

Strategies for Pre-training Graph Neural Networks 的复现与改进

范洛玮

摘要

近年来，图神经网络（GNN）蓬勃发展，其中预训练 GNN 是其中一个努力方向。在实际应用场景中，我们往往需要训练一个模型来对与训练数据截然不同的样本做出准确的预测，而训练期间特定于该类型任务的标签很少。应对这一挑战的有效方法是在数据丰富的相关任务上预训练模型，然后在感兴趣的下游任务上对其进行微调。虽然预训练在许多计算机视觉和自然语言处理领域都很有效，但如何在图数据集上有效地使用预训练仍然是一个悬而未决的问题。以往人们在整个图或单个节点的水平上预训练 GNN 的简单策略只能带来有限的性能提升，甚至可能导致许多下游任务的负迁移。因此，原论文提出了一个用于预训练 GNN 的策略。它能够成功地在单个节点以及整个图的层面上预训练一个富有表现力的 GNN，能使 GNN 可以同时学习有用的局部和全局表示。尽管原论文的预训练策略性能足够出色，但我系统地研究了多个图分类数据集的预训练。我发现，相较于原论文的方法：在节点上使用上下文预测任务或者属性屏蔽预测任务加上图级别的监督任务；我把上下文预测与属性屏蔽预测任务结合起来充当节点级别的预训练策略，然后加上图级别监督任务后构成完整的预训练策略；它对下游任务产生了更好的性能，实验证明了这一点。我分析其原因可能是它能学习到图上的更多信息。更多信息详见下文。

关键词：图神经网络；图预训练；图表示学习

1 引言

迁移学习是指最初在某些任务上训练的模型被重新用于不同但相关的任务的设置。迁移学习在计算机视觉和自然语言处理领域取得了巨大成功。尽管预训练是一种有效的迁移学习方法，但很少有研究将预训练推广到图数据。

分子可以自然地表示为图，其中原子为节点，化学键为边。因此，GNN 可用于处理分子图数据。预训练 GNN 有可能为图数据集学习所具有的两个基本挑战提供有吸引力的解决方案：首先，特定于任务的标记数据可能极其稀缺。在化学和生物学等科学领域的重要图数据集中，这个问题更加严重，其中数据标记是极其耗费时间和资源的工作；其次，来自现实世界的图数据通常包含分布外样本，这意味着训练集中的图在结构上与测试集中的图有很大不同。分布外预测在现实世界的图数据集中很常见，例如，当人们想要预测一种新合成的分子或者寻找自然界新的分子时，该分子的化学性质往往与迄今为止所发现的所有分子都不相同，因此也不同于所有训练集中的分子。然而，图数据集的预训练仍然是一个艰巨的挑战。几项重要研究表明，成功的迁移学习不仅仅是增加与下游任务相关领域的标记数据集的数量。更

重要的是，它需要大量的领域专业知识来仔细选择与感兴趣的下游任务相关的示例和目标标签。否则，将知识从相关预训练任务转移到新的下游任务可能会损害泛化能力，这被称为负转移，并显著地限制预训练模型的适用性和可靠性。

为了利用丰富的未标记分子，人们付出了巨大的努力来预训练 GNN 以进行分子图表示。原文是该主题的开创性工作，它首先在大规模未标记分子图数据集上使用 Masked Atoms Modeling (MLM) 式预训练任务来预训练 GNN。具体来说，我们随机屏蔽一定比例的原子，然后预训练模型来预测它们。属性屏蔽预测任务已成为一项基本的预训练任务，许多后续工作都将其作为预训练的子任务。然而，原文观察到仅使用属性屏蔽预测任务（节点级预训练任务）进行预训练有时会引发负迁移问题（即预训练模型落后于未预训练模型）。直观上，他们认为这种现象可以归因于缺乏图级预训练任务，因此他们引入了有监督的图级预训练策略。总的来说，原文的预训练策略包含有节点级和图级，其中，任务类型又可能划分为属性预测和上下文预测。在节点级别任务中，原文使用属性屏蔽预测或者图上下文预测任务。但我发现，仅仅使用其中一种任务充当节点级预训练时，虽然能达到出色的预测性能，但它并不是最佳的选择。我通过在 8 个数据集上的多轮实验证明了这一点。最优的设计应该是把这两种任务结合起来，因为两种任务学习到的信息侧重点是不一样的，属性屏蔽预测偏向于去学习图上的节点信息，而图上下文预测偏向于去学习图上的更为丰富的结构化信息，这种结构化信息通常涉及多个节点和边，不同的节点属性、数量，以及边的不同连接、数量等都会影响这个图上下文结构。具体来说，我重新设计了一种有效的预训练 GNN 策略。关键思想是使用易于访问的节点级信息，并鼓励 GNN 除了图级知识之外还捕获有关节点和边的特定领域知识。这有助于 GNN 在全局和局部级别学习有用的表示，并且能够对不同的下游任务来生成鲁棒且可转移的图级别表示。我把原先的时间划分为两部分，分别进行两种任务的预训练。在时间上，它与原先做法一样，并无增加时间开销。在性能上，它优于之前的策略。

总之，我的贡献是：（1）在原文基础上重新设计了节点级别预训练。相较于之前的策略，它能学习到图数据中更多有用信息，从而产生了更好的性能。（2）整理了原开源代码，使之可读性更强。

2 相关工作

近年来，图神经网络在图表示学习方面取得了显著的成功。虽然有效且普遍，但它们需要昂贵的标注，并且几乎无法推广到未见过的图数据 [12, 14]，这给实际应用带来了障碍。为了弥补这些缺陷，人们在预训练图神经网络方面付出了巨大的努力。最近的一些工作遵循对比学习范式 [6, 7, 21, 23]，其中 Qiu 等人设计了图对比编码 (GCC) 以捕获跨多个网络的通用网络拓扑属性，GCC 的预训练任务为网络内和跨网络的子图实例辨别，并利用对比学习使图神经网络能够学习可转移的结构表示 [7]。对于分子预训练，GraphCL [19] 及其变体 [11, 18] 将锚分子图的增强版本嵌入彼此靠近，并将其他分子的嵌入推开。此外，DGI [13] 和 InfoGraph [10] 被提出通过最大化不同粒度的图级表示和子结构级表示之间的互信息来获得图或节点的表达表示。此外，有些工作采用生成或预测先验任务，通常，GPT-GNN [3] 引入了属性图生成任务来预训练 GNN，以便它们能够捕获图的结构和语义属性。为了捕获分子图基序中的丰富信息，GROVER [8] 和 MGSSL [20] 提出预测或生成基序。考虑到 3D 几何信息在预测分子特性中起着至关重要的作用，最近的几项工作 [1, 5, 9, 22] 在具有 3D 几何信息的分子数据集上预训练

GNN 编码器。此外, Xia 等人提出一种新的预训练策略, 包括一种新颖的节点级预训练任务, 称为掩蔽原子建模 (MAM); 和用于图级预训练的三元组掩码对比学习 (TMCL); MAM 和 TMCL 构成了一种新颖的预训练框架 Mole-BERT [15]。我建议读者参考最近的一项调查 [16] 以获取更多相关文献。上述许多工作都是在 2020 年之后发表的, 而原文工作是在 2020 年发表的, 所以上述一些工作比原文工作更加优秀本身并不奇怪, 需要注意的是, Mole-BERT 是原文的进一步探索。

3 预备知识

图神经网络 (GNN) 是图数据建模的主要工具。图数据的结构指导局部邻域信息的聚合, 并为每个节点提供更具上下文的表示。此外, 我们可以采用图池化操作来获得整个图的表示。设 $G = (\mathcal{V}, \mathcal{E})$ 表示一个图, 其节点属性为 $x_v (v \in V)$ 和边属性为 $e_{uv} ((u, v) \in \mathcal{E})$ 。形式上, 假设 $h_v^{(l)}$ 是节点 v 在 GNN 第 l 层处的表示, $\mathcal{N}(v)$ 是节点 v 的所有邻居节点, GNN 从第 $(l - 1)$ 层到第 l 层的更新过程为:

$$h_v^{(l)} = \text{COMBINE} \left(h_v^{(l-1)}, \text{AGGREGATE} \left(\left\{ \left(h_u^{(l-1)}, e_{uv} \right) : u \in \mathcal{N}(v) \right\} \right) \right)$$

其中 e_{uv} 表示节点 u 和 v 之间的边。AGGREGATE (\cdot) 是邻域信息的聚合函数 (例如均值算子)。COMBINE (\cdot) 联合邻居和节点 v 的信息。经过 L 次消息传递迭代后, 最后一次迭代中的隐藏状态 $h_v^{(L)}$ 是 v 的嵌入。最后, 我们采用 READOUT (\cdot) 操作 (例如求平均、求和或图池化) 来获得整个图 G 的表示 \mathbf{h}_G :

$$\mathbf{h}_G = \text{READOUT} \left(\left\{ h_v^{(L)} \mid v \in \mathcal{V} \right\} \right)$$

4 本文方法

预训练策略的技术核心是在单个节点以及整个图的层面上预训练 GNN 的概念, 这个概念鼓励 GNN 在两个级别捕获特定于领域的语义。接下来, 我首先介绍节点级预训练方法, 然后介绍图级预训练方法。最后, 我将描述完整的预训练策略。

4.1 本文方法概述

我在原文方法的基础上做了一些改进。我重新设计了节点级预训练策略, 具体来说, 我把属性屏蔽预测任务和子图上下文预测任务结合起来充当节点级预训练策略; 这样做的理由有两点, 首先, 两种任务会使模型学习到更多知识, 因为使用单一任务时实验效果是非常好的, 所以很自然的想法就是组合起来的效果怎么样; 其次, 这样的新策略并不会额外增加计算量。在 8 个数据集上的实验结果表明了它能够带来性能上的提升。

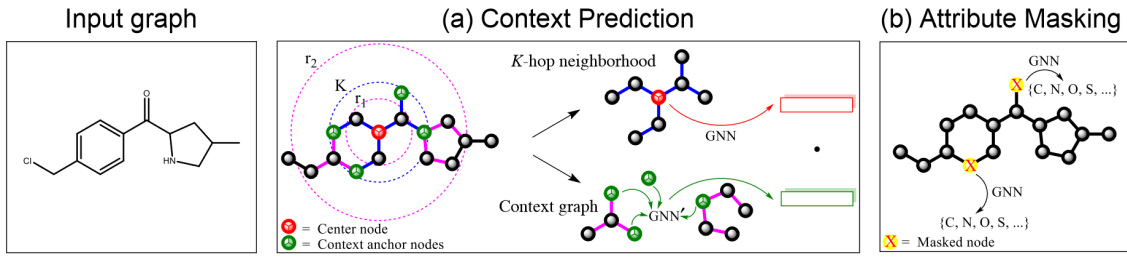


图 1. 节点级预训练策略示意图：上下文预测和属性屏蔽。(a) 在上下文预测中，邻域图是围绕选定中心节点的 K 跳邻域，其中 K 是 GNN 层数，图中设置为 2。上下文图被定义为从中心节点开始的 r_1 -hop 和 r_2 -hop 之间的周围图结构，其中图中使用 $r_1 = 1$ 和 $r_2 = 4$ 。(b) 在属性屏蔽中，输入节点/边属性被随机屏蔽，并要求 GNN 来预测它们。

4.2 节点级预训练

对于 GNN 的节点级预训练，我们使用易于访问的未标记数据来捕获图中特定领域的知识。在这里，我们使用到两种自监督方法：上下文预测和属性屏蔽预测。

4.2.1 上下文预测

在上下文预测中，我们使用子图来预测其周围的图结构。我们的目标是预训练 GNN，以便它将出现在相似上下文结构中的节点映射到附近的嵌入。

邻域图和上下文图。对于每个节点 v ，我们定义 v 的邻域图和上下文图如下。 v 的 K 跳邻域图包含图中距 v 最多 K 跳的所有节点和边。这是因为 K 层 GNN 聚合了 v 的 K 阶邻域的信息，因此节点嵌入 $h_v^{(K)}$ 取决于距离 v 最多 K 跳的节点。我们定义节点 v 的上下文图作为围绕 v 邻域的图结构。上下文图由两个超参数 r_1 和 r_2 描述，它表示距 v 的 r_1 跳和 r_2 跳之间的子图（即，它是宽度为 $r_2 - r_1$ 的环）。邻域图和上下文图的示例如图 1 (a) 所示。我们要求 $r_1 < K$ ，以便一些节点在邻域图和上下文图之间共享，我们将这些节点称为上下文锚节点。这些锚节点提供有关邻域图和上下文图如何相互连接的信息。

使用辅助 GNN 将上下文编码为固定向量。由于图的组合性质，直接预测上下文图是很困难的。这与自然语言处理不同，自然语言处理中的单词来自固定且有限的词汇表。为了实现上下文预测，我们将上下文图编码为固定长度的向量。为此，我们使用辅助 GNN，我们将其称为上下文 GNN。如图 1 (a) 所示，我们首先应用上下文 GNN（在图 1 (a) 中表示为 GNN'）来获得上下文图中的节点嵌入。然后，我们对上下文锚节点的嵌入进行平均以获得固定长度的上下文嵌入。对于图 G 中的节点 v ，我们将其相应的上下文嵌入表示为 c_v^G 。

通过负采样学习。然后，我们使用负采样来联合学习主 GNN 和上下文 GNN。主 GNN 对邻域图进行编码以获得节点嵌入。上下文 GNN 对上下文图进行编码以获得上下文嵌入。具体来说，上下文预测的学习目标是特定邻域图和特定上下文图是否属于同一节点的二元分类：

$$\sigma \left(h_v^{(K)\top} c_{v'}^{G'} \right) \approx \mathbf{1}\{v \text{ and } v' \text{ are the same nodes}\}$$

其中 $\sigma(\cdot)$ 是 sigmoid 函数， $\mathbf{1}(\cdot)$ 是指示函数。我们要么让 $v' = v$ 且 $G' = G$ （即，正邻域上下文对），要么从随机选择的图 G' （即负邻域上下文对）中随机采样 v' 。我们使用负采样率为 1（正负样本对一比一），并使用负对数似然作为损失函数。预训练后，保留主 GNN 作为我们的预训练模型。

4.2.2 属性屏蔽预测

在属性屏蔽预测中，我们的目标是通过学习分布在图结构上的节点/边属性的规律来捕获领域知识。

屏蔽节点和边属性。属性屏蔽预训练的工作原理如下：我们屏蔽节点/边属性，然后让 GNN 基于邻近结构预测这些属性。图 1 (b) 说明了我们提出的方法应用于分子图的情况。具体来说，我们通过用特殊的屏蔽指示符替换它们来随机屏蔽输入节点/边属性，例如分子图中的原子类型。然后，我们应用 GNN 来获得相应的节点/边嵌入（边嵌入可以用边的末端节点的节点嵌入之和来获得）。最后，在嵌入之上应用线性模型来预测屏蔽节点/边缘属性。需要注意的是，我们允许屏蔽边属性，而不仅仅是屏蔽节点属性。

我们的节点和边属性屏蔽方法对于科学领域中注释丰富的图特别有益。例如，(1) 在分子图中，节点属性对应于原子类型，捕获它们在图上的分布方式使 GNN 能够学习简单的化学规则（例如化合价）以及更复杂的化学现象（例如电子或官能团的空间性质）。类似地，(2) 在蛋白质-蛋白质相互作用 (PPI) 图中，边属性对应于一对蛋白质之间不同类型的相互作用。捕捉这些属性在 PPI 图中的分布方式使得 GNNs 能够学习不同相互作用如何关联和相互关系。

4.3 图级预训练

我们的目标是预训练 GNN 以生成有用的图嵌入，该图嵌入由通过上述节点级预训练的方法获得的有意义的节点嵌入组成。我们需要确保节点和图嵌入都是高质量的，以便图嵌入具有鲁棒性并且可以跨下游任务转移。此外，图级预训练有两种选择，对整个图的多个监督标签进行预测，或对图结构进行预测。但由于对图结构进行预测难度较大，找到真实图距离值是一个困难的问题，所以本文不作介绍。

4.3.1 有监督的图级属性预测

由于图级表示 h_G 直接用于下游预测任务的微调，因此需要将特定于域的信息直接编码到 h_G 中。

我们通过定义有监督的图级预测任务，将图级特定领域的知识注入到我们的预训练嵌入中。特别是，我们考虑了一种预训练图表示的实用方法：图级多任务监督预训练，以联合预测各个图的不同监督标签集。例如，在分子属性预测中，我们可以预训练 GNN 来预测迄今为止已通过实验测得的分子所有属性。在蛋白质功能预测中，目标是预测给定蛋白质是否具有给定功能，我们可以预训练 GNN 来预测迄今为止已验证的多种蛋白质功能的存在。在后面的实验中，我们准备了一组不同的监督任务（最多 5000 个任务）来模拟这些实际场景。为了联合预测许多图属性，其中每个属性对应于一个二元分类任务，我们在图表示之上应用线性分类器。

重要的是，单独执行广泛的多任务图级预训练可能无法给出可转移的图级表示，正如实验证明得那样。这是因为一些有监督的预训练任务可能与下游任务无关。甚至可能损害下游性能（负转移）。一种解决方案是选择“真正相关”的监督预训练任务，并仅针对这些任务预训练 GNN。然而，这种解决方案成本极高，因为选择相关任务需要大量的领域专业知识，并且需要针对不同的下游任务单独进行预训练。

为了缓解这个问题，我们的主要想法是多任务监督预训练仅提供图级监督；因此，创建

图级嵌入的节点嵌入可能没有意义。这种无用的节点嵌入会加剧负迁移问题，因为许多不同的预训练任务在节点嵌入空间中更容易相互干扰。受此启发，我们的预训练策略是，首先通过在节点级别上对 GNN 进行预训练，然后再进行图级预训练。

4.4 GNN 预训练策略

总而言之，我们的预训练策略是首先执行节点级自监督预训练，其中节点级预训练存在两个任务：属性屏蔽预测和上下文预测。然后进行图级多任务监督预训练。当 GNN 预训练完成后，我们在下游任务上对预训练的 GNN 模型进行微调。具体来说，我们在图级表示之上添加线性分类器来预测下游图标签。完整的模型，即预训练的 GNN 和下游线性分类器，随后以端到端的方式进行微调。正如实验证明的那样，组合策略产生了更多可转移的图表示，并在无需专家选择监督预训练任务的情况下稳健地提高了下游性能。

5 复现细节

5.1 与已有开源代码对比

原文已将源代码开源，故我在它所给的框架基础上进行了我的创新和实验。由于原文的源代码和实验都是针对单一的节点级预训练策略来展开的。但这样的做法可能是片面的，因为同一任务学习到的图数据知识侧重点是不一样的，如果节点级预训练任务能够丰富一些，可能会学习到更一般、更容易泛化的知识。最直接的想法就是增加一些有用任务以让模型学习到多方面知识，当然我并不鼓励随意增加任务，它需要基于一定的理由和有理论支持下来增加任务。因此，我重新设计了节点级预训练策略，具体来说，我把屏蔽预测任务和子图上下文预测任务结合起来充当节点级预训练策略；这样做的理由有两点，首先，两种任务会使模型学习到更多知识，因为单一使用时实验效果是非常好的，所以很自然的想法就是组合起来的效果怎么样；其次，这样的策略并不会额外增加计算量。基于这个策略我修改了原文源代码以适应我的策略设计。在 8 个数据集上的实验结果表明了它能够带来性能上的提升。

5.2 实验环境搭建

GNN 架构。我主要研究图同构网络 (GINs) [17]，这是用于图级预测任务的最具表现力和最先进的 GNN 架构。我们还尝试了其他表达能力较差的流行架构：GCN [4]、GAT [13] 和 GraphSAGE [2]。但是效果都不如 GINs 架构表现出色。我选择以下在验证集中的所有下游任务中表现良好的超参数：300 维隐藏单元、5 个 GNN 层 ($K = 5$) 以及 READOUT 平均池化函数。

预训练。对于图 1 (a) 所示的上下文预测，在分子图上，我通过设置内半径 $r_1 = 4$ 来定义上下文图。在直径通常小于 5 的 PPI 网络上，我们使用 $r_1 = 1$ ，这在经验上效果很好。对于分子图和 PPI 图，我让外半径 $r_2 = r_1 + 3$ ，并使用 3 层 GNN 来编码上下文结构。对于图 1 (b) 中所示的属性屏蔽，我随机屏蔽 15% 的节点（对于分子图）或边属性（对于 PPI 网络）进行预测。

深度学习环境。源码基于 Pytorch 框架，虽然图领域的环境比较好搭建，但是涉及到一些关键的依赖库时，它不仅需要版本的严格对应，还需要和其他库相匹配。因为它的源代码

是基于这些库来写的，所以，你得搭建支持源代码各函数的依赖库。下面是一些重要的依赖库，特别是 torch-geometric、torch-sparse 和 torch-cluster。

```
CUDAtoolkit=10.2
torch=1.10.0
torchaudio=0.10.0
scatter=2.1.2
sparse=0.6.13
torch-geometric=1.0.3
cluster=1.6.3
```

5.3 创新点

我在原文方法的基础上做了一些改进。我重新设计了节点级预训练策略，具体来说，我把属性屏蔽预测任务和子图上下文预测任务结合起来充当节点级预训练策略；这样做的理由有两点，首先，两种任务会使模型学习到更多知识，因为单一使用时实验效果是非常好的，所以很自然的想法就是组合起来的效果怎么样；其次，这样的策略并不会额外增加计算量。在 8 个数据集上的实验结果表明了它能够带来性能上的提升。

Dataset			BBBP	Tox21	ToxCast	Sider	ClinTox	MUV	HIV	BACE	Average
Molecules			2039	7831	8575	1427	1478	93087	41127	1513	/
Binary prediction tasks			1	12	617	27	2	17	1	1	/
Pre-training strategy			Out-of-distribution prediction(scaffold split)								
source	Graph-level	Node-level									
Paper	-	Mask*	64.3	76.7	64.2	61.0	71.8	74.7	77.2	79.3	71.1
Reproduction	-	Mask	63.3	70.1	54.9	52.5	51.5	70.9	72.8	69.6	63.2
Ours	-	Mask+Context	65.5	71.9	59.3	58.9	65.1	69.7	75.0	73.1	67.3
Paper	-	Context*	68.0	75.7	63.9	60.9	65.9	75.8	77.3	79.6	70.9
Reproduction	-	Context	70.1	74.7	63.0	61.7	65.7	73.1	75.7	79.7	70.4
Ours	-	Context+Mask	68.5	72.8	61.9	59.8	66.2	73.2	76.8	75.1	69.3
Paper	Supervised	Mask*	66.5	77.9	65.1	63.9	73.7	81.3	77.1	80.3	73.2
Reproduction	Supervised	Mask	65.5	76.3	63.5	62.5	72.5	76.2	75.5	77.0	71.1
Ours	Supervised	Mask+Context	66.5	74.9	63.5	61.4	68.5	70.8	75.4	80.4	70.2
Paper	Supervised	Context*	68.7	78.1	65.7	62.7	72.6	81.3	79.9	84.5	74.2
Reproduction	Supervised	Context	70.0	76.8	65.5	61.0	76.5	75.1	77.6	83.1	73.2
Ours	Supervised	Context+Mask	70.2	76.6	64.7	62.4	66.4	76.7	76.5	78.2	71.5

表 1. 基于 GIN 架构的不同预训练策略的 ROC-AUC (%) 性能。最右边的列是 8 个数据集测试性能的平均值。每个数据集的最佳结果以粗体显示。

6 实验结果分析

从表 1 可以看出，原文结合图级监督多任务预训练和节点级自监督预训练的预训练策略避免了下游数据集之间的负迁移，并实现了出色的性能。而我重新设计的预训练策略虽然并不总是最优的，但它的性能与我们复现原文的性能几乎相一致，而且最重要的是，新策略在各个数据集上的性能表现更加平缓，不会出现突好突坏的情况，这有助于我们将模型应用于更广泛的领域。

具体来说，我们从表 1 中看到，原文的 Context + Supervised 预训练策略表现出了最好的性能。但在有些数据集中，如 ToxCast、Sider 和 ClinTox，仅执行屏蔽预测任务预训练策略时，我似乎并不能复现出它所给的数据；不过，大部分复现出来的数据与原文相差不大。在仅在节点级别执行屏蔽预测任务时，新策略的平均性能比复现出来的数据高出 4.1%，其它新策略方法均对复现的结果产生具有竞争力的性能。

最后一点，一旦模型经过预训练，它就可以用于任意数量的下游任务，并在很少的训练时间内提高性能。

7 总结与展望

我重新设计了一种新的预训练图神经网络（GNN）策略。简单地，我把原文的两个任务关联在一起，充当节点级预训练策略；然后，节点级和图级预训练相结合，并采用富有表现力的 GNN 架构 GIN。这确保节点嵌入能够捕获本地邻域语义，而这些语义被汇总以生成有意义的图级表示，进而可用于下游任务。通过在多个数据集、不同下游任务和不同 GNN 架构上进行实验证明，虽然新预训练策略相比于原文预训练策略表现出的性能略差，但它在多个数据集、多个下游任务上的表现更加平缓，不会出现很差的情况，当然也不会出现很好的情况，这对以后的工作如何展开可能有新的启发。首先，原文的工作在图迁移学习方向上迈出了重要一步，并成功解决了先前研究中观察到的负迁移问题。最后，将原文的方法应用于其他领域（例如物理学、材料科学和结构生物学）将会十分有趣，因为许多问题可以通过表示原子、粒子和氨基酸等相互作用的图来定义。

由于原文出自 *ICLR* 2020。现在已经过去三四年时间了，虽说图这一领域是一个小众领域，但经过这么长的时间，现已出现了许多优秀的关于图预训练策略的文章。如 Xia 等人提出的 MOLE-BERT 是原文的进一步探索 [15]；Qiu 等人提出的 GCC 利用对比学习对网络内和跨网络的子图实例进行辨别，以使图神经网络能够学习可转移的结构表示 [7]。所以，如何把原文的思想、我改进原文所遇到的问题和启发和新文章思想结合起来，或者寻求不足之处是一个很有前景的发展方向。

参考文献

- [1] Xiaomin Fang, Lihang Liu, Jieqiong Lei, Donglong He, Shanzhuo Zhang, Jingbo Zhou, Fan Wang, Hua Wu, and Haifeng Wang. Geometry-enhanced molecular representation learning for property prediction. *Nature Machine Intelligence*, 4(2):127–134, 2022.

- [2] William L Hamilton, Rex Ying, and Jure Leskovec. Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584*, 2017.
- [3] Ziniu Hu, Yuxiao Dong, Kuansan Wang, Kai-Wei Chang, and Yizhou Sun. Gpt-gnn: Generative pre-training of graph neural networks. pages 1857–1867, 2020.
- [4] Thomas N Kipf and Max Welling. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*, 2016.
- [5] Shengchao Liu, Hanchen Wang, Weiyang Liu, Joan Lasenby, Hongyu Guo, and Jian Tang. Pre-training molecular graph representation with 3d geometry. *arXiv preprint arXiv:2110.07728*, 2021.
- [6] Yue Liu, Wenxuan Tu, Sihang Zhou, Xinwang Liu, Linxuan Song, Xihong Yang, and En Zhu. Deep graph clustering via dual correlation reduction. 36(7):7603–7611, 2022.
- [7] Jiezhong Qiu, Qibin Chen, Yuxiao Dong, Jing Zhang, Hongxia Yang, Ming Ding, Kuansan Wang, and Jie Tang. Gcc: Graph contrastive coding for graph neural network pre-training. pages 1150–1160, 2020.
- [8] Yu Rong, Yatao Bian, Tingyang Xu, Weiyang Xie, Ying Wei, Wenbing Huang, and Junzhou Huang. Self-supervised graph transformer on large-scale molecular data. *Advances in Neural Information Processing Systems*, 33:12559–12571, 2020.
- [9] Hannes Stärk, Dominique Beaini, Gabriele Corso, Prudencio Tossou, Christian Dallago, Stephan Günnemann, and Pietro Liò. 3d infomax improves gnns for molecular property prediction. pages 20479–20502, 2022.
- [10] Fan-Yun Sun, Jordan Hoffmann, Vikas Verma, and Jian Tang. Infograph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization. *arXiv preprint arXiv:1908.01000*, 2019.
- [11] Susheel Suresh, Pan Li, Cong Hao, and Jennifer Neville. Adversarial graph augmentation to improve graph contrastive learning. *Advances in Neural Information Processing Systems*, 34:15920–15933, 2021.
- [12] Cheng Tan, Jun Xia, Lirong Wu, and Stan Z Li. Co-learning: Learning from noisy labels with self-supervision. pages 1405–1413, 2021.
- [13] Petar Veličković, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. Deep graph infomax. *arXiv preprint arXiv:1809.10341*, 2018.
- [14] Jun Xia, Haitao Lin, Yongjie Xu, Lirong Wu, Zhangyang Gao, Siyuan Li, and Stan Z Li. Towards robust graph neural networks against label noise. 2020.

- [15] Jun Xia, Chengshuai Zhao, Bozhen Hu, Zhangyang Gao, Cheng Tan, Yue Liu, Siyuan Li, and Stan Z Li. Mole-bert: Rethinking pre-training graph neural networks for molecules. 2022.
- [16] Jun Xia, Yanqiao Zhu, Yuanqi Du, Yue Liu, and Stan Z Li. A systematic survey of molecular pre-trained models. *arXiv preprint arXiv:2210.16484*, 2022.
- [17] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.
- [18] Yuning You, Tianlong Chen, Yang Shen, and Zhangyang Wang. Graph contrastive learning automated. pages 12121–12132, 2021.
- [19] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. *Advances in neural information processing systems*, 33:5812–5823, 2020.
- [20] Zaixi Zhang, Qi Liu, Hao Wang, Chengqiang Lu, and Chee-Kong Lee. Motif-based graph self-supervised learning for molecular property prediction. *Advances in Neural Information Processing Systems*, 34:15870–15882, 2021.
- [21] Jiangbin Zheng, Yile Wang, Cheng Tan, Siyuan Li, Ge Wang, Jun Xia, Yidong Chen, and Stan Z Li. Cvt-slr: Contrastive visual-textual transformation for sign language recognition with variational alignment. pages 23141–23150, 2023.
- [22] Jinhua Zhu, Yingce Xia, Lijun Wu, Shufang Xie, Tao Qin, Wengang Zhou, Houqiang Li, and Tie-Yan Liu. Unified 2d and 3d pre-training of molecular representations. pages 2626–2636, 2022.
- [23] Yanqiao Zhu, Yichen Xu, Feng Yu, Qiang Liu, Shu Wu, and Liang Wang. Graph contrastive learning with adaptive augmentation. pages 2069–2080, 2021.