arXiv:2601.03793v1 [cs.LG] 7 Jan 2026

文本属性图中无需标注样本的零样本节点分类提示调优

Sethupathy Parameswaran*印度科学研究所，印度卡纳塔克邦班加罗尔
sethupathyp@iisc.ac.in

Suresh Sundaram 印度科学研究所，印度卡纳塔克邦班加罗尔
vssuresh@iisc.ac.in

Yuan Fang 新加坡管理大学，新加坡，新加坡
yfang@smu.edu.sg

摘要

节点分类是信息检索中的一项基础问题，在现实世界中有许多应用，例如社交网络中的社区检测、在线发表文章的分组以及电子商务中的产品分类。文本属性图（TAG）中的零样本节点分类面临着重大挑战，尤其是缺少带标签数据。本文提出一种新颖的零样本提示调优（ZPT）框架，通过利用通用双模态条件生成器（UBCG）来解决这一问题。我们的方法首先对图语言模型进行预训练，以捕捉图结构以及每个节点相关的文本描述。随后，训练条件生成模型，学习图与文本两种模态下节点的联合分布，从而仅根据类别名称为每个类别生成合成样本。接着，利用这些合成节点和文本嵌入进行连续提示调优，从而在零样本设置下有效地完成节点分类。此外，我们在多个基准数据集上进行了广泛实验，结果表明，该框架优于现有最先进的基线方法。我们还通过消融研究验证了双模态生成器的贡献。代码见：https://github.com/Sethup123/ZPT。

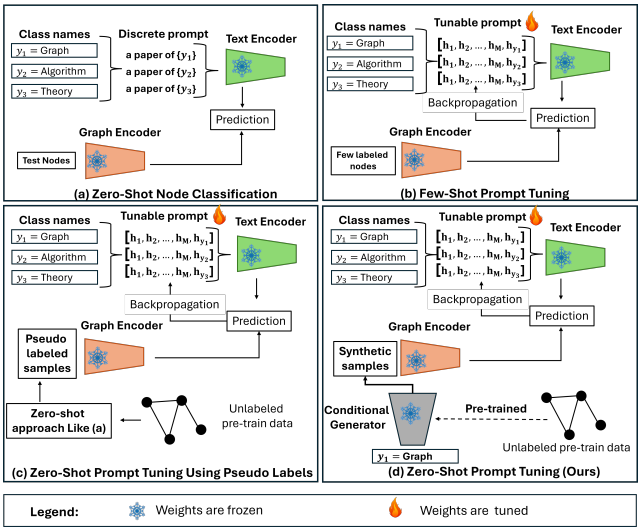


图1：不同提示学习技术的示意图。冻结符号表示参数已冻结；火焰符号表示参数经过调优。

CCS 概念

• **计算方法** → **学习设置**；• **信息系统** →**数据挖掘**。

关键词

零样本节点分类；提示调优；文本属性图；条件生成模型

ACM 引用格式：
Sethupathy Parameswaran、Suresh Sundaram 和 YuanFang。2026。无需标注样本的提示调优：文本属性图中的零样本节点分类。载于第十九届 *ACM* 万维网搜索与数据挖掘国际会议论文集（*WSDM '26*），*2026*年*2*月*22*日至*26*日，美国爱达荷州博伊西。ACM，美国纽约，11页。https: //doi.org/10.1145/3773966.3777999

*本工作的一部分是在新加坡管理大学担任研究工程师期间完成的

本作品采用知识共享署名 4.0 国际许可协议授权。*WSDM '26*，美国爱达荷州博伊西© 2026 版权所有归作者/所有者所有。ACM ISBN 979-8-4007-2292-9/2026/02 https://doi.org/10.1145/3773966.3777999

1 引言

图是万维网上许多重要应用的基础，例如社交网络中的用户关系 [5, 61, 65]，电子商务平台中的用户—商品交互 [28, 58]，以及超链接文章或科学论文中的引文 [14, 46]。在这些图中，节点分类 [47, 50, 52]——为用户或文章等节点分配有意义的标签——能够打造更智能、高效的系统。例如，在社交网络中，根据用户的兴趣或行为对其分类，可以提升内容推荐效果；而在引文图中，为文章分配主题则有助于趋势分析和知识发现。传统的监督式或半监督式节点分类模型通常存在两个缺点。首先，它们需要大量带标签节点进行训练，而获取人工标注标签可能十分困难或成本高昂。其次，这种监督式范式无法处理训练期间未见过的新类别，限制了模型的泛化能力。因此，我们需要能够针对任意类别执行节点分类、且不依赖带标签节点的方法。

研究问题。 在这项工作中，我们研究图上的零样本节点分类问题，即我们只能获取类别名称，而没有任何带标签节点。类别名称是用于描述类别或分类的简短文本短语，例如在计算机科学文章的引文图中，“理论”或“算法”。

更具体地说，我们的工作聚焦于文本属性图 （TAGs）[55, 62], 这是一种常见的图结构，其中每个节点都与一段文本描述相关联，为类别名称提供相关的语义信号。文本属性图自然地出现在各种应用中。在电子商务的用户-物品交互图中，每个物品都有标题和产品描述；而在引文图中，每篇文章都与标题以及摘要或正文相关联。

为了在文本属性图上进行零样本节点分类，预训练图语言模型 [48, 51] 已成为最先进的范式。这些方法首先在无标签文本属性图上进行自监督学习，以预训练图语言模型。随后，它们为预训练模型设计离散提示 [1, 32]，使其能够在零样本设置下推断节点类别归属，如图 1a 所示。然而，离散提示依赖人工设计，缺乏确保最佳性能的系统化方法。另一方面，连续提示 [19, 20] 是可学习的词元向量，无需人工设计。通过提示调优，可针对每项分类任务对其进行优化，但这要求每个类别至少有一个有标签样本，如图 1b 所示。这意味着连续提示调优仅适用于单样本或少样本分类，而不适用于零样本分类。

一种使零样本分类任务能够进行提示调优的简单变通方法，是使用现有的零样本方法（例如基于离散提示的方法），为预训练图中的无标签节点分配伪类别标签。如图 1c 所示，这些伪标签随后可用于提示调优，以模拟带标签节点的存在。然而，伪标签中的错误难以控制，并且在提示调优过程中可能会进一步放大。此外，当出现包含新类别的新分类任务时，必须重新生成伪标签，这使得扩展变得困难。

启示。为解决上述问题，我们针对文本属性图上的节点分类提出以下研究问题：如何在不依赖任何标签或伪标签的情况下进行零样本提示调优？众所周知，生成模型 [8, 9, 16] 能够学习底层数据分布，并且可以生成合成样本。因此，我们提出训练一个条件生成模型[24, 27],，使其能够在零样本分类任务中根据类别名称生成类别特定的合成节点。随后，这些合成节点将作为提示调优的标注样本。然而，条件生成模型的设计并非易事，需要满足两个关键要求。

首先，为确保可扩展性，图语言模型预训练完成后，该模型只需训练一次，便可普遍适用于所有下游分类任务，而无需针对每项任务重新训练。这意味着它只能访问无标签的预训练图，既没有带标签节点，也没有类别名称，因为它无法预知未来可能出现的任意分类任务。因此，在生成模型的训练过程中，我们不能直接将类别名称用作条件输入来重建节点样本。相反，我们提出用每个节点的文本描述作为其类别名称的替代，因为属于同一类别的节点，其文本描述很可能具有很强的语义相似度。基于这一替代方案，条件生成模型无需任何标注样本即可完成训练；在推理阶段，它仍可利用类别名称作为条件输入为下游分类任务中的每个类别生成合成节点样本，如图 1d 所示。

其次，生成模型必须是双模态的，能够生成同一节点的图表征和文本表示，以促进两种模态的融合，因为同时具备两种模态的合成节点可以提升提示调优的性能。为此，我们通过两种方式训练条件生成模型：不仅以节点的文本表示为条件来生成其图表征，还以其图表征为条件来生成其文本表示。因此，在推理过程中，以类别名称的文本表示为条件，我们首先为相应类别的节点生成合成图表征。然后，以合成图表征为条件，生成相应的合成文本表示。因此，两种模态的合成表征都将用于优化连续提示。

值得注意的是，我们的条件生成模型完全在潜在空间中运行，其中输入、条件和输出都表示为连续嵌入。这类潜在表征通常比离散的伪标签分配更加平滑，从而可以减弱标注错误的影响。

贡献。基于上述洞见，我们提出了一种用于文本属性图（TAGs）节点分类的零样本提示调优（ZPT）方法，无需任何标注样本。具体而言，我们首先针对图预训练一个现成的图语言模型，以获取每个节点的图表征和文本表征。接下来，作为我们的核心贡献，我们训练了一个通用双模态条件生成器（UBCG），以捕捉节点在图模态和文本模态中的底层分布。随后，对于给定的下游零样本节点分类任务（其中只有类别名称可用），我们使用训练好的 UBCG 为每个类别生成合成的图表征和文本表征。最后，我们利用这些多模态合成样本，为该分类任务进行连续提示调优。

总而言之，我们的工作作出了以下主要贡献。(1) 我们研究了一种新颖的文本属性图节点分类零样本提示调优设定，该设定在任何阶段都不需要标注样本。(2) 我们提出了一种新颖的条件生成模型 UBCG，它无需带标签节点或类别名称，只需训练一次，便可普遍适用于未来下游任务中可能出现的任何未见类别。此外，它还能生成图模态和文本模态的合成表征，进一步提升提示调优效果。(3) 实验结果表明，与多种最先进的方法相比，所提出的方法 ZPT 在四个基准数据集上均取得了更优的性能。

2 相关工作

语言预训练与提示。语言预训练与提示领域发展迅速，以 BERT [15], GPT [29], 和 T5 [30] 为代表的基础模型率先在大规模文本语料库上进行预训练。这些模型在预训练过程中掌握了丰富的语言与上下文知识，因此能够通过微调或提示有效迁移到各种下游任务。早期研究，例如 BERT 的掩码语言模型和 GPT 的自回归预训练，

展示了这一范式在文本分类等任务中的优势 [39, 56]。以提示为基础的学习由 GPT-3 [1], 等模型引入，通过使用提示来构建任务框架，进一步拓展了这一思路，使模型能够通过少样本乃至零样本学习来执行任务 [12, 45, 63]。这种方法减少了对特定任务数据和微调的需求，使其成为一种适用于广泛应用、泛化能力强的方法。

图预训练与提示。 图神经网络（GNN） [17, 18, 42]通过沿图结构传播信息来运行，其中借助消息传递机制聚合邻近节点的特征，从而计算图中每个节点的嵌入。然而，传统的监督式方法（如图卷积网络（GCN） [17], 、图注意力网络（GAT） [42], 和图同构网络（GIN） [53] ）需要大量标注样本进行训练。另一类方法是自监督方法 [10, 11, 43, 44, 59, 66],，它们利用定义明确的代理任务从数据中获取监督信号，而不需要标注样本。具体而言，深度图互信息（DGI） [43] 最大化局部节点表征与全局图级摘要之间的互信息，促使模型捕获有意义的结构和特征信息。GraphCL [59]通过最大化同一图不同增强视图之间的一致性，同时最小化不同图视图之间的一致性，来学习鲁棒的图嵌入。在 GraphMAE [11], 中，部分节点特征会被掩蔽，模型需要根据周围的图结构和其余节点特征重建被掩蔽的特征。尽管自监督方法能够学习鲁棒的嵌入，但它们仍需要标注样本来适配节点分类等下游任务。少样本节点分类方法 [3, 57, 64] 通过仅使用少量标注样本学习对图中的节点进行分类，从而解决对大量标注样本的需求。此外，已有多项研究采用预训练与提示调优范式，以应对图上少样本节点分类的挑战 [6, 22, 37, 38]。

零样本节点分类。 零样本节点分类作为一种无需任务专属带标签数据、借助预训练模型所编码的知识在未见类别间进行泛化的图节点分类方法，正受到越来越多的关注。尤其是，DGPN [49], DBiGCN [60], 和 Graph-CEN [13] 等方法，首先在一组带标签的基类（已见类别）上训练，并利用维基百科页面等辅助信息泛化到新的未见类别，从而实现零样本节点分类。近年来，OFA [21] 和 GraphGPT [40] 等基础模型也能够进行零样本节点分类。然而，与上述方法类似，OFA 也需要带标签的基类才能泛化到未见类别。另一方面，GraphGPT 在带标签的源数据集（如 arXiv）上训练，并以零样本方式泛化到不同的目标数据集（如 Cora）。不过，在进行零样本迁移之前，训练仍然需要标注样本。与上述方法不同，G2P2 [51] 和 Hound [48] 等方法利用文本属性图（TAG）中的无标签节点及其关联的文本描述进行预训练，并在下游任务中仅将类别名称用作离散提示来执行零样本节点分类。具体而言，G2P2 联合预训练图编码器和文本编码器，并通过若干对比损失对齐这两种模态。除了节点-文本对齐之外，Hound 还利用额外的自监督信号，即节点扰动、文本匹配和语义否定。在下游任务中，G2P2 和 Hound 均将类别名称用作离散提示来进行零样本分类。然而，它们至少需要一个标注样本来进行提示调优。在我们的工作中，我们提出一种新颖的零样本提示调优方法，无需任何标注样本，同时消除了对手工设计的离散提示的需求。

零样本学习也已在其他领域得到广泛研究，其中最著名的是图像分类。一种常见的方法是学习从视觉空间到类别描述语义空间的映射，随后根据图像视觉表征与语义表征之间的相似度对图像进行分类 [2, 35, 54]。另一种常见方法利用生成模型为未见类别合成样本，然后将这些样本与已见类别的训练数据结合起来训练分类器 [4, 25, 33, 34]。然而，需要注意的是，这些零样本图像分类方法要求从某些基础（即已见）类别中获取标注训练样本——要么用于学习从视觉空间到语义空间的映射，要么用于训练生成模型以生成合成数据。相比之下，本文研究的零样本节点分类设置并不假设存在任何标注样本。

3 预备知识

本节将介绍问题定义，并概述图语言模型的技术背景。

3.1 问题定义

文本属性图（TAG）。一个文本属性图 [55, 62] 可表示为 $\mathrm{G} = \{\mathrm{V}, \mathrm{E}, \mathbf{X}, \mathrm{T}\}$，其中 V 表示节点集合，E 表示节点之间的边集合，$\mathbf{X} \in \mathbb{R}^{|V| \times d}$ 是节点的特征矩阵，T 是节点的文本描述集合。例如，在引文图中，每个节点 $v \in \mathrm{V}$ 对应一篇文章，每条边 $e \in \mathrm{E}$ 表示文章之间的引用或被引关系。一个节点 v 同时关联着来自 $\mathbf{x}_v \mathbf{X}$ 的特征向量和来自 T 的文本描述，后者可以是文章的标题、摘要或全文。

零样本节点分类。在我们的工作中，我们采用一个用于零样本节点分类的预训练框架 [48, 51]，该框架包含两个阶段：预训练和下游适配。

首先，在预训练阶段，我们利用已有工作 [51] ，在给定的文本属性图上预训练一个图语言模型。该文本属性图没有标签，不包含任何与下游分类任务相关的信息。图语言模型的细节将在第 3.2 节中介绍。

随后，在下游阶段，我们调整预训练模型，使其能够处理零样本节点分类任务。在每个任务中，考虑图上节点所属的 N 个类别。每个类别仅由一个类别名称指定，该名称是一个简短的文本短语（即由自然语言词元组成的序列），例如在学术引文图中，“机器学习”或“理论”。请注意，对于任何类别，都没有提供带标签节点。目标是预测由无标签节点组成的查询集的类别，其中每个节点属于 N 个类别中的一个。这是 N类 K样本分类的一种特殊情况， [7], 其中 $K = 0$。

3.2 图语言模型

我们遵循 Wen 和 Fang [51]，在文本属性图上预训练图语言模型。其主要目标是在一个对齐空间中，基于节点与其文本描述之间的对应关系，联合训练图编码器 G_θ和文本编码器 T_ϕ。具体而言，我们使用图卷积网络 [17] 对图编码器进行建模，并使用 Transformer [41]对文本编码器进行建模。在开展分类任务之前，两个编码器都在一个无标签文本属性图上进行预训练。

预训练。图编码器生成的节点嵌入$\mathbf{V}$ 与文本编码器生成的相应文本嵌入$\mathbf{T}$ 通过以下三种对比损失进行对齐。

首先，通过最小化以下对比损失，将节点的表征与其对应文本描述的表征对齐。

$$\mathcal{L}_1 = \frac{1}{2}\left(\mathsf{CE}(\Lambda_1, \mathbf{y}) + \mathsf{CE}(\Lambda_1^\top, \mathbf{y})\right), \tag{1}$$

其中，Λ_1 是节点嵌入与相应文本嵌入之间的成对余弦相似度矩阵，$\mathbf{y} = [1, 2, ..n]^\top$是用于对比学习的标签向量，CE 表示交叉熵损失，并以逐行方式应用于 Λ_1 和 $\Lambda_1^\top$。具体而言，Λ_1 由 L_2 归一化节点嵌入矩阵$\tilde{\mathbf{V}}$ 和文本嵌入矩阵 $\tilde{\mathbf{T}}$计算得出：$\Lambda_1 = (\tilde{\mathbf{V}}\tilde{\mathbf{T}}^\top) \cdot \exp(\tau)$，其中 τ 是用于缩放相似度分数的可训练参数。

接下来，给定一个节点，根据图结构的定义，与其邻居节点相关联的文本描述可以视为该节点的摘要。节点 v 的文本摘要嵌入是其邻居节点文本嵌入的均值，$\mathbf{s}_v = \text{Mean}(\{\mathbf{t}_u : u \in \mathrm{N}_v\})$，其中 N_v 表示 v的邻居集合。令 $\tilde{\mathbf{S}}$ 表示所有节点的经 L_2 归一化的摘要嵌入矩阵。然后，我们可以使用以下对比损失函数对齐文本嵌入与摘要嵌入。

$$\mathcal{L}_2 = \frac{1}{2}\left(\mathsf{CE}(\Lambda_2, \mathbf{y}) + \mathsf{CE}(\Lambda_2^\top, \mathbf{y})\right), \tag{2}$$

其中，$\Lambda_2 = (\tilde{\mathbf{S}}\tilde{\mathbf{T}}^\top) \cdot \exp(\tau)$ 是文本 -摘要相似度矩阵。

类似地，我们也可以将节点嵌入与其对应的摘要嵌入对齐：

$$\mathcal{L}_3 = \frac{1}{2}\left(\mathsf{CE}(\Lambda_3, \mathbf{y}) + \mathsf{CE}, (\Lambda_3^\top, \mathbf{y})\right) \tag{3}$$

其中，$\Lambda_3 = (\tilde{\mathbf{V}}\tilde{\mathbf{S}}^\top) \cdot \exp(\tau)$ 是节点—摘要相似度矩阵。

因此，总体对齐损失为 $\mathrm{L}^{\text{align}} = \mathrm{L}_1 + \alpha(\mathrm{L}_2 + \mathrm{L}_3)$，其中 α 是一个超参数，用于平衡节点—摘要对齐损失和文本摘要对齐损失。

使用预训练模型进行分类。在N类零样本分类任务中，目标是在无法获取任何标注样本的情况下，将目标节点分配给其最相似的 N类别之一。如图1a所示，给定类别 $y \in \{1、2、.....、N\}$，我们可以仅使用类别名称或离散提示（例如“{class name}的一篇论文”）来构造类别描述。类别 y的分类权重$\mathbf{w}_y$可以通过使用预训练文本编码器 T_ϕ对其类别描述进行编码来获得，具体如下。

$$\mathbf{w}_y = T_\phi(\text{“a paper of \{class name\}”}). \tag{4}$$

节点 v 属于类别 y 的概率因此为

$$p(y|v) = \frac{\exp(\cos(\mathbf{w}_y, \mathbf{v}))}{\sum_{y'=1}^{N} \exp(\cos(\mathbf{w}_{y'}, \mathbf{v}))}, \tag{5}$$

其中，节点嵌入 $\mathbf{v}$ 通过预训练图编码器 G_θ获得。

在零样本设置下，提示是离散的，并且由人工设计。在少样本学习中，我们假设每个类别都有少量标注样本，此时提示可以建模为一系列 $M + 1$ 连续嵌入 $[\mathbf{h}_1, \mathbf{h}_2, ...\mathbf{h}_M、\ \mathbf{h}_{\{\text{class}\}}]$，其中$\mathbf{h}_1$、.....、$\mathbf{h}_M$是可训练向量，而$\mathbf{h}^{\{\text{class}\}}$是类别名称的嵌入。序列长度 M是一个超参数。随后，可以利用少量标注样本对连续嵌入进行调优，如图1b所示。

4 提出的方法

本节将详细介绍所提出的 ZPT 方法。我们首先介绍其总体框架，然后阐述其中的关键组件。

4.1 总体框架

我们的 ZPT 框架包含三个阶段：前两个阶段在接触任何下游分类任务之前进行预训练，第三个阶段则用于处理下游任务。总体框架如图 2 所示。

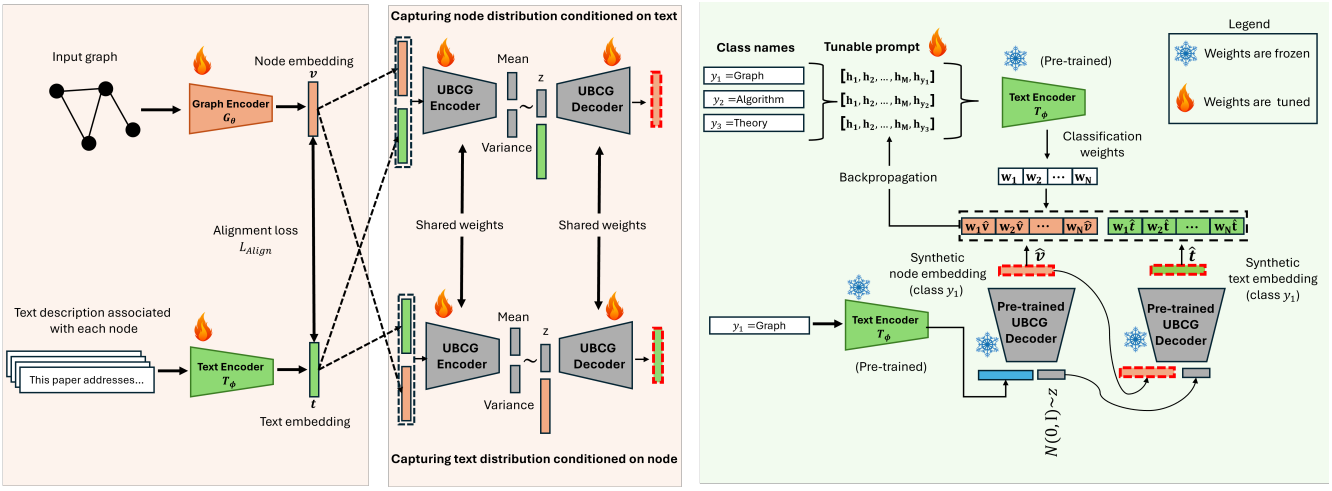
在图 2a 所示的第一阶段，我们按照第 3.2 节的描述对图语言模型 [51] 进行预训练。该模型联合训练图编码器和文本编码器，使图表示与文本表示对齐，从而确保两种模态之间的语义一致性。

在图 2b 所示的第二阶段，我们利用预训练编码器生成的对齐节点嵌入和文本嵌入来训练条件生成模型，以生成用于提示调优的特定类别合成样本。正如第 1 节所述，条件生成器必须能够普遍适用于所有分类任务，并能够生成双模态合成样本。因此，我们提出了通用双模态条件生成器（UBCG），它学习以文本为条件的底层节点分布，以及以节点为条件的文本分布。值得注意的是，UBCG 无需任何下游任务信息即可进行预训练，既不需要带标签节点，也不需要类别名称。

转向下游任务，图 2c 中的第三阶段是将类别名称输入预训练的 UBCG，以生成每个类别同时具有图和文本表示的合成样本。这些生成的合成样本随后用于连续提示调优，不仅提升了模型对零样本类别的适应能力，还通过免去手工设计离散提示的需求增强了鲁棒性。

4.2 通用双模态条件生成器

为实现无需依赖人工设计的离散提示的零样本提示调优，我们提出了通用双模态条件生成器（UBCG）。这是一种基于条件变分自编码器架构 [36] 的生成模型，用于学习图与文本两种模态各自的底层分布，并以另一种模态为条件。如图 2b 所示，UBCG 只需预训练一次，因为它仅利用预训练编码器中的节点嵌入和文本嵌入，而不需要任何下游类别信息。训练完成后，可在下游任务中以提供的类别名称为条件，利用 UBCG 在节点嵌入空间中生成特定类别的合成样本。随后，我们



(a) Graph-Language Model Pre-Training (b) Generative Model Pre-Training (c) Synthetic Sample Generation and Zero-Shot Prompt Tuning

图 2：我们的零样本提示调优（ZPT）总体框架。(a) 和 (b) 作为预训练步骤执行，而 (c) 表示下游分类阶段。

进一步以合成节点嵌入为条件，利用 UBCG 在文本嵌入空间中生成相应的样本。

捕获节点分布。用于捕获以文本为条件的节点底层分布的损失函数定义如下。

$$\mathcal{L}_{\text{gen}}^{\text{node}} = -\mathbb{E}_{\mathbf{z}\sim q_\zeta(\cdot|\mathbf{v},\mathbf{t})}[\log p_\psi(\mathbf{v}|\mathbf{z},\mathbf{t})] + \text{KL}(q_\zeta(\mathbf{z}|\mathbf{v},\mathbf{t})\|p_\psi(\mathbf{z}|\mathbf{t})), \quad (6)$$

其中，$\mathbf{v}$、$\mathbf{t}$ 分别表示节点在图上的预训练编码器生成的节点嵌入和对应的文本嵌入 v。此外，$\mathbf{z}$ 是一个潜变量，KL 表示 Kullback-Leibler（KL）散度。此处，$\mathbf{t}$ 充当条件向量，而$\mathbf{v}$ 是我们旨在对其分布进行建模的输入节点嵌入。具体而言，将 $\mathbf{t}$ 用作节点v 的条件输入，可将其作为类别名称的替代，因为$\mathbf{t}$ 蕴含了与 v的类别在语义上相关的表征。

在此损失函数中，后验分布 $q_\zeta(\mathbf{z}|\mathbf{v},\mathbf{t})$ 通过编码器网络 E 建模，其参数为 ζ；该网络将输入空间中的 $\mathbf{v}$ 映射到以 $\mathbf{t}$ 为条件的潜在空间。类似地，$p_\psi(\mathbf{v}|\mathbf{z},\mathbf{t})$ 可以通过解码器网络 D建模，其参数为 ψ；该网络将潜在空间映射到以 $\mathbf{t}$ 为条件的输入 $\mathbf{v}$。具体而言，损失函数中的第一项可视为重构损失，而第二项可视为最小化预测后验分布$q_\zeta(\mathbf{z}|\mathbf{v},\mathbf{t})$ 与先验分布 $p_\psi(\mathbf{z}|\mathbf{t})$ 之间的 KL 散度。后验分布被建模为多元正态分布，其均值和方差由编码器预测。先验分布取自标准多元正态分布 $N(\mathbf{0},\mathbf{I})$。

捕获文本分布。类似地，为捕获以节点为条件的文本底层分布，损失函数可表示如下。

$$\mathcal{L}_{\text{gen}}^{\text{text}} = -\mathbb{E}_{z\sim q_\zeta(\cdot|\mathbf{t},\mathbf{v})}[\log p_\psi(\mathbf{t}|\mathbf{z},\mathbf{v})] + \text{KL}(q_\zeta(\mathbf{z}|\mathbf{t},\mathbf{v})\|p_\psi(\mathbf{z}|\mathbf{v})). \quad (7)$$

与公式 (6) 不同，$\mathbf{v}$ 成为条件向量，而$\mathbf{t}$ 是我们旨在对其分布进行建模的输入嵌入。我们使用相同的编码器 E和解码器 D 对后验分布$q_\zeta(\mathbf{z}|\mathbf{t},\ \mathbf{v})$ 和 $p_\psi(\mathbf{t}|\mathbf{z},\ \mathbf{v})$ 进行建模。

因此，UBCG 的整体损失函数为

$$\mathcal{L}_{\text{gen}} = \mathcal{L}_{\text{gen}}^{\text{node}} + \mathcal{L}_{\text{gen}}^{\text{text}}. \quad (8)$$

4.3 使用合成样本进行提示调优

在下游阶段，考虑第 3.1 节中定义的零样本节点分类任务。为了进行连续提示调优，我们利用 UBCG 的预训练解码器 D 生成具有双模态嵌入的特定类别合成样本。

合成节点嵌入的生成。为了为给定类别生成样本，我们首先将其类别名称作为条件输入提供给 UBCG。类别名称是唯一可用的信息，并作为与该类别在语义上相关的文本描述。具体而言，我们首先从多元正态分布中采样潜变量 $\mathbf{z}\sim N(\mathbf{0},\mathbf{I})$。与此同时，我们将类别名称输入预训练文本编码器，并获得其嵌入。随后，将潜变量$\mathbf{z}$输入 UBCG 的预训练解码器 D，并以类别名称嵌入为条件，从而获得给定类别的合成节点嵌入 $\hat{\mathbf{v}}$。

合成文本嵌入的生成。获得合成节点嵌入 $\hat{\mathbf{v}}$后，我们接着生成相应的合成文本嵌入 $\hat{\mathbf{t}}$。具体而言，我们将生成 $\hat{\mathbf{v}}$时使用的相同潜变量$\mathbf{z}$ 与条件向量 $\hat{\mathbf{v}}$一同输入预训练解码器。

提示调优。我们将连续提示表述为一系列连续嵌入 $[\mathbf{h}_1,\mathbf{h}_2,\ldots,\mathbf{h}_M,\mathbf{h}_{\{\text{class}\}}]$，如第 3.2 节所述。将连续提示输入预训练文本编码器，以获得每个类别的分类权重 y:

$$\mathbf{w}_y = T_\phi([\mathbf{h}_1,\ldots,\mathbf{h}_M,\mathbf{h}_{\{\text{class}\}}]). \quad (9)$$

然后，类别概率计算如下

$$\hat{p}(y|\hat{\mathbf{v}},\hat{\mathbf{t}}) = \lambda\frac{\exp(\cos(\mathbf{w}_y,\hat{\mathbf{v}}))}{\sum_{y'=1}^{N}\exp(\cos(\mathbf{w}_{y'},\hat{\mathbf{v}}))} + (1-\lambda)\frac{\exp(\cos(\mathbf{w}_y,\hat{\mathbf{t}}))}{\sum_{y'=1}^{N}\exp(\cos(\mathbf{w}_{y'},\hat{\mathbf{t}}))}, \quad (10)$$

表1：数据集的统计描述。

数据集	Cora	Art	Industial	MI
节点数	25,120	1,615,902	1,260,053	905,453
边数	182,280	4,898,218	3,101,670	2,692,734
平均度	7.26	3.03	2.46	2.97
类别数	70	3,347	2,462	1,191

其中，第一项表示基于合成节点嵌入的概率，第二项表示基于合成文本嵌入的概率。$\lambda \in [0,1]$ 是一个用于平衡图与文本模态的超参数。随后，使用标准交叉熵损失对提示进行调优。

推理。给定测试节点 v 及其节点嵌入 $\mathbf{v}$ 和文本嵌入 $\mathbf{t}$，我们以与公式（10）相同的方式计算各类别的混合概率。然后，预测概率最大的类别。

计算开销。G2P2模型预训练阶段的复杂度为 $O(|V|(\eta+\beta))$，其中 $|V|$ 表示节点数，η 表示提示初始化中使用的邻居数，β 表示批量大小。相比之下，所提出的 UBCG模型引入的开销要小得多，为 $O(|V|/\beta)$。实测结果显示，在 Cora 数据集上，G2P2 的总运行时间为 1225 秒，而我们的方法为 1360 秒，增加幅度可以忽略不计。此外，所提出的 UBCG模型仅有 68,560 个参数，远少于 G2P2 的 6300 万个参数。

5 实验

本节通过与多种基线方法进行比较，并进一步分析模型，开展实验以评估我们的方法。

5.1 实验设置

5.1.1 数据集。本文考虑以下四个公开数据集，如表 1 所总结。Cora [23] 是一个引文网络，其中论文（节点）通过引文关系相连，摘要则作为每个节点的文本描述。每个节点都对应一个主题类别。其余三个数据集源自 Amazon 评论数据集 [26]，涵盖艺术、手工艺与缝纫（Art）、工业与科学（Industrial）以及乐器（MI）等类别。产品和用户均视为节点；如果用户评论过某个产品，则在该用户与产品之间建立一条边。类别标签对应产品类别，分类仅针对产品节点进行。关于数据集和预处理的更多细节，请参阅先前的论文 [51]。

5.1.2 任务与评估。我们的实验遵循先前研究中的设置 [48, 51]。具体而言，我们采用 5 类分类设置，每个任务通过从数据集中的全部可用类别中抽取五个类别来构建。在零样本学习场景下，每个任务仅包含查询集，不含任何标注样本。重要的是，数据集中的所有类别都会在不同任务中覆盖，但每个任务仅涉及其中五个类别的一个子集。测试时，我们在多个任务上评估性能，以确保全面覆盖各类别。报告的最终结果是每个数据集所有任务的平均性能。

5.1.3 基线。我们的实验广泛考虑了两类不同的基线，即预训练语言模型和联合预训练的图语言模型。

对于预训练语言模型，我们将**BERT**和**RoBERTa** 模型作为对比，它们均通过掩码语言建模进行预训练。此外，我们还在训练数据上对 BERT 和 RoBERTa 进行微调，分别记为 **BERT*** 和 **RoBERTa***，以缩小领域差距。这些模型通过比较与每个节点相关联的文本描述嵌入和下游任务中类别名称的嵌入之间的余弦相似度，来执行零样本分类。最后，除了使用类别名称外，我们还给出使用离散提示（如“类别为 {class name} 的研究论文”）的结果，以 **+ d** 表示。此外，我们还与大型语言模型 **Llama3-8B** 进行比较。

对于联合预训练的图语言模型，我们考虑**G2P2** 和 **Hound**。具体而言，G2P2 首先使用基于对齐的对比损失函数，在没有任何标注样本的情况下联合预训练图编码器和文本编码器。随后，在下游任务中，它首先将类别名称输入文本编码器，以获取类别嵌入。接着，通过比较图编码器生成的节点嵌入与类别嵌入之间的余弦相似度，对查询集中的测试节点进行分类。相比之下，Hound 引入了更多自监督信号，包括节点扰动、文本匹配和语义否定，从而能够联合预训练一个额外的负向文本编码器。在下游任务中，将类别名称输入文本编码器以获取正类嵌入，并输入负向文本编码器以获取负类嵌入。利用这些嵌入，计算图编码器为每个测试节点生成的节点嵌入对应的正向概率和负向概率。随后，将正向概率与负向概率的补数进行融合，得到最终预测。

5.1.4 实现细节。对于我们提出的 ZPT 方法，我们根据合成样本的生成方式提供了两个变体。首先，**ZPT** 表示仅使用类别名称作为条件输入来生成合成样本的变体。其次，**ZPT + 上下文** 表示使用额外上下文词作为条件输入的变体，例如“a paper of {class name}”，这与 G2P2 中使用的手工设计的离散提示类似。所提方法和基线模型的完整实现细节见附录 A.1 和 A.2。

5.2 结果

我们在表 2 中展示了各种基线模型和所提出的 ZPT 方法的结果。我们有以下几项观察。

首先，仅使用 BERT、RoBERTa 和 Llama-3-8B 等语言模型，表现并不理想。这是因为这些模型仅依赖与各节点相关的文本描述，而没有利用底层图结构。相比之下，G2P2 和 Hound 等图语言模型通过纳入图结构，表现优于纯语言模型。尤其是，Hound 通过引入额外的自监督机制，并同时利用正向和负向文本编码器，在零样本预测方面优于 G2P2。

表 2：我们提出的方法与基线方法的零样本节点分类性能。结果取五次运行的平均值。Hound 的结果取自先前的工作 [48]。最佳结果以粗体显示。

方法	Cora		Art		工业		MI	
	Acc	宏平均 F1	Acc	宏平均 F1	Acc	宏平均 F1	Acc	宏平均 F1
BERT	23.58±2.14	16.80±0.76	35.88±1.64	22.12±0.51	37.32±0.97	29.41±0.39	37.42±0.91	28.07±0.71
BERT*	23.38±2.24	15.83±0.96	54.27±2.11	40.24±0.56	56.02±1.39	47.50±0.83	50.19±0.82	41.12±0.28
BERT* + d	26.65±1.75	19.11±1.17	56.61±1.79	41.62±0.57	55.93±0.98	47.55±0.70	52.13±0.82	43.19±0.39
RoBERTa	30.46±2.29	22.40±2.16	42.80±1.07	30.73±0.68	42.89±1.11	33.30±0.41	36.40±1.37	27.05±0.77
RoBERTa*	39.58±1.43	32.83±1.26	34.77±0.74	21.98±0.16	37.78±0.36	27.85±0.30	32.17±0.78	23.09±0.28
RoBERTa* + d	45.53±1.35	39.11±1.47	36.11±0.66	23.65±0.48	39.40±1.24	28.06±0.58	37.65±0.34	29.45±0.11
Llama3-8B	43.49±1.77	38.10±1.71	58.03±0.90	46.77±0.86	57.78±0.28	51.42±0.38	51.83±0.57	43.67±0.32
G2P2	63.52±2.94	58.42±1.84	76.52±0.60	64.71±0.38	76.66±0.32	69.59±0.31	74.60±0.64	67.13±0.71
G2P2 + d	65.28±3.19	60.20±2.05	76.99±0.61	65.09±0.48	77.43±0.27	70.32±0.20	75.86±0.71	68.53±0.98
Hound +d	**69.21±1.35**	61.41±1.82	78.22±1.70	67.71±0.02	81.99±0.58	73.84±0.33	79.85±1.35	72.58±0.79
ZPT（我们的方法）	66.39±2.03	60.55±1.24	84.67±0.64	75.76±0.55	86.86±0.16	**81.91±0.20**	85.53±0.75	80.40±1.26
ZPT + 上下文（本文方法）	68.15±2.08	**62.26±1.90**	**84.76±0.62**	**75.90±0.53**	**86.86±0.34**	81.88±0.32	**85.54±0.74**	**80.49±1.21**

表3：合成样本生成和提示调优的效果。

L_{gen}^{node}	L_{gen}^{text}	下游训练	Cora		Art		Industial		MI	
			Acc	宏平均 F1	Acc	宏平均 F1	Acc	宏平均 F1	Acc	宏平均 F1
✓	✓	提示调优	**68.15±2.08**	**62.26±1.90**	**84.76±0.62**	**75.90±0.53**	**86.86±0.34**	**81.88±0.32**	**85.54±0.74**	**80.49±1.21**
✓	×	提示调优	64.41±2.54	58.74±2.07	77.87±0.61	66.24±0.31	78.57±0.25	71.58±0.21	77.51±0.57	70.57±0.83
×	×	离散提示	65.86±3.09	60.38±2.14	84.70±0.44	75.61±0.56	86.08±0.30	80.94±0.26	83.64±0.64	78.13±1.02
✓	✓	简单分类器	63.81±3.23	58.66±2.25	81.76±0.81	71.99±0.39	84.76±0.39	79.37±0.18	83.52±0.63	77.84±1.08

其次，我们提出的 ZPT 方法通过利用合成样本进行提示调优，始终表现出更优的性能。尽管在 Cora 数据集上的准确率不及 Hound，但其宏平均 F1 分数优于 Hound。值得注意的是，由于 Cora 存在类别不平衡，宏平均 F1 比准确率更适合作为评估指标。

第三，使用离散提示（例如“{class name} 的一篇论文”）的模型（记为 + d）通常优于仅依赖类别名称的模型。与离散提示类似，将“{class name} 的一篇论文”等上下文词纳入 ZPT 的条件输入（记为 + Context），而非仅使用类别名称，通常会带来进一步提升。此处报告了若干手工设计的离散提示或上下文词中的最佳表现，但选择最优提示或上下文词需要大量人工设计，而且往往缺乏有原则的策略。附录 B 提供了使用不同提示或上下文词的性能比较。尽管如此，值得注意的是，不包含任何手工设计上下文词的原始 ZPT 仍具有竞争力。

5.3 消融研究

我们进行了一系列消融研究，以评估不同组件对模型整体性能的影响。除非另有说明，在以下实验中，我们使用离散提示（+ d）或上下文词（+ Context）。

合成样本生成和提示调优的效果。表3展示了使用生成的合成样本进行提示调优的效果。具体而言，我们探讨以下问题。

(1) 训练 *UBCG* 模型以生成节点及其对应文本嵌入有何影响？通过比较前两行可以看出，训练 UBCG 模型以为两种模态都生成嵌入，效果优于仅生成节点嵌入的变体。

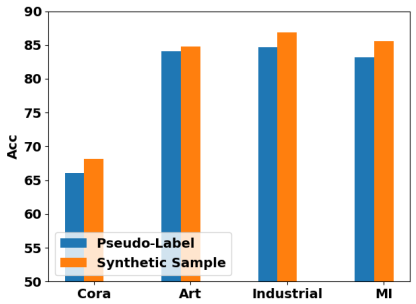
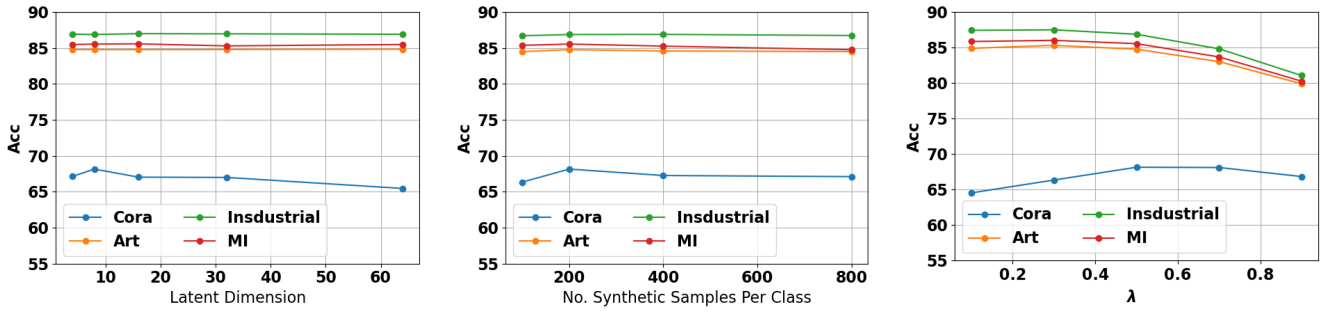


图3：伪标签与合成样本的性能对比。

(2) 与仅使用离散提示进行零样本节点分类相比，提示调优有何效果？比较第一行和第三行可以发现，使用合成样本进行提示调优的效果始终优于仅使用离散提示。

(3) 与直接在生成的合成样本上训练分类器相比，提示调优有何效果？比较第一行和最后一行可以发现，与提示调优相比，在生成的合成样本上训练简单分类器会显著降低性能。

伪标签与合成样本。如第1节所述，在零样本节点分类场景中，一种朴素的提示调优变通方案是采用现有的零样本方法（例如基于离散提示的方法），为下游任务中的无标签节点分配伪标签。因此，在本实验中，



(a) 潜在维度的影响 (b) 合成样本的影响 (c) λ的影响

图 4：模型参数的敏感性分析。

表4：在下游任务中使用不同嵌入时，G2P2 与 ZPT 的准确率比较。

	Cora	Art	工业	MI
G2P2-Node	65.28±3.19	76.99±0.61	77.43±0.27	75.86±0.71
ZPT-Node	**66.39±2.35**	**77.86±0.57**	**78.46±0.25**	**77.9±0.33**
G2P2-Text	61.00±2.19	**84.99±0.48**	86.68±0.26	83.78±0.53
ZPT-Text	**63.01±1.07**	84.45±0.59	**87.20±0.27**	**85.50±0.71**
G2P2-Hybrid	65.86±3.09	84.70±0.44	86.08±0.30	83.64±0.64
ZPT-Hybrid	**68.15±2.08**	**84.76±0.62**	**86.86±0.34**	**85.54±0.74**

我们比较了采用G2P2零样本模型分配伪标签的朴素方法与所提出的ZPT方法的性能。结果如图3所示。我们发现，所提出的ZPT方法优于朴素方法，这进一步凸显了合成样本的优势。伪标签方法的实现细节见附录A.3。

下游任务中不同嵌入的效果。在下游提示调优和推理期间，我们按照公式（10）所示，同时使用节点嵌入及其对应的文本嵌入。在本实验中，我们证明，在下游任务中融合这两种模态是有益的。表4比较了 G2P2 和 ZPT 的几种变体：仅使用节点嵌入（*-Node）、仅使用文本嵌入（*-Text），以及融合两种模态的混合方法（*-Hybrid）。对于 G2P2 和 ZPT，尽管节点嵌入和文本嵌入的相对表现因数据集而异，但混合方法通常能在不同数据集上展现出稳健的性能。此外，在三种变体的大多数情况下，我们提出的 ZPT 方法都优于 G2P2。因此，对于 ZPT，混合方法能够可靠地利用下游任务中的节点和文本信息。

5.4 敏感性分析

我们研究了各种超参数对所提出方法性能的影响。

UBCG 中的潜在维度。 UBCG 中潜变量$\mathbf{z}$ 的维度表示编码器的可压缩程度。我们将潜在维度的大小设为 {4，8，16，32，64}，并在图4a中展示结果。该模型通常对不同潜在维度取值具有鲁棒性。总体而言，我们在所有数据集中都使用潜在维度 8。

合成样本数量。使用训练好的 UBCG，我们可以根据下游任务的需要生成任意数量的样本。因此，我们研究了用于提示调优的合成样本数量所产生的影响。具体而言，我们将每个类别的合成样本数量设为 {100，200，400，800}，并在图4b中展示结果。总体而言，在四个数据集上，为每个类别生成 200 个合成样本都能取得更好的性能。

平衡参数 λ。在混合策略中，λ用于平衡节点嵌入和文本嵌入之间的权衡。这里，我们研究 λ对整体性能的影响。具体而言，我们将 λ设为 {0.1，0.3，0.5，0.7，0.9}，并在图4c中报告结果。Cora 和 Amazon 数据集上的性能趋势有所不同，反映出数据集特有的属性，以及对两种模态的侧重不同。尽管如此，设置 $\lambda = 0.5$（即为两种模态赋予相同权重）仍是稳健的选择。

5.5 合成样本的可视化

最后，我们通过可视化考察真实样本与合成样本之间的关系。在图5中，我们展示了 Cora 数据集上五个类别的真实样本和合成样本的 t-SNE嵌入结果。在节点和文本特征空间中，合成样本均形成清晰的簇，并且与真实数据分布保持一致，这表明 UBCG模型能够捕捉关键的类别特定结构。然而，我们观察到每个合成簇与其对应的真实簇之间存在轻微偏移，这是因为 UBCG 的训练过程中未使用任何类别信息或带标签节点作为条件。

6 结论

在本研究中，我们提出零样本提示调优（ZPT）方法，以应对文本属性图上的零样本节点分类挑战。我们利用通用双模态条件生成器（UBCG），根据类别名称合成图和文本表示，从而有效消除提示调优对标注节点或伪标签的需求。我们的生成模型在预训练期间仅训练一次，便可普遍应用于新的下游任务，确保其可扩展性和适应性。大量实验验证了所提出的 ZPT 方法在零样本节点分类中的有效性。

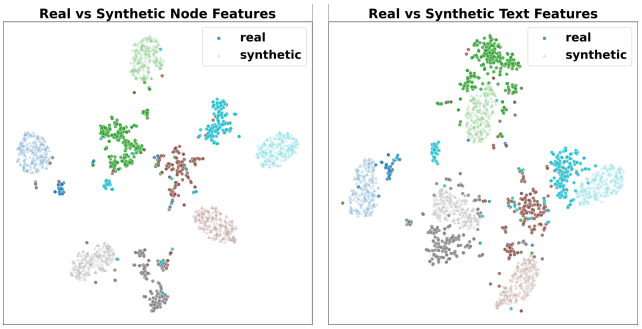


图5： *t*-SNE 可视化真实样本和合成样本，分别以深色和淡色标记表示。

致谢
本研究获新加坡教育部学术研究基金一级项目资助（22-SIS-SMU-054）。

附录
A 实现细节
A.1 所提出的方法
对于我们的图语言模型，我们遵循先前的工作 [51]。具体而言，文本编码器是一个 12 层、宽度为 512 的 Transformer 模型，具有 8 个注意力头。最大序列长度设为 128。图编码器是一个包含两层的图卷积网络，每层维度为 128，并采用 LeakyReLU 激活函数。图文预训练从头开始，权重随机初始化。基于摘要的对比损失函数的系数超参数，在 Cora 上设为 0.1，在三个亚马逊评论数据集上设为 10。我们使用 Adam优化器，以 2×10^{-5} 的学习率训练模型。对于 UBCG，编码器采用一个多层感知机（MLP）建模，包含两个隐藏层。输入维度为 256（输入嵌入与条件嵌入拼接），隐藏层维度为 128。输出为潜在空间中的均值和方差，维度为 8。同样，解码器采用包含一个隐藏层的 MLP 建模。隐藏层维度为 64，随后输出维度为 128。输入为采样得到的潜在向量与条件嵌入的拼接。所有层均使用 ReLU 激活函数。UBCG 使用 Adam优化器，以 10^{-3} 的学习率进行训练。对于下游任务，我们为每个类别生成 200 个合成节点和文本嵌入，并用它们来调整提示。所有数据集的融合参数 λ 均设为 0.5。

A.2 基线方法
对于预训练语言模型 BERT 和 RoBERTa，我们沿用先前工作的设置 [48, 51]。具体而言，模型以及掩码语言建模的实现均基于 Hugging Face 的 Transformers库。此外，我们还将预训练模型 Llama3-8B 作为基线。一般来说，预训练语言模型通过提供的类别名称获取类别嵌入，并获取与目标节点相关联的文本嵌入，从而执行零样本分类。然后，它们根据余弦相似度，将目标文本嵌入与类别嵌入最接近的类别分配给目标节点。然而，对于 Llama3-8B，我们发现，使用形式为“This sentence: a {text} means in one word:”的提示（其中 {text} 可以是类别名称，也可以是目标节点相关联的文本描述），并将最后一个隐藏状态作为嵌入，能够取得更好的分类性能。对于 G2P2，我们使用作者提供的代码复现结果。此外，第 5.3 节中的 G2P2-Hybrid 变体采用与公式（10）相同的融合策略，并将融合参数 λ 设为 0.5，这与我们提出的方法一致。对于 Hound，我们报告原始论文中的结果 [48]。

表 5：在 Cora 数据集上，G2P2 和 ZPT 在不同提示或上下文词下的准确率对比。

提示	G2P2	ZPT
{类别名称}	64.06±2.61	66.39±2.03
一个{Class Name}	65.86±3.09	68.15±2.08
一个{Class Name}	65.21±3.00	67.82±1.51
一篇关于{Class Name}的论文	62.94±3.10	65.84±2.25
一项关于{Class Name}的研究	61.68±2.31	65.10±1.41
{Class Name} 的一篇研究论文	62.02±1.77	63.87±2.06

A.3 使用伪标签进行提示调优
首先，我们使用覆盖数据集中所有类别的离散提示，以零样本方式应用 G2P2，为图中的所有节点分配伪标签。我们选择在每个数据集上表现最佳的离散提示，这与表 2 中 G2P2 +d 使用的提示一致。对于下游任务中的每个类别，我们随后最多抽取 200 个标记为该类别的节点，并将其用于提示调优。如果某个类别的伪标签样本少于 200 个，则改用所有可用样本。提示调优进行一轮，批量大小为 64，学习率为 2×10^{-5}。

B 离散提示与上下文的影响
我们在表5中探究了 G2P2 和 ZPT 分别使用的不同离散提示或上下文词的影响。不同提示或上下文的表现差异显著，因此通常需要投入大量人工精力进行设计。此外，结果表明 ZPT 具有鲁棒性，在各种提示和上下文下均持续优于 G2P2。这些发现还表明，我们提出的 UBCG 模型能够有效利用此类语义上下文来生成合成样本。

C 替代预训练文本编码器
在表6中，我们探究了将简单的 Transformer 文本编码器替换为预训练的 SentenceTransformer（SBERT）[31] 对 Cora 数据集的影响。以 SBERT 作为文本编码器的 G2P2 表现优于单独使用 SBERT，因为 SBERT 未考虑图结构。此外，我们提出的以 SBERT 作为文本编码器的 ZPT 模型取得了最佳表现。

表6：在 Cora 上使用 SBERT 作为文本编码器的效果。

	Acc	宏平均 F1
SBERT	77.86±1.61	73.29±1.88
G2P2-SBERT	80.19±1.83	75.34±2.04
ZPT-SBERT	**81.73±1.79**	**77.71±1.43**

伦理考量

我们的工作提出了一种无标签提示调优方法，用于文本属性图上的零样本节点分类。从社会角度来看，减少对带标签数据的依赖可以扩大低资源环境中人们获取机器学习模型的机会，从而可能惠及那些缺乏标注数据和领域专业知识的代表性不足群体。尽管我们的工作不会引发直接或具体的伦理问题，但图机器学习和人工智能技术的广泛应用仍可能伴随固有风险。

参考文献

[1] Tom B Brown。2020。语言模型是少样本学习器。arXiv预印本 arXiv:2005.14165（2020）。[2] Yannick Le Cacheux、Herve Le Borgne 和 Michel Crucianu。2019。在三元组损失中建模类间和类内关系以实现零样本学习。载于 IEEE/CVF国际计算机视觉大会论文集。10333–10342。[3] Kaize Ding、Jianling Wang、Jundong Li、Kai Shu、Chenghao Liu 和 Huan Liu。2020。用于属性网络少样本学习的图原型网络。载于第29届ACM信息与知识管理国际会议论文集。295–304。[4] Mohamed Elhoseiny、Kai Yi 和 Mohamed Elfeki。2021。Cizsl++：受创造力启发的生成式零样本学习。arXiv预印本 arXiv:2101.00173（2021）。[5] Wenqi Fan、Yao Ma、Qing Li、Yuan He、Eric Zhao、Jiliang Tang 和 Dawei Yin。2019。用于社交推荐的图神经网络。载于万维网大会。417–426。[6] Taoran Fang、Yunchao Zhang、Yang Yang、Chunping Wang 和 Lei Chen。2024。面向图神经网络的通用提示调优。神经信息处理系统进展36（2024）。[7] Chelsea Finn、Pieter Abbeel 和 Sergey Levine。2017。模型无关的元学习：实现深度网络的快速适应。载于国际机器学习大会。PMLR，1126–1135。[8] Ian Goodfellow、Jean Pouget-Abadie、Mehdi Mirza、Bing Xu、David Warde-Farley、Sherjil Ozair、Aaron Courville 和 Yoshua Bengio。2020。生成对抗网络。《ACM通讯》63卷11期（2020），139–144。[9] Jonathan Ho、Ajay Jain 和 Pieter Abbeel。2020。去噪扩散概率模型。神经信息处理系统进展33（2020），6840–6851。[10] Zhenyu Hou、Yufei He、Yukuo Cen、Xiao Liu、Yuxiao Dong、Evgeny Kharlamov 和 Jie Tang。2023。Graphmae2：一种解码增强的掩码自监督图学习器。载于2023年ACM Web会议论文集。737–746。[11] Zhenyu Hou、Xiao Liu、Yukuo Cen、Yuxiao Dong、Hongxia Yang、Chunjie Wang 和 Jie Tang。2022。Graphmae：自监督掩码图自编码器。载于第28届ACM SIGKDD知识发现与数据挖掘会议论文集。594–604。[12] Shengding Hu、Ning Ding、Huadong Wang、Zhiyuan Liu、Jingang Wang、Juanzi Li、Wei Wu 和 Maosong Sun。2022。知识增强提示调优：将知识融入文本分类的提示词表述器。载于第60届计算语言学协会年会论文集（第1卷：长篇论文）。2225–2240。[13] Wei Ju、Yifang Qin、Siyu Yi、Zhengyang Mao、Kangjie Zheng、Luchen Liu、Xiao Luo 和 Ming Zhang。2023。基于图对比嵌入网络的零样本节点分类。《机器学习研究汇刊》（2023）。[14] Anshul Kanakia、Zhihong Shen、Darrin Eide 和 Kuansan Wang。2019。面向 Microsoft Academic 的可扩展混合式研究论文推荐系统。载于万维网大会。2893–2899。[15] Jacob Devlin、Ming-Wei Chang、Kenton 和 Lee Kristina Toutanova。2019。BERT：用于语言理解的深度双向变换器预训练。载于 naacL-HLT 论文集，第1卷。美国明尼苏达州明尼阿波利斯，2。[16] Diederik P Kingma。2013。自编码变分贝叶斯。arXiv预印本 arXiv:1312.6114（2013）。[17] Thomas N Kipf 和 Max Welling。2016。基于图卷积网络的半监督分类。arXiv预印本 arXiv:1609.02907（2016）。[18] Thomas N Kipf 和 Max Welling。2016。变分图自编码器。arXiv预印本 arXiv:1611.07308（2016）。[19] Brian Lester、Rami Al-Rfou 和 Noah Constant。2021。参数高效提示调优的规模效应。载于2021年自然语言处理实证方法会议论文集。3045–3059。[20] Xiang Lisa Li 和 Percy Liang。2021。前缀调优：优化用于生成的连续提示。载于第59届计算语言学协会年会暨第11届自然语言处理国际联合会议论文集（第1卷：长篇论文）。4582–4597。[21] Hao Liu、Jiarui Feng、Lecheng Kong、Ningyue Liang、Dacheng Tao、Yixin Chen 和 Muhan Zhang。 [n. d.]。One For All：迈向训练一个图模型适用于所有分类任务。载于《第十二届国际学习表征会议》。[22] Zemin Liu、Xingtong Yu、Yuan Fang 和 Xinming Zhang。2023。Graphprompt：统一图神经网络的预训练与下游任务。载于《2023年ACM Web会议论文集》。417–428。[23] Andrew K. McCallum、Kamal Nigam、Jason Rennie 和 Kristie Seymore。2000。利用机器学习自动构建互联网门户。信息检索 3, 2 2000 127 163 10 1023 1009953814988 ( )，–。doi：。/A：[24] Mehdi Mirza。2014。条件生成对抗网络。arXiv预印本 arXiv:1411.1784（2014）。[25] Ashish Mishra、Shiva Krishna Reddy、Anurag Mittal 和 Hema A Murthy。2018。使用条件变分自编码器进行零样本学习的生成模型。载于《IEEE/CVF计算机视觉与模式识别会议研讨会论文集》。2188–2196。[26] Jianmo Ni、Jiacheng Li 和 Julian McAuley。2019。利用远程标注评论和细粒度方面为推荐提供依据。载于《2019年自然语言处理实证方法会议暨第九届自然语言处理国际联合会议（EMNLP-IJCNLP）论文集》。188–197。[27] Albert Pumarola、Stefan Popov、Francesc Moreno-Noguer 和 Vittorio Ferrari。2020。C-flow：用于图像和三维点云的条件生成流模型。载于《IEEE/CVF计算机视觉与模式识别会议论文集》。7949–7958。[28] Liang Qu、Ningzhi Tang、Ruiqi Zheng、Quoc Viet Hung Nguyen、Zi Huang、Yuhui Shi 和 Hongzhi Yin。2023。用于推荐的半去中心化联邦自我图学习。载于《2023年ACM Web会议论文集》。339–348。[29] Alec Radford。2018。通过生成式预训练提升语言理解。（2018）。[30] Colin Raffel、Noam Shazeer、Adam Roberts、Katherine Lee、Sharan Narang、Michael Matena、Yanqi Zhou、Wei Li 和 Peter J Liu。2020。探索统一文本到文本转换器的迁移学习极限。《机器学习研究期刊》21，140（2020），1–67。[31] Nils Reimers 和 Iryna Gurevych。2019。Sentence-BERT：使用孪生BERT网络的句子嵌入。载于《2019年自然语言处理实证方法会议论文集》。计算语言学协会。https://arxiv.org/abs/1908 10084 。[32] Timo Schick 和 Hinrich Schütze。2021。。重要的不只是规模：小型语言模型同样是少样本学习器。载于《2021年北美计算语言学协会年会：人类语言技术会议论文集》。2339–2352。[33] Edgar Schonfeld、Sayna Ebrahimi、Samarth Sinha、Trevor Darrell 和 Zeynep Akata。2019。通过对齐变分自编码器实现广义零样本与少样本学习。载于《IEEE/CVF计算机视觉与模式识别会议论文集》。8247–8255。[34] Yuming Shen、Jie Qin、Lei Huang、Li Liu、Fan Zhu 和 Ling Shao。2020。可逆零样本识别流。载于《计算机视觉—ECCV 2020：第十六届欧洲会议论文集》，第XVI部分16。614–631。[35] Ivan Skorokhodov 和 Mohamed Elhoseiny。 [未注明日期。]。面向（持续式？）广义零样本学习的类别归一化。载于《国际学习表征会议》。[36] Kihyuk Sohn、Honglak Lee 和 Xinchen Yan。2015。使用深度条件生成模型学习结构化输出表示。《神经信息处理系统进展》28（2015）。[37] Mingchen Sun、Kaixiong Zhou、Xin He、Ying Wang 和 Xin Wang。2022。GPPT：图预训练与提示调优，实现图神经网络的泛化。载于《第28届ACM SIGKDD知识发现与数据挖掘会议论文集》。1717–1727。[38] Xiangguo Sun、Hong Cheng、Jia Li、Bo Liu 和 Jihong Guan。2023。All in one：面向图神经网络的多任务提示。载于《第29届ACM SIGKDD知识发现与数据挖掘会议论文集》。2120–2131。[39] Yu Sun、Shuohuan Wang、Yukun Li、Shikun Feng、Xuyi Chen、Han Zhang、Xin Tian、Danxiang Zhu、Hao Tian 和 Hua Wu。2019。ERNIE：通过知识整合增强表征。arXiv预印本 arXiv:1904.09223（2019）。[40] Jiabin Tang、Yuhao Yang、Wei Wei、Lei Shi、Lixin Su、Suqi Cheng、Dawei Yin 和 Chao Huang。2024。GraphGPT：面向大语言模型的图指令调优。载于《第47届国际ACM SIGIR信息检索研究与发展会议论文集》。491–500。[41] A Vaswani。2017。注意力机制即一切。《神经信息处理系统进展》（2017）。[42] Petar Veličković、Guillem Cucurull、Arantxa Casanova、Adriana Romero、Pietro Lio 和 Yoshua Bengio。2017。图注意力网络。arXiv预印本 arXiv:1710.10903（2017）。[43] Petar Veličković、William Fedus、William L Hamilton、Pietro Liò、Yoshua Bengio 和 R Devon Hjelm。 [未注明日期。]。深度图互信息最大化。载于《国际学习表征会议》。[44] Chun Wang、Shirui Pan、Guodong Long、Xingquan Zhu 和 Jing Jiang。2017。MGAE：用于图聚类的边缘化图自编码器。载于《2017年ACM信息与知识管理会议论文集》。889–898。

[45] Jianing Wang、Chengyu Wang、Fuli Luo、Chuanqi Tan、Minghui Qiu、Fei Yang、Qihui Shi、Songfang Huang 和 Ming Gao。2022。面向少样本文本分类的统一提示调优。载于计算语言学协会论文集：*EMNLP 2022*。524–536。[46] Yifan Wang、Yiping Song、Shuai Li、Chaoran Cheng、Wei Ju、Ming Zhang 和 Sheng Wang。2022。Disencite：基于图的解耦表示学习，用于上下文特定引文生成。载于*AAAI*人工智能大会论文集，第36卷。11449–11458。[47] Yiwei Wang、Wei Wang、Yuxuan Liang、Yujun Cai、Juncheng Liu 和 Bryan Hooi。2020。Nodeaug：通过数据增强进行半监督节点分类。载于第*26*届*ACM SIGKDD*知识发现与数据挖掘国际会议论文集。207–217。[48] Yuxiang Wang、Xiao Yan、Shiyu Jin、Quanqing Xu、Chuanhui Yang、Yuanyuan Zhu、Chuang Hu、Bo Du 和 Jiawei Jiang。2024。Hound：为文本属性图上的少样本与零样本节点分类寻找监督信号。*arXiv*预印本 *arXiv:2409.00727*（2024）。[49] Zheng Wang、Jialong Wang、Yuchen Guo 和 Zhiguo Gong。2021。使用分解式图原型网络进行零样本节点分类。载于第*27*届*ACM SIGKDD*知识发现与数据挖掘会议论文集。1769–1779。[50] Xiaowen Wei、Xiuwen Gong、Yibing Zhan、Bo Du、Yong Luo 和 Wenbin Hu。2023。Clnode：用于节点分类的课程学习。载于第十六届*ACM*网络搜索与数据挖掘国际会议论文集。670–678。[51] Zhihao Wen 和 Yuan Fang。2023。利用图支撑预训练与提示增强低资源文本分类。载于第*46*届国际*ACM SIGIR*信息检索研究与发展会议论文集。506–516。[52] Qitian Wu、Wentao Zhao、Zenan Li、David P Wipf 和 Junchi Yan。2022。Node-former：用于节点分类的可扩展图结构学习Transformer。神经信息处理系统进展35（2022），27387–27401。[53] Keyulu Xu、Weihua Hu、Jure Leskovec 和 Stefanie Jegelka。2018。图神经网络有多强大？载于国际学习表征会议。[54] Wenjia Xu、Yongqin Xian、Jiuniu Wang、Bernt Schiele 和 Zeynep Akata。2020。用于零样本学习的属性原型网络。神经信息处理系统进展33（2020），21969–21980。[55] Hao Yan、Chaozhuo Li、Ruosong Long、Chao Yan、Jianan Zhao、Wenwen Zhuang、Jun Yin、Peiyan Zhang、Weihao Han、Hao Sun 等。2023。文本属性图综合研究：基准测试与重新思考。神经信息处理系统进展

信息处理系统 36（2023），17238–17264。[56] Zhilin Yang。2019。XLNet：用于语言理解的广义自回归预训练。*arXiv*预印本 *arXiv:1906.08237*（2019）。[57] Huaxiu Yao、Chuxu Zhang、Ying Wei、Meng Jiang、Suhang Wang、Junzhou Huang、Nitesh Chawla 和 Zhenhui Li。2020。通过知识迁移实现图少样本学习。载于*AAAI*人工智能大会论文集，第34卷，6656–6663。[58] Rex Ying、Ruining He、Kaifeng Chen、Pong Eksombatchai、William L Hamilton 和 Jure Leskovec。2018。面向Web规模推荐系统的图卷积神经网络。载于第*24*届*ACM SIGKDD*知识发现与数据挖掘国际会议论文集，974–983。[59] Yuning You、Tianlong Chen、Yongduo Sui、Ting Chen、Zhangyang Wang 和 Yang Shen。2020。具有增强的图对比学习。神经信息处理系统进展 33（2020），5812–5823。[60] Qin Yue、Jiye Liang、Junbiao Cui 和 Liang Bai。2022。用于零样本节点分类的双向图卷积网络。载于第*28*届*ACM SIGKDD*知识发现与数据挖掘会议论文集，2408–2417。[61] Yanfu Zhang、Hongchang Gao、Jian Pei 和 Heng Huang。2022。用于社交网络预测的鲁棒自监督结构图神经网络。载于*2022*年*ACM Web*会议论文集，1352–1361。[62] Jianan Zhao、Meng Qu、Chaozhuo Li、Hao Yan、Qian Liu、Rui Li、Xing Xie 和 Jian Tang。 [n. d.]。通过变分推断在大规模文本属性图上进行学习。载于第十一届国际学习表征会议。[63] Ruiqi Zhong、Kristy Lee、Zheng Zhang 和 Dan Klein。2021。通过在数据集和提示集合上进行元调优，使语言模型适应零样本学习。载于计算语言学协会论文集：*EMNLP 2021*论文集，2856–2878。[64] Fan Zhou、Chengtai Cao、Kunpeng Zhang、Goce Trajcevski、Ting Zhong 和 Ji Geng。2019。Meta-GNN：图元学习中的少样本节点分类。载于第*28*届*ACM*国际信息与知识管理会议论文集，2357–2360。[65] Zhilun Zhou、Yu Liu、Jingtao Ding、Depeng Jin 和 Yong Li。2023。层次化知识图学习支持基于位置的社交网络中的社会经济指标预测。载于*2023*年*ACM Web*会议论文集，122–132。[66] Yanqiao Zhu、Yichen Xu、Feng Yu、Qiang Liu、Shu Wu 和 Liang Wang。2021。采用自适应增强的图对比学习。载于*2021*年万维网会议论文集，2069–2080。