

# 一种双通道半监督网络表示学习模型

杜航原<sup>1</sup>, 谢富中<sup>1</sup>, 王文剑<sup>1</sup>, 白亮<sup>2</sup>

1. 山西大学计算机与信息技术学院, 山西 太原 030006;

2. 山西大学智能信息处理研究所, 山西 太原 030006

## 摘要

在半监督网络表示学习中, 节点标签对于网络在不同空间中映射关系的建立具有重要指导意义。然而在很多实际任务中, 可用标签信息往往比较有限或难以获取, 这导致在学习网络低维表示的过程中无法提供充分有效的监督。针对这一问题, 提出了一种双通道半监督网络表示学习模型, 该模型以自编码器为基本框架, 由自监督和半监督两个信息传递通道构成。自监督信号与标签信息分别在两个通道中对网络表示映射关系的建立提供指导, 同时二者之间形成信息互补与增强。考虑到两个通道间可能存在信息冗余, 在互信息视角下设计了冗余识别与消除机制。在此基础上, 构造了一体化优化模型, 实现自监督学习与半监督学习的协同, 使学习到的网络表示更好地捕捉和保持网络的结构和特性。在真实数据集上的实验结果表明, 提出的模型学习的网络表示在节点分类、聚类和可视化等任务中能够获得优于基线方法的性能。

## 关键词

半监督网络表示学习; 标签信息; 自监督学习; 互信息; 图神经网络

中图分类号: TP391

文献标志码: A

doi: 10.11959/j.issn.2096-0271.2024052

## *A dual channel semi-supervised network representation learning model*

DU Hangyuan<sup>1</sup>, XIE Fuzhong<sup>1</sup>, WANG Wenjian<sup>1</sup>, BAI Liang<sup>2</sup>

1. School of Computer and Information Technology, Shanxi University, Taiyuan 030006, China

2. Institute of Information Processing, Shanxi University, Taiyuan 030006, China

## *Abstract*

In semi-supervised network representation learning, node labels play an important role in guiding the establishment of network mapping relationships among different spaces. However, in many practical tasks, the available label information is usually limited or difficult to obtain, which makes it difficult to provide sufficient and effective supervision to the process of learning low-dimensional network representations. In order to solve this problem, a dual channel semi-supervised network representation learning model is proposed, which is composed of two information transmission channels, namely self-supervised and semi-supervised channels, with the framework of autoencoder. The self-supervised information and the label information provide guidance for the establishment of network representation mapping in the

two channels respectively, and they form such a sense of information complementation and enhancement for the learning process. Considering the possible information redundancy between the two channels, a redundancy recognition and elimination mechanism is designed in the perspective of mutual information. On this basis, an integrated optimization model is constructed to combine the self-supervised learning and the semi-supervised learning into a collaborative mechanism, which enables the learned representations to capture and preserve the structure and characteristics of the network. Experimental results on several real datasets show that the network representations learned by the proposed model can achieve better performance than baseline methods in node classification, clustering and visualization tasks.

### Key words

semi-supervised network representation learning, label information, self-supervised learning, mutual information, graph neural network

## 0 引言

网络数据在现实世界中普遍存在,许多复杂系统包含网络数据,如社交网络、引文网络、生物网络、化学网络和交通网络等。对网络数据的分析和挖掘,有助于进一步了解复杂系统的内部模式、形成机理及演化趋势。然而,网络中边的存在导致了计算爆炸以及传统机器学习方法无法在非欧空间中直接应用等问题。网络表示学习为这些问题提供了有效的解决方法。网络表示学习又称为网络嵌入,旨在将存在于高维空间中的网络(节点、边或子图)映射到低维隐空间中,表示稠密的向量形式,同时最大限度地保留原有网络中的结构、属性等特性<sup>[1-2]</sup>。

目前,网络表示学习已成为一个热点领域,产生了大量研究成果,其代表性方法大致可分为:基于矩阵分解的方法<sup>[3]</sup>、基于浅层网络的表示方法<sup>[4]</sup>、基于深层神经网络的方法<sup>[5]</sup>。在早期研究中,网络表示学习通常被视为一个无监督学习任务,主要依赖网络结构和节点属性学习网络的低维表示,在节点分类、链路预测等下游任务中难以确保表示结果的有效性。然而,在许多现实网络中,节点除了拓扑结构和节

点属性之外,还有节点标签信息。一般情况下,类别标签相同的节点往往具有相似的低维表示,因此,这些标签信息能在网络表示学习方面发挥关键的监督作用,提升网络表示在下游任务中的有效性。近年来,一些学者尝试融合标签信息,提出了一系列半监督网络表示学习方法,如:最大间隔化DeepWalk算法<sup>[6]</sup>在DeepWalk的基础上增加了一个最大间隔分类器,用来指导模型优化过程,使网络表示保留更多对下游任务有益的判别性信息;图卷积网络(graph convolutional networks, GCN)算法<sup>[7]</sup>将一阶切比雪夫多项式作为卷积核进行信息聚合,并使用一种基于边的标签传播规则实现半监督的网络表示学习;Yang等<sup>[8]</sup>设计了一种基于随机游走的采样方法,该方法将标签信息注入编码表示中,通过联合预测节点的邻居节点和类别标签进行半监督表示学习。上述半监督方法通过标签信息的指导监督提高了网络表示学习的有效性。但在实际应用中,标签信息获取难度较大,只能发挥较为有限的监督作用,这导致在一些下游任务中难以获得令人满意的决策结果。

针对上述问题,本文基于网络数据在原始空间和低维空间中的互信息度量构造了自监督信号,对输入网络有限的标签信息形成补充,从而提高半监督场景下网络

表示学习的有效性。在此基础上,提出了一种双通道半监督网络表示学习模型。该模型通过在自监督和半监督两个通道中对自监督信号和标签信息的传递及协同,使网络中的模式和特性在低维表示中得到更好的保持。本文主要贡献如下:①提出了一种双通道网络表示学习模型,通过构造自监督信号补充了半监督场景下网络表示学习的判别性信息;②实现了自监督信号和标签信息在建立空间映射关系中的融合与协同,并对二者之间的信息冗余进行识别和消除;③给出了双通道网络表示学习的一体化优化模型;④在真实数据集上,通过节点分类、聚类以及可视化等任务验证了本文模型的有效性。

## 1 相关工作

### 1.1 图神经网络

图神经网络(graph neural networks, GNN)是一种处理图结构数据的深度学习模型,通过网络中节点之间的信息传递获取图的结构信息<sup>[9]</sup>。GNN因其出色的非结构化数据处理能力,在多种任务中获得成功应用,受到了广泛关注。近年来,循环网络<sup>[10]</sup>、卷积网络<sup>[11]</sup>和深度自编码器<sup>[12]</sup>等经典深度模型被用于设计图神经网络,形成了大量具有代表性的模型与算法,例如递归图神经网络<sup>[13]</sup>、卷积图神经网络<sup>[7,14]</sup>、图自编码器<sup>[15]</sup>以及时空图神经网络<sup>[16]</sup>。在这些GNN模型中,卷积图神经网络具有强大的特征表征能力,且能够方便、有效地与其他神经网络组合,是GNN中最活跃的分支。依据定义图卷积算子的方式,卷积图神经网络可分为谱域方法<sup>[17]</sup>和空域方法。基于谱域的方法从图信号处理的角度引入滤波器来定义图卷积,根

据图谱理论和卷积定理,先将网络数据由空域转换到谱域进行信息聚合,再将结果返回空域得到网络表示;基于空域的方法继承了递归图神经网络的思想,从信息传播的角度定义图卷积,在一定范围内进行节点信息的传播和聚合,实现中心节点表示的递归更新<sup>[18]</sup>。基于空域的方法因效率高、灵活性强等优势,得到了迅速发展和广泛应用。

### 1.2 自监督网络表示学习

为了应对标签信息缺失或难以获取导致的表示学习有效性变差的问题,自监督学习(self-supervised learning, SSL)提供了在无标签数据上进行训练的新思路。SSL依据精心设计的代理任务基于无标签数据自身构造监督信息,再利用这些自监督信号指导模型学习,提升模型特征提取能力。近年来,由于SSL在计算机视觉和自然语言处理等领域获得成功应用,人们尝试将SSL应用于图数据的分析挖掘,提出了生成式、预测式和对比式等图自监督表示学习方法<sup>[19]</sup>,其中,对比式方法被笔者重点关注。该方法通过特定方式对每个样本进行数据增强,再在原有数据和增强数据上采样并构造正负样本对,最后通过最大化正样本对间的一致性以及最小化负样本对间的一致性来构造网络表示学习的优化模型。在这一过程中,样本间的一致性通常由互信息进行衡量。在对比式方法中,数据增强有多种实现方式,如特征增强、结构增强和结构-特征增强。对比式方法的正负样本对可能在节点、子图或整个图等多种尺度上进行构建,因此对比学习过程可能在不同尺度上进行。例如,节点-子图对比<sup>[20]</sup>、节点-图对比<sup>[21]</sup>以及子图-图对比<sup>[22]</sup>。

在对比式学习中,互信息是重要的一

致性衡量手段。由于需要同时获取样本对的联合分布和边缘分布,互信息的计算往往十分困难。为此,神经网络互信估计器(mutual information neural estimator, MINE)<sup>[23]</sup>模型可以实现高维连续随机变量之间的互信息估计,这得益于训练了一个判别输入数据是来自联合分布还是边缘分布乘积的分类器。MINE方法主要基于Kullback-Leibler散度(简称为KL散度)衡量两个分布之间的差异。在随后的工作中,深度信息最大化(deep infomax, DIM)<sup>[24]</sup>和深度图信息最大化(deep graph infomax, DGI)<sup>[21]</sup>使用JSD散度计算分布间差异,通过Jensen-Shannon估计器对互信息进行更加有效的估计。

## 2 双通道半监督网络表示学习模型

本文提出了一种双通道半监督网络表示学习(dual channel semi-supervised network representation learning, DCSNRL)模型,该模型采用对比式策略构造自监督信号,在双通道架构下与标签信息形成协同互补,使学习的网络表示能够保持输入网络的丰富结构和特性,进而提升网络表示在下游任务中的有效性。同时,针对两个通道中可能存在的冗余信息,设计了有效的识别和消除机制,避免对模型优化过程的过度干预。

### 2.1 模型总体架构

给定一个包含 $N$ 个节点的图 $G=(V,A)$ ,其中 $V=\{v_1,\dots,v_N\}$ 为图中节点构成的集合, $A\in R^{N\times N}$ 表示图中节点之间的关联关系, $A_{ij}=1$ 表示节点 $v_i$ 和 $v_j$ 之间有连边, $A_{ij}=0$ 表示节点 $v_i$ 和 $v_j$ 之间没有连

边。节点特征由 $X=\{x_1,\dots,x_N\}$ 表示, $G_i$ 表示由节点 $v_i$ 及其邻居节点构成的子图。 $\bar{V}=\{v_1,\dots,v_L\}$ 为图 $G$ 中具备类别标签的节点构成的集合, $Y=\{y_1,\dots,y_L\}$ 为 $\bar{V}$ 中节点的标签构成的集合,其中 $y_l$ 表示节点 $v_l$ 的类别标签,将节点类别数量记为 $K$ 。

DCSNRL模型的总体架构如图1所示,该模型由半监督通道、自监督通道以及冗余消除模块构成。半监督通道与自监督通道使用GCN构成的编码器,分别以半监督和自监督学习模式获得网络的低维表示。在半监督通道中,利用节点类别标签为编码过程提供判别性信息;在自监督通道中,通过数据增强构造正负样本对,采用对比式策略产生自监督信号,形成与半监督通道的信息互补,建立对网络表示学习的协同监督机制;冗余消除模块用于识别和消除两个通道中冗余监督信息,从而确保网络表示的有效性和可靠性。

### 2.2 模型构成

#### 2.2.1 半监督通道

该通道使用GCN构成的编码器学习输入网络 $G$ 的低维表示 $H$ ,将节点类别标签作为监督信息指导网络表示学习过程,通过使用具有Softmax函数和交叉熵损失函数的神经网络分类器作为互信息估计器<sup>[25]</sup>来确保标签相同的节点具有相近的低维表示。

GCN编码器学习网络低维表示的过程如式(1)所示。

$$H = \text{GCNs}(X, A) \quad (1)$$

其中, $H$ 为学习得到的向量表示,GCNs( $\cdot$ )由两层图卷积网络构成,其结构如式(2)所示。

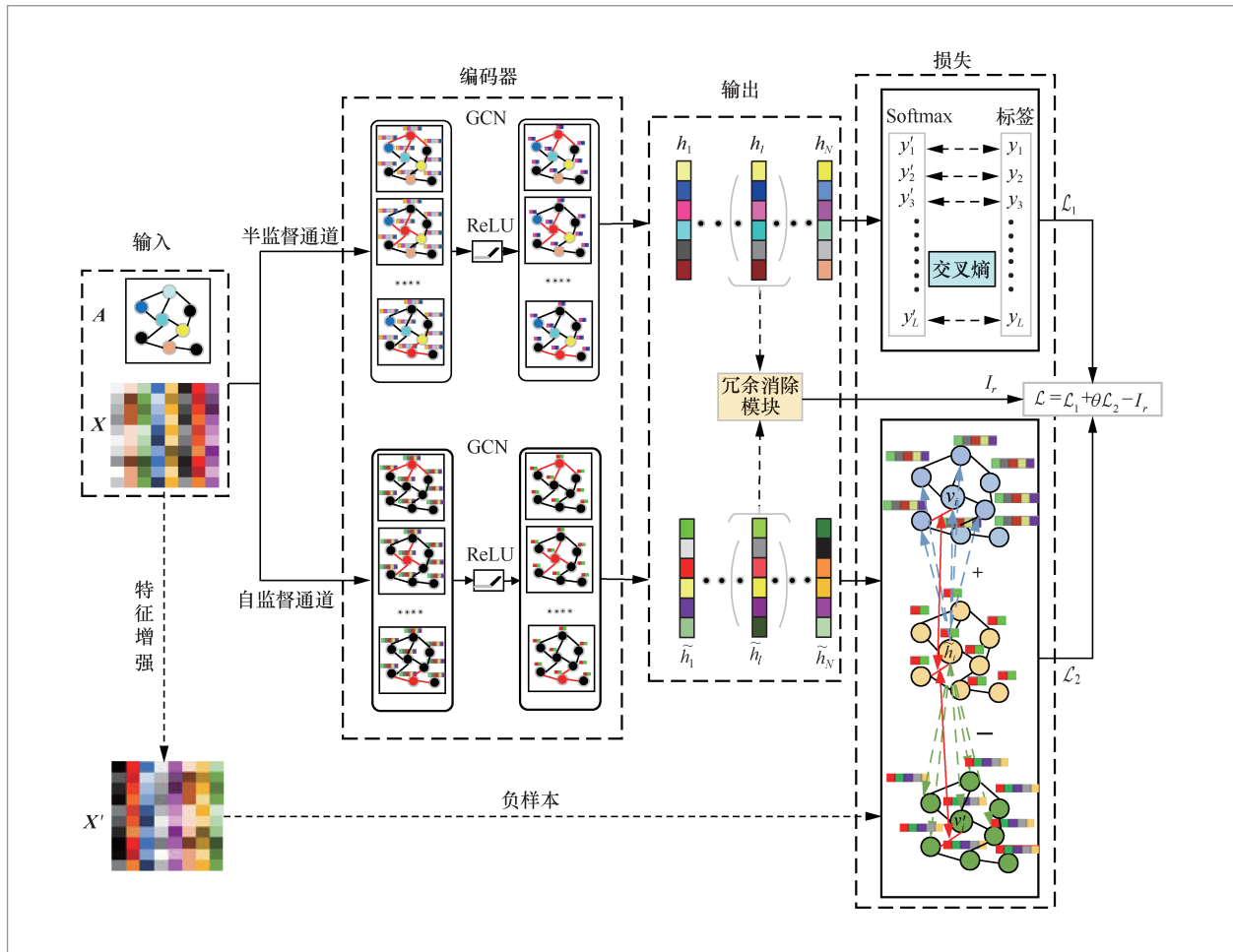


图1 双通道半监督网络表示学习模型框架

$$GCNs(X, A) = \hat{A} \text{Relu}(AXW_0)W_1 \quad (2)$$

其中,  $\hat{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ ,  $D$  为  $A$  的度矩阵,  $W_0$  和  $W_1$  是可学习的参数矩阵。

半监督通道使用Softmax层作为输出层, 进行节点类别预测, 并利用交叉熵损失函数衡量节点真实类别与预测类别之间的差异, 该通道的优化目标如式(3)所示。

$$\mathcal{L}_1 = E_{(\bar{v}, y)} \left[ h_{y_i} - \log \left( \sum_{k=1}^K \exp(h_{ik}) \right) \right] \approx \frac{1}{L} \sum_{v_i \in \bar{V}} \log \frac{\exp(h_{y_i})}{\sum_{k=1}^K \exp(h_{ik})} \quad (3)$$

其中,  $h_{y_i}$  为节点  $v_i$  学得的低维表示  $h_i$  中的第  $y_i$  个分量, 类似地,  $h_{ik}$  表示  $h_i$  的第  $k$  个分量。

### 2.2.2 自监督通道

在自监督通道中, 使用与半监督通道的编码器结构相同的GCN, 将输入网络映射为低维空间中的向量表示, 并通过构造自监督信号指导这一表示学习过程。

自监督学习模式的构造是通过设计代理任务实现的, 使得输入网络的拓扑结构和节点特性在表示结果中尽可能完整保

留。在输入图和网络表示之间分别构造正样本对和负样本对,通过鉴别器的判别过程建立节点-子图尺度的对比学习机制,进而指导表示学习的优化。具体而言,将节点表示  $\tilde{h}_i$  和  $G_i$  中的节点  $v_j$  作为正样本对,将通过特征增强方式得到子图节点  $v_j$  的负样本  $v'_j$  与节点表示  $\tilde{h}_i$  构成负样本对。使用互信息分别对正负样本对在子图范围内的一致性进行度量,如式(4)所示。

$$I(\tilde{h}_i; V_{G_i}) = \sum_{v_j \in G_i} w_{ij} I(\tilde{h}_i; v_j) \quad (4)$$

其中  $I(\tilde{h}_i; v_j)$  表示构成样本对的节点表示  $\tilde{h}_i$  与子图节点  $v_j$  间的互信息,由正样本对互信息和负样本对互信息两部分组成,如式(5)所示。

$$I(\tilde{h}_i; v_j) = -\log\left(1 + e^{-D_1(\tilde{h}_i, v_j)}\right) - E\left[\log\left(1 + e^{D_1(\tilde{h}_i, v'_j)}\right)\right] \quad (5)$$

其中,鉴别器  $D_1(\tilde{h}_i, v_j) = \sigma(\tilde{h}_i^T \Theta_1 v_j)$ ,  $\Theta_1$  是一个可训练的得分矩阵。

在式(4)中,  $w_{ij} = \sigma(\tilde{h}_i^T \tilde{h}_j)$  为网络表示中节点  $v_i$  和  $v_j$  的相关性,控制着局部互信息  $I(\tilde{h}_i; v_j)$  对全局互信息  $I(\tilde{h}_i; V_{G_i})$  的贡献。为使得到的互信息更好地反映节点特性和拓扑关系在表示结果中的保留情况,将网络表示中节点  $v_i$  和  $v_j$  的相关性  $w_{ij}$  与输入网络中边权值  $a_{ij}$  进行关联,节点  $v_i$  和  $v_j$  越靠近,即边权值  $a_{ij}$  越大,则  $w_{ij}$  越大。通过交叉熵损失函数计算  $w_{ij}$  和  $a_{ij}$  之间的相关性,如式(6)所示。

$$I(a_{ij}; w_{ij}) = a_{ij} \log w_{ij} + (1 - a_{ij}) \log(1 - w_{ij}) \quad (6)$$

在自监督学习模式下,利用代理任务确保输入网络的拓扑特性和节点特性在低维表示中得到完整保留。自监督通道的一体化优化目标如式(7)所示。

$$\mathcal{L}_2 = \sum_{v_i \in G} I(\tilde{h}_i; G_i) \quad (7)$$

其中,  $I(\tilde{h}_i; G_i)$  的计算如下。

$$I(\tilde{h}_i; G_i) = \sum_{v_j \in G_i} \left[ w_{ij} I(\tilde{h}_i; v_j) + I(a_{ij}; w_{ij}) \right] \quad (8)$$

### 2.2.3 信息冗余的识别

DCSNRL模型的两个信息传递通道,分别使用了节点标签信息和构造的自监督信号为网络表示学习过程提供监督指导。上述信息在不同视角下对输入网络的特性做出了描述,以互补方式确保网络特性在表示学习过程中得到完整保留。同时,由于这些信息源于对同一输入网络的抽取,它们在对输入网络的某些特性进行描述时不可避免地存在重合,即信息冗余。信息冗余可能导致网络中某些局部特性被过度关注,进而对模型性能产生不利影响。为解决这一问题,本文提出了如图2所示的信息冗余识别及消除机制。对于某一具备标签的网络节点  $v_i$ ,其在两个通道中分别依靠标签信息和自监督信号获得的低维表示为  $h_i$  和  $\tilde{h}_i$ ,利用二者之间的互信息进行信息冗余的度量,如式(9)所示。

$$I(h_i; \tilde{h}_i) = -\log\left(1 + e^{-D_2(h_i, \tilde{h}_i)}\right) - E\left[\log\left(1 + e^{D_2(h_i, \tilde{h}'_i)}\right)\right] \quad (9)$$

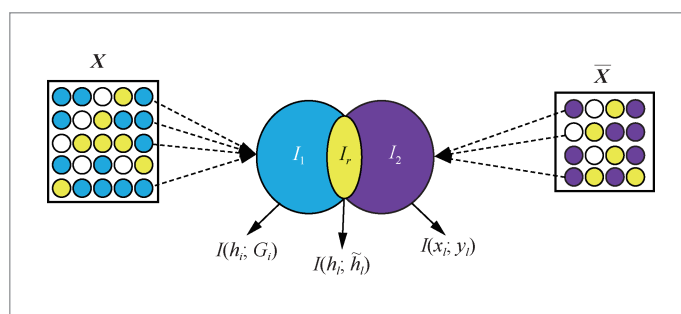


图2 信息冗余识别及消除机制

式(9)通过分别构造正负样本对计算两个通道间的信息冗余,其中 $\tilde{h}_i'$ 表示由 $\tilde{h}_i$ 经过特征增强生成的负样本,鉴别器 $\mathcal{D}_2(h_i, \tilde{h}_i) = \sigma(h_i^T \Theta_2 \tilde{h}_i)$ ,  $\Theta_2$ 是一个可训练的得分矩阵。

由此,输入网络在表示学习过程中存在的信息冗余可由式(10)计算,

$$I_r = \sum_{v_i \in \tilde{V}} I(h_i; \tilde{h}_i) \quad (10)$$

### 2.3 优化目标

在DCSNRL模型中,标签信息和自监督信号在不同通道中以互补方式为表示学习过程提供全面的监督指导,同时对通道间存在的信息冗余进行消除。将上述构件进行整合,得到模型总体优化目标,如下。

$$\mathcal{L} = \mathcal{L}_1 + \theta \mathcal{L}_2 - I_r \quad (11)$$

其中,  $\theta \in (0,1)$  为调节系数,用于调整不同通道在表示学习过程中的权重,其具体取值可依据网络中节点的数量和可靠性进行调整。

### 2.4 算法流程

在训练过程中,两个通道都采用互信的方式进行损失计算,并以迭代方式使用随机梯度下降(stochastic gradient descent, SGD)算法对两个通道中的神经网络参数和得分矩阵进行交替更新。算法1总结了DCSNRL模型的执行流程。

---

算法1: 双通道半监督网络表示学习方法

输入: 原始网络 $G$ , 节点特征 $X$ , 邻接矩阵 $A$ , 部分节点标签 $Y$ , 调节系数 $\theta$ , 最大迭代次数MaxIter

输出: 节点的嵌入向量

---



---

```

1: 随机初始化模型参数
2: for iter=1 to MaxIter
3:     根据式(1)分别输出半监督通道网络表示 $H$ 和自监督通道网络表示 $\tilde{H}$ 
4:     根据式(3)、式(7)、式(10)分别计算半监督通道损失 $\mathcal{L}_1$ 、自监督通道损失 $\mathcal{L}_2$ 和冗余信息 $I_r$ 
5:     根据式(11)计算总体损失 $\mathcal{L}$ 
6:     if iter % 2 == 0 then
7:         根据 $\mathcal{L}$ 反向传播更新神经网络参数 $W_0$ 、 $W_1$ 和得分矩阵 $\Theta_2$ 
8:     else
9:         根据 $\mathcal{L}$ 反向传播更新神经网络参数 $\tilde{W}_0$ 、 $\tilde{W}_1$ 和得分矩阵 $\Theta_1$ 、 $\Theta_2$ 
10:    return  $Z=(H+\tilde{H})/2$ 
11: end

```

---

## 3 实验与分析

### 3.1 数据集

在仿真实验中,本文在5个不同规模的网络数据集<sup>[1,7]</sup>上通过分类、聚类和可视化任务对DCSNRL模型的有效性进行验证。这些数据集包括Cora、Citeseer、Pubmed、Facebook和Blogcatalog,见表1。

### 3.2 对比方法与评价指标

将本文提出的DCSNRL方法与几种网络表示学习基线方法进行比较,包括对比式表示学习方法GMI<sup>[20]</sup>和DGI<sup>[21]</sup>,深度图表示学习方法GraphMLP<sup>[26]</sup>、GWNN<sup>[27]</sup>和GCN<sup>[7]</sup>,浅层表示学习方法DeepWalk和Node2vec。在实验结果的评价方面,本

文使用Micro-F1和Macro-F1作为节点分类结果的评价指标,并利用标准化互信息(normalized mutual information, NMI)以及调节兰德系数(adjusted rand index, ARI)进行聚类性能的评价。

表1 实验数据集

| 数据集         | 节点数    | 边数      | 特征维度  | 类别 |
|-------------|--------|---------|-------|----|
| Cora        | 2 708  | 5 429   | 1 433 | 7  |
| Citeseer    | 3 327  | 4 732   | 3 703 | 6  |
| Pubmed      | 19 717 | 44 338  | 500   | 3  |
| Facebook    | 22 470 | 171 002 | 4 714 | 4  |
| Blogcatalog | 5 196  | 171 743 | 8 189 | 6  |

### 3.3 实验结果与分析

对于每个数据集,首先设置数据集中节点的类别标记数量分别为5、10、20、30、40,从未标记节点中随机选取1 000个节点作为测试数据。对实验结果进行比较分析,其中最优结果使用粗体显示。

#### 3.3.1 节点分类

表2分别展示了当节点类别标记数量为20时,不同方法在5个数据集上进行节点分类获得的Micro-F1值和Macro-F1值。总体来看,与其他方法相比,DCSNRL方法在不同数据集上都获得了最优结果。这得益于该方法构造的自监督信号对标签信息形成了有效补充,在标签有限的情

况下为表示学习过程提供了更多的指导,提升了网络表示的有效性。GraphMLP方法在向前传播的过程中未利用节点间的关联关系进行特征信息的传递,而是仅将其作为结构监督信息,这加强了邻居节点间的关联,削弱了与其他节点的联系,导致分类结果难以令人满意。在对比方法中,DGI和GMI方法面向无监督场景,仅依靠构造的自监督信号进行训练,无法对已有的标签信息加以利用,使其结果始终次于DCSNRL方法。GWNN和GCN都是基于图神经网络的方法,它们使用标签信息作为监督信息对表示学习过程进行指导,而当标签信息有限时,其表示结果难以满足分类任务的要求。DeepWalk和

表2 不同数据集节点分类的实验结果

| 方法       | 算法       | 数据集            |                |                |                |                |                |                |          |                |                |         |
|----------|----------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------|----------------|----------------|---------|
|          |          | Cora           |                | Citeseer       |                | Pubmed         |                | Facebook       |          | Blogcatalog    |                |         |
|          |          | Micro-F1       | Macro-F1       | Micro-F1       | Macro-F1       | Micro-F1       | Macro-F1       | Micro-F1       | Macro-F1 | Micro-F1       | Macro-F1       |         |
| 无监督      | DeepWalk | 70.09%±        | 69.99%±        | 69.55%±        | 48.66%±        | 74.67%±        | 73.01%±        | 64.32%±        | 62.99%±  | 33.47%±        | 26.84%±        |         |
|          |          | 0.85%          | 0.93%          | 0.54%          | 1.94%          | 1.16%          | 1.17%          | 1.67%          | 1.79%    | 0.61%          | 0.56%          |         |
|          | Node2vec | 71.96%±        | 71.86%±        | 70.35%±        | 50.37%±        | 74.51%±        | 73.10%±        | 66.71%±        | 65.16%±  | 33.42%±        | 26.37%±        |         |
|          |          | 0.79%          | 0.80%          | 0.44%          | 1.12%          | 0.66%          | 0.63%          | 1.34%          | 1.53%    | 0.94%          | 0.86%          |         |
|          | DGI      | 81.47%±        | 79.96%±        | 72.11%±        | 67.33%±        | 77.13%±        | 77.36%±        | 71.29%±        | 69.88%±  | 44.12%±        | 25.53%±        |         |
|          |          | 0.20%          | 0.17%          | 0.38%          | 0.23%          | 0.18%          | 0.15%          | 0.14%          | 0.17%    | 0.63%          | 0.45%          |         |
|          | GMI      | 81.57%±        | 80.42%±        | <b>73.12%±</b> | 66.37%±        | 78.40%±        | 78.32%±        | 72.36%±        | 70.55%±  | 35.26%±        | 23.14%±        |         |
|          |          | 0.36%          | 0.45%          | <b>0.59%</b>   | 0.56%          | 0.30%          | 0.35%          | 0.26%          | 0.34%    | 1.67%          | 1.29%          |         |
|          | GCN      | 81.02%±        | 79.87%±        | 69.55%±        | 67.23%±        | 78.38%±        | 78.03%±        | 68.17%±        | 66.17%±  | 45.92%±        | 37.87%±        |         |
|          |          | 0.09%          | 0.07%          | 0.54%          | 0.42%          | 0.26%          | 0.23%          | 0.05%          | 0.07%    | 1.45%          | 0.68%          |         |
|          | 有监督      | GWNN           | 81.62%±        | 80.25%±        | 70.35%±        | 67.32%±        | 78.64%±        | 77.98%±        | 72.90%±  | <b>72.61%±</b> | 48.21%±        | 38.45%± |
|          |          |                | 0.86%          | 0.79%          | 0.44%          | 0.29%          | 0.23%          | 0.26%          | 0.42%    | <b>0.41%</b>   | 1.37%          | 1.17%   |
| GraphMLP |          | 79.90%±        | 78.04%±        | 72.11%±        | 68.16%±        | 79.20%±        | 78.66%±        | 61.05%±        | 59.93%±  | 46.86%±        | 36.77%±        |         |
|          |          | 0.43%          | 0.49%          | 0.38%          | 0.36%          | 0.45%          | 0.48%          | 0.91%          | 0.80%    | 0.77%          | 0.53%          |         |
| DCSNRL   |          | <b>83.57%±</b> | <b>82.27%±</b> | <b>73.12%±</b> | <b>68.73%±</b> | <b>80.85%±</b> | <b>80.23%±</b> | <b>73.41%±</b> | 72.22%±  | <b>53.37%±</b> | <b>40.77%±</b> |         |
|          |          | <b>0.34%</b>   | <b>0.37%</b>   | <b>0.59%</b>   | <b>0.59%</b>   | <b>0.18%</b>   | <b>0.16%</b>   | <b>0.70%</b>   | 0.73%    | <b>1.24%</b>   | <b>0.36%</b>   |         |

Node2vec方法侧重于在网络表示中保持拓扑结构信息,无法对节点标签信息进行充分利用,导致其分类结果相比于其他方法具有显著差异。

表3展示了当节点类别标记数量分别为5、10、20、30、40时,不同算法在5个数据集上进行节点分类的结果。总体来看,DCSNRL模型获得了优于其他算法的分类结果。当标记数量处于较低水平时,DCSNRL方法已经能够获得较高的Micro-F1和Macro-F1值,这表明模型构造的自监督信号在表示学习过程中对判别性信息的保留和捕获产生了积极作用。

### 3.3.2 聚类

在聚类任务中,本文通过在不同算法

得到的节点表示上执行k-means算法<sup>[28]</sup>来获得对数据集的簇划分,将10次随机初始化条件下执行结果的平均值作为各算法的最终结果。表4展示了当节点类别标记数量为20时,不同算法在5个数据集上的NMI值和ARI值。在Cora、Pubmed和Blogcatalog数据集上,DCSNRL模型相比于其他表示学习方法存在较为明显的优势。对于Citeseer和Facebook数据集,DCSNRL模型与部分对比方法的聚类结果较为相近,与其他对比方法存在较为明显的差距。表5展示了当节点类别标记数量为5、10、20、30、40时,不同算法在5个数据集上进行聚类的结果。由表5可知,DCSNRL模型在3个数据集上都获得了最优的结果,主要原因如下:一方面,该模型在自监督通道中使用对比式自监督学习

表3 不同节点标记数量下各数据集节点分类的实验结果

| 数据集              | 节点<br>标记<br>数量 | 方法           |                     |                     |                     |                     |                     |                     |                     |
|------------------|----------------|--------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                  |                | GCN          |                     | GWNN                |                     | GraphMLP            |                     | DCSNRL              |                     |
|                  |                | Micro-F1     | Macro-F1            | Micro-F1            | Macro-F1            | Micro-F1            | Macro-F1            | Micro-F1            | Macro-F1            |
| Cora             | 5              | 70.57%±1.03% | 68.29%±0.98%        | 74.31%±0.64%        | 72.78%±1.05%        | 71.42%±1.78%        | 70.52%±1.70%        | <b>76.38%±0.83%</b> | <b>74.69%±0.92%</b> |
|                  | 10             | 73.91%±0.17% | 72.96%±0.25%        | 75.50%±0.87%        | 74.70%±0.73%        | 74.48%±1.30%        | 72.71%±1.35%        | <b>77.61%±0.46%</b> | <b>76.45%±0.54%</b> |
|                  | 20             | 81.02%±0.09% | 79.87%±0.07%        | 81.62%±0.86%        | 80.25%±0.79%        | 79.90%±0.43%        | 78.04%±0.49%        | <b>83.57%±0.34%</b> | <b>82.27%±0.37%</b> |
|                  | 30             | 82.76%±0.16% | 81.74%±0.16%        | 82.96%±0.56%        | 81.82%±0.65%        | 81.64%±0.94%        | 80.07%±0.98%        | <b>84.63%±0.23%</b> | <b>83.52%±0.24%</b> |
|                  | 40             | 83.49%±0.24% | 82.51%±0.28%        | 83.40%±0.57%        | 82.44%±0.71%        | 81.94%±0.33%        | 81.93%±0.33%        | <b>84.98%±0.36%</b> | <b>83.84%±0.37%</b> |
| Citeseer         | 5              | 53.17%±0.22% | 48.52%±0.32%        | 54.77%±1.73%        | 48.62%±2.57%        | 61.19%±1.67%        | 57.69%±1.94%        | <b>65.88%±1.69%</b> | <b>61.16%±1.39%</b> |
|                  | 10             | 66.51%±0.34% | 63.68%±0.27%        | 65.92%±1.07%        | 62.62%±0.99%        | 65.70%±0.91%        | 62.25%±0.81%        | <b>68.73%±0.54%</b> | <b>64.91%±0.46%</b> |
|                  | 20             | 69.55%±0.54% | 67.23%±0.42%        | 70.35%±0.44%        | 67.32%±0.29%        | 72.11%±0.38%        | 68.16%±0.36%        | <b>73.12%±0.59%</b> | <b>68.73%±0.59%</b> |
|                  | 30             | 72.17%±0.34% | 69.37%±0.26%        | 72.56%±0.63%        | 69.73%±0.48%        | 71.14%±0.40%        | 67.56%±0.41%        | <b>73.87%±0.48%</b> | <b>69.85%±0.59%</b> |
|                  | 40             | 72.36%±0.27% | 69.57%±0.32%        | 72.78%±0.28%        | 69.81%±0.78%        | 72.32%±0.65%        | 68.42%±0.61%        | <b>73.91%±0.41%</b> | <b>70.04%±0.45%</b> |
| Pubmed           | 5              | 67.84%±0.19% | 68.70%±0.27%        | 71.84%±0.05%        | 71.96%±0.05%        | 70.14%±2.24%        | 70.02%±2.11%        | <b>75.01%±0.82%</b> | <b>75.45%±0.73%</b> |
|                  | 10             | 71.43%±1.49% | 71.97%±1.61%        | 72.76%±0.16%        | 72.72%±0.15%        | 72.06%±1.65%        | 72.43%±1.81%        | <b>76.24%±0.67%</b> | <b>76.34%±0.65%</b> |
|                  | 20             | 78.38%±0.26% | 78.03%±0.23%        | 78.64%±0.23%        | 77.98%±0.26%        | 79.20%±0.45%        | 78.66%±0.48%        | <b>80.85%±0.18%</b> | <b>80.23%±0.16%</b> |
|                  | 30             | 80.40%±0.23% | 79.65%±0.24%        | 79.14%±0.13%        | 78.44%±0.12%        | 79.34%±0.49%        | 78.89%±0.51%        | <b>82.35%±0.61%</b> | <b>81.53%±0.45%</b> |
|                  | 40             | 80.85%±0.52% | 79.86%±0.51%        | 79.45%±0.12%        | 78.79%±0.08%        | 80.05%±0.48%        | 79.34%±0.57%        | <b>82.55%±0.34%</b> | <b>81.76%±0.27%</b> |
| Facebook         | 5              | 57.80%±0.08% | 55.81%±0.08%        | 62.30%±0.35%        | 60.79%±0.35%        | 47.04%±3.04%        | 46.12%±2.80%        | <b>63.27%±1.20%</b> | <b>61.55%±1.03%</b> |
|                  | 10             | 63.71%±0.24% | 60.51%±0.22%        | <b>69.21%±0.31%</b> | <b>67.48%±0.24%</b> | 55.50%±1.93%        | 54.25%±1.87%        | 68.97%±1.16%        | 67.10%±1.08%        |
|                  | 20             | 68.17%±0.05% | 66.17%±0.07%        | 72.90%±0.42%        | <b>72.61%±0.41%</b> | 61.05%±0.91%        | 59.93%±0.80%        | <b>73.41%±0.70%</b> | 72.22%±0.73%        |
|                  | 30             | 70.43%±0.47% | 69.13%±0.07%        | <b>74.62%±0.27%</b> | 73.34%±0.27%        | 62.48%±1.99%        | 61.22%±1.89%        | 74.37%±0.73%        | <b>73.53%±0.65%</b> |
|                  | 40             | 75.14%±0.58% | 74.15%±0.54%        | 78.20%±0.49%        | 77.43%±0.43%        | 66.71%±1.71%        | 65.43%±1.69%        | <b>78.26%±0.67%</b> | <b>77.54%±0.60%</b> |
| Blogc-<br>atalog | 5              | 38.83%±1.63% | 30.08%±1.24%        | 40.12%±2.19%        | 32.21%±1.53%        | <b>42.00%±1.57%</b> | <b>33.02%±1.45%</b> | 39.73%±0.98%        | 30.24%±1.09%        |
|                  | 10             | 44.01%±1.69% | <b>36.62%±0.98%</b> | 45.60%±2.27%        | 35.18%±0.69%        | 43.94%±1.01%        | 34.61%±0.52%        | <b>47.80%±0.17%</b> | 34.62%±0.46%        |
|                  | 20             | 45.92%±1.45% | 37.87%±0.68%        | 48.21%±1.37%        | 38.45%±1.17%        | 46.86%±0.77%        | 36.77%±0.53%        | <b>53.37%±1.24%</b> | <b>40.77%±0.36%</b> |
|                  | 30             | 48.64%±1.36% | 39.33%±0.91%        | 51.21%±2.24%        | 39.76%±1.21%        | 48.04%±0.87%        | 37.51%±0.61%        | <b>58.32%±1.45%</b> | <b>41.37%±1.65%</b> |
|                  | 40             | 48.83%±1.32% | 40.44%±0.59%        | 53.10%±2.28%        | 40.98%±0.67%        | 48.81%±0.30%        | 38.65%±0.36%        | <b>59.81%±1.10%</b> | <b>41.53%±0.64%</b> |

表 4 不同数据集聚类的实验结果

| 方法       | 算法             | 数据集            |                |                |                |                |                |                |                |                |              |
|----------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|--------------|
|          |                | Cora           |                | Citeseer       |                | Pubmed         |                | Facebook       |                | Blogcatalog    |              |
|          |                | NMI            | ARI            | NMI            | ARI            | NMI            | ARI            | NMI            | ARI            | NMI            | ARI          |
| 无监督      | DeepWalk       | 44.05%±        | 33.76%±        | 24.15%±        | 11.22%±        | 29.78%±        | 29.83%±        | 25.33%±        | 17.92%±        | 24.57%±        | 16.67%±      |
|          |                | 2.43%          | 4.01%          | 1.35%          | 0.45%          | 0.23%          | 0.36%          | 1.59%          | 1.94%          | 0.63%          | 0.93%        |
|          | Node2vec       | 44.55%±        | 34.32%±        | 24.91%±        | 10.97%±        | 29.84%±        | 29.87%±        | 25.24%±        | 17.94%±        | 24.39%±        | 16.16%±      |
|          |                | 2.64%          | 4.12%          | 0.41%          | 0.24%          | 0.23%          | 0.28%          | 1.31%          | 1.56%          | 0.53%          | 1.19%        |
|          |                | DGI            | 58.99%±        | 55.54%±        | 42.85%±        | 40.71%±        | 38.58%±        | 36.77%±        | 38.59%±        | 34.19%±        | 33.13%±      |
| 有监督      | GMI            | 0.51%          | 1.19%          | 0.19%          | 0.29%          | 0.21%          | 0.22%          | 0.29%          | 0.60%          | 0.15%          | 0.19%        |
|          |                | 58.26%±        | 55.24%±        | 41.44%±        | 40.41%±        | 33.03%±        | 36.95%±        | 40.67%±        | 36.20%±        | 28.17%±        | 17.41%±      |
|          | GCN            | 1.12%          | 1.54%          | 0.58%          | 0.83%          | 2.66%          | 2.74%          | 1.51%          | 1.67%          | 1.56%          | 1.44%        |
|          |                | 41.04%±        | 23.91%±        | 37.97%±        | 29.53%±        | 34.16%±        | 32.57%±        | 36.77%±        | 34.93%±        | 28.94%±        | 15.51%±      |
|          |                | 0.68%          | 0.78%          | 0.36%          | 0.41%          | 0.22%          | 0.52%          | 0.46%          | 0.54%          | 0.80%          | 0.69%        |
| GWNN     | 44.19%±        | 22.85%±        | 36.43%±        | 33.89%±        | 37.38%±        | 38.57%±        | <b>42.92%±</b> | <b>41.66%±</b> | 28.57%±        | 19.03%±        |              |
|          | 1.09%          | 1.19%          | 1.29%          | 2.65%          | 0.18%          | 0.25%          | <b>0.57%</b>   | <b>0.25%</b>   | 0.57%          | 0.53%          |              |
| GraphMLP | 50.12%±        | 37.12%±        | 41.96%±        | 38.03%±        | 36.04%±        | 36.91%±        | 25.54%±        | 19.82%±        | 39.96%±        | 31.08%±        |              |
|          | 0.51%          | 1.02%          | 0.83%          | 1.71%          | 0.69%          | 1.36%          | 2.55%          | 3.15%          | 0.81%          | 0.38%          |              |
| DCSNRL   | <b>62.65%±</b> | <b>62.99%±</b> | <b>43.41%±</b> | <b>43.78%±</b> | <b>41.36%±</b> | <b>46.32%±</b> | 42.17%±        | 41.23%±        | <b>42.85%±</b> | <b>36.51%±</b> |              |
|          |                | <b>0.26%</b>   | <b>0.69%</b>   | <b>0.26%</b>   | <b>0.28%</b>   | <b>0.57%</b>   | <b>0.46%</b>   | 0.60%          | 1.57%          | <b>0.16%</b>   | <b>0.79%</b> |

表 5 不同节点标记数量下各数据集聚类的实验结果

| 数据集         | 节点标记数量 | 方法           |              |                     |                     |                     |              |                     |                     |
|-------------|--------|--------------|--------------|---------------------|---------------------|---------------------|--------------|---------------------|---------------------|
|             |        | GCN          |              | GWNN                |                     | GraphMLP            |              | DCSNRL              |                     |
|             |        | NMI          | ARI          | NMI                 | ARI                 | NMI                 | ARI          | NMI                 | ARI                 |
| Cora        | 5      | 29.11%±0.11% | 14.23%±0.18% | 32.17%±0.72%        | 13.49%±1.04%        | 46.71%±1.06%        | 34.64%±1.04% | <b>57.50%±0.47%</b> | <b>57.00%±0.90%</b> |
|             | 10     | 34.77%±0.10% | 20.88%±0.28% | 38.90%±1.26%        | 19.29%±0.69%        | 47.80%±1.26%        | 35.12%±0.87% | <b>58.04%±0.66%</b> | <b>58.09%±0.71%</b> |
|             | 20     | 41.04%±0.68% | 23.91%±0.78% | 44.19%±1.09%        | 22.85%±1.19%        | 50.12%±0.51%        | 37.12%±1.02% | <b>62.65%±0.26%</b> | <b>62.99%±0.69%</b> |
|             | 30     | 43.82%±1.84% | 29.53%±1.98% | 48.11%±1.01%        | 27.81%±0.28%        | 49.55%±0.44%        | 35.45%±1.02% | <b>63.57%±0.35%</b> | <b>64.48%±0.56%</b> |
|             | 40     | 45.31%±0.32% | 27.54%±0.66% | 60.08%±1.13%        | 42.81%±1.47%        | 50.18%±2.09%        | 39.32%±3.32% | <b>65.26%±0.44%</b> | <b>66.15%±0.98%</b> |
| Citeseer    | 5      | 27.85%±0.13% | 20.87%±0.27% | 25.19%±0.73%        | 17.49%±1.06%        | 36.65%±0.70%        | 32.32%±1.22% | <b>38.77%±0.92%</b> | <b>38.50%±0.83%</b> |
|             | 10     | 32.49%±0.20% | 23.29%±0.93% | 33.21%±2.14%        | 30.10%±2.46%        | 39.54%±2.12%        | 35.71%±3.26% | <b>41.98%±0.48%</b> | <b>41.97%±0.78%</b> |
|             | 20     | 37.97%±0.36% | 29.53%±0.41% | 36.43%±1.29%        | 33.89%±2.65%        | 41.96%±0.83%        | 38.03%±1.71% | <b>43.41%±0.26%</b> | <b>43.78%±0.28%</b> |
|             | 30     | 40.08%±0.07% | 30.95%±0.16% | 36.02%±1.05%        | 34.07%±1.87%        | 40.96%±0.79%        | 37.44%±0.83% | <b>44.65%±0.99%</b> | <b>45.62%±1.71%</b> |
|             | 40     | 41.87%±0.16% | 35.31%±0.34% | 39.48%±0.78%        | 34.94%±2.46%        | 42.66%±0.57%        | 38.71%±0.84% | <b>44.76%±0.16%</b> | <b>45.70%±0.36%</b> |
| Pubmed      | 5      | 30.39%±1.17% | 30.03%±1.34% | 31.41%±0.07%        | 31.69%±0.04%        | 30.27%±4.53%        | 25.67%±4.77% | <b>34.94%±1.11%</b> | <b>38.06%±1.34%</b> |
|             | 10     | 32.97%±0.70% | 34.48%±0.54% | 34.38%±0.16%        | 34.19%±0.13%        | 33.12%±1.24%        | 31.09%±1.80% | <b>36.47%±0.94%</b> | <b>38.87%±1.02%</b> |
|             | 20     | 34.16%±0.22% | 32.57%±0.52% | 37.38%±0.18%        | 38.57%±0.25%        | 36.04%±0.69%        | 36.91%±1.36% | <b>41.36%±0.57%</b> | <b>46.32%±0.46%</b> |
|             | 30     | 39.07%±0.31% | 39.54%±0.65% | 38.35%±0.15%        | 38.96%±0.07%        | 38.24%±0.73%        | 40.79%±0.89% | <b>44.48%±0.67%</b> | <b>49.53%±0.57%</b> |
|             | 40     | 40.16%±0.58% | 41.79%±1.22% | 38.42%±0.05%        | 39.27%±0.11%        | 39.89%±0.59%        | 41.82%±1.86% | <b>44.87%±0.45%</b> | <b>51.02%±0.89%</b> |
| Facebook    | 5      | 19.33%±0.78% | 10.54%±1.03% | 14.00%±0.71%        | 21.70%±0.35%        | 14.17%±2.91%        | 9.11%±2.54%  | <b>29.72%±0.85%</b> | <b>23.09%±0.59%</b> |
|             | 10     | 30.15%±0.69% | 25.14%±0.78% | 35.97%±0.63%        | <b>35.52%±0.40%</b> | 18.96%±3.46%        | 13.06%±2.21% | <b>36.26%±1.77%</b> | 34.03%±0.87%        |
|             | 20     | 36.77%±0.46% | 34.93%±0.54% | <b>42.92%±0.57%</b> | <b>41.66%±0.25%</b> | 25.54%±2.55%        | 19.82%±3.15% | 42.17%±0.60%        | 41.23%±1.57%        |
|             | 30     | 39.03%±1.03% | 37.64%±0.79% | <b>45.94%±0.74%</b> | 45.06%±0.37%        | 30.06%±2.17%        | 26.63%±3.83% | 45.07%±0.32%        | <b>45.50%±0.66%</b> |
|             | 40     | 47.49%±0.82% | 45.58%±1.21% | <b>51.16%±0.91%</b> | <b>51.79%±1.02%</b> | 33.60%±2.64%        | 30.65%±2.61% | 50.15%±0.41%        | 50.28%±1.14%        |
| Blogcatalog | 5      | 27.55%±0.83% | 14.00%±1.27% | 27.31%±0.45%        | 15.01%±1.03%        | <b>33.89%±2.53%</b> | 24.59%±3.10% | 30.62%±2.37%        | <b>25.13%±2.48%</b> |
|             | 10     | 27.71%±1.27% | 14.81%±1.54% | 27.18%±0.26%        | 15.05%±0.54%        | 34.34%±2.56%        | 24.85%±2.53% | <b>38.72%±0.52%</b> | <b>27.03%±0.94%</b> |
|             | 20     | 28.94%±0.80% | 13.07%±0.71% | 28.57%±0.57%        | 17.81%±0.61%        | 39.96%±0.81%        | 30.72%±1.34% | <b>42.85%±0.16%</b> | <b>33.69%±1.35%</b> |
|             | 30     | 28.95%±0.58% | 13.17%±1.36% | 29.23%±0.34%        | 18.73%±0.64%        | 41.43%±1.53%        | 29.09%±1.46% | <b>44.22%±0.16%</b> | <b>36.22%±0.72%</b> |
|             | 40     | 29.96%±0.53% | 15.51%±0.69% | 30.15%±0.41%        | 19.03%±0.53%        | 42.04%±1.12%        | 31.08%±0.38% | <b>44.31%±0.27%</b> | <b>36.51%±0.79%</b> |

方法,能够使相邻节点的表示彼此接近,不相邻节点的表示彼此远离,使判别性信息在表示结果中得以保持;另一方面,DCSNRL模型能够在半监督通道中将有限的监督信息用于指导表示学习过程,确保相同类别的节点具有相似的网络表示。在自监督信号和标签信息的共同作用下,节点表示的有效性得到了提升。DGI和GMI算法通过构造自监督信号在一定程度上提高了模型捕捉网络特性的能力,获得的聚类效果相比于DeepWalk和Node2vec这类无监督模型有显著提升,但只依赖自监督信号的可靠性不足,导致其聚类效果相比于DCSNRL模型仍有差距。GraphMLP、GWNN和GCN算法在学习网络表示的过程中能够使标签相同的节点有相似表示,但对节点标签有较大程度的依赖,当标签信息较少时,学习得到的表示对簇内相似性和簇间差异性的刻画不足,导致聚类效果的有效性受到影响。

### 3.3.3 消融实验

为了验证DCSNRL模型中冗余消除模块的有效性,在Cora和Citeseer数据集上进行了节点分类和聚类任务的消融实验。将去除冗余消除模块后的模型记作DCSNRL\_1,在不同节点标记数量下与DCSNRL进行比较,实验结果见表6。从表6可知,DCSNRL在2个数据集上都获得了优于DCSNRL\_1的分类和聚类结果。这表明冗余消除模块对两个通道间存在的信息冗余进行识别和剔除,能够防止表示学习模型陷入局部最优,进而提升下游任务表现。

### 3.3.4 可视化测试

将学习得到的网络表示进行可视化展示,有助于对不同算法的表示学习能力进

表6 不同节点标记数量下消融实验结果

| 数据集      | 评价指标     | 模型                  | 节点标记数量              |                     |                     |                     |                     |
|----------|----------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|          |          |                     | 5                   | 10                  | 20                  | 30                  | 40                  |
| Cora     | Micro-F1 | DCSNRL_1            | 75.03%±1.12%        | 76.12%±0.78%        | 81.93%±0.83%        | 82.87%±0.69%        | 83.21%±0.84%        |
|          |          | DCSNRL              | <b>76.38%±0.83%</b> | <b>77.61%±0.46%</b> | <b>83.57%±0.34%</b> | <b>84.63%±0.23%</b> | <b>84.98%±0.36%</b> |
|          | Macro-F1 | DCSNRL_1            | 73.03%±1.30%        | 74.82%±0.93%        | 80.72%±0.81%        | 82.02%±0.53%        | 82.55%±0.71%        |
|          |          | DCSNRL              | <b>74.69%±0.92%</b> | <b>76.45%±0.54%</b> | <b>82.27%±0.37%</b> | <b>83.52%±0.24%</b> | <b>83.84%±0.37%</b> |
|          | NMI      | DCSNRL_1            | 54.67%±0.85%        | 55.43%±0.76%        | 60.19%±0.69%        | 61.12%±0.77%        | 62.35%±0.91%        |
|          |          | DCSNRL              | <b>57.50%±0.47%</b> | <b>58.04%±0.66%</b> | <b>62.65%±0.26%</b> | <b>63.57%±0.35%</b> | <b>65.26%±0.44%</b> |
| ARI      | DCSNRL_1 | 52.59%±1.27%        | 53.48%±1.03%        | 59.76%±0.85%        | 61.23%±0.91%        | 62.98%±0.82%        |                     |
|          | DCSNRL   | <b>57.00%±0.90%</b> | <b>58.09%±0.71%</b> | <b>62.99%±0.69%</b> | <b>64.48%±0.56%</b> | <b>66.15%±0.98%</b> |                     |
| Citeseer | Micro-F1 | DCSNRL_1            | 64.46%±1.71%        | 67.30%±1.14%        | 70.22%±0.89%        | 71.65%±0.77%        | 72.00%±0.86%        |
|          |          | DCSNRL              | <b>65.88%±1.69%</b> | <b>68.73%±0.54%</b> | <b>73.12%±0.59%</b> | <b>73.87%±0.48%</b> | <b>73.91%±0.41%</b> |
|          | Macro-F1 | DCSNRL_1            | 60.65%±1.41%        | 63.23%±0.91%        | 66.34%±0.79%        | 67.59%±0.87%        | 68.11%±0.92%        |
|          |          | DCSNRL              | <b>61.16%±1.39%</b> | <b>64.91%±0.46%</b> | <b>68.73%±0.59%</b> | <b>69.85%±0.59%</b> | <b>70.04%±0.45%</b> |
|          | NMI      | DCSNRL_1            | 36.08%±1.06%        | 39.37%±0.88%        | 41.93%±1.12%        | 42.41%±1.31%        | 43.15%±0.94%        |
|          |          | DCSNRL              | <b>38.77%±0.92%</b> | <b>41.98%±0.48%</b> | <b>43.41%±0.26%</b> | <b>44.65%±0.99%</b> | <b>44.76%±0.16%</b> |
| ARI      | DCSNRL_1 | 35.67%±1.45%        | 39.21%±1.23%        | 40.89%±1.02%        | 43.00%±1.59%        | 43.74%±0.87%        |                     |
|          | DCSNRL   | <b>38.50%±0.83%</b> | <b>41.97%±0.78%</b> | <b>43.78%±0.28%</b> | <b>45.62%±1.71%</b> | <b>45.70%±0.36%</b> |                     |

行直观比较。为此,在Cora和Citeseer数据集上,使用t-SNE工具将节点类别标记数量为20时获得的网络表示映射到二维空间中,以便直观展示输入网络的群落结构,节点类别通过不同颜色进行区分。

表7展示了DeepWalk、GMI、GCN以及DCSNRL等算法在Cora和Citeseer数据集上的可视化结果。与其他方法相比,DCSNRL模型获得的节点表示能够更有效地反映网络的内在结构。例如,在Cora数据集上,DCSNRL模型的节点表示结果聚集为7个较为清晰的簇团。可视化结果表明,DCSNRL模型在学习低维表示时能对网络中的结构信息和关键特性进行有效捕捉和保留。这得益于双通道模式下节点标签和自监督信号的协作为表示学习过程提供了更丰富的判别性信息。

### 3.3.5 参数敏感性分析

为了探究参数 $\theta$ 对模型性能的影响,在不同数据集上测试节点分类任务,设定节点类别标记数量为20,分类结果的Micro-F1值和Macro-F1值如图3所示。Micro-F1值和Macro-F1值随着 $\theta$ 的增大呈现出一定程度的提升;当 $\theta$ 值达到一定水平后,分类结果维持在一定水平。这是因为



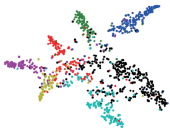
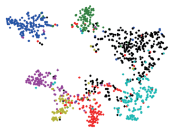



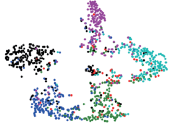
当 $\theta$ 值较小时,节点标签作为表示学习过程的主要监督信息来源;随着 $\theta$ 值的增加,补充了的自监督信号越来越多,节点分类结果得到了改善;当自监督信号达到相对充足的状态时,节点分类结果也趋于稳定。

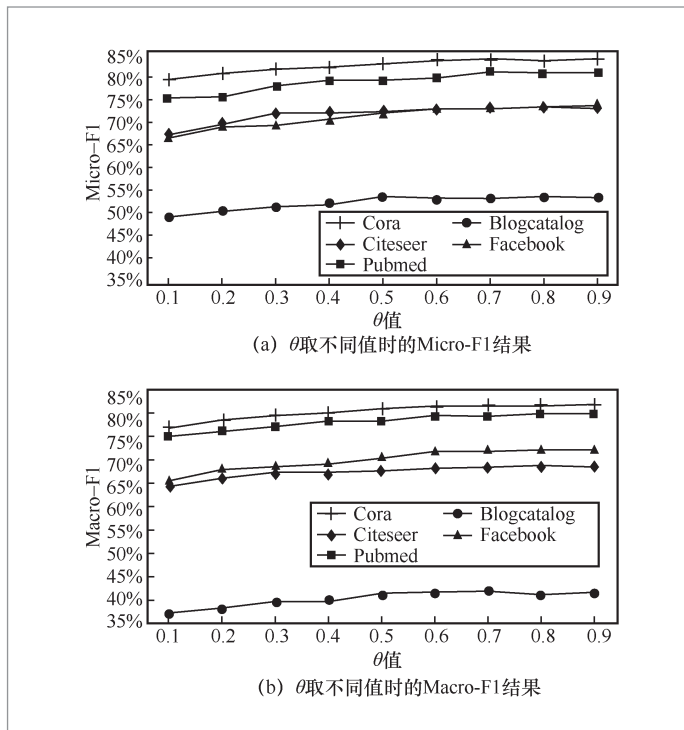
## 4 结束语

本文提出了一种双通道半监督网络表示学习模型DCSNRL。该模型在两个信息传递通道下分别利用节点标签和自监督信号形成对网络表示学习过程的互补性指导,使学习得到的网络表示更好地捕获和保留网络的关键结构和特性。此外,设计了冗余消除策略,用于克服两个通道间信息冗余导致的不利影响。DCSNRL改善了在监督信息有效性不足场景下的模型性能,提高了网络表示结果在下游任务中的有效性。在真实网络数据集上,通过节点分类、聚类以及可视化测试等任务验证了该模型的有效性和优越性。

本文为半监督网络表示学习的研究提供了一种新的思路和视角,在未来的工作中,将围绕以下方面开展研究:一是探索DCSNRL模型在多标记和偏标记等复杂场景中的应用;二是优化模型执行机制和

表7 Cora和Citeseer数据集上的可视化结果

| 数据集      | DeepWalk  | GMI   | GCN  | DCSNRL  |
|----------|---|---|--|---|
| Cora     |  |  |  |  |
| Citeseer |  |  |  |  |

图3 参数 $\theta$ 的影响分析

训练策略,使其适应超大规模网络的效率要求;三是探索网络中不同成分在网络形成和演化过程中的互作用关系及其在表示学习过程中的识别和保留。

## 参考文献:

- [1] 涂存超, 杨成, 刘知远, 等. 网络表示学习综述[J]. 中国科学: 信息科学, 2017, 47(8): 980-996.  
TU C C, YANG C, LIU Z Y, et al. Network representation learning: an overview[J]. Scientia Sinica (Informationis), 2017, 47(8): 980-996.
- [2] 戴筠. 基于双曲空间图嵌入的科研热点预测[J]. 大数据, 2022, 8(6): 94-104.  
DAI J. Emerging scientific topic prediction based on Poincare graph embedding[J]. Big Data Research, 2022, 8(6): 94-104.
- [3] CAO S S, LU W, XU Q K. GraRep: learning graph representations with global structural

information[C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. New York: ACM, 2015: 891-900.

- [4] PEROZZI B, AL-RFOU R, SKIENA S. DeepWalk: online learning of social representations[C]//Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. New York: ACM, 2014: 701-710.
- [5] WANG D X, CUI P, ZHU W W. Structural deep network embedding[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, 2016: 1225-1234.
- [6] TU C C, ZHANG W C, LIU Z Y, et al. Max-margin deepwalk: discriminative learning of network representation[C]// Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence. New York: ACM, 2016: 3889-3895.
- [7] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[EB]. arXiv preprint, 2016, arXiv: 1609.02907.
- [8] YANG Z L, COHEN W W, SALAKHUTDINOV R. Revisiting semi-supervised learning with graph embeddings[C]//Proceedings of the 33rd International Conference on International Conference on Machine Learning. New York: ACM, 2016: 40-48.
- [9] FAN S H, WANG X, SHI C, et al. Debaised graph neural networks with agnostic label selection bias[J]. IEEE Transactions on Neural Networks and Learning Systems, 2022, 1-12.
- [10] ZHANG H G, WANG Z S, LIU D R. A comprehensive review of stability analysis of continuous-time recurrent neural networks[J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25(7): 1229-1262.
- [11] CHEN L Y, LI S B, BAI Q, et al. Review of image classification algorithms based on convolutional neural networks[J]. Remote Sensing, 2021, 13(22): 4712.
- [12] ZHANG G J, LIU Y, JIN X N. A survey

- of autoencoder-based recommender systems[J]. *Frontiers of Computer Science*, 2020, 14(2): 430–450.
- [13] RUIZ L, GAMA F, RIBEIRO A. Gated graph recurrent neural networks[J]. *IEEE Transactions on Signal Processing*, 2020, 68: 6303–6318.
- [14] GAO H Y, WANG Z Y, JI S W. Large-scale learnable graph convolutional networks[C]//*Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. New York: ACM, 2018: 1416–1424.
- [15] WENG Z Q, ZHANG W Y, DOU W. Adversarial attention-based variational graph autoencoder[J]. *IEEE Access*, 2020, 8: 152637–152645.
- [16] CHENG D W, WANG X Y, ZHANG Y, et al. Graph neural network for fraud detection via spatial-temporal attention[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2022, 34(8): 3800–3813.
- [17] ZHOU J, CUI G Q, HU S D, et al. Graph neural networks: a review of methods and applications[J]. *AI Open*, 2020, 1: 57–81.
- [18] WU Z H, PAN S R, CHEN F W, et al. A comprehensive survey on graph neural networks[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, 32(1): 4–24.
- [19] WU L R, LIN H T, TAN C, et al. Self-supervised learning on graphs: contrastive, generative, or predictive[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(4): 4216–4235.
- [20] PENG Z, HUANG W B, LUO M N, et al. Graph representation learning via graphical mutual information maximization[C]//*Proceedings of The Web Conference 2020*. New York: ACM, 2020: 259–270.
- [21] VELIČKOVIĆ P, FEDUS W, HAMILTON W L, et al. Deep graph infomax[C]//*Proceedings of the 7th International Conference on Learning Representations*. arXiv preprint, 2019, arXiv: 1809.10341.
- [22] SUN F, HOFFMANN J, VERMA V, et al. InfoGraph: unsupervised and semi-supervised graph-level representation learning via mutual information maximization[EB]. arXiv preprint, 2019, arXiv: 1908.01000.
- [23] BELGHAZI M I, BARATIN A, RAJESWAR S, et al. MINE: mutual information neural estimation[EB]. arXiv preprint, 2018, arXiv: 1801.04062.
- [24] HJELM R D, FEDOROV A, LAVOIE-MARCHILDON S, et al. Learning deep representations by mutual information estimation and maximization[EB]. arXiv preprint, 2018, arXiv: 1808.06670.
- [25] QIN Z, KIM D, GEDEON T. Rethinking softmax with cross-entropy: neural network classifier as mutual information estimator[EB]. arXiv preprint, 2018, arXiv: 1911.10688.
- [26] HU Y, YOU H X, WANG Z C, et al. Graph-MLP: node classification without message passing in graph[EB]. arXiv preprint, 2021, arXiv: 2106.04051.
- [27] XU B B, SHEN H W, CAO Q, et al. Graph wavelet neural network[EB]. arXiv preprint, 2019, arXiv:1904.07785.
- [28] KANG Z, LIN Z P, ZHU X F, et al. Structured graph learning for scalable subspace clustering: from single view to multiview[J]. *IEEE Transactions on Cybernetics*, 2022, 52(9): 8976–8986.

## 作者简介



杜航原 (1985- ), 男, 博士, 山西大学计算机与信息技术学院副教授, 主要研究方向为图机器学习理论及应用。



谢富中(1998- ),男,山西大学计算机与信息技术学院硕士生,主要研究方向为机器学习、网络数据挖掘。



王文剑(1968- ),女,博士,山西大学计算机与信息技术学院教授、院长,主要研究方向为机器学习、数据挖掘、计算智能。



白亮(1982- ),男,博士,山西大学智能信息处理研究所教授、所长,主要研究方向为机器学习与数据挖掘。

收稿日期: 2023-07-17

基金项目: 国家自然科学基金项目(No.U21A20513, No.62076154, No.62276159, No.62276161), 山西省重点研发计划项目(No.202302010101007, No.202202020101003), 山西省基础研究计划项目(No.202303021221055)

**Foundation Items:** The National Natural Science Foundation of China(No.U21A20513, No.62076154, No.62276159, No.62276161), The Key R&D Program of Shanxi Province(No. 202302010101007, No.202202020101003), The Fundamental Research Program of Shanxi Province(No.202303021221055)