计 算 机 研 究 与 发 展
Journal of Computer Research and Development
https://crad.ict.ac.cn
63(1): 176−188, 2026

面向非同配图的非对称自监督学习方法

秦佳雯[1] 黄鹏峰[1] 孙庆赟[1] 周号益[2] 李建欣[1]
[1](北京航空航天大学计算机学院 北京 100191)
[2](北京航空航天大学软件学院 北京 100191)
(qinjw@buaa.edu.cn)

Asymmetric Self-Supervised Learning Method on Non-Homophilic Graphs

Qin Jiawen[1], Huang Pengfeng[1], Sun Qingyun[1], Zhou Haoyi[2], and Li Jianxin[1]
[1](*School of Computer Science and Engineering*, *Beihang University*, *Beijing* 100191)
[2](*School of Software*, *Beihang University*, *Beijing* 100191)

Abstract Self-supervised learning has emerged as a promising approach in addressing the issues of label dependency and poor generalization performance in traditional graph neural networks (GNNs). This method leverages the inherent structure and properties of the data to generate supervisory information without relying on extensive labeled datasets. However, most existing self-supervised learning methods hold the assumption that the graphs are homophilic and consequently fail to generalize well to graphs with high heterophily, where connected nodes may have different class labels and dissimilar features. In this paper, we study the problem by developing an asymmetric self-supervised learning on non-homophilic graphs framework (MIRROR), which is easy to implement and does not rely on random graph augmentations and homophilic assumptions. Inspired by the designs of existing heterophilic graph neural networks, MIRROR learns the node representations by capturing one-hop local neighborhood information and informative distant neighbors. Such two properties are obtained through carefully designed pretext tasks that are optimized based on predicted neighborhood context and estimated high-order mutual information. Extensive experiments on various types of graph benchmarks demonstrate that our proposed framework can achieve better performance compared with competitive baselines on non-homophilic graphs. The superiority in multiple downstream tasks also demonstrates the generalization capability of MIRROR.

Key words graph neural networks; self-supervised learning; deep learning; graph representation learning; non-homophilic graph

摘 要 自监督学习已逐渐成为解决传统图神经网络模型因为过度依赖标签而导致模型泛化性能差的一种新的学习范式,该方法利用数据的固有结构和属性来生成监督信息,而不依赖于标记数据。然而,大多数现有的自监督学习方法的前提假设是图具有同配性,不能较好地推广到异配性强的图,即连接的节点具有不同的类别和不同的特征。研究非同配图的自监督学习,不依赖图的同配性假设,设计了一种非对称自监督学习框架 MIRROR,通过捕获节点 1 阶邻域信息和自适应选择高阶邻域信息来学习节点的自监督信息。根据预测邻域上下文信息和估计的高阶互信息进行联合优化。模型在多个同配图数据集和非

收稿日期: 2024-08-01; **修回日期**: 2025-05-19
基金项目: 国家杰出青年科学基金项目(62225202); 国家自然科学基金青年科学基金项目(62302023); 中国人工智能学会-华为 MindSpore 学术奖励基金
This work was supported by the National Science Foundation of China for Distinguished Young Scholars (62225202), the National Natural Science Foundation of China for Young Scientists (62302023), and the CAAI-Huawei MindSpore Open Fund.
通信作者: 孙庆赟(sunqy@buaa.edu.cn)

同配图数据集上进行了大量实验，与最新的基线相比都取得了较优的效果，在多个下游任务上的优越性也表明了提出的框架具有较好的泛化性能。

关键词 图神经网络；自监督学习；深度学习；图表示学习；非同配图

中图法分类号 TP391.1

DOI: 10.7544/issn1000-1239.202440660 **CSTR:** 32373.14.issn1000-1239.202440660

图神经网络(graph neural network, GNN)已被证明在图数据上具有强大的节点表示学习的能力[1-3]，并且应用于各种下游任务，如节点分类[4]和节点聚类等[5]。通常，GNN以监督学习的方式使用带标签的数据进行训练，然而，在许多需要领域知识的应用中，如药物学和化学[6-7]，收集有标签的数据是昂贵的。此外，在任务标签稀缺的情况下，监督学习可能会出现过拟合、泛化性差等问题[8-9]。

针对监督学习的问题，自监督学习(self-supervised learning, SSL)[10-12]通过从数据中提取监督信号提供了一种新的学习范式，减少了对人工标签的依赖。与计算机视觉领域、自然语言处理领域的数据相比，图数据领域中的自监督学习存在不同的挑战，图数据中的节点是相互关联而不是完全独立的，并且图的拓扑结构至关重要，单独考虑节点特征严重影响下游任务的性能。目前的图自监督学习方法可以概括为以下2种方案：1)按照传统的网络嵌入目标重构节点邻接关系[13-16]，该方案的前提假设是语义相似的节点比语义不相似的节点更容易有连接，连接的节点在潜在空间中具有相似的表征[15,17]，即同配性(homophily)假设[18]；2)对比学习(contractive learning)方法通过随机增广算法构造多个视图获得正负样本，利用信息最大化原则保证正样本对之间的一致性学习节点表征[9,19-24]。虽然可以捕获结构信息而不强调节点邻近性，但它们的性能依赖于随机数据增广，引入了额外的噪声[25]。在非同配图(non-homophilic graphs)中，连接的节点不一定属于同一个类别[26-27]，图中节点的邻域分布存在多样性，不应该简单地在潜在空间中拉近图上邻近节点的距离。由于目前的相关研究主要集中于同配图数据，对于非同配图数据的研究较少，本文的研究问题是如何设计有效的自监督方案用于非同配图中的节点表示学习。

本文通过在非同配图上利用节点的不同邻域策略设计自监督学习目标解决这个问题。通常，具有相似邻域模式的节点应该具有相似的标签和表征[28]。例如，常见的交友网络中的用户节点性别类别分布通常呈现高异配性，具有相似邻域分布的节点很可能有相同的性别[29]。因此，本文利用不同的邻域分布区分不同的节点，主要面对以下2个挑战：1)以自监督学习的方式捕捉局部邻域分布，非同配图中的邻域分布存在多样性，节点可能与其他类别节点相连，标签的缺失阻碍了对不同节点邻域分布的建模；2)在异配图中语义高度相似性的节点可能在图中彼此距离较远，需要以自监督学习的方式捕获长距离的语义关系。一个直接的方法是使用更深的GNN编码器进行编码，然而，直接堆叠多个GNN层会引入额外的训练参数，加剧模型过平滑(over-smoothing)问题[30]。

为应对上述挑战，本文提出了一个简单有效的自监督学习框架MIRROR(asymmetric self-supervised learning on non-homophilic graphs)，提出2种新的代理任务：节点邻域信息重构任务以及高阶结构对比任务，使得该框架能够以更简单和少参数的主干编码器(如多层感知机或单层神经网络)适用于非同配图数据。为了在无标签的情况下建模节点邻域分布，通过非对称预测器预测目标节点1阶邻域信息来学习节点表征。非对称设计直接预测每个节点的原始邻域信息，允许相邻节点有不同的语义，而不是限制邻域信息与目标节点具有相似的表征。为了在不过度平滑的情况下捕获长距离的语义关系，高阶结构对比学习算法首先基于不同的特征传播尺度，即多尺度特征传播，得到节点的高阶邻域特征，然后通过最大化节点表征与其自适应邻域表征之间的互信息来选择与节点有关联的高阶邻域信息，提高节点与其有效高阶邻域信息之间的一致性，获得有辨别力的节点表征。通过联合优化节点邻域信息重构以及高阶结构对比损失，预训练的编码器可以在不使用任何人工标注数据的情况下，面对各种下游任务学习有表达能力的节点表征。

本文的主要贡献有3点：

1)提出了面向非同配图的非对称自监督学习方法，通过使用简单且更少参数的骨干编码器，缓解了直接堆叠多个图卷积层导致的过平滑问题，学习具有表达能力的节点表示，适用于多个下游任务。

2)设计了2种辅助任务,节点邻域信息重构任务以及高阶结构对比任务,分别用于捕获节点局部邻域分布以及长距离语义依存关系,实现了利用有用的局部结构和全局语义信息学习适用于同配图和异配图的节点表征。

3)在多个同配性不等的公开图数据集上对模型进行训练和验证,并与多种现有的主流自监督模型进行对比。实验结果表明,模型在数据集上均达到了较好的效果,在异配性较强的数据集上,与现有模型相比取得了最优的效果。

1 相关工作

图表示学习旨在学习一个能够将图数据映射到低维空间的映射函数,使得在低维空间中的表征能够反映原始图的拓扑结构和语义信息[31]。本文介绍基于图神经网络的异配图表示学习方法和自监督图表示学习方法。

1.1 异配图表示学习

图经常用来表示包括社会科学、化学、生物学在内的很多不同领域的数据。图表示学习可以自动地从图上学习特征,并且可以灵活地适用于不同的下游任务。经典的图表示学习方法受词向量方法的启发,将节点看作词,把在图中通过随机游走生成的节点序列当作句子,以此类比词向量方法从而学习出节点的表示。例如DeepWalk[32]等借鉴自然语言处理任务中word2vec的方法。近年来,图神经网络[1-2,4]的出现极大地促进了各种图上计算任务的发展,也促进了图表示学习在现实场景中的广泛应用。传统的GNNs大多都是在同配性假设下设计的,即假设图中拓扑距离接近的节点具有相似的特征。然而,在现实场景中,拓扑距离接近的节点可能具有不同的类别,图的异配性也很常见[27,33]。异配图表示学习是近年来研究的一个活跃领域,根据相关方法可以将异配图表示学习分为扩展潜在邻居[26,34-36]和改进消息传播机制[37-40]2类。

1)扩展潜在邻居。Geom-GCN[26]提出了一种利用结构信息为节点扩展新邻居的方法,利用传统的网络表示方法学习每个节点的低维向量表示,选择在潜在空间中具有相似结构信息的节点作为新的邻居,增加图的同配性。WRGCN[35]提出了一种局部同配性度量方法,计算以节点为中心的混合模式同配性,并利用节点特征邻域相似性和节点在多个尺度上的局部结构相似性,将输入图转换为具有更高同配性的新图,在新的图结构上获得最终的节点表征。U-GCN[34]通过实验发现了1阶邻域、2阶邻域、k-NN邻域分别对目标节点的表征存在不同影响,设计了一个通用的GCN框架,针对给定的学习目标聚合3种不同邻居节点的信息,自动学习每个部分的重要性。NLGCN[36]提出了一种简单而有效的非局部聚合算法,利用每个节点局部信息的注意力分数排序,使用深度学习中经典的局部聚合算子实现有相似分数的非局部节点聚合。

2)改进消息传递机制。MixHop[37]相比于常规的GNN只能聚合目标节点的邻居节点(即1阶邻域)的信息,设计了新的聚合方式,直接混合节点不同距离(k阶邻域)的邻居表征。H2GCN[38]描述了GNNs在异配图上学习的局限性,针对性设计了改善方法,分开聚合节点和邻居的表征、聚合高阶邻域信息,以及在最后一层组合每个节点的中间表示。FAGCN[39]从谱域的角度分析,发现GNNs的低通滤波特性,提取低频信号可以有效平滑邻居节点的表示,在同配图上的效果较好,而在异配图上的高频信号较为重要。因此设计了分别抽取低频、高频信息的卷积核,然后通过数学推导将谱域卷积转换为带有可学习注意力参数的空域图卷积。GPR-GNN[40]将经过特征变换之后的特征矩阵根据图拓扑传播k步分别得到不同步数传播后的结果。然后把得到的不同步数的结果按照不同的权重进行线性组合,得到最终的输出结果。GPR-GNN在数学上与多项式的图滤波器等价,学习最优权重等同于学习最优多项式滤波器。

这些方法都基于半监督设置,尽管在特定下游任务中都取得了一定的效果,但是在实际中标签往往是有限的、昂贵的。在本文中,主要研究在不依赖标签的情况下学习有效的节点表征,能够适用于不同的下游任务。

1.2 自监督图表示学习

监督图表示学习需要大量有标签的数据,数据标注耗费大量资源[6,41]。为了充分利用数据中的大量未标记数据,近年来的工作开始研究结合自监督学习的图表示学习,利用图数据本身固有的潜在联系作为模型监督信号,预测输入数据的其他部分数据。自监督图表示学习可以分为2个类别:生成式图表示学习和对比式图表示学习[23,42-43]。

1)生成式自监督图表示学习。该类方法从图的节点特征和结构特征2个角度对图数据进行重构,并将输入数据作为监督信号。经典的机器学习方法node2vec[14]基于有偏的随机游走算法提出邻域搜索

策略，生成不同语义的节点序列信息，拥有相同局部结构的节点通常有相似的表征。LINE[15] 设计了损失函数来分别保持节点 1 阶或 2 阶邻域的接近性。在优化损失函数之后，将这些表征连接起来作为节点表征。深度生成模型 VGAE[13] 基于变分自动编码器设计了一种图结构数据的无监督学习框架，该方法使用 1 个 GCN 编码器和 1 个简单的内积解码器来重构原始图的结构。SIG-VAE[44] 提出了半隐式图变分自动编码扩展 VGAE 对图数据建模的灵活性，采用分层变分框架，使相邻节点共享能够更好生成原始图结构。GALA[45] 对于节点特征的重构构建了一个完全对称的自动编码器，采用拉普拉斯平滑，解码器使用拉普拉斯锐化，该自动编码器体系结构的整个过程都可以利用图结构。这些生成式自监督图表示学习方法大多是学习 1 个编码器，将原始图结构中节点之间的拓扑距离映射到在潜在空间中的节点表征上[15]。它们的关键前提假设是节点与周围邻居节点具有相似的语义，但是这个假设在非同配性的图上并不是一直成立的。

2）对比式自监督图表示学习。该类方法通过构建不同视图下的节点表征，并最大化它们之间的一致性，根据模型设计的对比不同视图的尺度，可以将现有方法归为节点与图对比、节点与节点的对比、图与图的对比 3 个子类。DGI[17] 通过打乱节点特征矩阵顺序，即按整行进行重排的方式生成一个新的图，分别用同一个编码器得到节点表征和图表征，利用最大化目标节点所生成的子图表征和相应图表征之间的互信息来学习最终的节点表征。GMI[46] 基于 MINE 估计策略对图卷积神经网络输入端和输出端之间节点表征的互信息进行优化，最大化节点表征与其 1 阶邻域的原始表征之间的互信息。最大化节点表征和图级表征之间的互信息，旨在同时捕获局部和全局信息。GCA[47] 借助先验信息从拓扑和语义角度引入自适应的增强策略，采用基于节点的中心性来衡量连接结构的重要性的方式设计增强模式，同时给不重要的节点特征加入更多噪声来扰动节点特征，强化模型识别潜在语义信息。GRACE[48] 通过设计了 2 种随机图数据增广方法，即随机去除图中的边和屏蔽节点属性特征，得到一个图的多个视图，使用对比损失训练模型，拉近同一节点表征在不同增广的距离，增大与其他节点表征之间的距离。为了保证数据增广的扰动不变性，GraphCL[19] 对于不同领域图数据选择不同的数据增广策略，提取不同视图下的图级表征，设计对比损失函数，最大化相关图表征之间的互信息。与之前对比学习方法不一样的是 BGRL[21] 设计了无需负样本的对比学习方法，沿用 GRACE 的数据增广策略，将图对比任务变为样本预测的问题，设计了非对称损失，通过在线编码器预测目标编码器节点相似度，减少节点对之间的距离。由于上述工作基于同配性假设，对于非同配图的表现性能不佳。近期，相关研究开始关注解决非同配图上的自监督学习方法。HGRL[49] 利用随机特征增广和结构学习方法在图中选择特征相似的节点重构原始图结构作为数据增广方式，在图中选择节点特征相似的构建连接，然后设计保持原始语义特征和高阶邻域信息 2 种目标损失函数。DSSL[50] 基于先验假设非同配图的生成过程，联合局部语义偏移和全局语义聚类算法构造了生成式非同配图自监督学习框架。SP-GCL[51] 基于节点表征集中特性设计了单通道图对比学习方法，在对比学习目标与矩阵分解之间建立了联系。

虽然这些对比方法可以捕获结构信息而不直接保证节点与邻域的相似性，但它们的性能依赖于不同的数据增广方式[9]，最近证明了这些基于随机数据增广的图对比学习方法大多基于同配性假设，编码器捕获的是低频信息而不是高频信息[20]，但是高频信息对于异配图表示学习有着重要的作用。另一方面，这些随机扰动原始图结构和节点特征的数据增广方式在异配图上的应用是相对困难的，因为连接的节点可能是不相似的，具有高相似语义的节点可能彼此距离较远，随机数据增广破坏了原来的异配图性质，所以使用这些方法在异配图上的性能普遍下降。相比之下，本文提出了一种非随机数据增广和非同配性假设的非对称自监督学习框架。

2 相关定义和问题描述

在本节中，首先介绍一些相关定义，然后给出针对非同配图自监督学习中节点分类问题的形式化描述。

定义 1. 消息传递神经网络。 令 $\mathcal{G}=(\mathcal{V},\mathcal{E})$ 表示一个输入图，其中 $\mathcal{V}=\{v_1,v_2,\cdots,v_{|\mathcal{V}|}\}$ 是节点的集合，$\mathcal{E}\subseteq\mathcal{V}\times\mathcal{V}$ 是边的集合，令 $\boldsymbol{X}=(\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_{|\mathcal{V}|})\in\mathbb{R}^{|\mathcal{V}|\times F}$ 表示节点特征矩阵，每个节点都是 1 个 F 维的特征向量。$\boldsymbol{A}\in[0,1]^{|\mathcal{V}|\times|\mathcal{V}|}$ 表示图的邻接矩阵，如果 $e_{i,j}\in\mathcal{E}$ 则 $A_{i,j}=1$，否则，$A_{i,j}=0$。在消息传递神经网络中，邻居通常被定义为以节点为中心的 1 阶（跳）或多阶（跳）邻居。通过某种聚合和组合方式将信息传递给不同节点，迭代更新每个节点的表征。消息传递神经网

络的第 l 层可定义为

$$\boldsymbol{h}_i^{(l)} = Combine^{(l)}\left(\boldsymbol{h}_i^{(l)}, Aggr\left(\left\{\boldsymbol{h}_i^{(l-1)} : i \in N_i\right\}\right)\right), \quad (1)$$

其中 $\boldsymbol{h}_i^{(l)}$ 是第 l 层节点 v_i 的表征向量，$\boldsymbol{h}_i^{(0)}$ 是节点 v_i 的原始特征向量，N_i 是 v_i 节点的 1 阶(跳)邻居节点集合。不同的函数 $Combine(\cdot)$ 和 $Aggr(\cdot)$ 可以组合成不同的图神经网络模型。

定义 2. 图的同异配性。给定图 $\mathcal{G}=(\mathcal{V},\mathcal{E})$ 和节点标签向量，计算所有边中 2 边连接相同类别节点的边的占比，用于度量一个图整体的同配程度，具体定义为

$$\alpha = \frac{\left|(v_i, v_j) : e_{i,j} = 1 \wedge y_i = y_j\right|}{|\mathcal{E}|}, \quad (2)$$

其中 y_i 表示节点 v_i 的标签，$|\mathcal{E}|$ 是图中的边总数，根据定义得 $\alpha \in [0,1]$，α 的值越趋近于 1，图中有更多的边越趋近于连接相同类别节点，该图的同配性越强，反之，α 的值越小趋近于 0，该图的异配性越强。

非同配图自监督表示学习的目标是学习一个编码器 $f_{\boldsymbol{\theta}}$，计算节点的表征矩阵 $f_{\boldsymbol{\theta}}(\boldsymbol{X},\boldsymbol{A}) = \boldsymbol{H} \in \mathbb{R}^{|\mathcal{V}|\times d}(d \ll F)$，其中 d 表示向量维度，第 i 行表示节点 v_i 的低维表征 $\boldsymbol{h}_i$，这些表征可以捕获图中包含的拓扑结构、语义信息，从而能够用于完成不同的下游任务，如节点分类、节点聚类等。

3 MIRROR 模型

为了解决现有自监督学习方法在非同配图上面临的问题，本文提出面向非同配图的非对称自监督学习方法 MIRROR，方法总体框架如图 1 所示。本节先概述模型的总体流程，然后分别详述模型的 3 个主要模块：高阶邻域依赖关系学习、节点邻域信息重构算法、损失联合优化。

3.1 高阶邻域依赖关系学习

在自监督图表示方法中，由于节点本身没有监督信号，之前的一些自监督学习中基于同配性假设，节点和 1 阶邻居节点属于同一个类别，最大化锚节点与 1 阶邻域的信息，保证锚点节点与其周围子图之间的强相关性。然而，在非同配图中，锚节点的有效监督信息更可能分布在高阶邻域中。直接堆叠多个 GNN 层获取锚节点的高阶依赖关系易导致模型的过平滑问题，本文提出了高阶邻域依赖关系学习算法，通过对多尺度特征增广矩阵进行对比学习，使一个简单的神经网络能够自适应从高阶邻域中选择有效监督信息。

3.1.1 多尺度特征增广矩阵

为了获得锚节点的高阶语义依赖关系而不受过平滑问题的影响，本文采用多尺度特征传播策略从结构角度对输入图进行增广，相比于任意修改图结构的增广方式比如随机删减边、屏蔽节点特征，特征传播不仅能够保持锚节点与高阶邻域的节点交互信息，而且可以缓解对原始图增加额外的噪声。具体地，本文利用节点特征在图中不同尺度的消息传播过程生成 L 个特征增广矩阵，其中以每个特征矩阵

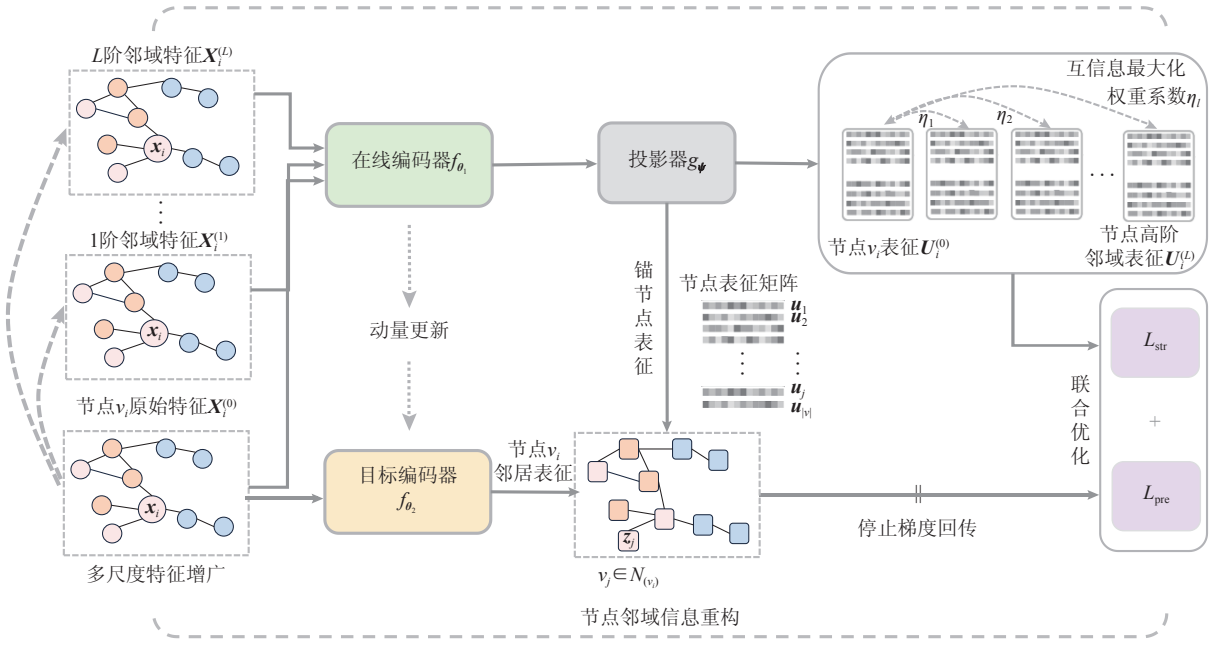


Fig. 1 Schematic diagram of MIRROR framework

图 1 MIRROR 框架示意图

为节点与 l 阶邻居的依赖关系定义为

$$\tilde{\boldsymbol{X}}^{(l)}=\boldsymbol{T}^{l}\boldsymbol{X}, \tag{3}$$

其中 $\boldsymbol{T}=\tilde{\boldsymbol{D}}^{-1/2}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-1/2}$ 为广义转移矩阵，$\tilde{\boldsymbol{A}}=\boldsymbol{A}+\boldsymbol{I}$ 表示带有自环的邻接矩阵，$\tilde{\boldsymbol{D}}$ 是 $\tilde{\boldsymbol{A}}$ 的节点度矩阵，$\tilde{\boldsymbol{D}}^{-1/2}\tilde{\boldsymbol{A}}\tilde{\boldsymbol{D}}^{-1/2}$ 表示归一化对称邻接矩阵，$\{\tilde{\boldsymbol{X}}^{(l)}\}_{l=1}^{L}$ 为生成的节点 $1\sim l$ 阶邻域多尺度特征增广矩阵。每个特征增广矩阵由 1 个 f_{θ_1} 编码器编码，相应的节点表征计算如式(4)所示。

$$\boldsymbol{H}^{(l)}=f_{\theta_1}(\tilde{\boldsymbol{X}}^{(l)})=ReLU(\tilde{\boldsymbol{X}}^{(l)}\boldsymbol{\theta}_1), \tag{4}$$

其中 $ReLU$ 为非线性激活函数，$\boldsymbol{\theta}_1$ 为可训练的转换矩阵，$\boldsymbol{H}^{(0)}$ 为原始节点特征编码后的节点表征矩阵，$\{\boldsymbol{H}^{(l)}\}_{l=1}^{L}$ 为节点与不同尺度邻域交互后学习到的特征矩阵。$\boldsymbol{H}^{(l)}$ 为 $\boldsymbol{H}^{(l-1)}$ 聚合邻居信息进行更新后的节点特征矩阵。这样将消息传递和转换步骤分开的操作，相比于标准 GNNs 在每层中将 2 个步骤耦合，降低了模型在学习过程过平滑的风险。非对称自监督学习架构中的在线编码器 f_{θ_1} 可以是 1 层神经网络也可以是多层感知机(multi-layer perceptron, MLP)，本文使用 2 层的多层感知机作为主干编码器获得节点表征，不需要堆叠多层 GNNs。

3.1.2 自适应高阶邻域学习

由于在非同配图中，连接的节点不一定属于同一个类别，语义高度相似性的节点可能彼此在图中距离较远，图中节点的有效监督信号分布在节点的多阶邻域中，所以需要在自监督学习中利用节点全局语义信息。本文设计了自适应高阶邻域机制，将传统的以锚节点为中心的 1 阶邻域扩展到 l 阶高阶邻域来提取对于锚节点学习有用的信息，不同的锚节点需要自适应从相应的高阶邻域中获取信息。在 N2N[52] 的基础上，本文提出使用 InfoNCE 来最大化锚节点表征和自适应邻居表征之间的互信息的下界，获取有辨别力的节点表征。通过对齐锚节点表征与不同邻域表征，从不同范围的上下文信息中提炼出有效监督信号。本文不直接使用在线编码器的输出 f_{θ_1}，而是按照之前的研究使用映射函数 $g_{\psi}(\cdot)$ 投影节点表征，其定义为

$$\boldsymbol{U}^{(l)}=g_{\psi}(\boldsymbol{H}^{(l)})=\sigma\left(\boldsymbol{H}^{(l)}\boldsymbol{W}_1\right)\boldsymbol{W}_2, \tag{5}$$

其中 $g_{\psi}(\cdot)$ 为 1 个 MLP，$\sigma(\cdot)$ 为非线性激活函数 sigmoid。$\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2$ 为关于节点特征矩阵 $\boldsymbol{H}^{(l)}$ 的可训练转换矩阵。

得到锚节点 v_i 不同邻域范围节点表征后，通过区分来自不同邻域表征的重要性，节点可以自适应提取不同邻域信息作为监督信号，计算公式如式(6)~(8)所示。

$$L_{\mathrm{str}}=-\sum_{l=1}^{L}E_{v_i\in\mathcal{V}}\left[\mathrm{lb}\frac{\exp(sim(\gamma_i\boldsymbol{u}_i^{l},\boldsymbol{u}_i)/\tau)}{\sum_{v_k\in\mathcal{V}}\exp(sim(\boldsymbol{u}_k,\boldsymbol{u}_i)/\tau)}\right], \tag{6}$$

$$\gamma_l=softmax\left(\boldsymbol{u}_i^{l}\right), \tag{7}$$

$$sim(\boldsymbol{u}_k,\boldsymbol{u}_i)=\frac{\boldsymbol{u}_k\cdot\boldsymbol{u}_i}{\|\boldsymbol{u}_k\|\times\|\boldsymbol{u}_i\|}, \tag{8}$$

其中 $sim(\cdot,\cdot)$ 方程计算余弦相似度，$\exp(\cdot)$ 为指数函数。$\boldsymbol{u}_i$ 为节点 v_i 的原始特征经由在线编码器和映射函数投影后得到的节点表征，$\boldsymbol{u}_i^{l}$ 为节点 v_i 的 l 阶邻域表征，γ_l 为节点 v_i 的 l 阶邻域的可学习权重系数，τ 为温度系数，随着 τ 的减少会更关注困难负样本，在训练过程中调节对负样本的关注度。此处锚节点 v_i 的正样本分布在以锚节点为中心的子图结构中，负样本为除了锚节点以外的图中其他节点。

通过提出的高阶邻域依赖关系学习算法，编码器 f_{θ} 在学习过程中可以自适应在高阶邻域中选择对锚节点有效的高阶邻域信息，提高了节点在非同配中的可判别性。

3.2 节点邻域信息重构

由于在非同配图中的邻域分布存在多样性，节点可能与其他类别节点相连，与之前的基于同配性假设的自监督学习方案不同，MIRROR 不限制邻居节点具有相似的表示。在已有的关于异配图表征学习研究中观察到在图中具有相似邻域分布的节点往往倾向于属于同一个类别，但是在无标签的场景下直接建模节点周围邻域分布比较困难。MIRROR 通过设计节点邻域重构任务以及非对称编码器优化架构避免过于依赖同配性假设，具体地，MIRROR 直接通过一个目标编码器预测并重构锚节点的原始邻域信息，该编码器的参数基于指数级平均滑动机制，根据另一个任务中编码器参数进行动态更新。这种非对称设计允许相邻节点具有不同的表征，但仍然可以捕获 1 阶邻域上下文信息，间接重构了节点邻域分布，提高了模型在嵌入空间中对齐具有类似特征节点的能力。

具体地，本文提出节点邻域信息重构任务，通过预测每个节点 v_i 的 1 阶邻域信息来学习节点的有效监督信号。由于原始节点特征可能存在噪声，将节点邻域投影到潜在的表征空间中，通过锚节点的表征预测周围邻域表征。节点邻域信息重构任务损失函数的计算如式(9)(10)所示。

$$L_{\mathrm{pre}}=\frac{1}{|\mathcal{V}|}\sum_{v_i\in\mathcal{V}}\frac{1}{|N(v_i)|}\sum_{j\in N(v_i)}\left\|\boldsymbol{u}_i-\boldsymbol{z}_j\right\|_2^2, \tag{9}$$

$$\boldsymbol{z}_j = f_{\theta_2}(\boldsymbol{X}), \tag{10}$$

其中$\boldsymbol{u}_i$为节点v_i经过编码器f_{θ_1}编码后得到的表征，$N(v_i)$为节点v_i的1阶邻居节点，$\boldsymbol{z}_j$为经过编码器f_{θ_2}编码后得到的节点v_j表征。预测器$g_\psi(\cdot)$将节点v_i在潜在空间的表征$\boldsymbol{h}_i$映射为节点邻居表征，每个节点被视为特定的邻居上下文，并假设邻居上下文相似的时候节点是相似的。从另一个角度上，可以通过邻域上下文进一步选择正样本，增强节点与正样本之间的对齐特性。由于从模型框架设计中不直接限制节点$\boldsymbol{h}_i$和周围邻居分布相似，式(9)中的预测任务可以在不基于同配性假设的情况下保留节点表征的周围邻居分布。

然而，随机更新所有预测器参数存在一个问题，目标存在平凡解，在优化过程中对所有节点输出相同的表示。为了防止编码器$g_\psi(\cdot)$退化为恒等函数，受文献[28]工作的启发，设计了一种非对称编码器结构，包括在线编码器f_{θ_1}和目标编码器f_{θ_2}。具体来说，锚节点和邻域节点表征分别从2个解耦的编码器f_{θ_1}和f_{θ_2}中提取。优化损失函数计算的梯度只用于更新在线编码器f_{θ_1}的参数，目标编码器的参数通过在线编码器参数的指数级平均滑动(exponential moving average, EMA)来更新。目标编码器f_{θ_2}参数更新的计算如式(11)所示。

$$\boldsymbol{\theta}_2 \leftarrow \lambda\boldsymbol{\theta}_2 + (1-\lambda)\boldsymbol{\theta}_1, \tag{11}$$

其中$\lambda \in [0,1]$为目标衰减率。这种更新的形式引入了2个编码器之间的不对称性，防止了模型坍塌，得到平凡解。

与之前的基于随机数据增广的对比损失函数相比，如BGRL[21]和BYOL[12]也设计了预测任务，本文的动机为基于节点周围邻域分布相似则有相似的节点特征，文献[12, 21]的方法依赖于随机数据增广算法，并且默认增广算法可以保留样本的原始语义，增广得到的样本与原始样本有相同的语义标签。然而，与图像数据不同的是在不改变原始语义的情况下设计图数据增广算法是困难的，目前的基于增广算法的图对比学习算法往往通过编码器提取的是数据的低频信息而不是高频信息，不利于对非同配中有效自监督信号的提取。

3.3 模型联合优化学习

在提出的MIRROR模型学习框架下，本文的目标是在没有任何语义标签的情况下，学习有辨别力的节点表征，能够保留非同配图中有价值的结构和节点邻域分布。本节将介绍整体的损失函数，以及所提出的框架在网络参数方面的优化。

本文提出的MIRROR总体采用非对称编码器架构，根据非同配图数据的属性特点，设计了2种辅助任务，即高阶邻域依赖关系学习以及节点邻域信息重构。为了以端到端方式训练模型，对模型参数进行优化并学习编码器f_{θ_1}，本文将节点自适应高阶邻域互信息最大化与节点邻域信息重构误差最小化作为目标函数进行自监督联合优化。模型总体的损失函数定义如式(12)所示。

$$L = \beta L_{\text{str}} + (1-\beta)L_{\text{pre}}, \tag{12}$$

其中参数β可控制对锚节点高阶结构信息和周围邻域上下文信息的侧重程度。在模型训练过程中，首先通过最小化高阶邻域依赖关系优化目标来预训练在线编码器和目标编码器，用预训练的权重初始化模型参数。然后，对模型进行联合优化训练学习。

4 实验结果与分析

4.1 数据集

为了验证所提模型的有效性，本文采用了7个真实世界数据集进行实验，包含了4个异配性强和3个同配性强的图数据集，关于数据集节点数、边数等统计信息如表1所示。

Table 1 Parameters of Dataset

表1 数据集参数

数据集	节点数	边数	类别数	特征数	同配比
Texas	183	309	5	1 793	0.11
Wisconsin	251	466	5	1 703	0.21
Cornell	183	295	5	1 703	0.30
Actor	7 600	33 544	5	931	0.22
Cora	2 708	5 278	7	1 433	0.81
Citeseer	3 327	4 552	6	3 703	0.74
Pubmed	19 717	44 324	3	500	0.80

下面分别介绍本文实验中使用的7个数据集：

1)Texas, Wisconsin, Cornell来自3个大学计算机科学系手机的网页数据集，都具有强异配性。其中，每个节点是1个网页，边是网页之间的超链接，节点特征是网页的词袋表示，根据网页类型如学生、教师等给节点分配标签。

2)Actor是一个演员共现网络数据集，其中每个节点是1个演员，边表示2个演员在同一部电影中共同出演，节点特征表示维基百科页面中关键词的词

袋表示，根据不同演员属性分配标签。

3）Cora, Citeseer, Pubmed 是 3 个论文引文网络数据集，都具有强同配性。其中，每个节点是 1 个出版物，边是引文链接，根据研究领域的不同给每个节点分配标签，以及出版物摘要的词袋表示得到每个节点特征。

4.2 基线模型

本文与以下 11 个基线模型在节点分类任务上进行模型性能的比较，其中包括传统的无监督模型 node2vec[14], LINE[15]，基于同配性假设的自监督图神经网络模型 GAE[13], VGAE[13], DGI[17], GCA[19], GRACE[47], BGRL[21]，以及非同配图自监督图神经网络模型 HGRL[49], DSSL[50], SP-GCL[51]。

为了分析模型在不同任务上的泛化性，本文与以下 9 个基线模型在节点聚类任务上进行模型性能的比较，其中包括传统的聚类方法 AE[53], Node2vec[14], LINE[15]，属性图聚类方法 GAE[13], VGAE[13]，基于同配性假设的自监督方法 DGI[17], BGRL[21]，以及非同配图自监督图神经网络模型 HGRL[49], DSSL[50]。

4.3 实验设置和评价指标

本文参照文献 [26] 中的实验设置，对所有的数据集进行 10 次随机分割。对于非同配数据集，将每个数据集中 48% 的节点作为训练集，32% 的节点作为验证集，20% 的节点作为测试集。对于同配数据集，按照公共的划分方式进行实验。为了公平比较整体性能，所有基线模型都使用了 10 次随机分割，对于没有使用标准公共数据分割的基线模型，使用官方代码重现结果，基线模型的参数按照原文中使用的参数进行设置，实验结果均基于 10 次实验的平均值和方差。

本文使用 2 个下游任务来评估模型学习到的节点表征的有效性和泛化性：1）节点分类任务，本文采用准确率（accuracy, *ACC*）作为评价指标，遵循线性评估标准[17]，保存训练后模型的参数，在学习到的节点表征上使用一个逻辑回归分类器，测试模型分类的准确率；2）节点聚类任务，本文采用 2 个评价指标，聚类准确率（cluster accuracy, *C-ACC*）和归一化互信息（normalized mutual information, *NMI*）。将每个模型训练后的节点表征进行 *k*-means 聚类，并将聚类的数量设置为每个数据集中真实的类别个数[49] 来测试模型聚类效果。为了公平比较模型性能，对于所有模型，在节点分类任务上固定节点表征维数为 512，在节点聚类任务上固定表征维数为 16。

本文实验环境的操作系统为 Ubuntu 20.04.3 LTS，服务器处理器为 Intel® Xeon® Platinum 8358 CPU @ 2.60 GHz，本文实验使用 NVIDIA GeForce RTX 3090 GPUs 显卡进行训练，其 CUDA 版本为 10.1。本文采用深度学习框架 PyTorch 1.9.1 和深度图框架 PyTorch Geometric(PyG)2.0.1 实现模型并完成实验。

本文为了优化目标函数，使用 Adam 优化器优化模型参数。对于本文提出的优化过程中涉及的超参，采用网格搜索来选择每个数据集的超参。其中，设置动量更新参数 λ 范围为 {0, 0.90, 0.95, 0.99, 0.999, 1}，编码器嵌入维度范围为 {256, 512, 1 024, 2048, 4 096}，温度系数范围为 {0.25, 0.5, 0.75, 0.99, 1}，学习率范围为 {0.000 1, 0.000 5, 0.001}，权重衰减范围为 {0, 0.000 1, 0.000 3, 0.000 001}，联合训练参数 β 范围为 {0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.7, 0.9}。为了防止模型训练过拟合，本文设置初始训练轮数为 1 000，并根据训练集损失值的变化情况采用训练早停策略防止模型退化及过拟合。本文中默认设置编码器嵌入维度即节点特征维度为 2 048，高阶邻居阶数 L=7，λ=0.99。

4.4 实验结果和分析

4.4.1 节点分类

本文在 7 个真实数据集上进行了实验，比较了不同基线模型的节点分类性能，使用分类准确率作为评价指标。如表 2 所示，表中各个模型的实验结果表明，本文提出的 MIRROR 模型在 7 个同配性不等的数据集上取得不错的实验结果，在 4 个强异配数据集上取得最优的效果，在 3 个强同配数据集上也取得相对较好的效果。以下是详细的实验结果分析。

1）与不同数据集上表现最佳的基线模型相比，在节点分类任务中，本文提出的模型 MIRROR 在大部分数据集上都取得了最佳的性能，在 4 个非同配图上相对于最佳基线模型性能提升了 4.1~14.2 个百分点，表明了 MIRROR 在非同配图上的有效性，验证了模型中学习节点高阶邻域依赖和重构节点上下文信息这 2 种辅助任务能够有利于模型获取异配图中更多有效的自监督信号，使得学习到的节点表征更有辨别力。另外，MIRROR 在同配图数据集上也具有良好的性能，表明模型可以在一般属性图中获取有辨别力的节点表征信息。

2）与传统的无监督模型 Node2ve, LINE 相比，MIRROR 的表现始终显著优于它们，主要因为这 2 种经典模型都基于同配性假设，并且难以兼顾从不同邻域中获取锚节点自监督信号。MIRROR 相较于一般的自监督学习基线方法如 GAE, VGAE, DGI, GCA, GRACE, BGRL，也一直保持着最优的效果且分类准

Table 2 Accuracy Results of Node Classification on Various Datasets

表 2 各数据集上的节点分类的准确率结果 %

模型	非同配图				同配图		
	Texas	Wisconsin	Cornell	Actor	Cora	Citeseer	Pubmed
node2vec	41.92±0.6	37.45±0.5	42.94±1.3	28.28±0.1	71.24±0.9	47.64±0.8	66.46±1.0
LINE	48.69±1.4	37.45±2.5	43.68±2.2	27.55±0.3	68.25±0.3	43.92±0.5	66.29±0.6
GAE	58.64±0.3	52.55±0.4	58.85±0.1	28.03±0.2	71.07±0.4	65.22±0.4	71.71±0.9
VGAE	50.27±2.2	55.67±1.4	48.73±4.2	26.99±1.6	76.30±0.2	66.80±0.2	75.80±0.4
DGI	60.53±3.0	55.21±1.0	52.33±6.1	28.30±0.8	82.29±0.6	70.89±0.5	76.80±0.6
GCA	52.92±0.5	59.55±0.8	52.31±1.1	28.77±0.3	80.39±0.4	71.81±0.4	80.37±0.8
GRACE	57.57±5.7	50.00±5.8	54.86±7.0	29.01±0.2	80.08±0.5	71.11±0.2	80.15±0.3
BGRL	59.19±2.0	52.23±1.2	57.33±2.3	28.80±0.5	81.08±0.2	71.10±0.4	79.60±0.5
HGRL	61.83±0.7	63.90±0.6	58.78±1.0	27.95±0.3	80.66±0.4	68.56±1.1	79.83±0.3
DSSL	62.11±1.5	62.25±0.6	60.15±1.3	28.36±0.7	82.05±0.3	**73.20±0.5**	81.25±0.3
SP-GCL	59.81±1.3	60.12±0.4	59.29±1.2	28.94±0.7	**83.16±0.1**	71.96±0.8	79.16±0.7
MIRROR（本文）	**74.08±3.8**	**78.14±1.0**	**71.59±3.7**	**33.15±0.5**	82.34±0.6	72.00±1.0	**81.58±0.5**

注：“±”前表示准确率平均值，“±”后表示标准差，黑体数值表示最优结果，下划线数值表示次优结果。

确率提升显著。主要是因为这些模型基于同配性假设，设计了不同的数据增广算法，例如随机去除图中的边、随机扰动节点、屏蔽节点属性特征等，使它们在 Cora 这类具有强同配性的数据集上有效，但难以适用于强异配图数据集。这也证明了这些基于随机数据增广的图对比学习方法隐含了同配性假设，通过编码器捕获的是低频信息而不是高频信息。而本文提出的在非同配图上利用节点的不同邻域策略设计自监督学习目标，不限制节点近邻分布的一致性，能够从节点不同邻域中提取有效的自监督信号。

3)与现有的非同配图自监督学习方法 HGRL，DSSL，SP-GCL 相比，MIRROR 始终优于它们。HGRL 基于不同的数据增广方法设计了自监督学习方法，虽然保留了原始节点特征语义和节点周围不同邻域信息，但是随机特征和结构学习增广算法隐含了同配性假设。DSSL 假设图的生成过程构造自监督学习框架，联合局部语义偏移和全局语义聚类算法，但是对于真实世界的图来说可能并不总是成立。SP-GCL 设计了单通道图对比学习方法，较为依赖节点周围正样本的选择。所以相较于本文中节点邻域信息重构的非同配性假设，未限制周围邻域特征的相似性，可以从高阶邻域中自适应提取有效自监督信号，总体上提升了在非同配图上的性能。

4.4.2 节点聚类

本文在 4 个强异配图数据集上与不同基线模型比较了节点聚类性能，使用 *C-ACC*，*NMI* 作为评价指标，*C-ACC* 和 *NMI* 的值越高代表模型学到的节点表征聚类效果越好。实验结果如表 3 所示，与其他基线模型相比，本文提出的 MIRROR 模型在强异配图数据集上性能表现较优，平均排名值为 1.3，排名第 1，以下是详细的实验结果分析。

Table 3 Results of Node Clustering on Various Datasets

表 3 各数据集上的节点聚类结果

模型	Texas		Cornell		Wisconsin		Actor		平均排名值
	C-ACC/%	*NMI*/%	*C-ACC*/%	*NMI*/%	*C-ACC*/%	*NMI*/%	*C-ACC*/%	*NMI*/%	
AE	50.49	16.63	52.19	17.08	58.61	30.92	24.19	0.97	5.5
node2vec	48.80	12.58	50.98	15.84	41.39	4.23	25.02	0.09	7.4
LINE	49.40	16.90	34.10	12.85	39.40	9.70	22.70	0.04	8.6
GAE	42.02	8.49	44.72	15.11	37.81	9.19	23.45	0.18	8.8
VGAE	50.27	11.73	43.39	17.87	40.00	9.87	23.30	0.21	7.5
DGI	55.74	8.73	44.10	5.79	44.58	10.72	24.26	1.38	7.0
BGRL	58.68	21.95	55.08	19.74	57.66	31.74	28.20	1.84	3.6
HGRL	61.97	**34.54**	60.56	**35.03**	68.36	30.06	29.79	3.80	2.0
DSSL	57.43	18.60	44.70	20.36	58.97	32.45	26.15	2.76	3.8
MIRROR（本文）	**67.83**	27.80	**64.71**	31.80	**70.54**	**37.89**	**30.54**	**4.04**	**1.3**

注：黑体数值表示最优结果，下划线数值表示次优结果。

与不同数据集上表现最佳的基线模型相比，MIRROR 在大部分数据集上都具有较好的性能，表明了模型学到的节点表征在节点聚类任务中有一定的泛化性。相较于传统的无监督聚类算法和基于同配性假设的自监督学习模型，MIRROR 在不同评价指标下的节点聚类效果一直优于它们。MIRROR 与 HGRL 在 Texas 和 Wisconsin 数据集上表现性能相当，

而相较于 DSSL 模型，MIRROR 模型一直表现更优。MIRROR 在所有数据集上 *C-ACC* 指标最高，效果最优。*C-ACC* 指标的改进表明了 MIRROR 学习到的节点表示可以更加正确地分配到 *k*-means 聚类模型中各自的簇中。在 Wisconsin 数据集上的 *NMI* 相对于次优基线改进了 16.8%，体现了 MIRROR 学习到的节点表征的聚类结果与真实标签更有相似性，能够提取数据集中更多有效自监督信号。

MIRROR 总体在节点聚类任务上的性能表现表明了模型中学习节点高阶邻域依赖和重构节点上下文信息这 2 种辅助任务能够针对非同配图提取到更有效的自监督信号，模型也对不同任务有较好的泛化性能。

4.4.3 消融实验

为了检验 MIRROR 框架中各组成部分的重要性，评估不同部分对模型的影响，此处设计模型的各种变体进行了消融实验，主要分为以下变体：1）w/o L_{str}，该变体去除了高阶邻域依赖关系学习辅助任务、只用节点邻域信息重构辅任务训练模型；2）w/o L_{pre}，该变体去除了节点邻域信息重构辅助任务，只用高阶邻域依赖关系学习辅助任务训练模型；3）w/o asy，该变体去除了非对称编码器架构设计，使用相同的编码器和参数训练模型。完整模型的结果表示为 MIRROR，在不同数据集上的消融实验结果具体如图 2 所示。根据消融实验有以下观察：

1）MIRROR 完整模型性能在 4 个数据集上都优于其他变体模型，表明了节点邻域信息重构任务、高阶依赖关系学习任务以及非对称编码器架构的设计都有助于提升模型性能，都是模型重要的组成部分。

2）在具有强异配性质的图数据集 Texas, Wisconsin, Cornell 上，节点邻域信息重构任务相较于高阶邻域依赖关系学习任务都不同程度影响了模型的性能，

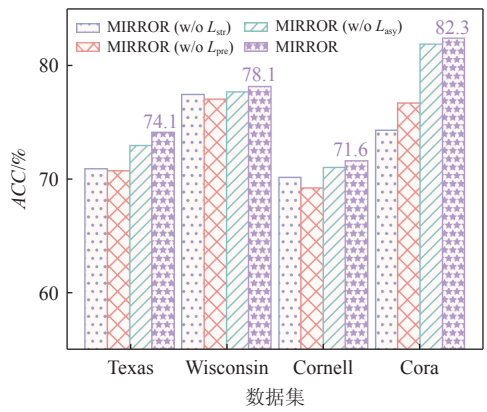


Fig. 2 Ablation study results

图 2 消融实验结果

其中节点邻域信息重构任务对模型总体性能影响会更为明显，表明强异配图数据周围邻域的上下文信息能够有助于模型提取更多的自监督信号。而在具有强同配性质的数据集 Cora 上，高阶邻域信息的捕捉对模型的影响更大，节点周围高阶邻域信息有助于在不同同配比的数据集上提取节点训练中需要的自监督信息。

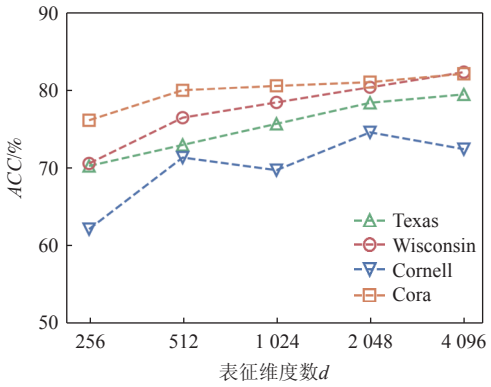
4.4.4 超参数敏感性分析

本节对模型的主要参数进行敏感性分析，其中包括节点表征维度数 d、节点高阶邻域阶数 L、目标编码器动量更新参数 λ。本文通过在不同超参数设置下观察节点 *ACC* 的变化来进行分析。不同参数敏感性实验结果如图 3 所示。根据实验结果有以下观察：

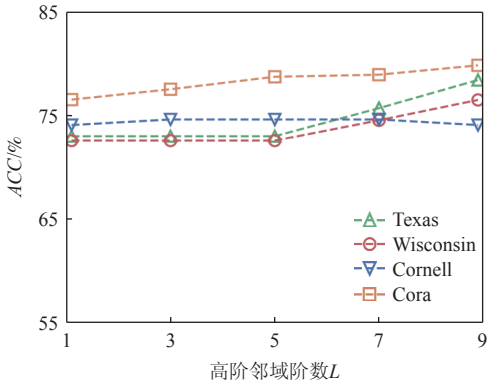
1）节点表征维度数 d。节点通过编码器编码得到节点表征，节点表征维度的大小影响了模型的复杂度以及表征包含的信息量。为了探索 MIRROR 在不同节点表征维度数下的性能，本文实验设定了维度 $d \in \{256, 512, 1\,024, 2\,048, 4\,096\}$ 分别在不同数据集上测试了模型的性能。如图 3（a）所示，随着维度的增加，模型在非同配数据集和同配数据集上的性能总体上呈现上升趋势，这是由于可学习参数增加，模型拟合数据集能力提升。此外，可以观察到相较于同配图，非同配图更受节点表征维度的影响。当维度过高时，模型可能存在过拟合问题，限制了其泛化性能，因此对某些图设置超大维度的训练可能会导致性能的略微下降。

2）节点高阶邻域阶数 L。通过改变特征传播步长，能够控制节点从不同的高阶邻域范围提取自监督信号。本文实验设定了 $L \in \{1,3,5,7,9\}$ 在不同数据集上测试模型的性能，如图 3（b）所示，较大的 L 允许在较大的邻域内进行消息传递来学习节点表示。随着 L 增加，MIRROR 通过在更大的邻域中提取自监督信息进行节点表征学习，总体上提高了性能。同时，由于编码器的选择，模型也不存在因为过度平滑而导致的性能严重退化问题。

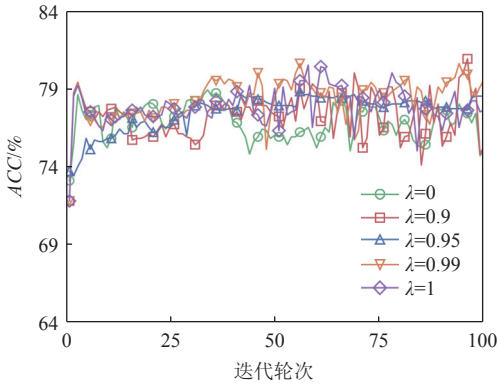
3）动量更新参数 λ。通过改变动量更新参数 λ 来控制目标编码器参数更新情况，本文实验设定了 $\lambda \in \{0.0, 0.9, 0.95, 0.99, 1\}$ 在不同数据集上测试模型的性能。如图 3（c）所示，λ 一定程度上影响了模型的性能，具有较大的 λ 通常可以提升模型的性能。而当 $\lambda = 0$ 时，在线编码器和目标编码器参数更新完全一致，MIRROR 的性能相较于其他设置表现出退化。因此，模型在训练的过程中，需要权衡目标编码器是选择频繁更新还是选择缓慢更新。当 $\lambda = 1$ 时，即使



(a) 4个数据集上的表征维度数分析



(b) 4个数据集上的高阶邻域阶数分析



(c) Cora数据集上动量更新参数的分析

Fig. 3 Hyper-parameters sensitivity analysis

图 3 超参数敏感性分析

用一个随机的目标编码器，也可以在数据集上获得良好的性能，表明仅仅使用简单的停止梯度操作有时也可以防止模型坍塌。

5 结 论

本文研究面向非同配图的非对称自监督学习方法，提出了一种非对称自监督学习框架 MIRROR，设计了 2 种辅助任务，即节点邻域重构任务以及高阶结构对比任务，分别用于获取锚节点局部邻域分布以及高阶语义依存关系来学习节点的自监督信息。通过预测邻域上下文信息和估计的高阶互信息进行联合优化和训练模型。在多个同配图数据集和非同配图数据集上进行的大量实验表明，与最优的基线方法相比，MIRROR 可以在不同数据集上均达到较好的效果，并且在不同的任务中也具有较好的泛化性能。

作者贡献声明：秦佳雯负责方案整体设计并撰写论文；黄鹏峰负责实验验证；孙庆赟、周号益对研究思路与论文撰写提出具体优化建议；李建欣对研究方向进行了总体把控，并提供指导性建议与工作支持。

参 考 文 献

[1] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint, arXiv: 1609.02907, 2016

[2] Hamilton W, Ying Zhitao, Leskovec J. Inductive representation learning on large graphs[C]//Proc of the 31st Int Conf on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc, 2017: 1025−1035

[3] Xu Keyulu, Li Chengtao, Tian Yonglong, et al. Representation learning on graphs with jumping knowledge networks[C]//Proc of the 35th Int Conf on Machine Learning. New York: ACM, 2018: 5453–5462

[4] Veličković P, Cucurull G, Casanova A, et al. Graph attention networks[C/OL]//Proc of the 6th Int Conf on Learning Representations. [2024-07-29]. https://openreview.net/pdf?id=rJXMpikCZ

[5] Wang Chun, Pan Shirui, Hu Ruiqi, et al. Attributed graph clustering: A deep attentional embedding approach[C]//Proc of the 28th Association for the Advancement of Artificial Intelligence (AAAI) Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2019: 3670–3676

[6] Hu Weihua, Liu Bowen, Gomes J, et al. Strategies for pre-training graph neural networks[C/OL]//Proc of the 8th Int Conf on Leaning Representations. [2024-07-29]. https://openreview.net/forum?id=HJlWWJSFDH&no

[7] Saifuddin K M, Bumgardner B, Tanvir F, et al. HyGNN: Drug-drug interaction prediction via hypergraph neural network[C]//Proc of the 39th IEEE Int Conf on Data Engineering. Piscataway, NJ: IEEE, 2023: 1503–1516

[8] Ericsson L, Gouk H, Hospedales T M. How well do self-supervised models transfer?[C]//Proc of the 38th IEEE/CVF Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 5410–5419

[9] Trivedi P, Lubana E S, Yan Yujun, et al. Augmentations in graph contrastive learning: Current methodological flaws & towards better practices[C]//Proc of the 31st Int World Wide Web Conf. New York:

ACM, 2022: 1538–15499

[10] Gidaris S, Singh P, Komodakis N. Unsupervised representation learning by predicting image rotations[C/OL]//Proc of the 6th Int Conf on Leaning Representations. [2024-07-29]. https://openreview.net/forum?id=S1v4N2l0-

[11] Chen Ting, Kornblith S, Norouzi M, et al. A simple framework for contrastive learning of visual representations[C]//Proc of the 37th Int Conf on Machine Learning. New York: ACM, 2020: 1597−1607

[12] Grill J B, Strub F, Altché F, et al. Bootstrap your own latent-a new approach to self-supervised learning[C]//Proc of the 34th Int Conf on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc, 2020: 21271−21284

[13] Kipf T N, Welling M. Variational graph auto-encoders[J]. arXiv preprint, arXiv: 1611.07308, 2016

[14] Grover A, Leskovec J. Node2vec: Scalable feature learning for networks[C]//Proc of the 22nd ACM SIGKDD Conf on Knowledge Discovery and Data Mining. New York: ACM, 2016: 855–864

[15] Tang Jian, Qu Meng, Wang Mingzhe, et al. LINE: Large-scale information network embedding[C]//Proc of the 24th Int World Wide Web Conf. New York: ACM, 2015: 1067–1077

[16] Suresh S, Li Pan, Hao Cong, et al. Adversarial graph augmentation to improve graph contrastive learning[C]//Proc of the 35th Int Conf on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc, 2021: 15920−15933

[17] Veličković P, Fedus W, Hamilton W L, et al. Deep graph infomax[C/OL]//Proc of the 7th Int Conf on Leaning Representations. [2024-07-29]. https://openreview.net/forum?id=rklz9iAcKQ

[18] Zhou Zhengyang, Huang Qihe, Lin Gengyu, et al. Greto: Remedying dynamic graph topology-task discordance via target homophily[C/OL]//Proc of the 11th Int Conf on Leaning Representations. [2024-07-29]. https://openreview.net/pdf?id=8duT3mi_5n

[19] You Yuning, Chen Tianlong, Sui Yongduo, et al. Graph contrastive learning with augmentations[C]//Proc of the 34th Int Conf on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc, 2020: 5812−5823

[20] Liu Nian, Wang Xiao, Bo Deyu, et al. Revisiting graph contrastive learning from the perspective of graph spectrum[C]//Proc of the 36th Int Conf on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc, 2022: 2972−2983

[21] Thakoor S, Tallec C, Azar M G, et al. Large-scale representation learning on graphs via bootstrapping[C/OL]//Proc of the 10th Int Conf on Leaning Representations. [2024-07-29]. https://openreview.net/pdf?id=0UXT6PpRpW

[22] Trivedi P, Lubana E S, Heimann M, et al. Analyzing data-centric properties for graph contrastive learning[C]//Proc of the 36th Int Conf on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc, 2022: 14030−14043

[23] Wang Luzhi, Zheng Yizhen, Jin Di, et al. Contrastive graph similarity networks[J]. ACM Transactions on the Web, 2024, 18(2): 1−20

[24] Wu Junran, Chen Xueyuan, Shi Bowen, et al. SEGA: Structural entropy guided anchor view for graph contrastive learning[C]//Proc of the 40th Int Conf on Machine Learning. New York: ACM, 2023: 37293−37312

[25] Zhu Yanqiao, Xu Yichen, Liu Qiang, et al. An empirical study of graph contrastive learning[J]. arXiv preprint, arXiv: 2109.01116, 2021

[26] Pei Hongbin, Wei Bingzhe, Chang K C-C, et al. Geom-GCN: Geometric graph convolutional networks[C/OL]//Proc of the 8th Int Conf on Leaning Representations. [2024-07-29]. https://openreview.net/forum?id=S1e2agrFvS

[27] Zheng Xin, Wang Yi, Liu Yixin, et al. Graph neural networks for graphs with heterophily: A survey[J]. arXiv preprint, arXiv: 2202.07082, 2022

[28] Ma Yao, Liu Xiaorui, Shah N, et al. Is homophily a necessity for graph neural networks?[C/OL]//Proc of the 10th Int Conf on Leaning Representations. [2024-07-29]. https://openreview.net/forum?id=ucASPPD9GKN

[29] Altenburger K M, Ugander J. Monophily in social networks introduces similarity among friends-of-friends[J]. Nature Human Behaviour, 2018, 2(4): 284−290

[30] Chen Deli, Lin Yankai, Li Wei, et al. Measuring and relieving the over-smoothing problem for graph neural networks from the topological view[C]//Proc of the 29th Association for the Advancement of Artificial Intelligence (AAAI) Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020: 3438–3445

[31] Wu Zonghan, Pan Shirui, Chen Fengwen, et al. A comprehensive survey on graph neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(1): 4−24

[32] Perozzi B, Al-Rfou R, Skiena S. DeepWalk: Online learning of social representations[C]//Proc of the 20th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2014: 701–710

[33] Qin Jiawen, Yuan Haonan, Sun Qingyun, et al. IGL-Bench: Establishing the comprehensive benchmark for imbalanced graph learning[J]. arXiv preprint, arXiv: 2406.09870, 2024

[34] Jin Di, Yu Zhizhi, Huo Cuiying, et al. Universal graph convolutional networks[C]//Proc of the 35th Int Conf on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc, 2021: 10654−10664

[35] Suresh S, Budde V, Neville J, et al. Breaking the limit of graph neural networks by improving the assortativity of graphs with local mixing patterns[C]//Proc of the 27th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2021: 1541–1551

[36] Liu Meng, Wang Zhengyang, Ji Shuiwang. Non-local graph neural networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(12): 10270−10276

[37] Abu-El-Haija S, Perozzi B, Kapoor A, et al. MixHop: Higher-order graph convolutional architectures via sparsified neighborhood mixing[C]//Proc of the 36th Int Conf on Machine Learning. New York: ACM, 2019: 21–29

[38] Zhu Jiong, Yan Yujun, Zhao Lingxiao, et al. Beyond homophily in graph neural networks: Current limitations and effective designs[C]//Proc of the 34th Int Conf on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc, 2020: 7793−7804

[39] Bo Deyu, Wang Xiao, Shi Chuan, et al. Beyond low-frequency information in graph convolutional networks[C]//Proc of the 29th

Association for the Advancement of Artificial Intelligence (AAAI) Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2020: 3950−3957

[40] Chien E, Peng Jianhao, Li Pan, et al. Adaptive universal generalized pagerank graph neural network[C/OL]//Proc of the 9th Int Conf on Leaning Representations. [2024-07-29]. https://openreview.net/forum?id=n6jl7fLxrP

[41] Jiao Pengfei, Liu Huan, Lü Le, et al. Globally enhanced heterogeneous temporal graph neural networks based on contrastive learning[J]. Journal of Computer Research and Development, 2023, 60(8): 1808−1821 (in Chinese)
(焦鹏飞, 刘欢, 吕乐, 等. 基于对比学习的全局增强动态异质图神经网络[J]. 计算机研究与发展, 2023, 60(8): 1808−1821)

[42] Liu Yixin, Jin Ming, Pan Shirui, et al. Graph self-supervised learning: A survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(6): 5879−5900

[43] Liu Yue, Yang Xihong, Zhou Sihong, et al. Simple contrastive graph clustering[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 35(10): 13789−13800

[44] Hasanzadeh A, Hajiramezanali E, Narayanan K, et al. Semi-implicit graph variational auto-encoders[J]. arXiv preprint, arXiv: 1908.07078, 2019

[45] Park J, Lee M, Chang H J, et al. Symmetric graph convolutional autoencoder for unsupervised graph representation learning[C]//Proc of the 36th IEEE/CVF Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2019: 5410–5419

[46] Peng Zhen, Huang Wenbing, Luo Minnan, et al. Graph representation learning via graphical mutual information maximization[C]//Proc of the 29th Int World Wide Web Conf. New York: ACM, 2020: 259–270

[47] Zhu Yanqiao, Xu Yichen, Yu Feng, et al. Graph contrastive learning with adaptive augmentation[C]//Proc of the 30th Int World Wide Web Conf. New York: ACM, 2021: 2069–2080

[48] Zhu Yanqiao, Xu Yichen, Yu Feng, et al. Deep graph contrastive representation learning[J]. arXiv preprint, arXiv: 2006.04131, 2020

[49] Chen Jingfan, Zhu Guanghui, Qi Yifan, et al. Towards self-supervised learning on graphs with heterophily[C]//Proc of the 31st ACM Int Conf on Information & Knowledge Management. New York: ACM, 2022: 201–211

[50] Xiao Teng, Chen Zhengyu, Guo Zhimeng, et al. Decoupled self-supervised learning for non-homophilous graphs[C]//Proc of the 36th Int Conf on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc, 2022: 620−634

[51] Wang Haonan, Zhang Jieyu, Zhu Qi, et al. Single-pass contrastive learning can work for both homophilic and heterophilic graph[J/OL]. Transactions on Machine Learning Research. [2024-07-29]. https://openreview.net/forum?id=244KePn09i

[52] Dong Wei, Wu Junsheng, Luo Yi, et al. Node representation learning in graph via node-to-neighborhood mutual information maximization[C]//Proc of the 39th IEEE/CVF Int Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2022: 16599–16608

[53] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks[J]. Science, 2006, 313(5786): 504−507



Qin Jiawen, born in 1995. PhD candidate. Member of CCF. Her main research interests include graph machine learning and network big data analysis.

秦佳雯, 1995 年生。博士研究生。CCF 会员。主要研究方向为图机器学习、网络大数据分析。



Huang Pengfeng, born in 1999. Master. His main research interest includes graph machine learning.

黄鹏峰, 1999 年生。硕士。主要研究方向为图机器学习。



Sun Qingyun, born in 1996. PhD, assistant professor. Member of CCF. Her main research interests include graph machine learning, network big data analysis, and data mining.

孙庆赟, 1996 年生。博士, 助理教授。CCF 会员。主要研究方向为图机器学习、网络大数据分析、数据挖掘。



Zhou Haoyi, born in 1991. PhD, associate professor. Member of CCF. His main research interests include time-series forecasting, task decision making, and AI safety.

周号益, 1991 年生。博士, 副教授。CCF 会员。主要研究方向为时间序列预测、任务决策、人工智能安全。



Li Jianxin, born in 1979. PhD, professor. Distinguished member of CCF. His main research interests include big data analysis and processing, network big data analysis, and scientific computing.

李建欣, 1979 年生。博士, 教授。CCF 杰出会员。主要研究方向为大数据分析与处理、网络大数据分析、科学计算。