	87.92	62.03	91.75	85.08	95.35	96.57
GCN2MLP	75.95	72.93	86.25	62.39	91.10	85.41	95.29	96.46
链接DistMLP	87.58	75.25	88.79	69.53	93.83	89.44	95.68	96.91
协同工作	87.54	75.77	89.53	69.83	94.12	88.85	95.74	96.87
GCN	86.03	73.37	84.79	68.97	93.09	90.08	92.27	95.96
链接地址	88.24	74.72	88.86	69.87	93.75	89.49	95.66	96.87
协同工作	87.89	75.79	89.58	70.32	94.36	89.42	95.80	97.05

验证集的准确性得分作为被评估方法的得分。我们在每个数据集上运行每个方法10次，并在表2、表3和表4中记录平均得分。

从这些表格中，我们可以看到，LinkDistMLP在几乎所有的8个数据集上都能以较大的幅度超过同结构MLP的表现。虽然LinkDistMLP只用中心节点的特征进行预测，但它在归纳设置中仍然与GCN相匹配，在其他两个设置中甚至超过了GCN。此外，利用消息传递（LinkDist）和对比训练（CoLinkDistMLP，CoLinkDist）都可以持续提高准确率。这表明从节点对中提炼自我知识是成功的，而且LinkDists在节点分类任务中是与GNN竞争的方法。

我们还注意到，GCN2MLP在半监督环境下表现得非常好，特别是在归纳环境下，LinkDistMLP和GCN2MLP之间仍有差距，如表2和表3所示。然而，如表4所示，在完全监督环境中，GCN2MLP的准确性几乎与MLP相同。它比GCN和LinkDists差得多。这是因为半监督设置中的GCN2MLP从训练集之外的节点获得了额外的知识，而在完全监督设置中，所有的节点要么在训练集中，要么在评估集中。没有太多的额外知识需要提炼。

我们倾向于用PPI数据集[8]进行实验，训练集是由几个完整的图组成的，训练的方法需要对完全未见过的图上的节点进行分类，但发现LinkDist中的超参数 α 变成了0，LinkDist退化成了一个普通的MLP。换句话说，LinkDist在完全未见过的图上诱导可能会有问题。

5 结论

在这项工作中，我们提出了一种名为LinkDist的新方法，在不需要传递信息的情况下对图上的节点进行分类。

半监督和全监督的节点分类实验表明，即使没有相邻节点的信息，LinkDist也能以相当的准确度预测GNN。

LinkDist可以非常实用，因为它纠正了消息传递的三个问题：

- LinkDist以小批量的边缘进行学习。与使用邻接矩阵相比，它需要更少的空间，而且对于大型图来说，可以很容易地实现。
- LinkDist可以部署到资源有限的设备上，像MLP一样快速预测，但又像GNN一样准确，不需要存储相邻节点及其特征。
- 当新的节点或新的边被插入时，LinkDist通过优化其可学习参数来更新现有节点的预测，而不是将信息传播给所有节点。

虽然消息传递模式下的LinkDist（LinkDist和CoLinkDist）可以产生与GNN非常有竞争力的准确度，但如果实验归纳设置，非消息传递模式下的LinkDist（LinkDistMLP和CoLinkDistMLP）与GCN2MLP之间仍有差距。这显示了像MLPs这样简单的网络的力量，并指出了LinkDist的潜力。在未来，我们可能会研究找到更有效的解决方案，将知识提炼成MLPs，以缩小差距。

更广泛的影响

这项工作解决了图神经网络中广泛采用的消息传递的问题。提出的LinkDist是一种通用的方

法。它可以降低现有应用中从图形数据中挖掘信息的成本。就我们所知，它没有可预见的负面社会影响。

参考文献

[1] Ba, L. J., Kiros, J. R., and Hinton, G. E. (2016). Layer normalization. *CoRR*, abs/1607.06450.

- [2] Bishop, C. M. (2007). *模式识别和机器学习, 第五版*. 信息科学和统计学. Springer.
- [3] Bojchevski, A. and Günnemann, S. (2018). 图的深度高斯嵌入: 通过排名进行无监督的归纳学习. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.
- [4] Chen, J., Ma, T., and Xiao, C. (2018). Fastgcn: 通过重要性采样实现图卷积网络的快速学习. In *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*. OpenReview.net.
- [5] Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., and Dahl, G. E. (2017). 量子化学的神经信息传递. *CoRR*, abs/1704.01212.
- [6] Gou, J., Yu, B., Maybank, S. J., and Tao, D. (2020). 知识提炼: A survey. *CoRR*, abs/2006.05525.
- [7] Grover, A. and Leskovec, J. (2016). node2vec: 网络的可扩展特征学习. In Krishnapuram, B., Shah, M., Smola, A. J., Aggarwal, C. C., Shen, D., and Rastogi, R., editors, *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pages 855-864. ACM.
- [8] Hamilton, W. L., Ying, Z., and Leskovec, J. (2017). 大图上的归纳表征学习. In Guyon, I., von Luxburg, U., Bengio, S., Wallach, H. M., Fergus, R., Vishwanathan, S. V. N., and Garnett, R., editors, *Proances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pages 1024-1034.
- [9] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2012). 通过防止特征检测器的共适配来改进神经网络. *CoRR*, abs/1207.0580.
- [10] Ioffe, S. and Szegedy, C. (2015). 批量归一化: 通过减少内部协变量偏移来加速深度网络训练. In Bach, F. R. and Blei, D. M., editors, *Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015*, volume 37 of *JMLR Workshop and Conference Proceedings*, pages 448-456. JMLR.org.
- [11] Kingma, D. P. and Ba, J. (2015). Adam: 一种随机优化的方法. In Bengio, Y. and LeCun, Y., editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, Conference Track Proceedings*.
- [12] Kipf, T. N. and Welling, M. (2016). 用图卷积网络进行半监督性分类. *CoRR*, abs/1609.02907.
- [13] Liu, X., Zhang, F., Hou, Z., Wang, Z., Mian, L., Zhang, J., and Tang, J. (2020). 自我监督的学习: 生成性或对比性. *CoRR*, abs/2006.08218.
- [14] McAuley, J. J., Targett, C., Shi, Q., and van den Hengel, A. (2015). 基于图像的风格和替代物推荐. In Baeza-Yates, R., Lalmas, M., Moffat, A., and Ribeiro-Neto, B. A., editors, *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, Santiago, Chile, August 9-13, 2015*, pages 43-52. ACM.
- [15] Sen, P., Namata, G., Bilgic, M., Getoor, L., Gallagher, B., and Eliassi-Rad, T. (2008). 网络数据中的集体分类. *AI Mag.*, 29(3):93-106.
- [16] Wang, K., Shen, Z., Huang, C., Wu, C., Dong, Y., and Kanakia, A. (2020). 微软学术图: 当专家是不够的. *Quant.Sci. Stud.*, 1(1):396-413.
- [17] Wang, M., Zheng, D., Ye, Z., Gan, Q., Li, M., Song, X., Zhou, J., Ma, C., Yu, L., Gai, Y., Xiao, T., He, T., Karypis, G., Li, J., and Zhang, Z. (2019). 深度图库: 一个以图为中心的、高性能的图神经网络包. *arXiv预印本 arXiv:1909.01315*.
- [18] Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., and Yu, P. S. (2021). 图形神经网络的综合调查. *IEEE Trans. Neural Networks Learn. Syst.*, 32(1):4-24.
- [19] Xu, B., Wang, N., Chen, T., and Li, M. (2015). 卷积网络中整流激活的经验性评估. *CoRR*, abs/1505.00853.

- [20] Yun, S., Park, J., Lee, K., and Shin, J. (2020).通过自知之明提炼的常规化类预测。In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 13873-13882.IEEE.
- [21] Zhang, L., Song, J., Gao, A., Chen, J., Bao, C., and Ma, K. (2019).做你自己的老师：通过自我提炼提高卷积神经网络的性能。

检查清单

检查表在参考文献之后。请仔细阅读检查表指南，以了解如何回答这些问题的信息。对于每个问题，将默认的[TODO]改为[是]、[否]、或[不适用]。我们强烈建议你**在答案中加入一个理由**，可以参考你论文中适当的部分，或者提供一个简短的内联描述。比如说：

- 你是否包括代码和数据集的许可证？ [是]见第？
- 你是否包括代码和数据集的许可证？ [没有] 代码和数据是专有的。
- 你是否包括代码和数据集的许可证？ [不适用]

请不要修改问题，只使用提供的宏来回答。请注意，检查表部分不计入页数限制。在你的论文中，请删除这个说明部分，只保留上面的检查表部分标题和下面的问题/答案。

1. 对于所有的作者...
 - (a) 摘要和引言中的主要主张是否准确反映了论文的贡献和范围？ [是]。
 - (b) 你是否描述了你工作的局限性？ [是的] GCN2MLP在归纳设置上优于我们的工作 2.我们的工作完全未见过的图上诱导有其问题 4.3.
 - (c) 你是否讨论过你的工作可能带来的任何负面社会影响？ [是] 见第5节
 - (d) 你是否阅读过伦理审查指南并确保你的论文符合这些指南？ [是]。
2. 如果你包括理论上的结果...
 - (a) 你是否说明了所有理论结果的全套假设？ [不适用] 这项工作不包含任何理论结果。
 - (b) 你是否包括所有理论结果的完整证明？ [不适用] 这项工作不包含任何理论结果。
3. 如果你做实验...
 - (a) 你是否包括了重现主要经验-心理结果所需的代码、数据和说明（在补充材料中或作为一个URL）？ [是] 我们在补充材料和网页 <https://github.com/cf020031308/LinkDist>，包括了足以重现我们工作的代码和说明。
 - (b) 你是否指定了所有的训练细节（例如，数据分割，超参数，它们是如何选择的）？ [是]请参见第4节。
 - (c) 你是否报告了误差条（例如，在多次运行体验后，与随机种子有关）？ [是]我们在附录A的表5、表6和表7中列出了标准偏差。
 - (d) 你是否包括计算总量和使用的资源类型（例如，GPU的类型，内部集群，或云供应商）？ [没有]我们认为硬软件并不重要。即使用CPU也可以重现我们的工作，我们使用的是11GB内存的GPU。
4. 如果你正在使用现有的资产（如代码、数据、模型）或策划/发布新的资产...
 - (a) 如果你的工作使用了现有的资产，你是否引用了创作者的名字？ [是]在本文开头引用了他们的名字1，并在实验的第4节再次引用。
 - (b) 你有没有提到资产的许可证？ [没有]我们无法找到我们使用的数据集的许可证

-
- (c) 你是否在补充材料中包括任何新的资产，或作为一个URL？[是]我在补充材料中包括我的代码。
- (d) 你是否讨论过是否以及如何从你正在使用/收集数据的人那里获得同意？[不适用]

表5：半监督和归纳式节点分类的标准偏差（%）。方法分为两组，按方法是否在评估阶段聚集来自相邻节点的信息来划分。我们对每组中的每个数据集的最高准确率进行了加粗。

方法	Cora	景点介绍	公共医学杂志	科拉满	亚马逊照片	亚马逊电脑	共同作者CS	合著者物理学
MLP	0.87	1.10	1.05	0.98	2.97	2.42	0.93	1.14
GCN2MLP	0.98	0.91	1.76	1.92	2.10	1.47	0.40	0.95
链接DistMLP	0.90	0.72	2.03	1.81	1.73	1.95	1.35	1.05
协同工作	0.62	1.04	0.47	1.67	1.90	2.41	0.84	0.90
GCN	0.91	0.64	1.61	1.79	1.82	2.25	1.19	1.09
链接地址	0.79	0.91	1.20	1.74	1.91	1.23	0.76	1.01
协同工作	0.58	0.77	0.68	2.08	2.23	2.12	0.83	1.08

表6：半监督和过渡性节点分类的标准偏差（%）。按方法在评估阶段是否聚集相邻节点的信息，将方法分为两组。我们对每组中的每一个数据集都大胆采用了最高的准确性。

方法	Cora	景点介绍	公共医学杂志	科拉满	亚马逊照片	亚马逊电脑	共同作者CS	合著者物理学
MLP	0.87	1.10	1.05	0.98	2.97	2.42	0.93	1.14
GCN2MLP	0.83	1.74	1.68	2.14	1.23	1.73	0.89	0.79
链接DistMLP	0.43	0.70	1.35	1.43	2.07	2.49	1.11	1.10
协同工作	0.61	0.79	0.38	1.83	1.88	2.05	0.68	1.65
GCN	1.12	1.24	1.39	1.68	1.53	2.55	1.79	0.86
链接地址	0.85	0.75	0.85	2.05	1.78	2.51	1.08	1.06
协同工作	0.46	0.74	1.02	1.43	1.88	2.52	0.98	1.42

(e) 你是否讨论过你正在使用/收集的数据是否包含个人信息或攻击性内容？ [不适用]

5. 如果你使用众包或进行以人为对象的研究...

- (a) 你是否包括给参与者的指示全文和屏幕截图（如果适用）？ [不适用]
- (b) 你是否描述了任何潜在的参与者风险，如果适用的话，请附上机构审查委员会（IRB）批准的链接？ [不适用]
- (c) 你是否包括支付给参与者的估计小时工资和用于参与者报酬的总金额？ [不适用]

A 附录

表7: **全监督和过渡性**节点分类的标准偏差 (%)。按方法在评估阶段是否聚集相邻节点的信息, 将方法分为两组。我们对每组中的每一个数据集的最高准确率进行了加粗。

方法	Cora	景点介绍	公共医学杂志	科拉满	亚马逊照片	亚马逊电脑	共同作者CS	合著者物理学
MLP	1.80	2.23	0.52	0.90	0.82	0.77	0.22	0.19
GCN2MLP	1.70	1.55	0.64	0.61	0.55	0.63	0.34	0.24
链接DistMLP	1.24	2.00	0.56	0.67	0.41	0.63	0.25	0.20
协同工作	1.61	1.35	0.37	0.61	0.48	0.88	0.51	0.20
GCN	1.67	1.59	0.82	0.76	0.64	0.54	0.45	0.23
链接地址	1.23	1.90	0.44	0.66	0.30	0.51	0.32	0.22
协同工作	1.24	1.45	0.26	0.62	0.43	0.72	0.34	0.17