

# 基于重要性采样的超图网络高效表示方法<sup>\*</sup>

邵豪, 王伦文, 朱然刚, 刘辉



(国防科技大学 电子对抗学院, 安徽 合肥 230037)

通信作者: 王伦文, E-mail: eei\_wanglunwen@163.com

**摘要:** 现有的超图网络表示方法需要分析全批量节点和超边以实现跨层递归扩展邻域, 这会带来巨大的计算开销, 且因过度扩展导致更低的泛化精度. 为解决这一问题, 提出一种基于重要性采样的超图表示方法. 首先, 它将节点和超边看作是两组符合特定概率测度的独立同分布样本, 用积分形式解释超图的结构特征交互; 其次, 设计带可学习参数的邻域重要性采样规则, 根据节点和超边的物理关系和特征计算采样概率, 逐层递归采集固定数目的对象, 构造一个更小的采样邻接矩阵; 最终, 利用蒙特卡洛方法近似估计整个超图的空间特征. 此外, 借鉴 PINN 的优势, 将需要缩减的方差作为物理约束加入到超图神经网络中, 以获取更具泛化能力的采样规则. 多个数据集上的广泛实验表明, 所提出的方法能够获得更准确的超图表示结果, 同时具有更快的收敛速度.

**关键词:** 复杂网络; 超图表示学习; 重要性采样; 蒙特卡洛估计; 物理信息神经网络

**中图法分类号:** TP18

中文引用格式: 邵豪, 王伦文, 朱然刚, 刘辉. 基于重要性采样的超图网络高效表示方法. 软件学报. <http://www.jos.org.cn/1000-9825/6970.htm>

英文引用格式: Shao H, Wang LW, Zhu RG, Liu H. Importance Sampling Based Efficient Representation for Hypergraph Networks. Ruan Jian Xue Bao/Journal of Software (in Chinese). <http://www.jos.org.cn/1000-9825/6970.htm>

## Importance Sampling Based Efficient Representation for Hypergraph Networks

SHAO Hao, WANG Lun-Wen, ZHU Ran-Gang, LIU Hui

(College of Electronic Engineering, National University of Defense Technology, Hefei 230037, China)

**Abstract:** Existing hypergraph network representation methods need to analyze the full batch nodes and hyperedges to recursively extend the neighbors across layers, which brings huge computational costs and leads to lower generalization accuracy due to over-expansion. To solve this problem, this study proposes a hypergraph network representation method based on importance sampling. First, the method treats nodes and hyperedges as two sets of independent identically distributed samples that satisfy specific probability measures and interprets the structural feature interactions of the hypergraph in an integral form. Second, it designs a neighbor importance sampling rule with learnable parameters and calculates sampling probabilities based on the physical relations and features of nodes and hyperedges. A fixed number of objects are recursively acquired layer by layer to construct a smaller sampled adjacency matrix. Finally, the spatial features of the entire hypergraph are approximated using Monte Carlo methods. In addition, with the advantage of physically informed neural networks, the sampling variance that needs to be reduced is added to the hypergraph neural network as a physical constraint to obtain sampling rules with better generalization capability. Extensive experiments on multiple datasets show that the method proposed in this study can obtain more accurate hypergraph representation results with a faster convergence rate.

**Key words:** complex network; hypergraph representation learning; importance sampling; Monte Carlo estimation; physical information neural network

## 1 引言

随着网络科学的发展, 现实世界中的复杂系统可以表示成由边和节点组成的网络. 目前, 这些复杂交互的对象

\* 基金项目: 国家自然科学基金 (61602491); 安徽省自然科学基金 (2008085QF326)

收稿时间: 2023-01-19; 修改时间: 2023-03-14; 采用时间: 2023-04-27; jos 在线出版时间: 2023-10-25

大多被简化为成对网络,即每条链路仅连接一对节点.然而,现实很多场景中还存在大量一般简单图难以表达的非成对关系.例如,在学术网络中,一篇论文的合作者数并不固定<sup>[1]</sup>;在物质反应中,代表一组可以反应的化合物的关系本质上也是高阶的<sup>[2]</sup>.超图提供了一种建模这种复杂关系的自然方式,其一条边内包含任意个数的节点,能够有效捕捉对象间复杂的高阶交互,得到了非常广泛的应用<sup>[3]</sup>.

网络表示学习是网络数据挖掘的一个重要研究领域<sup>[4]</sup>,其目的是根据节点在网络中的结构特征去构造一个低维、稠密的向量.网络表示学习可以用在多种网络分析任务中,例如时序分析<sup>[5]</sup>、链路预测<sup>[6]</sup>、社团分析<sup>[7]</sup>等.常见的超图表示学习方法大多借鉴了成对网络的表示学习方法.成对网络的表示学习主要从空间维度和图谱维度两个角度出发,前者方法采集每个节点相邻的邻居,并聚合邻居节点的特征,后者方法借助图的拉普拉斯矩阵的特征向量来研究图的性质.在此基础上,研究人员提出了许多超图网络表示学习方法.一些分析超图的典型方法是将其展开为成对网络,例如,团式展开把每个超边展开为节点两两相连形成的全连接结构<sup>[8]</sup>,星式展开为每条超边引入一个连接所有原始节点的虚拟节点<sup>[9]</sup>,再利用成对网络的表示学习方法实现超图的表示学习.但是这些基于展开的方法不可避免地会丢失超图中部分重要的高阶特征.为保证超图特征的完整性,多种适用于超图的拉普拉斯矩阵被设计出来,以实现超图特征的傅里叶变换与超图卷积.Saha等人<sup>[10]</sup>设计了超图拉普拉斯矩阵的推广,计算谱半径边界并保留超图数据的几何概率分布.Javidian等人<sup>[11]</sup>把LWF chain graphs引入到超图的概率矩阵中,提出基于因果解释的超图因子分解方法.这些非展开式的超图分析方法直接构建超图的特征矩阵,拥有较强的可解释性,但灵活性很差.受图卷积神经网络的启发,Feng等人<sup>[12]</sup>设计了一个新颖的超图拉普拉斯矩阵,在使用二阶切比雪夫不等式近似求得超图卷积结果的基础上,提出了超图神经网络(hypergraph neural networks, HGNN).但是, HGNN依然存在以下缺陷,跨层递归扩展邻域以聚合远端节点的特征,这会带来巨大的计算开销.特别是对于较密集的超图而言,单个节点的邻域扩展在很短时间内也可能会占据大部分超图.这样的全批量操作限制了其在大规模超图上的广泛应用.

为解决这一问题,本文将FastGCN<sup>[13]</sup>的层采样思想扩展到超图中,提出基于重要性采样的超图表示方法(importance sampling-based efficient representation for hypergraph networks, IERHN).它首先将超图中的节点和超边分别看作是两组符合相应概率测度的独立同分布的样本,将超图卷积操作解释为表示函数在概率度量下的积分变换;然后基于节点和超边的物理关系和特征计算采样概率,采样固定数目的邻域节点和超边,从而限制每一层的样本范围;最后,利用蒙特卡洛估计近似计算超图卷积,利用超图的空间特征完成学习表示.

IERHN与最先进的超图网络表示学习方法一样,基于超图拉普拉斯矩阵实现超图卷积变换,从而有效保持了超图中高阶特征的传递. IERHN改变了现有超图表示学习方法全批量节点的传递过程,逐层采样合适数量的节点和超边,将采样对象的规模控制在一个预先定义的范围之内,避免了邻域特征的过度扩展.本文的主要贡献可概括如下.

(1) 提出了适用于超图网络的层采样模型,设计了带可学习参数的重要性采样规则,逐层获取与原始分布相一致的采样节点与超边,降低了复杂度,加快了模型的训练速度和收敛速度.据我们所知,本文是第1个在超图网络上研究层采样的嵌入方法.

(2) 利用物理信息神经网络(physical information neural networks, PINN)的优势,将缩减方差加入模型需要训练的目标中,模型不仅学习到训练数据样本的分布规律,而且超图网络中的结构特征与物理信息,以进一步获得更优的采样策略.

(3) 通过逐层采集固定数目的节点与超边,避免了邻域特征的过度扩展和传递. IERHN作为归纳式模型,只利用一部分训练节点来构成超图学习模型,不需要整个超图参与训练,从而缓解了传统直推式超图表示模型过度扩展带来的泛化精度下降的问题.

(4) 与最先进的基线方法相比,本文的方法在降低计算开销的同时,能够获得更好的超图表示学习性能.

本文第2节介绍本文所需的基础知识,包括超图表示学习和蒙特卡罗积分.第3节介绍本文构建的基于重要性采样的超图网络高效表示模型.第4节通过对比实验验证了所提模型的有效性和高效性.第5节总结全文.

## 2 相关工作

本节简要介绍超图的定义和常见的超图神经网络算法.

### 2.1 超图定义

如图 1 所示, 超图是一种特殊的复杂网络, 其中一条链路可以包含任意数目节点<sup>[14]</sup>. 在形式上, 一个超图可以表示为  $G = (V, E)$ ,  $V$  是所有节点的集合,  $E$  是所有超边的集合.  $E_\alpha \in E$  表示超图中的一条超边,  $|E_\alpha|$  是超边的长度, 表示  $E_\alpha$  中包含节点的数目. 当  $G$  中所有的超边都满足  $|E_\alpha| = 2$  时, 则  $G$  退化为普通的成对图<sup>[15]</sup>. 超图的连接状态可以由关联矩阵  $H \in \mathbb{R}^{|V| \times |E|}$  表示,  $H$  定义如公式 (1) 所示:

$$H(i, \alpha) = \begin{cases} 1, & i \in E_\alpha \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

此外,  $W \in \mathbb{R}^{|E| \times |E|}$  是超图的权重矩阵, 其元素表示对应超边间的权重. 超图的邻接矩阵  $A \in \mathbb{R}^{|V| \times |V|}$  被定义为  $A = HWH^T$ , 其中,  $A(i, j)$  表示节点  $i$  和  $j$  共享超边的加权数.

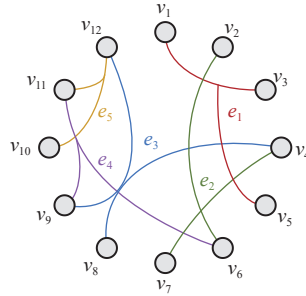


图 1 简单的超图模型

### 2.2 超图表示学习

超图表示<sup>[16]</sup>旨在学习一个从  $d_v$  维节点空间到  $d$  维向量空间的映射函数  $f: V \rightarrow \mathbb{R}^d$ , 将每个节点  $v \in V$  投影成向量  $h_v \in \mathbb{R}^d$ , 且  $d \ll d_v$ . 它将超图节点投影到同一个低维空间中方便进行下游计算, 并保留节点在原始空间的结构特征和物理信息, 通过结构属性可以深入挖掘节点之间的拓扑关系. 此外, 它能用更低的内存开销存储超图结构, 从而为使用并行和分布式计算来处理大规模网络数据提供可能<sup>[17]</sup>.

HGNN 是一种新颖的超图学习表示方法, 在结合超图拉普拉斯矩阵和传统图神经网络的基础上, 利用二阶切比雪夫多项式近似计算超图卷积核. 超图上的卷积运算可表示为:

$$Y = D_v^{-1/2} H W D_e^{-1} H^T D_v^{-1/2} X \theta \quad (2)$$

其中,  $X \in \mathbb{R}^{n \times C_1}$  表示节点的特征矩阵,  $n$  是节点数,  $C_1$  是节点特征的维度,  $W = \text{diag}(w_1, w_2, \dots, w_n)$  表示超边的权重,  $\theta \in \mathbb{R}^{C_1 \times C_2}$  表示模型需要训练的参数矩阵,  $Y \in \mathbb{R}^{n \times C_2}$  表示输出的表示向量. 事实上, HGNN 中的  $D_e^{-1} H^T D_v^{-1/2} X \theta$  表示节点特征映射到超边上,  $D_v^{-1/2} H W$  表示将获得的超边特征重新分配到节点中.

### 2.3 蒙特卡洛积分与重要性采样

蒙特卡罗 (Monte Carlo) 方法是一种利用随机性解决确定性问题的方法<sup>[18]</sup>, 它生成大量的随机样本, 去近似拟合一个系统确定性的值. 蒙特卡洛积分是对矩形法近似计算定积分的一种泛化, 它根据一个特定的概率密度函数 (probability density function, PDF)<sup>[19]</sup> 对被积函数进行采样, 每个采样点近似代表被积函数在各个长为  $1/\text{PDF}$  的区间上的平均值, 其对应的矩形面积和的数学期望看作是定积分的值.

在实际中, 实际概率分布可能比较复杂, 不易直接从实际概率分布进行采样. 重要性采样是蒙特卡洛积分中一种重要的方差缩减策略, 通过引入一个新的容易采样的概率分布进行采样, 如公式 (3) 所示:

$$E_{x \sim p}[f(x)] = \int p(x)f(x)dx = E_{x \sim q}[(p(x)/q(x))f(x)] \approx \frac{1}{N} \sum_{i=1}^N \left( \frac{p(x_i)}{q(x_i)} \right) f(x_i) \quad (3)$$

其中,  $p(x)/q(x)$  是实际概率分布和提议分布的比值, 被称为重要性权重<sup>[20]</sup>. 当  $p(x) = q(x)$  时, 蒙特卡洛积分的方差达到最小值.

为了使估计值方差更小, 重要性采样需要使用一个与被积函数尽可能接近的 PDF 计算蒙特卡洛积分<sup>[21]</sup>. 在被积函数的函数值较小的区间上, 采样频率较低, 每个采样点代表的区间范围较大; 而在被积函数的函数值较大的区间上, 采样频率较高, 每个采样点代表的区间范围较小<sup>[22]</sup>.

### 3 基于重要性采样的超图网络高效表示方法

IERHN 本质上是一个基于层采样的超图表示学习模型. 图 2 展示了 IERHN 的整体框架. IERHN 的每一层在统一的重要性采样框架下处理节点表示. 为了最小化采样产生的方差, 还利用 PINN 显式训练方差以获得最优的采样策略. 以下介绍 IERHN 的组成部分和采样思想.

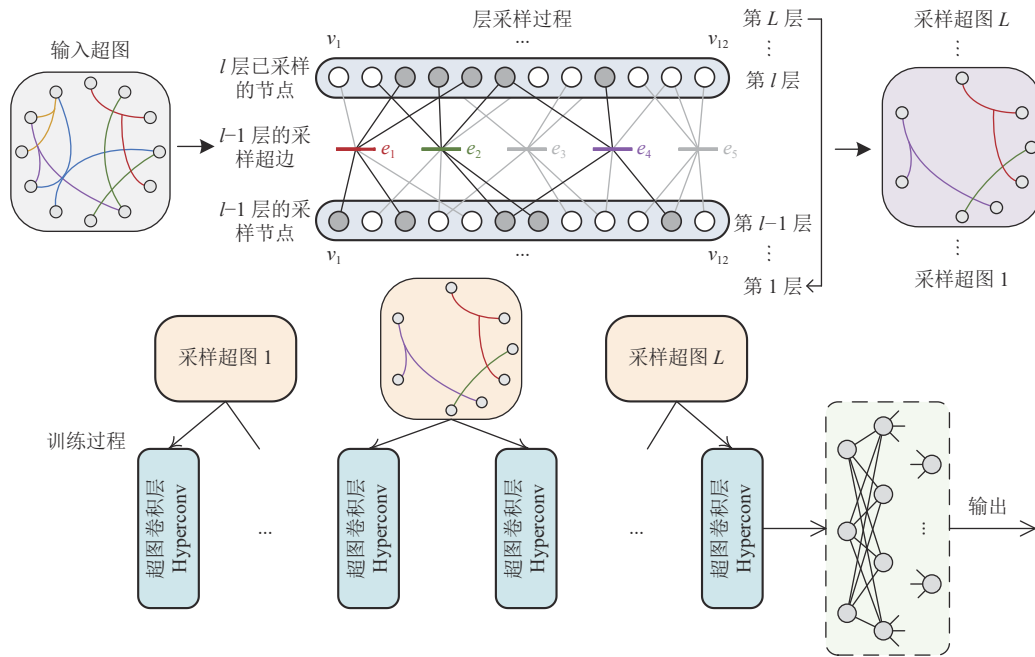


图 2 IERHN 的整体框架

#### 3.1 超图卷积的积分形式

与 FastGCN 类似, 一个超图  $G$  被看作是一个为无限大超图  $G'$  的子图.  $G'$  定义了一个概率空间  $(V', E', P, Q)$ ,  $V'$  表示无限个服从  $P$  概率分布的节点组成的集合,  $E'$  表示无限条服从  $Q$  概率分布的超边组成的集合. 超图  $G$  中的  $V$  和  $E$  被分别被看作是  $V', E'$  的子集, 也就是说, 超图的节点和超边分别被解释为符合特定概率测度的对象. 将节点特征和超图卷积运算推广为与节点表示函数相关的积分, 利用蒙特卡罗方法计算此积分从而获得一个近似的超图卷积值. 在传统的超图表示学习模型中, 特征向量在层间的传递如公式 (4) 所示:

$$X^{(l+1)} = D_v^{-1/2} H W D_e^{-1} H^T D_v^{-1/2} X^{(l)} \theta^{(l)} \quad (4)$$

其中,  $X^{(l)}$  表示在第  $l$  层节点的特征矩阵,  $X^{(0)}$  是节点的初始特征向量.  $l$  值的范围表示模型的深度, 也是超图中节点结构特征交互的距离.

现在, 将超图中的节点和超边看作是符合相应概率测度的样本, 节点特征向量的传递可以用超图测度空间上

的积分进行表示. 值得注意的是, 这里的积分并不是严格意义上的黎曼积分, 只是借助了它的表达形式. 因此, 公式 (4) 可以由积分的形式表示为:

$$X^{(l+1)}(v, :) = \iint D_v^{-1/2} H(v, e) W D_e^{-1} H^T(e, u) D_v^{-1/2} X^{(l)}(u, :) \theta^{(l)} dP(u) dQ(e) \quad (5)$$

其中,  $u, v$  分别表示连续两层的节点样本, 且  $u, v \sim P$ ;  $e$  是超边样本, 且  $e \sim Q$ . 为简化公式 (5), 设定  $\widehat{H}(v, u, e) = D_v^{-1/2} H(v, e) W D_e^{-1} H^T(e, u) D_v^{-1/2}$ .

### 3.2 超图的重要性采样

根据蒙特卡洛积分, 利用一系列随机变量采样公式 (5) 中的被积函数. 在第  $l$  层中, 采样  $c_l$  个节点,  $k_l$  条超边. 根据强大数定理<sup>[23]</sup>, 独立同分布的随机变量的均值以概率为 1 收敛到分布的均值. 当采样足够多的样本数目时, 公式 (5) 中原始积分的结果近似为:

$$\widetilde{X}^{(l+1)}(v, :) = \frac{n \times e}{c_l \times k_l} \sum_{i=1}^{c_l} \sum_{j=1}^{k_l} \widehat{H}(v, u_i^{(l)}, e_j^{(l)}) \widetilde{X}^{(l)}(u_i^{(l)}, :) \theta^{(l)} \quad (6)$$

其中,  $u_i^{(l)}$  和  $e_j^{(l)}$  分别表示在第  $l$  层中, 节点和超边的采样样本. 经过采样后, 用  $E_v$  表示节点  $v$  的估计值的均值如公式 (7) 所示:

$$E_v = E\left(\widetilde{X}^{(l+1)}(v)\right) = \frac{1}{c_{l+1}} \sum_{m=1}^{c_{l+1}} \left( \frac{1}{c_l \times k_l} \sum_{i=1}^{c_l} \sum_{j=1}^{k_l} \widehat{H}(v_m, u_i^{(l)}, e_j^{(l)}) \widetilde{X}^{(l)}(u_i^{(l)}, \theta^{(l)}) \right) \quad (7)$$

为了探究一个合理的样本采样策略, 以下计算  $E_v$  的方差并缩减方差值. 为更清晰地介绍方差的计算过程, 首先计算节点在原始分布下特征的条件均值 (用  $R(v)$  表示) 和方差:

$$E(x^{(l+1)}(v)|v) = \iint \widehat{H}(v, u, e) x^{(l)}(u) \theta^{(l)} dP(u) dQ(e) = R(v) \quad (8)$$

$$\begin{aligned} D(x^{(l+1)}(v)|v) &= E\left(\left(x^{(l+1)}(v)|v\right)^2\right) - \left(E\left(x^{(l+1)}(v)|v\right)\right)^2 \\ &= \iint \widehat{H}(v, u, e)^2 \left(x^{(l)}(u)\right)^2 \left(\theta^{(l)}\right)^2 dP(u) dQ(e) - R(v)^2 \end{aligned} \quad (9)$$

以下计算  $E_v$  的方差. 根据总方差定律 (law of total variance)<sup>[24]</sup>,  $E_v$  的方差可表示为:

$$D(E_v) = D(E(E_v|v_1, \dots, v_m, \dots, v_{c_{l+1}})) + E(D(E_v|v_1, \dots, v_m, \dots, v_{c_{l+1}})) \quad (10)$$

其中, 第 1 部分表示样本内差异的均值, 第 2 部分表示样本间的差异. 本文认为在一个合理的采样策略后, 得到的采样节点符合原始的概率测度, 即  $v_1, v_2, \dots, v_{c_{l+1}} \sim P$ , 因此可以得到:

$$D(E(E_v|v_1, \dots, v_{c_{l+1}})) = \frac{1}{c_{l+1}} D(R(v)) = \frac{1}{c_{l+1}} \int R(v)^2 dP(v) - \frac{1}{c_{l+1}} \left( \int R(v) dP(v) \right)^2 \quad (11)$$

$$\begin{aligned} E(D(E_v|v_1, \dots, v_m, \dots, v_{c_{l+1}})) &= E\left(\frac{1}{c_{l+1}} \times D(x^{(l+1)}(v)|v)\right) \\ &= \frac{1}{c_{l+1}} \iiint \widehat{H}(v, u, e)^2 \left(x^{(l)}(u)\right)^2 \left(\theta^{(l)}\right)^2 dP(u) dQ(e) dP(v) - \frac{1}{c_{l+1}} \int R(v)^2 dP(v) \end{aligned} \quad (12)$$

$D(E_v)$  的值由公式 (11)、公式 (12) 的和确定. 为了降低  $E_v$  的方差值, 利用  $S(v)$ ,  $D(e)$  作为新的概率测度用来抽取节点和超边样本, 在保持原有样本期望不变的情况下, 改变抽样重心, 使对最终结果贡献更大的抽样样本出现的概率更大, 获得更有效的抽样结果<sup>[25]</sup>. 为了分析一个合理的  $S(v)$ ,  $D(e)$ , 计算在新的概率测度下, 采样节点的均值方差. 以下, 计算新的节点特征均值  $E_{SD}^{(l+1)}$  如公式 (13) 所示:

$$E(x_{SD}^{(l+1)}(v)|v) = \iint \widehat{H}(v, u, e) x^{(l)}(u) \theta^{(l)} \frac{dP(u) dQ(e)}{dS(v) dD(e)} dS(v) dD(e) = E(x^{(l+1)}(v)|v) = R(v) \quad (13)$$

对比公式 (8), 公式 (13),  $E(x_{SD}^{(l+1)}(v)|v) = E(x^{(l+1)}(v)|v) = R(v)$ , 这说明不同的节点和超边采样策略不会改变  $R(v)$ . 因此, 不同采样方法造成的  $D(E_v)$  的变化来源于公式 (12) 中  $\iiint \widehat{H}(v, u, e)^2 \left(x^{(l)}(u)\right)^2 \left(\theta^{(l)}\right)^2 dP(u) dQ(e) dP(v)$  值的改变. 在节点采样测度为  $S(v)$ , 超边采样测度为  $D(e)$  的条件下,  $\iiint \widehat{H}(v, u, e)^2 \left(x^{(l)}(u)\right)^2 \left(\theta^{(l)}\right)^2 dP(u) dQ(e) dP(v)$

被更新如下 (命名为  $T_{SD}$ ):

$$T_{SD} = \iiint \widehat{H}(v, u, e)^2 (x^{(l)}(u))^2 (\theta^{(l)})^2 \frac{dP(u)^2 dQ(e)^2 dP(v)}{dS(u) dD(e)} \quad (14)$$

令  $H_e = D_v^{-1/2} H(v, e) W$ ,  $H_u = D_e^{-1} H^T(e, u) D_v^{-1/2}$ ,  $T_{SD}$  的计算方法更新为公式 (15) 所示:

$$T_{SD} = \iint H_e^2 \frac{dP(v) dQ(e)^2}{dD(e)} \int H_u^2 (x^{(l)}(u))^2 (\theta^{(l)})^2 \frac{dP(u)^2}{dS(u)} \quad (15)$$

在公式 (15) 中,  $T_{SD}$  被划分成两个部分, 分别是节点采样和超边采样. 考虑到节点和超边是互相独立的同分布样本, 以下分别考虑合理的节点和超边采样策略. 在考虑节点采样策略的同时, 假设采集超边是一个全采样的操作, 即所有超边都参与特征交互. 根据重要性采样的性质<sup>[26]</sup>, 一个合理的  $dS(u)$  如公式 (16) 所示:

$$dS(u) = \frac{H_u x^{(l)}(u) \theta^{(l)} dP(u)}{\int H_u x^{(l)}(u) \theta^{(l)} dP(u)} \quad (16)$$

由公式 (16) 可得,  $l+1$  层的节点采样概率与  $H_u$ ,  $x^{(l)}(u)$ ,  $\theta^{(l)}$  有关,  $x^{(l)}(u)$ ,  $\theta^{(l)}$  与节点在  $l$  层的特征有关. 但是在实际中, 这样的采样策略在层采样模型中是不可行的. 这是因为每一层的采样样本是从顶层到底层逐层确定的. 在确实采样节点和超边时, 需要划分一定比例的节点作为一个 batch 的样本并作为输出层, 随后根据采样概率从输出层到输入层逐层采样对象, 这与公式 (16) 中利用前一层的特征确定后一层的采样节点是矛盾的. 为解决这一矛盾, FastGCN 简单地将  $H_u x^{(l)}(u) \theta^{(l)}$  近似成是节点  $u$  在邻接矩阵上的分布, 但是这一方法将所有节点的特征看作是平等的, 忽略节点特征对采样概率的影响, 造成采样准确性的偏差<sup>[27]</sup>. 在本文中, 为了保持节点特征的差异,  $x^{(l)}(u)$  被看作是初始特征  $x(u)$  的一个线性变换, 即  $x^{(l)}(u) \theta^{(l)} \approx x(u) W_x$ . 最终, IERHN 中节点的逐层采样策略为:

$$s(u_i) = \frac{\sum_{j=1}^{|E|} H_u(e_j, u_i) |x(u_i) W_x|}{\sum_{i=1}^{|V|} \sum_{j=1}^{|E|} H_u(e_j, u_i) |x(u_i) W_x|} \quad (17)$$

其中,  $W_x \in \mathbb{R}^{C_l \times 1}$ . 同理, 超边的逐层采样策略如公式 (18) 所示:

$$d(e_j) = \frac{\sum_{u=1}^{|V|} H_e(e_j, u)}{\sum_{j=1}^{|E|} \sum_{u=1}^{|V|} H_e(e_j, u)} \quad (18)$$

本文设计的超图层采样策略有以下优势: (1) 它保留了 FastGCN 中多层共享一个采样策略的长处. 每一层的节点和超边的采样分布只与节点的初始特征与连接矩阵有关, 不需要为不同的层专门设计  $s(u_i)$ ,  $d(e_j)$ . 在每次迭代更新采样策略的参数时, 只需要更新一个通用的向量  $W_x$ , 从而降低采样的复杂性, 提高模型的效率. (2) 通过构造特征映射向量  $W_x$ , 保留了节点特征在采样策略中的影响, 使采样到的节点能更完整地揭示整个超图的特征, 获得更好的采样性能, 从而得到更准确的超图网络表示学习结果.

### 3.3 邻域采样

在前文中, 超图中的节点和超边分别被看成是符合特定概率密度的样本, 超图卷积被解释成为对应概率测度下的嵌入积分. 通过蒙特卡罗方法进行一致地积分估计, 并利用重要性采样缩减估计方差. 这样的层采样策略能够使限制批量训练的规模, 在保持准确性的同时获得训练效率的巨大提升.

但是这样的采样仍会带来以下问题: 对于超图而言, 在不同层中独立采样的节点和超边并不一定是连接的, 甚至可能会产生完全断开的连通分量, 这会导致超图特征无法得到正常传递. 以图 3 为例, 灰色的节点 ( $v_3, v_6, v_{11}$ ) 表示上一层的采样节点, 带颜色的超边 ( $e_2, e_3, e_4$ ) 是此层的采样超边, 带红框的节点 ( $v_4, v_5, v_{11}$ ) 是此层的采样节点. 可以发现, 部分上层的采样节点和采样超边之间是断开的,  $e_3$  无法接收到来自于 ( $v_3, v_6, v_{11}$ ) 的特征; 此外, 部分采样节点和采样超边之间是断开的,  $e_2, e_3, e_4$  的特征无法输入到  $v_5$  中. 这两种情况都会导致超图特征传递路径的断裂, 采样到的节点和超边之间不存在链接关系, 上一层的特征无法传递给下一层的节点或超边中. 这不仅会使影响超图特征的交互, 导致模型在训练时无法收敛, 并且会带来大量无用的计算开销.

为解决这一问题, 本文设计了邻域采样方法, 控制每一层的采样样本限制在上一层采样样本的邻域中. 图 4 展示了基于邻域的采样过程. 在获取上一层采样节点 ( $v_3, v_6, v_{11}$ ) 的基础上, 忽略那些与上层采样节点不连接的超边, 只提取所有包含上层采样节点的超边作为待采样超边, 用星号标识 ( $e_1, e_2, e_4, e_5$ ), 在该集合上进行采样操作, 得到采样超边 ( $e_2, e_4, e_5$ ). 可以看到, 采样得到的超边 ( $e_2, e_4, e_5$ ) 与上层节点 ( $v_3, v_6, v_{11}$ ) 之间存在依赖关系, 从而有效避免特征在传递的过程中因为链接断裂而消失. 由此, 公式 (18) 的超边采样策略  $d(e_j)$  可更新为:

$$d(e_j) = \frac{\gamma \sum_{u=1}^{|V|} H_e(e_j, u_i)}{\sum_{j=1}^{|E|} \gamma \sum_{u=1}^{|V|} H_e(e_j, u_i)}, \text{ if } \gamma = \begin{cases} 0, & e_j \notin \Omega_e \\ 1, & e_j \in \Omega_e \end{cases} \quad (19)$$

其中,  $\Omega_e$  表示待采样超边集合, 包含与上层采样节点有关系的超边.

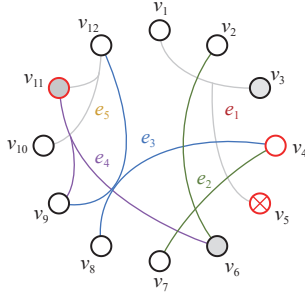


图 3 传统的非邻域采样

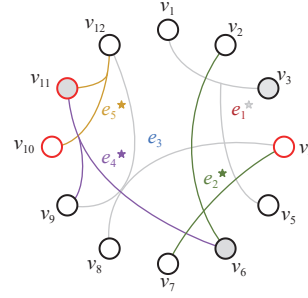


图 4 本文设计的邻域采样

同理, 在获取采样超边的基础上, 提取所有采样超边中的节点作为待采样节点集合, 在该集合上采样节点. 公式 (19) 节点采样策略  $s(u_i)$  可更新为:

$$s(u_i) = \frac{\gamma \sum_{j=1}^{|E|} H_u(e_j, u_i) |x(u_i) W_x|}{\sum_{i=1}^{|V|} \gamma \sum_{j=1}^{|E|} H_u(e_j, u_i) |x(u_i) W_x|}, \text{ if } \gamma = \begin{cases} 0, & e_j \notin \Omega_u \\ 1, & e_j \in \Omega_u \end{cases} \quad (20)$$

其中,  $\Omega_u$  表示待采样节点集合, 包含所有组成  $\Omega_e$  的节点.

在本文中, 基于邻域的采样策略有以下优势: 把采样范围限制在与上一步采样样本的邻域中, 将采样概率分配给那些与样本存在关联的对象. 根据上层中生成的采样节点生成采样超边, 再由采样超边生成此层的采样节点, 并且只聚合互相连接的样本的特征, 从而避免了无效的采样操作, 进一步提高了模型的训练效率和表征能力. 模型完整的层采样过程如算法 1 所示.

**算法 1.** 基于邻域的层重要性采样方法.

输入: 超图  $G = (V, E)$ , 节点特征矩阵  $F$ , 每层节点采样数  $\{s_l^{\text{node}}\}_{l=1}^L$ , 每层超边采样数  $\{s_l^{\text{link}}\}_{l=1}^L$ , 每 batch 的节点  $V_{\text{batch}}$ ; 输出: 采样节点集合  $\{V_l\}_{l=1}^L$ , 采样超边集合  $\{E_l\}_{l=1}^L$ .

1. **for**  $l \in [L, L-1, \dots, 1]$  **then**
2.  $\Omega_e \leftarrow$  the hyperedges connected to  $V_{l+1}$  ( $V_{L+1} = V_{\text{batch}}$ )
3. **for**  $e \in E$  **then**
4. calculate sampling probability  $d(e)$  by Formula (19)
5.  $E_l \leftarrow$  importance sampling hyperedges by  $\{d(e_j)\}_{j=1}^{|E|}$  and  $s_l^{\text{link}}$
6.  $\Omega_u \leftarrow$  the nodes connected to  $E_l^{\text{node}}$
7. **for**  $u \in V$  **then**
8. calculate sampling probability  $s(u)$  by Formula (20)

9.  $V_i \leftarrow$  importance sampling nodes by  $\{s(u_i)\}_{i=1}^{|V|}$  and  $s_i^{\text{node}}$

10. return  $\{V_i\}_{i=1}^L, \{E_i\}_{i=1}^L$

在不同的数据集中, 超图的规模是不同的, 因此针对不同的数据集设计需要不同的 batch 大小. 与其他深度学习模型类似, 如果 batch 过小, 此时训练时间更多, 梯度震荡严重, 不利于模型收敛; 如果 batch 过大, 模型参数容易陷入局部极小值. IERHN 中设定一个 batch 的节点数目是整张超图训练节点数目的 20%. 此时, 模型能够获得令人满意的超图表示学习结果.

一个合理的采样数量对于模型的性能有很大的影响. 如果采样对象的数量过多, 此时节点会获得过多的邻域信息, 这会带来两个问题: (1) 密集超图的表示学习性能可能会因为过平滑而降低. 每个节点使用来自其邻域的信息来更新其嵌入, 更多的采样数目意味着感受野的扩大, 使节点能够获得更丰富的邻居结构特征. 但是, 过多的采样数量会导致节点汇聚到近乎整个连通分量的特征. 此时, 同一个连通分量内节点的隐层表征可能会趋向于同一个值, 造成过平滑的问题. (2) 过多的采样数目会带来巨大的内存开销和计算复杂度. 在 IERHN 模型中, 每一层超边的采样数量为一个 batch 节点数目的两倍, 每一层超边采样数目为此层采样节点连接超边数的 80%. 这样的采样数目是通过网格搜索得到的, 在这样的采样数目下, 模型能同时获得较好的表示学习准确性和较快的训练速度. 在后续的实验中, 将验证设定这一采样数量的合理性.

### 3.4 损失函数

假设在某个 batch 中需要训练一小批节点记为  $V_{\text{batch}}$ , 其对应的标签为  $Z_{\text{batch}}$ . 利用公式 (19) 和公式 (20) 逐层递归采样超边和节点, 直到这个采样过程扩展到输入层. 随后构造一个采样节点和超边组成的邻接矩阵, 传递超图的特征, 在顶层获得  $V_{\text{batch}}$  的节点表示向量  $X_{\text{batch}}$ . 将  $X_{\text{batch}}$  馈入到全连接层中, 输出节点的预测标签  $\tilde{Z}_{\text{batch}}$ . 通过最小化  $Z_{\text{batch}}$  和  $\tilde{Z}_{\text{batch}}$  间的交叉熵损失优化模型参数.

在 IERHN 中, 利用  $W_x$  控制采样策略, 一个合理的  $W_x$  能够最小化采样的方差. 但是获取最优的  $W_x$  无法仅靠节点标签结果. PINN 是一类用于解决有监督学习任务的神经网络<sup>[28]</sup>, 它将物理方程作为限制加入到神经网络之中, 其不仅能够像传统神经网络一样学习到训练数据样本的分布规律, 而且能够学习到数学方程描述的物理定律. 受 PINN 的启发, 如图 5 所示, 本文计算输出层采样节点嵌入的平均方差, 并将其作为损失函数的一部分. 通过最小化节点表示的平均方差, 使模型获得最优的采样策略. 假设 IERHN 的深度为  $L$ , 定义一个 batch 节点在最后一层的采样方差  $std$  为:

$$std = \text{mean} \left( \sum_i s(i) \times \|X_{\Omega_{\text{batch}}}(i) - \text{mean}(X_{\Omega_{\text{batch}}})\|_{L_2}^2 \right) \quad (21)$$

其中,  $X_{\Omega_{\text{batch}}}$  表示一个 batch 中最后一层待采样节点的表示向量,  $s(i)$  表示每个待采样节点的采样概率.

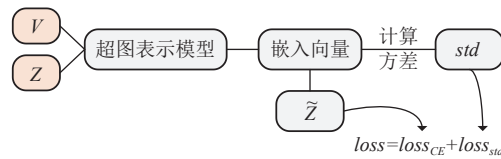


图 5 IERHN 中的 PINN 思想

采样方差  $std$  包含了采样规律的物理特征. 模型中的每一层共享一个相同的采样规则, 最小化  $std$  可以获得最优的采样规则, 它可以扩展到所有层中. 至此, 模型的损失函数为交叉熵损失以及采样方差的加权和:

$$Loss = \text{CrossEntropy}(Z_{\text{batch}}, \tilde{Z}_{\text{batch}}) + \xi std \quad (22)$$

其中,  $\xi$  表示采样方差的权重.  $\xi$  作为超参数, 是在模型学习之前预先设置的参数. 借鉴现有研究工作中混合损失参数设计的思想, 通过网格搜索在  $[0, 1]$  的区间内搜索合适的  $\xi$  值. 当  $\xi = 0.5$  时, 模型在各个数据集上能获得最优



的下游任务性能, 因此设定  $\xi = 0.5$ .

总的说来, 一个好的估计器应该减少由采样过程引起的方差. 因此, IERHN 的目标不仅需要学习训练节点的标签值, 还需要尽可能地缩减估计方差. 在损失函数中引入最小化采样方差, 这会对实验结果带来以下优势: (1) 通过训练显式地最小化方差, 能够自动学习到合理的节点和超边采样策略, 能够在一定数量的采样对象基础上增加超图卷积估计的准确性, 获得更准确的节点特征交互值, 得到优异的超图表示学习结果. (2) 加入最小化采样方差不仅可以利用训练样本的标签来学习超图的特征, 还可以学习到采样方差所表示的物理信息. 与纯数据驱动的超图表示学习模型相比, IERHN 在训练过程中施加了采样方差的约束, 因而能用更少的训练数据学习到更具泛化能力的模型.

综上, 超图表示学习模型 IERHN 如算法 2 所示.

---

**算法 2.** 超图表示学习模型 IERHN (per epoch).

---

输入: 超图  $G = (V, E)$ , 关联矩阵  $H$ , 节点特征矩阵  $F$ ;

输出: 超图表示矩阵  $\Phi$ .

---

初始化: 模型参数  $W$ ,  $X^{(0)} \leftarrow F$

1. **for** each batch **then**
  2.   sample node and hyperedge  $\{V_l\}_{l=1}^L, \{E_l\}_{l=1}^L$  based on Algorithm 1
  3.   calculate sampled adjacency matrix  $H^{(l)}, W^{(l)}$  by  $V_l, E_l$
  4.   **for**  $l \in [1, 2, \dots, L]$  **then**
  5.      $X_{\text{batch}}^{(l+1)} = D_v^{-1/2} H^{(l)} W^{(l)} D_e^{-1} H^{(l)T} D_v^{-1/2} X_{\text{batch}}^{(l)} \theta^{(l)}$
  6.   **end for**
  7.   output labels  $\tilde{Z}_{\text{batch}} \leftarrow X_{\text{batch}}^{(L)}$
  8.    $W \leftarrow W - \eta \nabla \text{Loss}$
  9.   **end for**
  10.  $\Phi \leftarrow X^{(L)}$
  11. **return**  $\Phi$
- 

值得注意的是, 本文提出的 IERHN 模型能够缓解现有超图表示学习泛化精度低的缺陷. 这是因为 IERHN 是归纳式学习. 传统的超图卷积神经网络是直推式学习. 在传统的超图表示学习模型中, 训练数据和测试数据都在同一个超图中. 在模型训练时, 将整个超图作为输入, 仅隐藏了测试节点的标签, 所有节点的特征和连接信息对于直推模型都是可见的. 但在实际中, 超图可能随着新节点的出现不断扩展, 直推模型学习到的隐藏特征难以直接扩展到未知超图中. IERHN 作为归纳式学习模型, 区分开了训练超图和测试超图, 训练和测试在两个独立的超图上进行, 只用一部分训练节点来训练模型, 不需要整张超图参加训练, 从而能很好地推广到有任何扩充的超图中, 从而有效改善传统超图表示学习方法泛化能力弱的缺陷.

### 3.5 复杂度分析

以下分析 IERHN 的复杂度. 为说明 IERHN 在复杂度上的优势, 作为对比, 以下还分析了 HGNN<sup>[12]</sup>, 这是一个被广泛应用的基于全批量操作的超图表示方法. 此外, 为了简化分析, 假设模型中每一层的表示向量的维度是相同的, 都为  $K$ ; IERHN 每一层采样的节点数和超边数都相同, 分别为  $s_v$  以及  $s_e$ .

#### 3.5.1 空间复杂度

IERHN 和 HGNN 的空间复杂度都由两部分组成, 分别是模型权重矩阵和嵌入向量的矩阵. 针对第 1 部分, 对于一个深度为  $L$  的模型, 权重矩阵的空间复杂度为  $O(LK^2)$ . 针对第 2 部分, HGNN 每一层都要保存一个形状为  $|V| \times K$  的嵌入向量矩阵, 因此 HGNN 的总空间复杂度为  $O(LK^2 + L|V|K)$ . 假设在 IERHN 中每一层采样的节点数

为  $s_v$ , 则每一层要保存一个形状为  $s_v \times K$  的嵌入矩阵, 因此 IERHN 的总空间复杂度为  $O(LK^2 + Ls_v K)$ . 值得注意的是,  $s_v \ll |V|$ . 因此, IERHN 的空间复杂度更低.

### 3.5.2 计算复杂度

IERHN 和 HGNN 的计算复杂度主要有 3 个步骤, 是邻接矩阵构造, 特征传递以及权重变换. 邻接矩阵构造表示节点和超边生成一个从节点到节点的链接矩阵, 对应稀疏矩阵的乘法  $\tilde{H} = D_v^{-1/2} H W D_e^{-1} H^T D_v^{-1/2}$ . 这一部分, HGNN 的计算复杂度为  $O(\|H\|_0 |V|)$ ,  $\|*\|_0$  表示矩阵中非零元素的个数; IERHN 只选取采样的节点和超边组成一个新的链接矩阵, 它的计算复杂度为  $O(s_v^2 s_e)$ . 特征传递表示不同层的特征相互交互聚合操作, 对应  $Q = \tilde{H} X^{(l)}$ , 这一部分, HGNN 的计算复杂度为  $O(|V|^2 K)$ , IERHN 的计算复杂度为  $O(s_v^2 K)$ . 权重变换表示嵌入向量与权重矩阵间的线性变换, 对应  $Q\theta^{(l)}$ , 这一部分, HGNN 的计算复杂度为  $O(|V|K^2)$ , IERHN 的计算复杂度为  $O(s_v K^2)$ . 综上, 对于一个深度为  $L$  的模型, HGNN 的计算复杂度为  $O(LH_0 |V| + L|V|^2 K + L|V|K^2)$ , IERHN 的计算复杂度为  $O(Ls_v^2 s_e + Ls_v^2 K + Ls_v K^2)$ . 因为  $s_v \ll |V|$ ,  $s_e \ll |E|$ , IERHN 的计算复杂度更低.

## 3.6 过平滑分析

随着网络模型层数的增加, 相较于现有的超图表示学习模型以及其他采样方法, 本文提出的 IERHN 拥有更好的抗过平滑性能表现. 为了能够更清晰地展示分析过程, 以下从超图神经网络过平滑, IERHN 的抗过平滑能力, IERHN 超图卷积结果的准确性 3 个方面进行阐述. 此外, 还在后续实验中验证了 IERHN 在不同层数下的性能表现与抗过平滑能力.

首先, 分析网络层数较多时传统模型会出现的过平滑问题. 直观来看, 传统的超图神经网络的本质是聚合邻居节点的特征. 对于一个节点而言, 每更新一次节点的特征, 代表节点聚合了更高一阶的邻域信息. 模型的深度意味着节点的特征聚合半径. 随着网络层数的增加, 节点特征聚合半径也会增加. 当达到某一阈值后, 一个节点聚合到的特征会覆盖大部分超图, 此时不同节点的特征会难以区分, 从而导致过平滑问题, 此时模型表示学习能力下降. 因此, 在现有的超图神经网络模型中, 模型的层数大多被设定为 2-4 层, 此时模型有最好的表征能力. 当网络层数增加后, 传统模型的表示学习能力会因为过平滑问题而发生不同程度的下降.

其次, 分析 IERHN 的抗过平滑能力. 当模型深度较深时, IERHN 能够利用层采样的方法, 控制邻域特征的过度扩展. 当传统模型发生过平滑问题时, IERHN 仅采样了大量远端无用节点中的小部分有用节点, 从而避免对目标节点的特征聚合产生干扰, 从而实现了很好的抗过平滑能力, 提升了模型的表示学习性能.

最后, IERHN 能够在训练时仅依靠采样节点和超边就获得准确的超图卷积结果. 这是因为两个原因: (1) 设计了重要性采样策略, 通过假设超图中的节点和超边是独立同分布的样本, 以积分的形式表示超图卷积, 通过蒙特卡罗方法尽可能减小采样方差, 保证采样部分对象就可以获得尽可能准确的超图卷积结果. (2) 根据节点和超边的采样策略, 在不同轮次的 epoch 中, 被采样到的节点和超边不是相同的, 目标节点会聚合不同的邻居特征. 因此, 通过 epoch 的迭代, 目标节点会获得足够的邻域信息, 避免因为层采样而丢失超图的正面特征.

相比于其他的采样方法, 邻域重要性采样方法能够避免采样到无连接的节点而造成特征传递的断裂, 在固定数目的有限采样样本中, 获得更好的特征聚合结果, 得到优异的性能表现.

综上所述, IERHN 能够有较好的抗过平滑能力, 能够在不同层数的模型中都能够保持节点间的特征差异性, 从而拥有优秀的表示学习能力, 鲁棒性更强, 有很好的现实应用场景.

## 4 实验分析

在本节中, IERHN 与多个基线模型在 4 个数据集上进行广泛的对比实验, 以衡量不同超图表示学习模型的性能, 以验证本文模型的优势.

### 4.1 实验数据

本文的实验数据来自于公开的超图数据集, 分别为 ModelNet40、NTU. ModelNet40、NTU 是基于 3D 点云图

像的超图数据集<sup>[12]</sup>, 通过衡量不同对象间的欧几里得距离以生成一个概率图, 获得多模态融合的超图结构. 利用不同的特征提取方法 MVCNN<sup>[29]</sup>和 GVCNN<sup>[30]</sup>预处理原始数据, 得到 4 个新的超图数据集, 分别命名为 MNetM、MNetG、NTUM、NTUG. 4 个超图数据集的参数如表 1 所示.

表 1 4 个公开的超图数据集参数

数据集	$ V $	$ E $	Features	Classes
MNetM	12311	12311	4 096	40
MNetG	12311	12311	2 048	40
NTUM	2 012	2 012	4 096	67
NTUG	2 012	2 012	2 048	67

## 4.2 基线方法

(1) GCN<sup>[31]</sup>. GCN 是一种基于拉普拉斯矩阵来实现谱图卷积的图学习表示方法, 它通过切比雪夫多项式近似捕捉中心节点特定距离内的特征. 在利用 GCN 处理超图网络之前, 利用星式展开和团式展开将超图转化为普通的成对图, 并分别命名为 CEGCN, SEGCN.

(2) Deep hyper-network embedding (DHNE)<sup>[32]</sup>. DHNE 提出超图节点间的一阶相似性和二阶相似性的计算函数, 利用自编码器<sup>[33]</sup>保留节点的局部和全局特征, 从而实现不可分解超图的表示学习.

(3) Hyper2vec<sup>[34]</sup>. Hyper2vec 在 Skip-gram<sup>[35]</sup>的框架下将简单图上带偏置的随机游走方法扩展到超图上. 它定义了从节点到超边和超边到节点的游走规则. 通过得到一个合理的节点序列, 捕获超图的空间特征, 结合 DeepWalk<sup>[36]</sup>获得超图网络的表示学习结果.

(4) HyperGCN<sup>[37]</sup>. HyperGCN 借鉴了 GCN 的优势, 在展开超图的基础上, 利用 BTR 和 Mediator 两种方法筛选二元连边, 从而过滤一部分对超图特征聚合产生负面影响的噪声, 促进有效特征的相互交互.

(5) HGNN<sup>[12]</sup>. HGNN 定义了一个适用于超图网络的拉普拉斯矩阵, 实现了超图上的频域卷积. 它通过节点-超边-节点间的特征转换, 捕获了节点间高阶的数据关系, 获得高阶数据的隐藏层表示.

## 4.3 模型参数

在本文实验中, 模型训练的 epoch 设置为 160, 隐藏层数为 2, 隐藏层维度是 128, 权重衰减是 0.000 5, dropout 是 0.5. 对于一个训练节点数目为  $V_{\text{train}}$  的超图网络, 每一个 batch 的节点数目  $V_{\text{batch}} = 20\% \times V_{\text{train}}$ , 每一层采样的节点记为  $V_l$ , 其数目是  $2V_{\text{batch}}$ , 每一层采样的超边数目是与  $V_l$  相连的超边数目的 80%. 利用 Adam 优化器<sup>[38]</sup>优化模型的参数, 优化器的初始学习率是 0.001, 在 epoch 分别达到 [80, 120] 后, 动态调整学习率为之前学习率的 0.9 倍. 模型运行在 PyTorch 框架下. 此外, 基线模型的设定参数遵循默认值.

## 4.4 实验结果

### 4.4.1 训练过程

IERHN 在 4 个数据集上的训练过程如后文图 6 所示. 由图中可以看到, 在训练的开始阶段, 训练集和验证集的损失函数快速下降, 这说明模型参数得到有效的优化. 当 epoch 大于 80 时, 训练集的损失函数的曲线开始变得平缓, 此时验证集的损失函数也处在较低值, 这表明模型有良好的泛化能力.

### 4.4.2 超图表示学习性能

当损失函数收敛后, 选择最优的验证集结果对应的参数作为模型参数, 并在不同的数据集上利用测试集分析模型的性能. 为分析超图表示学习的性能, 表 2 展示了各个模型的节点分类的指标结果. 此外, 以 1:5 的比例进行负采样, 利用得到的采样节点对和训练得到的节点表示向量构造一个二分类模型进行链路预测, 链路预测的性能指标如表 3 所示. 在表 2 和表 3 中, 粗体数字表示最优的性能.

从表 2 中可以看出, IERHN 在所有网络上都优于 5 种基线方法. 关于节点分类任务, 相比于基线方法, IERHN 在 Macro-F1 指标上分别提升 5.96%, 3.52%, 3.09%, 2.61%, 1.94%, 在 Micro-F1 指标上分别提升 6.20%, 2.57%,

2.99%, 2.46%, 1.78%. 在链路预测任务上, 比于基线方法, IERHN 在 AUC 指标上分别提升 11.15%, 8.6%, 6.3%, 3.8%, 1.38%, 在 ACC 指标上分别提升 12.85%, 10.82%, 7.13%, 5.56%, 1.41%. 总的来看, CEGCN 和和 SEGCN 的表现不尽如人意, 这是因为它们需要将超图网络转换成普通的成对图后进行处理, CEGCN 会忽略两个节点在不同超边中的区别, SEGCN 将连接同一超边节点简单地看作是隐式关联, 这两个方法都会忽略从超图中复杂的高阶特征, 造成表示学习的偏差. DHNE、Hyper2vec 和 HGNN 从不同角度考虑了节点间的高阶特征, 因此性能有所提升. 但是 DHNE 主要适用于均匀超图, 在不均匀超图中的表现不佳. 然而包括实验数据集在内的常见超图网络大多都是非均匀的, 因此性能提升的幅度较为有限. Hper2vec 通过带偏置随机游走获得包含超图特征的节点序列, 但是一个最优的节点跳转策略需要预先分析且很难轻易获得, 因此难以获得最优的表示学习性能. HGNN 利用节点与超边的特征传递获得较好的表示结果. 本文提出的 IERHN 在 HGNN 的基础上, 利用重要性采样获得采样节点和超边, 从实验结果看, 本文利用更少的节点和超边获得更准确的表示学习结果, 这说明 IERHN 利用带可学习参数的采样策略可以收集到有效的超图特征, 并排除负面特征的干扰, 提升泛化能力, 得到最佳的表示学习结果, 从而证明了 IERHN 利用重要性采样完成超图表示学习的有效性.

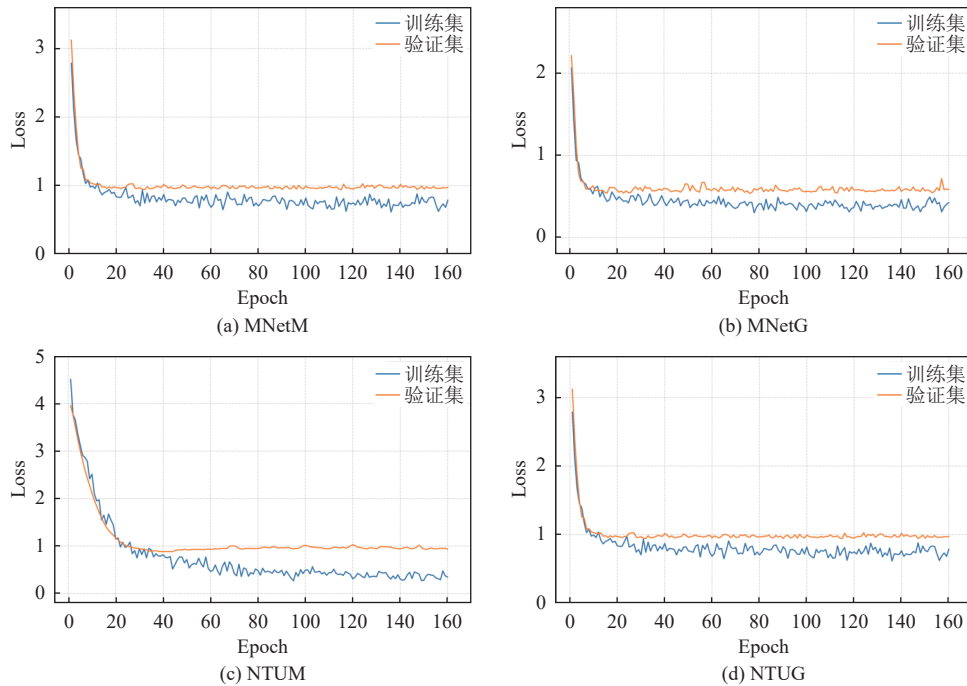


图 6 IERHN 在不同数据集上的损失函数

表 2 不同方法的节点分类性能比较 (%)

Dataset	Metric	CEGCN	SEGCN	DHNE	Hyper2vec	HGNN	IERHN
MNetM	Macro-F1	86.07	87.14	86.93	87.54	87.62	<b>88.73</b>
	Micro-F1	88.73	89.11	89.52	90.55	90.60	<b>91.63</b>
MNetG	Macro-F1	89.68	89.39	89.35	89.54	89.59	<b>91.27</b>
	Micro-F1	90.34	91.05	91.82	92.10	92.18	<b>94.31</b>
NTUM	Macro-F1	66.44	69.89	70.46	70.38	70.23	<b>71.21</b>
	Micro-F1	71.85	75.07	76.68	75.87	76.68	<b>77.48</b>
NTUG	Macro-F1	73.38	80.33	78.67	79.63	81.84	<b>83.94</b>
	Micro-F1	80.69	84.18	82.84	83.91	85.25	<b>87.96</b>

表3 不同方法的链路预测性能比较 (%)

Dataset	Metric	CEGCN	SEGCN	DHNE	Hyper2vec	HGNN	IERHN
MNetM	AUC	83.54	83.69	85.24	86.38	88.91	<b>90.12</b>
	ACC	81.68	82.45	85.06	84.06	89.72	<b>89.93</b>
MNetG	AUC	82.70	85.60	86.98	89.05	90.65	<b>90.78</b>
	ACC	83.76	84.15	87.69	87.43	91.09	<b>92.35</b>
NTUM	AUC	72.25	76.84	78.22	80.80	81.63	<b>82.98</b>
	ACC	69.62	73.56	74.19	78.66	79.77	<b>81.40</b>
NTUG	AUC	75.59	75.30	77.78	79.84	83.14	<b>85.23</b>
	ACC	71.95	72.47	76.46	78.07	81.08	<b>82.79</b>

#### 4.4.3 训练时间对比

为了体现 IERHN 利用重要性的层采样方法在效率上的优势, 图 7 记录了 IERHN 在 4 个数据集上每一个 epoch 的训练时间.

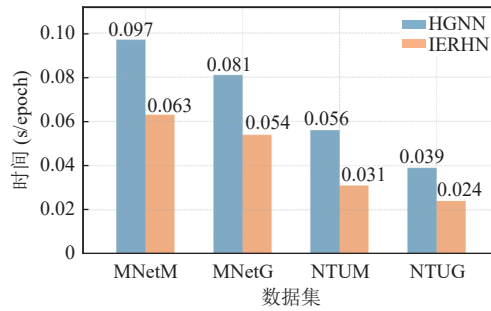


图 7 不同模型的训练时间对比

如图 7 所示, IERHN 在所有数据集上一个 epoch 的训练时间都低于全采样模型, 在 4 个数据集上花费的时间分别降低 35.05%, 33.33%, 44.28% 以及 38.46%. 这说明 IERHN 的层节点和超边采样策略, 能够显著限制递归邻域扩展的范围大小, 将特征传递的对象控制在一定比例的关键节点和超边上. 通过使用采样得到的节点和超边构造一个比原始超图更小的采样邻接矩阵, 可以有效降低计算成本, 大大提高模型的训练效率, 说明利用分层采样方法能够带来 IERHN 在训练时间上的优势.

#### 4.4.4 采样策略对比

IERHN 利用邻域采样控制每一层传递节点和特征的范围. 为了进一步阐述基于层间的邻域采样的优势, 将 IERHN 中的邻域重要性采样分别替换为全域重要性采样 (命名为 FIS)、全域均匀采样 (命名为 FUS)、邻域均匀采样 (命名为 NUS). 在应用不同的采样策略, 重新训练模型进行实验, 以分析不同的采样方法对层采样模型性能的影响. 表 4 记录了不同采样策略获得的超图节点表示向量的性能.

表 4 不同采样策略的节点分类性能对比 (%)

Dataset	Metric	NUS	FIS	FUS	IERHN
MNetM	Macro-F1	87.24	86.72	86.17	<b>88.73</b>
	Micro-F1	90.59	89.18	89.14	<b>91.63</b>
MNetG	Macro-F1	90.67	89.88	89.35	<b>91.27</b>
	Micro-F1	93.63	92.50	91.82	<b>94.31</b>
NTUM	Macro-F1	70.23	69.04	69.31	<b>71.21</b>
	Micro-F1	76.75	75.87	76.14	<b>77.48</b>
NTUG	Macro-F1	81.14	79.17	78.56	<b>83.94</b>
	Micro-F1	85.59	83.38	83.65	<b>87.96</b>

从表 4 中, 采用邻域重要性采样的 IERHN 模型在各个数据集上都获得了最优的性能. 相较于 NUS、FIS、FUS 采样方法, Micro-F1 指标分别提高 1.76%, 3.13%, 3.52%, Micro-F1 指标分别提高 1.37%, 2.96%, 3.00%. 当采样方式变为 FIS、FUS 时, 节点分类性能的下降最为明显, 这表明全域采样对超图表示学习带来的负面影响最大. 这是由于全域采样不可避免地会采样到与上一层对象无关的节点或超边, 导致节点特征传递发生中断, 目标节点无法获取到足够的空间结构特征. NUS 和 IERHN 采用邻域采样策略, 将节点采样的范围控制在上一层对象的邻域中, 从而有效保证不同层间的节点特征的相互传递, 提高网络表示的能力. 此外, FIS、IERHN 的性能分别高于 FUS 和 NIS, 这表明相比于均匀采样, 重要性采样能够有效估计节点原始分布性质和空间特征, 更有利于目标节点聚合尽可能多的全图特征. 本文提出的 IERHN 模型利用邻域重要采样策略采样每一层节点和超边样本, 能够有效促进不同节点的特征交互和空间特征提取, 获得最佳的超图表示结果.

为进一步揭示邻域重要性采样的优势, 图 8 表示了不同的采样策略下, 测试集的准确率随 epoch 值的变化结果. 考虑到图 6 中模型在 epoch 等于 80 时已基本收敛, 因此将图 8 中横坐标的范围限定在 80 以内.

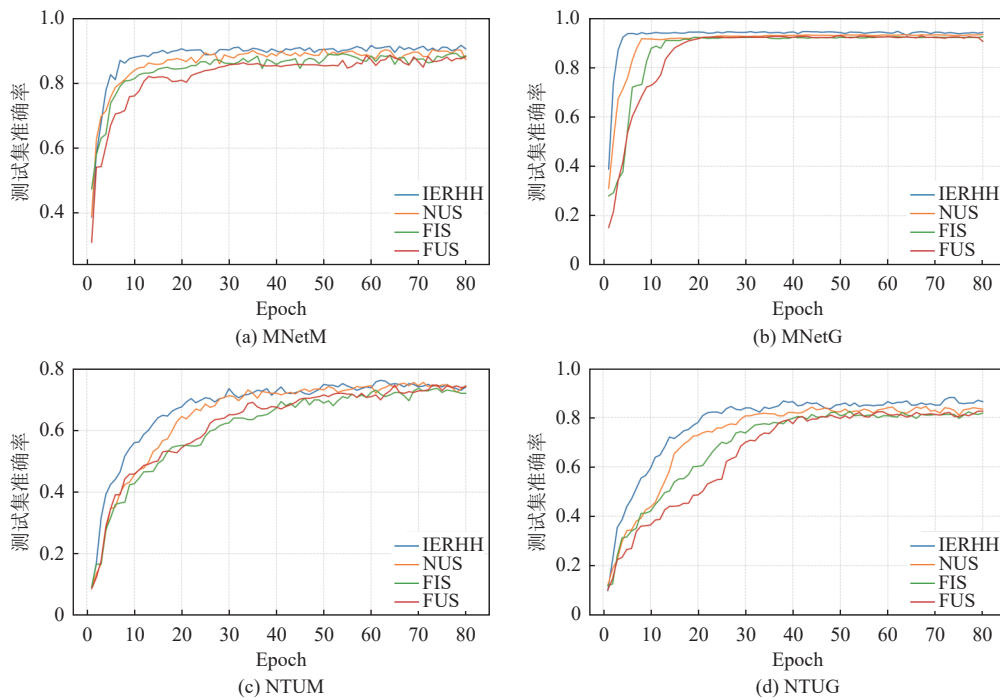


图 8 不同采样策略的测试集准确性曲线

从图 8 中, 初始时各个采样策略对应的测试集准确率都较低, 随着 epoch 的增加, 模型得到有效训练, 准确率开始上升, 最后准确性曲线收敛, 基本保持不变. 当收敛后, IERHN 的准确性比 3 种采样策略都高, 这与表 4 所展示的结果相一致. 值得注意的是, IERHN 在提高表示学习结果准确性的同时, 大大加快了收敛了速度. 可以看到, FUS 的收敛速度最慢, 其次为 FIS, NUS 拥有次优的收敛速度, IERHN 最快收敛. 这进一步说明了 IERHN 利用邻域重要性采样的方法, 不仅避免了不同层间无关特征的传递, 而且加快了有效特征的交互, 能够利用局部少量的节点和超边近似准确估计完整的超图全局空间特征, 证明了 IERHN 利用邻域重要性采样获取超图特征的有效性 & 高效性.

#### 4.4.5 超参数对模型的影响

本节分析 IERHN 中超参数设置的合理性以及对超图表示学习性能的影响. 节点和超边的采样数目会直接影响模型的性能, 在 IERHN 中, 节点采样比例与超边采样比例被设置为 [2, 0.8]. 为验证这一采样数目的优势, 在节点

采样比例为  $[1, 4]$ , 步长为 1 的范围内以及超边采样比例为  $[0.4, 0.8]$ , 步长为 0.2 的范围内进行网格搜索. 表 5 记录了 MNetM 数据集中不同采样节点和采样超边数目时模型的节点分类性能和训练时间. 其中, ACC 表示测试集节点标签被正确预测的比例, Time 表示一个 epoch 的平均训练时间.

表 5 不同采样数目在 MNetM 数据集上的性能对比

Metric	[1,0.4]	[1,0.6]	[1,0.8]	[2,0.4]	[2,0.6]	[2,0.8]	[3,0.4]	[3,0.6]	[3,0.8]	[4,0.4]	[4,0.6]	[4,0.8]
ACC (%)	81.43	85.16	88.14	88.79	90.08	91.63	89.23	91.41	91.59	90.04	90.85	91.42
Time (s)	0.052	0.055	0.056	0.060	0.061	0.063	0.105	0.107	0.108	0.153	0.155	0.161

从表 5 中, 模型的训练时间与采样节点的数目呈正相关. 随着采样节点数目的增加, IERHN 训练时间上升, 这是因为更多的采样节点意味着需要构造一个更大的采样邻接矩阵去实现超图的特征传递, 带来了更高的计算复杂度, 这与第 3.5.2 节中计算复杂度的结果是一致的. 采样超边数目的增加对模型训练时间的影响较小, 我们认为这是因为在层采样中, 由于一条超边上可能包含了多个节点, 因此采样节点数目要大于采样超边数目, 此时采样节点的数目对模型计算复杂度的影响更大. 当采样数目较少时, 目标节点无法获取到足够的邻域信息完成特征聚合. 适当提高节点和超边的采样数目能够提升模型的表示学习准确性. 当采样数目到达一定阈值后, IERHN 的性能提升有限, 甚至因为干扰信息的增加造成表示学习能力的下降. 当节点和超边的采样比例为  $[2, 0.8]$  时, 可以看到 IERHN 在表示学习准确性和训练时间上都有很好的表现, 能够用较少的训练时间获得令人满意的节点分类结果, 说明这一采样数目能够有效平衡表示学习性能和计算开销, 证明了将采样节点数目设置为 batch 大小的两倍, 采样超边比例设置为 80% 是合理的.

以下, 分析损失函数中方差权重  $\xi$  对超图表示学习性能的影响. 图 9 记录了在 4 个数据集下, 不同  $\xi$  对节点分类任务的影响.

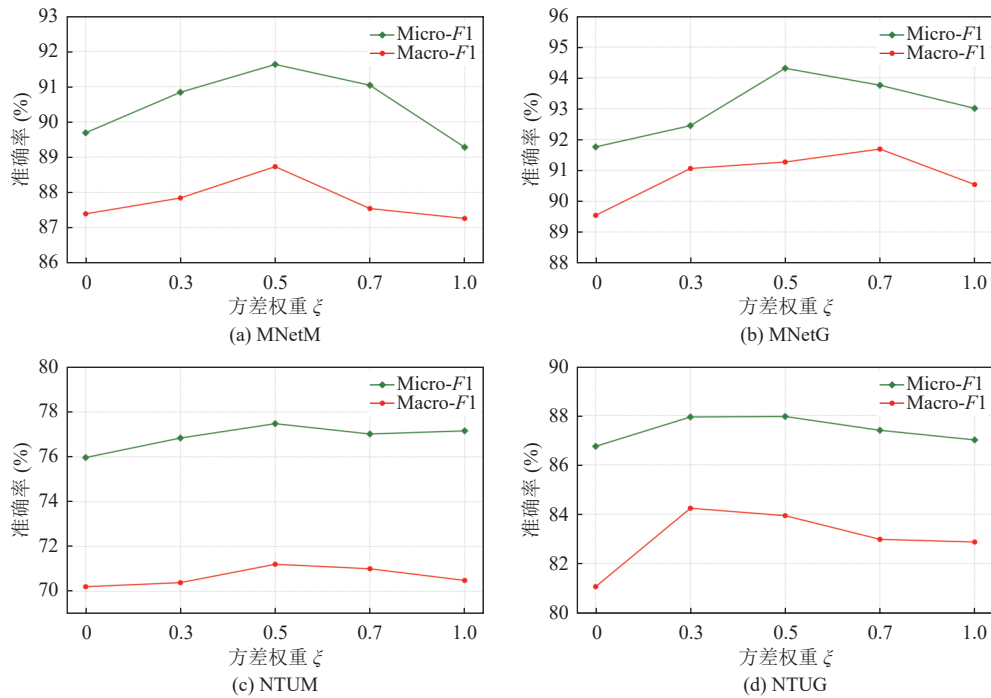


图 9 损失函数中不同采样方差权重  $\xi$  对表示学习性能的影响

从图 9 中, IERHN 模型中所有  $\xi > 0$  对应的超图表示学习结果都优于  $\xi = 0$  对应的表示学习结果. 这说明在损失函数中引入采样方差, 通过反向传播显式地最小化采样方差, 能够增加超图卷积估计的准确性, 从而自动学习到

更加合理的节点和超边采样策略,有效提高超图表示学习的性能.此外,当 $\xi = 0.5$ 时,IERHN能够在绝大部分数据集上获得最优的表示学习结果.因此,将超参数 $\xi$ 设定为0.5是合理的.

为了分析不同隐藏层深度对IERHN表示学习性能的影响以及IERHN的抗过平滑能力,图10展示了在MNetM数据集中,不同深度的IERHN和HGNN的超图表示学习准确性.

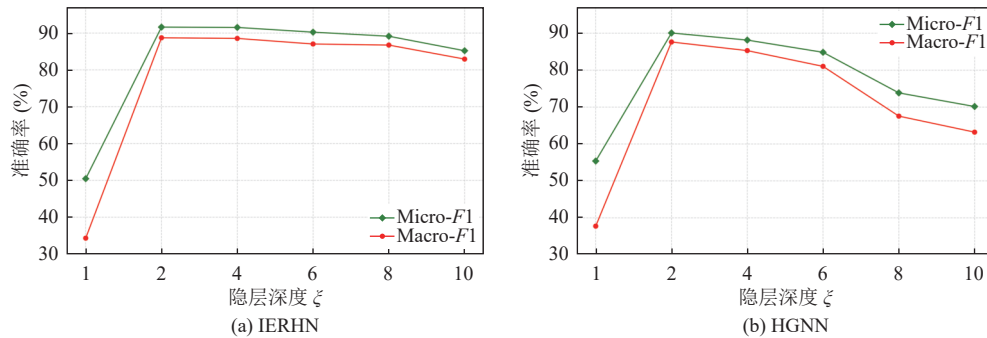


图10 MNetM数据集中不同隐藏层深度对表示学习性能的影响

从图10中,隐藏层的深度越深,模型可以聚合更远处的结构特征.当IERHN与HGNN的隐藏层为1层时,表示学习的性能非常差,这是正常的,因为1层的模型只能考虑非常有限的局部结构特征.随着隐藏层数的增加,HGNN的性能上升,当隐层数大于4后,由于发生了过平滑问题,模型的性能开始下降.IERHN利用层采样的方式采集节点和超边,因此不会聚合到过量的远端无用特征,从而保持节点间的差异性,有效避免了过平滑的问题.IERHN在较深的层数时,也能有很好的性能表现.

综上所述,经过在公开数据集上的广泛实验,IERHN相较于基线方法能获得最优的超图网络表示学习性能,在节点分类和链路预测任务上都能获得准确的结果.此外,IERHN利用重要性采样的优势,训练时间短,收敛速度快,具有很高的效率.

## 5 总结

在本文中,我们提出了一个高效的超图网络的表示学习模型IERHN,用于挖掘超图网络中复杂的高阶空间结构特征.IERHN设计了一个基于重要性采样的超图层采样策略,它根据一个统一的采样概率分别对每一层的节点和超边进行采样,构造更小的采样邻接矩阵,利用蒙特卡洛估计近似计算超图卷积的特征交互结果.此外还借助PINN的优势,将需要缩减的方差与目标函数相结合,使采样到的对象能更准确地揭示超图的结构特征.最终,IERHN不仅能够解决传统模型递归邻域扩展带来巨大计算开销的问题,而且能够抑制过量特征聚合带来的负面影响,提高泛化能力.

在真实数据集上进行的实验表明,IERHN能够获得高质量的超图网络表示学习结果.与现有的方法相比,IERHN在节点分类与链路预测两个任务上都有最优的性能表现.此外,IERHN计算复杂度和空间复杂度更低,训练效率和收敛速度都高于基线模型.目前,异质超图得到很多研究人员的关注,在未来,我们将研究IERHN在异质超图网络中的应用,分析不同类型节点和超边的采样方法,进一步扩展IERHN在大规模复杂网络中的应用.

## References:

- [1] Alvarez-Rodriguez U, Battiston F, de Arruda GF, Moreno Y, Perc M, Latora V. Evolutionary dynamics of higher-order interactions in social networks. *Nature Human Behaviour*, 2021, 5(5): 586–595. [doi: 10.1038/s41562-020-01024-1]
- [2] Battiston F, Amico E, Barrat A, Bianconi G, de Arruda GF, Franceschiello B, Iacopini I, Kéfi S, Latora V, Moreno Y, Murray MM, Peixoto TP, Vaccarino F, Petri G. The physics of higher-order interactions in complex systems. *Nature Physics*, 2021, 17(10): 1093–1098. [doi: 10.1038/s41567-021-01371-4]
- [3] Hu BD, Wang XG, Wang XY, Song ML, Chen C. Survey on hypergraph learning: Algorithm classification and application analysis.



- Ruan Jian Xue Bao/Journal of Software, 2022, 33(2): 498–523 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6353.htm> [doi: 10.13328/j.cnki.jos.006353]
- [4] Gao Y, Zhang ZZ, Lin HJ, Zhao XB, Du SY, Zou CQ. Hypergraph learning: Methods and practices. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2022, 44(5): 2548–2566. [doi: 10.1109/TPAMI.2020.3039374]
- [5] Wu AB, Yuan Y, Ma YL, Wang GR. Node embedding research over temporal graph. *Ruan Jian Xue Bao/Journal of Software*, 2021, 32(3): 650–668 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/6173.htm> [doi: 10.13327/j.cnki.jos.006173]
- [6] Ran YJ, Liu TY, Jia T, Xu XK. A novel similarity measure for mining missing links in long-path networks. *Chinese Physics B*, 2022, 31(6): 068902. [doi: 10.1088/1674-1056/ac4483]
- [7] Zhang L, Liu Q, Yang SS, Yang HP, Cheng F, Ma HP. A dual representation-based multi-objective evolutionary algorithm for overlapping community detection. *Acta Electronica Sinica*, 2021, 49(11): 2101–2107 (in Chinese with English abstract). [doi: 10.12263/DZXB.20201094]
- [8] Song GJ, Zhang YZ, Xu LJ, Lu HB. Domain adaptive network embedding. *IEEE Trans. on Big Data*, 2022, 8(5): 1220–1232. [doi: 10.1109/TBDATA.2020.3034201]
- [9] Yin N, Feng FL, Luo ZG, Zhang X, Wang WJ, Luo X, Chen C, Hua XS. Dynamic hypergraph convolutional network. In: *Proc. of the 38th IEEE Int'l Conf. on Data Engineering*. Kuala Lumpur: IEEE, 2022. 1621–1634. [doi: 10.1109/ICDE53745.2022.00167]
- [10] Saha SS, Sharma K, Panda SK. On the Laplacian spectrum of k-uniform hypergraphs. *Linear Algebra and Its Applications*, 2022, 655: 1–27. [doi: 10.1016/j.laa.2022.09.004]
- [11] Javidian MA, Wang ZY, Lu LY, Valtorta M. On a hypergraph probabilistic graphical model. *Annals of Mathematics and Artificial Intelligence*, 2020, 88(9): 1003–1033. [doi: 10.1007/s10472-020-09701-7]
- [12] Feng YF, You HX, Zhang ZZ, Ji RR, Gao Y. Hypergraph neural networks. In: *Proc. of the 33rd AAAI Conf. on Artificial Intelligence*. Honolulu: AAAI Press, 2019. 3558–3565. [doi: 10.1609/aaai.v33i01.33013558]
- [13] Chen J, Ma TF, Xiao C. FastGCN: Fast learning with graph convolutional networks via importance sampling. In: *Proc. of the 6th Int'l Conf. on Learning Representations*. Vancouver: OpenReview.net, 2018.
- [14] Maleki S, Saless D, Wall DP, Pingali K. HyperNetVec: Fast and scalable hierarchical embedding for hypergraphs. In: *Proc. of the 7th Int'l Conf. on Network Science*. Porto: Springer, 2022. 169–183. [doi: 10.1007/978-3-030-97240-0\_13]
- [15] Yang YZ, Hu M, Huang TY. Influential nodes identification in complex networks based on global and local information. *Chinese Physics B*, 2020, 29(8): 088903. [doi: 10.1088/1674-1056/ab969f]
- [16] Bai S, Zhang FH, Torr PHS. Hypergraph convolution and hypergraph attention. *Pattern Recognition*, 2021, 110: 107637. [doi: 10.1016/j.patcog.2020.107637]
- [17] Deng Y. Recommender systems based on graph embedding techniques: A review. *IEEE Access*, 2022, 10: 51587–51633. [doi: 10.1109/ACCESS.2022.3174197]
- [18] Kahalé N. On the effective dimension and multilevel Monte Carlo. *Operations Research Letters*, 2022, 50(4): 415–421. [doi: 10.1016/j.orl.2022.06.001]
- [19] Wu Y, Yu W, Wang XJ, Shen AT. The rate of complete consistency for recursive probability density estimator under strong mixing samples. *Statistics & Probability Letters*, 2021, 176: 13. [doi: 10.1016/J.SPL.2021.109130]
- [20] Tabandeh A, Jia GF, Gardoni P. A review and assessment of importance sampling methods for reliability analysis. *Structural Safety*, 2022, 97: 102216. [doi: 10.1016/J.STRUSAFE.2022.102216]
- [21] Shi CY, Zhang Q, Chu TG. Effect of observation time on source identification of diffusion in complex networks. *Chinese Physics B*, 2022, 31(7): 070203. [doi: 10.1088/1674-1056/ac5985]
- [22] Ökten G, Liu YN. Randomized quasi-Monte Carlo methods in global sensitivity analysis. *Reliability Engineering & System Safety*, 2021, 210: 107520. [doi: 10.1016/j.res.2021.107520]
- [23] Chibisov DM. Bernoulli's law of large numbers and the strong law of large numbers. *Theory of Probability & its Applications*, 2016, 60(2): 318–319. [doi: 10.1137/S0040585X97T987696]
- [24] Yu XJ, Nott DJ, Tran MN, Klein N. Assessment and adjustment of approximate inference algorithms using the law of total variance. *Journal of Computational and Graphical Statistics*, 2021, 30(4): 977–990. [doi: 10.1080/10618600.2021.1880921]
- [25] Krauth W. Event-chain Monte Carlo: Foundations, applications, and prospects. *Frontiers in Physics*, 2021, 9: 663457. [doi: 10.3389/FPHY.2021.663457]
- [26] Craiu RV, Gustafson P, Rosenthal JS. Reflections on Bayesian inference and Markov chain monte carlo. *Canadian Journal of Statistics*, 2022, 50(4): 1213–1227. [doi: 10.1002/cjs.11707]
- [27] Huang WB, Zhang T, Rong Y, Huang JZ. Adaptive sampling towards fast graph representation learning. In: *Proc. of the 32nd Int'l Conf.*

- on Neural Information Processing Systems. Montréal: Curran Associates Inc., 2018. 4563–4572.
- [28] Peng P, Pan JG, Xu H, Feng XL. RPINNs: Rectified-physics informed neural networks for solving stationary partial differential equations. *Computers & Fluids*, 2022, 245: 105583. [doi: [10.1016/J.COMPFLUID.2022.105583](https://doi.org/10.1016/J.COMPFLUID.2022.105583)]
- [29] Angrish A, Bharadwaj A, Starly B. MVCNN+: Computer-aided design model shape classification and retrieval using multi-view convolutional neural networks. *Journal of Computing and Information Science in Engineering*, 2021, 21(1): 011001. [doi: [10.1115/1.4047486](https://doi.org/10.1115/1.4047486)]
- [30] Feng YF, Zhang ZZ, Zhao XB, Ji RR, Gao Y. GVCNN: Group-view convolutional neural networks for 3D shape recognition. In: Proc. of the 31st IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 264–272. [doi: [10.1109/CVPR.2018.00035](https://doi.org/10.1109/CVPR.2018.00035)]
- [31] Wu ZH, Pan SR, Chen FW, Long GD, Zhang CQ, Yu PS. A comprehensive survey on graph neural networks. *IEEE Trans. on Neural Networks and Learning Systems*, 2021, 32(1): 4–24. [doi: [10.1109/TNNLS.2020.2978386](https://doi.org/10.1109/TNNLS.2020.2978386)]
- [32] Tu K, Cui P, Wang X, Wang F, Zhu WW. Structural deep embedding for hyper-networks. In: Proc. of the 32nd AAAI Conf. on Artificial Intelligence, the 30th Innovative Applications of Artificial Intelligence Conf. and the 8th AAAI Symp. on Educational Advances in Artificial Intelligence. New Orleans: AAAI Press, 2018. 53. [doi: [10.1609/aaai.v32i1.11266](https://doi.org/10.1609/aaai.v32i1.11266)]
- [33] dE Albuquerque Filho JE, Brandão LCP, Fernandes BJT, Maciel AMA. A review of neural networks for anomaly detection. *IEEE Access*, 2022, 10: 112342–112367. [doi: [10.1109/ACCESS.2022.3216007](https://doi.org/10.1109/ACCESS.2022.3216007)]
- [34] Huang J, Chen C, Ye FH, Wu JJ, Zheng ZB, Ling GH. Hyper2vec: Biased random walk for hyper-network embedding. In: Proc. of the 24th Int'l Conf. on Database Systems for Advanced Applications. Chiang Mai: Springer, 2019. 273–277. [doi: [10.1007/978-3-030-18590-9\\_27](https://doi.org/10.1007/978-3-030-18590-9_27)]
- [35] Du XB, Yan JC, Zhang R, Zha HY. Cross-network skip-gram embedding for joint network alignment and link prediction. *IEEE Trans. on Knowledge and Data Engineering*, 2022, 34(3): 1080–1095. [doi: [10.1109/TKDE.2020.2997861](https://doi.org/10.1109/TKDE.2020.2997861)]
- [36] Jin TS, Dai HQ, Cao LJ, Zhang BC, Huang FY, Gao Y, Ji RR. Deepwalk-aware graph convolutional networks. *Science China Information Sciences*, 2022, 65(5): 152104. [doi: [10.1007/s11432-020-3318-5](https://doi.org/10.1007/s11432-020-3318-5)]
- [37] Yadati N, Nimishakavi M, Yadav P, Nitin V, Louis A, Talukdar P. HyperGCN: A new method of training graph convolutional networks on hypergraphs. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver, 2019. 135.
- [38] Yan XQ, Hu SZ, Mao YQ, Ye YD, Yu H. Deep multi-view learning methods: A review. *Neurocomputing*, 2021, 448: 106–129. [doi: [10.1016/j.neucom.2021.03.090](https://doi.org/10.1016/j.neucom.2021.03.090)]

#### 附中文参考文献:

- [3] 胡秉德, 王新根, 王新宇, 宋明黎, 陈纯. 超图学习综述: 算法分类与应用分析. *软件学报*, 2022, 33(2): 498–523. <http://www.jos.org.cn/1000-9825/6353.htm> [doi: [10.13328/j.cnki.jos.006353](https://doi.org/10.13328/j.cnki.jos.006353)]
- [5] 吴安彪, 袁野, 马玉亮, 王国仁. 时序图节点嵌入策略的研究. *软件学报*, 2021, 32(3): 650–668. <http://www.jos.org.cn/1000-9825/6173.htm> [doi: [10.13327/j.cnki.jos.006173](https://doi.org/10.13327/j.cnki.jos.006173)]
- [7] 张磊, 刘庆, 杨尚尚, 杨海鹏, 程凡, 马海平. 基于双编码的重叠社团检测多目标优化方法. *电子学报*, 2021, 49(11): 2101–2107. [doi: [10.12263/DZXB.20201094](https://doi.org/10.12263/DZXB.20201094)]



邵豪(1995—), 男, 博士生, 主要研究领域为深度学习, 图嵌入, 数据挖掘.



朱然刚(1979—), 男, 博士, 讲师, 主要研究领域为数据挖掘, 网络结构分析.



王伦文(1966—), 男, 博士, 教授, 博士生导师, 主要研究领域为机器学习, 智能信息处理.



刘辉(1983—), 男, 博士, 讲师, 主要研究领域为自然语言处理.