

# 对于图谱神经网络进行节点分类来说，嗜好性是一个真正的噩梦吗？

栾思涛<sup>1,2</sup>, 华晨青<sup>1</sup>, 吕钦诚<sup>1</sup>, 朱家琪<sup>1</sup>, 赵明德<sup>1,2</sup>, 张淑媛<sup>1,2</sup>, 张晓文<sup>1</sup>, Doina Precup<sup>1,2,3</sup>

{sitao.luan@mail, chenqing.hua@mail, qincheng.lu@mail, jiaqi.zhu@mail, mingde.zhao@mail, shuyuan.zhang@mail, chang@cs, dprecup@cs}. mcgill.ca

<sup>1</sup>麦吉尔大学; <sup>2</sup> Mila; <sup>3</sup> DeepMind

## 摘要

图形神经网络（GNN）通过使用基于关系归纳偏见（同质性假设）的图形结构来扩展基本神经网络（NN）。尽管GNNs被认为在现实世界的任务中比NNs更出色，但GNNs相对于图-agnostic NNs的性能优势似乎并不普遍令人满意。嗜好被认为是一个主要原因，许多工作已经被提出来解决这个问题。在本文中，我们首先表明，并非所有的嗜好性情况都是有害的。对具有聚合操作的GNN有害。然后，我们提出了基于相似性矩阵的新指标，该指标考虑了图结构和输入特征对GNNs的影响。通过对合成图的测试，这些指标显示出比常用的同质性指标更多的优势。从这些指标和观察中，我们发现一些有害的异质性情况可以通过多样化操作来解决。利用这一事实和滤波器库的知识，我们提出了自适应通道混合（ACM）框架，以自适应地利用GNN各层中的聚合、多样化和身份通道来解决有害的异质性。我们用10个真实的基线验证了ACM增强的基线。世界节点分类任务。他们一直取得显著的性能在大多数任务上获得并超过了最先进的GNNs，而不产生重大的计算负担。

## 1 简介

深度神经网络（NN）[21]已经彻底改变了许多机器学习领域，包括图像识别[20]、语音识别[13]和自然语言处理[2]等。一个主要的优势是他们从欧几里得数据中学习潜在表示的能力和有效性。最近，人们开始关注它在非欧几里得数据上的应用[5]，例如，关系数据或图。结合图信号处理和卷积神经网络[22]，许多图神经网络（GNNs）[34, 9, 15, 35, 18, 28]已经被提出，根据经验，在基于图的机器学习任务上，例如节点分类、图分类、链接预测和图生成等，表现优于传统神经网络。GNNs建立在同质性假设[31]上，即连接节点往往彼此共享相似的属性[14]，除了节点特征，它提供额外信息。这种关系归纳偏见[3]被认为是导致GNNs在许多任务中比NNs性能更优越的一个关键因素。

然而，越来越多的证据表明，在处理关系型数据时，GNNs并不总是比传统的NNs更有优势。在某些情况下，即使是简单的多层感知器（MLPs）也能以较大的优势胜过GNNs [40, 26, 7]。性能下降的一个重要原因是

<sup>1</sup>一般来说，有害的异质性是指会使图感知模型的性能低于其相应的图诊断模型的异质性结构。

预印本。正在审查中。

被认为是异质性问题，也就是说，连接的节点往往有不同的标签，这使得同质性假设失效。最近，异质性挑战受到关注，有越来越多的模型被提出来解决这个问题[40, 26, 7, 39, 38]。

**贡献** 在本文中，我们首先证明了并非所有的异质图都对基于聚合的GNN有害，现有的同亲度指标不足以决定聚合操作是否会使节点的区分度降低。通过从反向传播分析中构建相似性矩阵，我们得出了新的指标来描述GNNs受图结构和节点特征的影响程度。我们通过比较两个基线GNN在不同同质性水平的合成图上的性能特征的能力，显示了我们的指标比现有指标的优势。从相似性矩阵中，我们发现多样化操作能够解决一些有害的异质性情况，在此基础上我们提出了自适应通道混合（ACM）GNN框架。在合成数据集、消融研究和真实世界数据集上的实验一致表明，由ACM框架增强的基线GNN能够在异质图的节点分类任务上获得显著的性能提升。

本文的其余部分主要组织如下：在第2节，我们介绍了符号和背景知识。在第3节中，我们对异质性进行了节点分析，得出了基于相似性矩阵的新的同质性指标，并进行了实验以显示它们比现有同质性指标的优势。在第4节中，我们展示了多样化操作在解决某些有害的嗜好问题上的能力，并提出了ACM-GNN框架，以适应性地利用来自不同过滤库渠道的信息来解决嗜好问题。在第5节中，我们讨论了相关的工作，并阐明了与我们方法的不同之处。在第6节中，我们提供了对ACM框架的实证评估，包括消融研究和在10个国家的测试。现实世界中的节点分类任务。

## 2 预备工作

我们将在本节中介绍相关的符号和背景知识。我们用**粗体**字表示向量（如 $\mathbf{v}$ ）。假设我们有一个无向连接图 $G = (V, E, A)$ ，其中 $V$ 是节点集， $|V| = N$ ； $E$ 是没有自环的边集； $A \in \mathbb{R}^{N \times N}$ 是对称邻接矩阵， $A_{i,j} = 1$  iff  $e_{ij} \in E$ ，否则 $A_{i,j} = 0$ 。我们用 $D$ 表示 $G$ 的对角线程度矩阵，即 $D_{i,i} = d_i = \sum_j A_{i,j}$ ，用 $N_i$ 表示节点 $i$ 的邻域集合，即 $N_i = \{j : e_{ij} \in E\}$ 。图信号是定义在 $V$ 上的向量 $\mathbf{x} \in \mathbb{R}^N$ ，其中 $\mathbf{x}_i$ 是定义在节点 $i$ 上的。我们还有一个特征矩阵 $X \in \mathbb{R}^{N \times F}$ ，其列是图信号，其第 $i$ 行 $\mathbf{x}_{i,:}$ 是节点 $i$ 的特征向量。我们用 $Z \in \mathbb{R}^{N \times C}$ 表示标签编码矩阵，其第 $i$ 行 $\mathbf{z}_{i,:}$ 是节点 $i$ 的标签单热编码。

### 2.1 图拉普拉斯、亲和矩阵及其变体

组合）图拉普拉斯定义为 $L = D - A$ ，它是对称正半定式（SPSD）[8]。它的特征分解给出了 $L = U \Lambda U^T$ ，其中 $U \in \mathbb{R}^{N \times N}$ 的列 $\mathbf{u}_i$ 是正交特征向量，即图傅里叶基， $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$ ， $\lambda_1 \leq \dots \leq \lambda_N$ ，这些特征值也被称为频率。图傅里叶变换图形信号 $\mathbf{x}$ 的定义为 $\mathbf{x}_F = U^{-1} \mathbf{x} = U^T \mathbf{x} = [\mathbf{u}_1^T \mathbf{x}, \dots, \mathbf{u}_N^T \mathbf{x}]^T$ ，其中 $\mathbf{u}_i^T \mathbf{x}$ 是 $\mathbf{x}$ 在 $\mathbf{u}_i$ 方向的分量 $i$ 。

除了 $L$ 之外，一些变体也常用，例如，对称归一化拉普拉斯 $L_{\text{sym}} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} A D^{-1/2}$ 和随机行走归一化拉普拉斯 $L_{\text{rw}} = D^{-1} L = I - D^{-1} A$ 。亲和（过渡）矩阵可以从拉普拉斯得出，例如， $A_{\text{rw}} = I - L_{\text{rw}} = D^{-1} A$ ， $A_{\text{sym}} = I - L_{\text{sym}} = D^{-1/2} A D^{-1/2}$ ，并被认为是低通滤波器[30]。它们的特征值满足 $\lambda_i(A_{\text{rw}}) = \lambda_i(A_{\text{sym}}) = 1 - \lambda_i(L_{\text{sym}}) = 1 - \lambda_i(L_{\text{rw}}) \in (-1, 1]$ 。对亲和矩阵和拉普拉斯矩阵分别应用重正化技巧[18]，可以得到 $A_{\text{sym}} = \tilde{A} \tilde{D}^{-1}$ 和 $L_{\text{sym}} = \tilde{D} - \tilde{A}$ ，其中 $\tilde{A} \equiv A + I$ ， $\tilde{D} \equiv D + I$ 。重正化的亲和矩阵本质上是给图中的每个节点增加一个自环，并被广泛用于图卷积网络（GCN）[18]，具体如下、

$$Y = \text{softmax}(A_{\text{sym}} \text{ReLU}(A_{\text{sym}} X W_0) W_1) \quad (1)$$

其中,  $W_0 \in \mathbb{R}^{F \times F}$  和  $W_1 \in \mathbb{R}^{F \times O}$  是可学习参数矩阵。GCN可以通过最小化以下交叉熵损失进行训练

$$L = -\text{trace}(Z^T \log Y) \quad (2)$$

其中  $\log(\cdot)$  是一个分量上的对数运算。随机漫步重正化矩阵

$A_{rw} = D^{-1}A$  与  $A_{sym}$  共享相同的特征值, 也可以应用于GCN。该相应的拉普拉斯被定义为  $L_{rw} = I - A_{rw}$ 。  $A_{rw}$  本质上是一个随机行走矩阵, 而表现为基于空间的GNNs[15, 14]中应用的平均聚合器。为了连接光谱和空间方法, 我们在本文中使用  $A_{rw}$ 。

## 2.2 同质性的衡量标准

同质的度量是通过考虑节点标签和由邻接矩阵定义的图结构之间的不同关系来定义的。有三个常用的同感度量: 边同感[1, 40], 节点同感[33], 和类同感[24]。定义如下:

$$H_{edge}(G) = \frac{|\{e_{uv} \mid e_{uv} \in E, Z_{u,:} = Z_{v,:}\}|}{|E|}, H_{node}(G) = \frac{1}{|V|} \sum_{v \in V} \frac{|\{u \mid u \in N_v, Z_{u,:} = Z_{v,:}\}|}{d_v},$$

$$H_{class}(G) = \frac{1}{C} \sum_{k=1}^C h_k, h_k = \frac{|\{v \mid Z_{v,k} = 1\}|}{N}, h_k = \frac{|\{u \mid Z_{v,k} = 1, u \in N_v, Z_{u,:} = Z_{v,:}\}|}{\sum_{v \in \{v \mid Z_{v,k} = 1\}} d_v} \quad (3)$$

其中  $[a]_+ = \max(a, 0)$ ;  $h_k$  是类的同质性指标[24]。它们都在[0, 1]的范围内, 接近1的值对应于强同质性, 而接近0的值表示强异质性。  $H_{edge}(G)$  衡量连接同一类别的两个节点的边缘比例;  $H_{node}(G)$  评估所有节点的边缘标签一致性的平均比例;  $H_{class}(G)$  试图避免对不平衡类别的敏感性, 这可能导致  $H_{edge}$  的误导性大。上述的

这些定义都是基于图标的一致性, 并暗示不一致将对GNN的性能造成有害影响。考虑到这一点, 我们将用一个反例来说明上述度量的不足, 并在下一节提出新的度量。

## 3 异质性的分析

### 3.1 激励和聚集的同质性

Heterophily is believed to be harmful for message-passing based GNNs [40, 33, 7] because intuitively features of nodes in different classes will be falsely mixed and this will lead nodes indistinguishable [40]. Nevertheless, it is not always the case, e.g., the bipartite graph shown in Figure 1 is highly heterophilous according to the homophily metrics in (3), but after mean aggregation, the nodes in classes 1 and 2 only exchange colors and are still distinguishable. [7]中的作者也指出了  $H_{node}$  的不足之处, 通过实例说明具有相同  $H_{node}$  的不同图类型可以携带不同的标签信息。

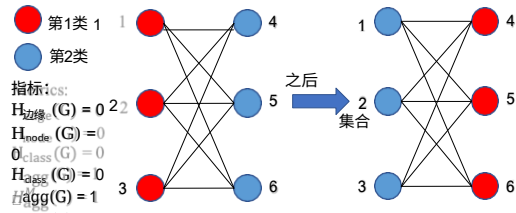


图1: 无害的异食癖的例子

为了分析图结构能在多大程度上影响GNN的输出, 我们首先通过去除其非线性来简化GCN, 如[36]。让  $\hat{A} \in \mathbb{R}^{N \times N}$  表示一个一般的聚合算子。那么, 方程 (1) 可以简化为、

$$Y = \text{softmax}(\hat{A}W) = \text{softmax}(Y') \quad (4)$$

<sup>2</sup>[24]中的作者并没有为这种同质性指标命名。我们根据它的定义将其命名为类亲和性。

在每个梯度衰减步骤  $\Delta W = \gamma \frac{dL}{dW}$ ，其中  $\gamma$  是学习率， $Y$  的更新将是 (见附录B的推导)、

$$\Delta Y = \gamma A^T X \frac{dL}{dW} \propto A^T X \frac{dL}{dW} = \hat{a} x^T a^t (z - y) = s(\hat{a}, x)(z - y) \quad (5)$$

其中  $s(\hat{a}, x) = A^T X$  是聚集后的节点相似性矩阵， $z - y$  是预测值误差矩阵。节点  $i$  的更新方向实质上是预测误差的加权和，即、

$$\Delta(Y)_{i,:} = \sum_{j \in V} h s(\hat{A})_{i,j} (Z - Y)_{j,:}.$$

为了研究异食癖的影响，我们首先定义聚集相似度得分如下。

**定义1. 聚合相似度得分**

$$s_{\text{agg}}(\hat{A}, x) = \frac{\text{平均}_u \{s(\hat{A}, x)_{v,u} / Z_{u,:} \mid Z_{u,:} \neq Z_{v,:}\}}{|V|} \quad (6)$$

其中  $\text{Mean}_u(\{-\})$  取给定的多组数值或变量的  $u$  的平均值。

$s_{\text{agg}}(s(\hat{A}, x))$  衡量节点  $v \in V$  的比例，这些节点在聚合后会将相对较大的相似性权重放在同一类别的节点上，而不是其他类别的节点。很容易看出， $s_{\text{agg}}(s(\hat{A}, x)) \in [0, 1]$ 。但是在实践中，我们观察到在大多数数据集中，我们会有  $s_{\text{agg}}(s(\hat{A}, x)) \geq 0.5$ 。基于这一观察，我们将(6)重新划分为以下修改后的 aggregation 相似性，以供实际使用、

$$s_{\text{agg}}^h(s(\hat{A}, x)) = h s_{\text{agg}}(s(\hat{A}, x)) - \frac{1}{2} \quad (7)$$

为了在不考虑节点特征的情况下衡量标签和图结构之间的一致性，并与现有的同质性指标进行公平的比较，在 (3) 中我们定义了图

(G) 聚集( $\hat{A}$ )的亲合性，其修改版为

$$h_{\text{agg}}(G) = s_{\text{agg}}(s(\hat{A}, z)), \quad h^M(G) = s^M(s(\hat{A}, z)) \quad (8)$$

在实践中，我们只在  $H^M(G) = 0$  时检查  $H_{\text{agg}}(G)$ 。如图1所示，当  $\hat{A} = A_w$  时， $H_{\text{agg}}(G) = H^M(G) = 1$ 。因此，这个新指标反映了1类和2类的节点在聚合后仍有很高的区分度，而之前提到的其他指标未能捕捉到这一信息，并误导性地给出了0值。这显示了  $H_{\text{agg}}(G)$  和  $H^M(G)$  通过额外考虑来自聚合算子  $\hat{A}$  和相似性矩阵的信息的优势。

为了全面比较  $H^M(G)$  和 (3) 中的指标如何揭示图结构对GNN性能的影响，我们在下一小节中生成合成图并对SGC [36]和GCN [18]进行评估。

### 3.2 对合成图的评估和比较

**数据生成和实验设置** 我们首先生成280个图，有28个边缘同质性水平，从0.005到0.95不等，每个图对应10个图。对于每个生成的图，我们有5个类，每个类有400个节点。对于每个节点，我们随机生成2条类内边和  $[2 - 2]$  条类间边（见附录C中关于数据生成过程的细节）。每个类中的节点特征都是从基础数据集的相应类中的节点特征中抽取的。节点被随机分成60%/20%/20%，用于训练/验证/测试。我们在合成数据上训练1跳SGC ( $sgc-1$ ) [36]和GCN[18]（见附录A.1的超参数

搜索范围）。对于  $H_{\text{edge}}(G)$  的每个值，我们取10个生成的图形的平均测试精度和运行的标准偏差。对于每个生成的图，我们还计算其  $H_{\text{node}}(G)$ ,  $H_{\text{class}}(G)$  和  $H^M(G)$ 。图2显示了模型在不

同源性值方面的表现。

**同源性指标的比较** SGC-1和GCN的性能预计会在适当的、信息丰富的同源性指标下呈单调增长。然而，图2(a)(b)(c)

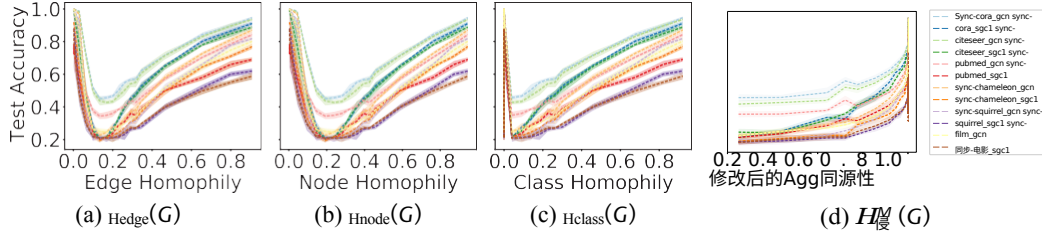


图2：不同亲缘度量下的基线性能比较。

表明 $H_{\text{edge}}(G)$ 、 $H_{\text{node}}(G)$ 和 $H_{\text{class}}(G)$ 下的性能曲线是U型的。<sup>3</sup>而图2(d)显示了一条几乎是单调的曲线，在1附近有一点扰动。这表明 $H^M(G)$ 可以比现有的指标更恰当和充分地描述图结构对SGC-1和GCN性能的影响。

此外，我们注意到在图2(a)中，当 $H_{\text{edge}}(G)$ 大约在0.1和0.2之间时，SGC-1和GCN在所有数据集上得到最差的性能。这个有趣的现象可以用以下基于相似性矩阵的定理来解释，它可以验证 $H^M(G)$ 的作用。

**Theorem 1.** (See Appendix D for proof). Suppose there are  $C$  classes in the graph  $G$  and  $G$  is a  $d$ -regular graph (each node has  $d$  neighbors). Given  $d$ , edges for each node are *i.i.d.* generated, such that each edge of any node has probability  $h$  to connect with nodes in the same class and probability  $1 - h$  to connect with nodes in different classes. Let the aggregation operator  $\hat{A} = \hat{A}_{\text{rw}}$ . Then, for nodes  $v$ ,  $u_1$  and  $u_2$ , where  $Z_{u_1,:} = Z_{v,:}$  and  $Z_{u_2,:} \neq Z_{v,:}$ , we have

$$g(h) \equiv \mathbb{E}_{\mathbf{z}} \left[ \frac{\|\hat{\mathbf{A}} \mathbf{z} - \mathbf{z}\|_{v,u}^2}{\|\mathbf{z}\|_{v,u}^2} \right] = \frac{(C-1)(hd+1) - (1-h)d^2}{(C-1)(d+1)} \quad (9)$$

而 $g(h)$ 的最小值是在以下位置达到的

$$h = \frac{d+1-C}{Cd} = \frac{d_{\text{intra}}/h + 1 - C}{C(d_{\text{intra}}/h)} \Rightarrow h = \frac{d_{\text{intra}}}{Cd_{\text{intra}} + C - 1}$$

其中， $d_{\text{intra}} = dh$ ，这是一个节点的邻居与该节点有相同标签的预期数量。

(9)中的 $g(h)$ 值是指与 $v$ 同级的节点与其他级的节点之间的相似性值的预期差异。 $g(h)$ 与聚集同质性的定义密切相关，其最小值可能意味着 $H_{\text{agg}}(G)$ 的最差值。在合成经验中，我们有 $d_{\text{intra}} = 2$ ， $C = 5$ ， $g(h)$ 的最小值在 $h = 1/7 \approx 0.14$ 时达到，这与图2(a)中性能曲线的最低点相对应。换句话说，SGC-1和GCN表现最差的 $h$ 是 $g(h)$ 得到最小值的地方，而不是边缘同质性值最小的点 $h=0$ 。这再次揭示了利用相似性矩阵， $H_{\text{agg}}(G)$ 比 $H_{\text{edge}}(G)$ 的优势。

## 4 自适应通道混合 (ACM) 框架

除了新的同质性指标，在这一节中，我们还将根据方程 (5) 中提出的相似性矩阵，弄清多样化操作（高通滤波器）是如何潜在地解决一些有害的异质性情况的。通过分析，我们认为低通滤波器和高通滤波器应该结合在一起进行特征提取，这使我们在第4.2小节提出了滤波器库方法。我们在第4.3小节中概括了滤波器库方法，并提出了ACM框架。

<sup>3</sup>在[40]中发现了一个类似的J型曲线，尽管使用了不同的数据生成过程。它并没有提到边缘同质性的不足。

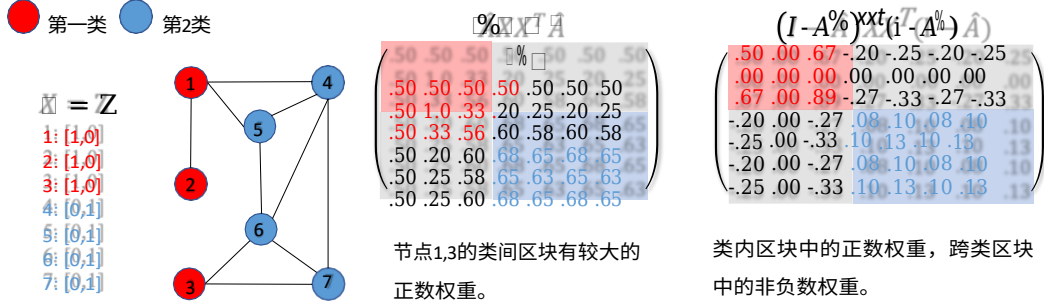


图3：HP过滤器如何解决有害的异食癖的例子

#### 4.1 多样化经营如何帮助解决有害的异质性问题

我们首先考虑图3中的例子。从 $s(\hat{A}, x)$ 来看，节点1,3在聚合后给第2类的节点分配了相对较大的正权，这将使节点1,3难以与第2类的节点区分开来。尽管如此，我们仍然可以通过考虑节点1,3和4,5,6,7的邻域差异来区分它们：节点1,3与大部分邻域不同，而节点4,5,6,7与大部分邻域相似。这表明，在某些情况下，尽管一些节点在聚合后变得相似，但它们仍然可以通过其周围的差异性而被区分出来。这导致我们使用多样化操作，即高通（HP）滤波器 $I - \hat{A}$  [10]（将在下一小节介绍）来提取邻域差异的信息并解决有害的异质性。如图3中的 $s(I - \hat{A}, x)$ 所示，节点1,3在多样化操作后将给节点4,5,6,7分配负的权重，即节点1,3将节点4,5,6,7为负样本，在反向传播过程中会远离它们。在此基础上，我们首先提出多样化的可区分性，以衡量多样化操作有潜在帮助的节点的比例。

**定义2.** 基于 $s(I - \hat{A}, x)$ 的多样化可区分性 (DD)。

给定 $s(I - \hat{A}, x)$ ，如果同时满足以下两个条件，则一个节点 $v$ 是可多样化区分的。

1. 平均 $_u \{s(I - \hat{A}, x)_{v,u} \mid u \in V \wedge Z_{u,:} = Z_{v,:}\} \geq 0$ ;
2. 平均 $_u \{s(I - \hat{A}, x)_{v,u} \mid u \in V \wedge Z_{u,:} \neq Z_{v,:}\} \leq 0$

那么，图的多样化可区分性值被定义为

$$DD_{\hat{A}, x}(G) = \frac{1}{|V|} |\{v \mid v \text{ is diversification distinguishable}\}| \quad (11)$$

我们可以看到， $DD_{\hat{A}, x}(G) \in [0, 1]$ 。根据定义2，多样化操作的有效性可以在某些条件下被证明用于二元分类问题，从而使我们得出：

**定理2.** (证明见附录E)。假设 $x=Z$ ， $\hat{A}=A_w$ 。那么，对于二元分类问题， $\ell_2 C = 2$ ，所有的节点都是可区分的， $DD_{\hat{A}, Z}(G) = 1$ 。

定理2从理论上证明了多样化操作对于提取图形信号高频信息的重要性[10]。结合聚合操作，也就是低通滤波器[10, 30]，我们可以得到一个滤波器库，它同时使用聚合和多样化操作，从图形信号中明显地提取低频和高频信息。我们将在下一小节中介绍滤波器库。



## 4.2 谱系和空间形式的滤波器库

**滤波器库** 对于定义在G上的图形信号 $\mathbf{x}$ ，一个双通道线性（分析）滤波器库<sup>4</sup>包括一对低通(LP)和高通(HP)滤波器 $H_{LP}, H_{HP}$ ，其中 $H_{LP}$ 和 $H_{HP}$ 分别保留 $\mathbf{x}$ 的低频和低频内容。

大多数现有的GNN是在单通道滤波架构下[18, 35, 15]，其通道为 $H_{LP}$ 或 $H_{HP}$ ，只保留了部分输入信息。与单通道结构不同，具有 $H_{LP} + H_{HP} = I$ 的滤波器库不会丢失输入信号的任何信息，~~即~~完美重建特性[10]。

一般来说，拉普拉斯矩阵 $(L_{sym}, L_{rw}, L_{sym}, L_w)$ 可以被视为HP滤波器[10]，亲和矩阵 $(A_{sym}, A_{rw}, A_{sym}, A_w)$ 可以被视为LP滤波器[30, 14]。此外，MLPs可以被认为是欠一个特殊的身份滤波器库，其矩阵 $I$ 满足 $H_{LP} + H_{HP} = I + 0 = I$ 。

**空间形式的滤波器库** 滤波器库方法也可以扩展到空间GNN。形式上，在节点层面上，在 $\mathbf{x}$ 上左乘 $H_{LP}$ 和 $H_{HP}$ ，分别执行聚合和分散操作。例如，假设 $H_{LP} = A^{\wedge}$ ， $H_{HP} = I - A^{\wedge}$ ，那么对于节点 $i$ 我们有

$$(H_{LP} \mathbf{x})_i = \sum_{j \in \{N(i)\}} A^{\wedge}_{ij} \mathbf{x}_j, (H_{HP} \mathbf{x})_i = \mathbf{x}_i - \sum_{j \in \{N(i)\}} A^{\wedge}_{ij} \mathbf{x}_j \quad (12)$$

其中 $A^{\wedge}_{ij}$ 是两个节点之间的连接权重。为了利用GNN中的HP和身份通道，我们在下面的小节中提出了自适应通道混合（ACM）架构。

## 4.3 自适应通道混合(ACM) GNN框架

ACM框架可以应用于很多基线GNN，在本小节中，我们以GCN为例，以矩阵形式介绍ACM框架。我们用 $H_{LP}$ 和 $H_{HP}$ 来表示一般的LP和HP滤波器。ACM框架包括以下3个步骤、

**步骤1.每个通道的特征提取：**

选项1：  $H'_L = \text{ReLU}(H H W_{LP}^{l-1/l-1})$ ，  $H'_H = \text{ReLU}(H H W_{HP}^{l-1/l-1})$ ，  $H'_I = \text{ReLU}(H W^{l-1/l-1})$ ；

选项2：  $H'_L = H_{LP} \text{ReLU}(H W^{l-1/l-1})$ ，  $H'_H = H_{HP} \text{ReLU}(H W^{l-1/l-1})$ ，  $H'_I = I \text{ReLU}(H W^{l-1/l-1})$ ；

$W_L^{l-1}, W_H^{l-1}, W_I^{l-1} \in \mathbb{R}^{F_{l-1} \times F_l}$ ；

**第2步。基于特征的权重学习**

$$\alpha_L^l = \sigma(H_L^l W_L^l), \alpha_H^l = \sigma(H_H^l W_H^l), \alpha_I^l = \sigma(H_I^l W_I^l), \tilde{W}_L^{l-1} W_L^{l-1} W_H^{l-1} W_H^{l-1} \in \mathbb{R}^{F \times 1}$$

$$\alpha_L^l, \alpha_H^l, \alpha_I^l \in \text{Softmax}, \tilde{q}_L^l, \tilde{q}_H^l, \tilde{q}_I^l \in \mathbb{R}^{3 \times 3}, T \in \mathbb{R} \text{ 是温度；}$$

**第3步。节点式通道混合：**

$$H^l = \text{diag}(\alpha_L^l) H_L^l + \text{diag}(\alpha_H^l) H_H^l + \text{diag}(\alpha_I^l) H_I^l \quad (13)$$

步骤1中选择1的框架是ACM框架，选择2的框架是ACMII框架。ACM(II)-GCN首先分别对3个通道实现不同的特征提取。在亲通过一组滤波库，得到3个滤波成分 $H'_L, H'_H, H'_I$ 。不同的节点可以

对3个通道的信息有不同的需求，例如，在图3中，节点1、3需要高频信息，而节点2只需要低频信息。为了自适应地利用来自不同通道的信息，ACM(II)-GCN学习了逐行（逐节点）的特征条件权重来组合3个通道。ACM(II)可以很容易地插入到空间GNN中，如(12)所示，通过聚合和分散操作取代 $H_{LP}$ 和 $H_{HP}$ 。关于合成数据集上的模型比较的详细讨论见附录F。

**复杂性** ACM(II)-GCN的第 $l$ 层的可学习参数数量是 $3F_{l-1}(F_l + 1) + 9$ ，而GCN中是 $F F_{l-1}$ 。步骤1-3的计算需要 $NF_l(8 + 6F_{l-1}) + 2F_l(\text{nnz}(H_{LP}) + \text{nnz}(H_{HP})) + 18N$  flops，而GCN层需要 $2NF_{l-1} + 2F_l(\text{nnz}(H_{LP}))$  flops，其中 $\text{nnz}(-)$ 是非零元素的数量。在第6.1节中对运行时间进行了详细比较。

<sup>4</sup>在图形信号处理中，需要一个额外的合成滤波器[10]来形成2通道的滤波器库。

但在我们的框架中不需要合成滤波器，所以我们在本文中不介绍它。

**局限性** 多样化操作在所有有害的异质性情况下都不能很好地工作。例如，考虑一个不平衡的数据集，其中几个具有独特标签的小集群与一个大集群紧密相连。在这种情况下，小聚类中的节点的周边差异是相似的，*也就是说*，周边差异主要来自于它们与同一个大聚类的连接，这可能使多样化操作无法区分它们。见附录G中更详细的演示和讨论。

## 5 先前的工作

**GNNs解决异质性问题** 我们在这一部分讨论GNNs解决异质性挑战的相关工作。[1]认识到在具有弱同质性的图上学习的困难，并提出MixHop，从多跳邻域中提取特征以获得更多信息。Geom-GCN[33]预先计算了无监督的节点嵌入，并使用由嵌入空间中的几何关系定义的图结构来定义双级聚合过程。[16]提出了基于特征平滑度和标签平滑度的测量方法，可能有助于指导GNN处理异质图。H<sub>2</sub> GCN [40] 结合了3个关键设计以解决异质性问题：（1）自我和邻居嵌入分离；（2）高阶邻域；（3）中间表征的组合。CPGNN[39]通过兼容性矩阵对标签相关性进行建模，这对异质性设置有利，并通过兼容性矩阵将先验信念估计传播到GNN中。FBGNN[29]首次提出使用滤波器库来解决异质性问题，但是它并没有完全解释HP滤波器背后的见解，也没有包含身份通道和节点通道混合机制。FAGCN[4]像GAT[35]一样学习边缘级聚合权重，但允许权重为负数，这使得网络能够捕捉到图信号中的高频成分。GPRGNN[7]使用可学习的权重进行特征传播，它使GRPGNN能够适应图的异质结构，并且能够处理图信号中的高频和低频部分。

**带有滤波库的GNN** 以前，有一些几何散射网络[12, 32]，应用滤波库来解决过度平滑[23]问题。散射结构捕捉了来自节点特征或标签的不同变化渠道。在几何学习和图信号处理中，带通滤波操作提取平滑信号之外的几何信息，因此认为滤波器库可以缓解GNN的过度平滑。在ACM框架中，我们的目标是在滤波器库的帮助下设计一个框架，以适应性地利用不同的通道来解决异质图上的学习挑战。我们处理的问题与[12, 32]不同。

## 6 在真实世界的数据集上进行实验

在本节中，我们在真实世界的数据集上评估ACM(II)框架。我们首先在第6.1小节进行消融研究，以验证不同的组件。然后，我们在第6.2小节中与最先进的（SOTA）模型进行比较。

### 6.1 消融研究和效率

我们通过消融研究来研究在ACM(II)框架中加入HP、身份通道和自适应混合机制的效果和效率。具体来说，我们将上述组件分别应用于SGC-1和GCN，在每个数据集上运行10次，用60%/20%/20%的随机分割进行训练/验证/测试，并报告平均测试精度以及标准偏差。我们还记录了每个epoch的平均运行时间（以毫秒为单位）以比较效率。我们设定（13）中的温度 $\tau$ 为3。（超参数搜索范围见附录A）。

从结果中我们可以看出，在大多数数据集上，额外的HP和身份通道是有帮助的，甚至在强同源性数据集上，如Cora、CiteSeer和PubMed。自适应混合机制也显示出它比直接将三个通道加在一起的方法更有优势。这说明了学习为不同节点自适应地定制通道使用的必要性。至于效率，我们可以看到，在ACM(II)框架下的运行时间比原来的模型大约增加了一倍。

模型	LP HP 身份混合	误差±标准	误差±标准	误差±标准	误差±标准	误差±标准	误差±标准	误差±标准	误差±标准	误差±标准	
SGC-1 w/	✓	70.98 ± 8.39	70.38 ± 2.85	83.28 ± 5.43	25.26 ± 1.18	64.86 ± 1.81	47.62 ± 1.27	85.12 ± 1.64	79.66 ± 0.75	85.5 ± 0.76	12.89
	✓ ✓	83.28 ± 5.81	91.88 ± 1.61	90.98 ± 2.46	36.76 ± 1.01	65.27 ± 1.9	47.27 ± 1.37	86.8 ± 1.08	80.98 ± 1.68	87.21 ± 0.42	10.44
	✓ ✓ ✓	93.93 ± 3.6	95.25 ± 1.84	93.93 ± 2.54	38.38 ± 1.13	63.83 ± 2.07	46.79 ± 0.75	86.73 ± 1.28	80.57 ± 0.99	87.8 ± 0.58	9.44
	✓ ✓ ✓ ✓	88.2 ± 4.39	93.5 ± 2.95	92.95 ± 2.94	37.19 ± 0.87	62.82 ± 1.84	44.94 ± 0.93	85.22 ± 1.35	80.75 ± 1.68	88.11 ± 0.21	11.00
	✓ ✓ ✓ ✓ ✓	93.77 ± 1.91	93.25 ± 2.92	93.61 ± 1.55	39.33 ± 1.25	63.68 ± 1.62	46.4 ± 1.13	86.63 ± 1.13	80.96 ± 0.93	87.75 ± 0.88	10.00
ACM-GCN w/	✓	82.46 ± 3.11	75.5 ± 2.92	83.11 ± 3.2	35.51 ± 0.99	64.18 ± 2.62	44.76 ± 1.39	87.78 ± 0.96	81.39 ± 1.23	88.9 ± 0.32	11.44
	✓ ✓	82.13 ± 2.59	86.62 ± 4.61	89.19 ± 3.04	38.06 ± 1.35	<b>69.21 ± 1.68</b>	57.2 ± 1.01	88.93 ± 1.55	<b>81.96 ± 0.91</b>	90.01 ± 0.8	7.22
	✓ ✓ ✓	94.26 ± 2.23	96.13 ± 2.2	94.1 ± 2.95	41.51 ± 0.99	67.44 ± 2.14	53.97 ± 1.39	88.95 ± 0.9	81.72 ± 1.22	90.88 ± 0.55	4.44
	✓ ✓ ✓ ✓	91.64 ± 2	95.37 ± 3.31	<b>95.25 ± 2.37</b>	40.47 ± 1.49	68.93 ± 2.04	54.78 ± 1.27	<b>89.13 ± 1.77</b>	<b>81.96 ± 2.03</b>	<b>91.01 ± 0.7</b>	3.11
	✓ ✓ ✓ ✓ ✓	94.75 ± 2.62	<b>96.75 ± 1.6</b>	95.08 ± 3.2	41.62 ± 1.15	69.04 ± 1.74	<b>58.02 ± 1.86</b>	88.95 ± 1.3	81.80 ± 1.26	90.69 ± 0.53	<b>2.78</b>
ACM-GCN w/	✓	3.67	3.74	3.59	4.86	4.96	6.41	4.24	4.18	5.08	
	✓ ✓	6.63	8.06	7.89	8.11	7.8	9.39	7.82	7.38	8.74	
	✓ ✓ ✓	5.73	5.91	5.93	6.86	6.35	7.15	7.34	6.65	6.8	
	✓ ✓ ✓ ✓	5.16	5.25	5.2	5.93	5.64	8.02	5.73	5.65	6.16	
	✓ ✓ ✓ ✓ ✓	8.25	8.11	7.89	7.97	8.41	11.9	8.84	8.38	8.63	
ACMII-GCN	✓ ✓	6.62	7.35	7.39	7.62	7.33	9.69	7.49	7.58	7.97	
	✓ ✓ ✓	6.3	6.05	6.26	6.87	6.44	6.5	6.14	7.21	6.6	
	✓ ✓ ✓ ✓	5.24	5.27	5.46	5.72	5.65	7.87	5.48	5.65	6.33	
	✓ ✓ ✓ ✓ ✓	7.59	8.28	8.06	8.85	8	10	8.27	8.5	8.68	

表1：对9个真实世界数据集的消融研究[33]。带有Cmeans的单元格被应用于基线模型。最好的测试结果被突出显示。

GNN 在 10 个数据集集中的8个实现了SOTA性能。总的来说，这表明 ACM(II) 框架可以帮助 GNNs 在异质图的节点分类任务中更好地泛化。

## 6.2 与最先进的模型比较

**数据集和实验设置** 在本节中，我们实现了1跳和2跳的SGC[36]（SGC-1，SGC-2）、GCNII[6]、GCNII\*[6]、GCN[18]以及2层和3层的雪球网络（雪球-2，雪球-3），并在ACM或ACMII框架中应用它们（详见附录A.5）：我们使用 $\hat{A}_{rw}$ 作为LP滤波器，相应的HP滤波器可以从（12）中得到。我们将它们与几个基线和SOTA模型进行比较：有2层的MLP（MLP-2），GAT[35]，APPNP[19]，GPRGNN[7]，H<sub>2</sub> GCN[40]，MixHop[1]，GCN+JK[18, 37, 24]，GAT+JK[35, 37, 24]，FAGCN[4]，GraphSAGE[15]和Geom-GCN[33]。除了[33]中使用的9个基准数据集，我们还进一步在一个新的基准数据集Deezer-Europe上测试上述模型，该数据集是在[24]中提出的。我们在Cornell, Wisconsin, Texas, Film, Chameleon, Squirrel, Cora, Citeseer和Pubmed上测试这些模型10次，采用相同的早期停止策略，相同的随机数据分割方法<sup>5</sup>和GPRGNN[7]中使用的Adam[17]优化器。对于Deezer-Europe，我们按照相同的早期停止策略、相同的固定拆分和[24]中使用的AdamW[27]对上述模型进行了5次测试。超参数搜索范围和最佳超参数的细节在附录A.3和A.4中报告。

这组实验的主要结果与数据集的统计结果总结在表2中，其中我们报告了平均准确率和标准偏差。我们可以看到，在ACM(II)框架中应用后，基线模型的性能几乎在所有任务上都得到了提升。特别是，ACMII-GCN在所有数据集的平均排名（3.40）方面表现最好，ACM(II)-

7.44	✓	✓	✓	82.46 ± 3.03	91.00 ± 1.75	90.33 ± 2.69	38.39 ± 0.75	67.59 ± 2.14	53.67 ± 1.71	89.13 ± 1.14	81.75 ± 0.85	89.87 ± 0.39	6.67
4.67	✓	✓	✓	94.26 ± 2.57	96.00 ± 2.15	94.26 ± 2.96	40.96 ± 1.2	66.35 ± 1.76	50.78 ± 2.07	89.06 ± 1.07	81.86 ± 1.22	90.71 ± 0.67	2.78
ACMIL-GCN w/	✓	✓	✓	91.48 ± 1.43	96.25 ± 2.09	93.71 ± 1.07	40.52 ± 0.65	66.52 ± 2.65	52.9 ± 1.64	88.83 ± 1.16	81.54 ± 0.95	90.6 ± 0.47	
	✓	✓	✓	-2.53	2.83	2.5	3.18	3.48	4.65	3.47	3.43	4.04	
	✓	✓	✓	4.01	4.57	4.24	4.55	4.76	5.09	5.39	4.69	4.75	
SGC-1 w/	✓	✓	✓	3.88	4.01	4.04	4.43	4.06	4.5	4.38	3.82	4.16	
	✓	✓	✓	3.31	3.49	3.18	3.77	3.53	4.8	4.9	3.87	4.24	
	✓	✓	✓	5.52	5.96	5.43	5.21	5.41	5.96	5.9	5.9	5.04	

见 <https://github.com/jianhao2016/GPRGNN/blob/f4aaad6ca28c83d3121338a4c4fe5d162edfa9a2/src/utils.py#L16> 中的拆分开放源代码。见附录

一中的表9，与几个SOTA模型在[33]提供的固定48%/32%/20%分割上的性能比较。

	康奈尔 大学	威斯康星 州	德州	电影	变色龙	松鼠	Deezer-Europe	Cora	CiteSeer	医学博士	
#nodes	183	251	183	7,600	2,277	5,201	28,281	2,708	3,327	19,717	
# 边缘	295	499	309	33,544	36,101	217,073	92,752	5,429	4,732	44,338	
#特征	1,703	1,703	1,703	931	2,325	2,089	31,241	1,433	3,703	500	
#类	5	5	5	5	5	5	2	7	6	3	
H_edge	0.5669	0.4480	0.4106	0.3750	0.2795	0.2416	0.5251	0.8100	0.7362	0.8024	
H_node	0.3855	0.1498	0.0968	0.2210	0.2470	0.2156	0.5299	0.8252	0.7175	0.7924	
H_class	0.0468	0.0941	0.0013	0.0110	0.0620	0.0254	0.0304	0.7657	0.6270	0.6641	
数据拆分	60%/20%/20%	60%/20%/20%	60%/20%/20%	60%/20%/20%	60%/20%/20%	60%/20%/20%	50%/25%/25%	60%/20%/20%	60%/20%/20%	60%/20%/20%	
H <sub>avg</sub> M(G)	0.8032	0.7768	0.694	0.6822	0.61	0.3566	0.5790	0.9904	0.9826	0.9432	
最先进模型、GNN基线模型和ACM-GNN模型的测试准确率 (%)。											
MLP-2*	91.30 ± 0.70	93.87 ± 3.33	92.26 ± 0.71	38.58 ± 0.25	46.72 ± 0.46	31.28 ± 0.27	66.55 ± 0.72	76.44 ± 0.30	76.25 ± 0.28	86.43 ± 0.13	18.60
GAT*	76.00 ± 1.01	71.01 ± 4.66	78.87 ± 0.86	35.98 ± 0.23	63.9 ± 0.46	42.72 ± 0.33	61.09 ± 0.77	76.70 ± 0.42	67.20 ± 0.46	83.28 ± 0.12	21.40
APNP*	91.80 ± 0.63	92.00 ± 3.59	91.18 ± 0.70	38.86 ± 0.24	51.91 ± 0.56	34.77 ± 0.34	67.21 ± 0.56	79.41 ± 0.38	68.59 ± 0.30	85.02 ± 0.09	18.00
GPRGN*	91.36 ± 0.70	93.75 ± 2.37	92.92 ± 0.61	39.30 ± 0.27	67.48 ± 0.40	49.93 ± 0.53	66.90 ± 0.50	79.51pm ± 0.36	67.63 ± 0.38	85.07 ± 0.09	14.40
H2GCN	86.23 ± 4.71	87.5 ± 1.77	85.90 ± 3.53	38.85 ± 1.17	52.30 ± 0.48	30.39 ± 1.22	<b>67.22 ± 0.90</b>	87.52 ± 0.61	79.97 ± 0.69	87.78 ± 0.28	17.00
MixHop	60.33 ± 28.53	77.25 ± 7.80	76.39 ± 7.66	33.13 ± 2.40	36.28 ± 10.22	24.55 ± 2.60	66.80 ± 0.58	65.65 ± 11.31	49.52 ± 13.35	87.04 ± 4.10	23.50
GCN+JK	66.56 ± 13.82	62.50 ± 15.75	80.66 ± 1.91	32.72 ± 2.62	64.68 ± 2.85	53.40 ± 1.90	60.99 ± 0.14	86.90 ± 1.51	73.77 ± 1.85	90.09 ± 0.68	18.80
GAT+JK	74.43 ± 10.24	69.50 ± 3.12	75.41 ± 7.18	35.41 ± 0.97	68.14 ± 1.18	52.28 ± 3.61	59.66 ± 0.92	89.52 ± 0.43	74.49 ± 2.76	89.15 ± 0.87	16.70
FAGCN	88.03 ± 5.6	89.75 ± 6.37	88.85 ± 4.39	31.59 ± 1.37	49.47 ± 2.84	42.24 ± 1.2	66.86 p, ± 0.53	88.85 ± 1.36	<b>82.37 ± 1.46</b>	89.98 ± 0.54	14.10
图谱	71.41 ± 1.24	64.85 ± 5.14	79.03 ± 1.20	36.37 ± 0.21	62.15 ± 0.42	41.26 ± 0.26	OOM	86.58 ± 0.26	78.24 ± 0.30	86.85 ± 0.11	20.89
Geom. <sup>GCN+</sup>	60.81	64.12	67.57	31.63	60.9	38.14	这是对的。	85.27	77.99	90.05	22.67
SGC-1	70.98 ± 8.39	70.38 ± 2.85	83.28 ± 5.43	25.26 ± 1.18	64.86 ± 1.81	47.62 ± 1.27	59.73 ± 0.12	85.12 ± 1.64	79.66 ± 0.75	85.5 ± 0.76	20.10
SGC-2	72.62 ± 9.92	74.75 ± 2.89	81.31 ± 3.3	28.81 ± 1.11	62.67 ± 2.41	41.25 ± 1.4	61.56 ± 0.51	85.48 ± 1.48	80.75 ± 1.15	85.36 ± 0.52	20.70
GCNII	89.18 ± 3.96	83.25 ± 2.69	82.46 ± 4.58	40.82 ± 1.79	60.35 ± 2.7	38.81 ± 1.97	66.38 ± 0.45	88.98 ± 1.33	81.58 ± 1.3	89.8 ± 0.3	14.80
GCNII*	90.49 ± 4.45	89.12 ± 3.06	88.52 ± 3.02	41.54 ± 0.99	62.8 ± 2.87	38.31 ± 1.3	66.42 ± 0.56	88.93 ± 1.37	81.83 ± 1.78	89.98 ± 0.52	12.30
GCN	82.46 ± 3.11	75.5 ± 2.92	83.11 ± 3.2	35.51 ± 0.99	64.18 ± 2.62	44.76 ± 1.39	62.23 ± 0.53	87.78 ± 0.96	81.39 ± 1.23	88.9 ± 0.32	16.30
雪球-2	82.62 ± 2.34	74.88 ± 3.42	83.11 ± 3.2	35.97 ± 0.66	64.99 ± 2.39	47.88 ± 1.23	OOM	88.64 ± 1.15	81.53 ± 1.71	89.04 ± 0.49	15.22
雪球-3	82.95 ± 2.1	69.5 ± 5.01	83.11 ± 3.2	36.00 ± 1.36	65.49 ± 1.64	48.25 ± 0.94	OOM	89.33 ± 1.3	80.93 ± 1.32	88.08 ± 0.82	14.78
ACM-SGC-1	93.77 ± 1.91	93.25 ± 2.92	93.61 ± 1.55	39.33 ± 1.25	63.68 ± 1.62	46.4 ± 1.13	66.67 ± 0.56	86.63 ± 1.13	80.96 ± 0.93	87.75 ± 0.88	12.60
ACM-SGC-2	93.77 ± 2.17	94.00 ± 2.61	93.44 ± 2.54	40.13 ± 1.21	60.48 ± 1.55	40.91 ± 1.39	66.53 ± 0.57	87.64 ± 0.99	80.93 ± 1.16	88.79 ± 0.5	13.40
ACM-GCNII	92.62 ± 3.13	94.63 ± 2.96	92.46 ± 1.97	41.37 ± 1.37	58.73 ± 2.52	40.9 ± 1.58	66.39 ± 0.56	89.1 ± 1.61	82.28 ± 1.12	90.12 ± 0.4	10.40
ACM-GCNII*	93.44 ± 2.74	94.37 ± 2.81	93.28 ± 2.79	41.27 ± 1.24	61.66 ± 2.29	38.32 ± 1.5	66.6 ± 0.57	89.00 ± 1.35	81.69 ± 1.25	90.18 ± 0.51	10.10
ACM-GCN	94.75 ± 3.8	95.75 ± 2.03	94.92 ± 2.88	41.62 ± 1.15	69.04 ± 1.74	<b>58.02 ± 1.86</b>	67.01 ± 0.38	88.62 ± 1.22	81.68 ± 0.97	90.66 ± 0.47	4.80
ACM-雪球-2	95.08 ± 3.11	96.38 ± 2.59	<b>95.74 ± 2.22</b>	41.4 ± 1.23	<b>68.51 ± 1.75</b>	55.97 ± 2.03	OOM	88.83 ± 1.49	81.58 ± 1.23	90.81 ± 0.52	4.44
ACM-雪球-3	94.26 ± 2.57	96.62 ± 1.86	94.75 ± 2.41	41.27 ± 0.8	68.4 ± 2.05	55.73 ± 2.39	OOM	<b>89.59 ± 1.58</b>	81.32 ± 0.97	<b>91.44 ± 0.5</b>	4.44
ACMII-GCN	<b>95.9 ± 1.83</b>	96.62 ± 2.44	95.08 ± 2.07	<b>41.84 ± 1.15</b>	68.38 ± 1.36	54.53 ± 2.09	67.15 ± 0.41	89.00 ± 0.72	81.79 ± 0.95	90.74 ± 0.5	<b>3.4</b>
ACMII-雪球	2 95.25 ± 1.55	96.63 ± 2.24	95.25 ± 1.55	41.1 ± 0.75	67.83 ± 2.63	53.48 ± 0.6	OOM	88.95 ± 1.04	82.07 ± 1.04	90.56 ± 0.39	4.78
ACMII-雪球	3 95.61 ± 2.79	<b>97.00 ± 2.63</b>	94.75 ± 3.09	40.31 ± 1.6	67.53 ± 2.83	52.31 ± 1.57	OOM	89.36 ± 1.26	81.36 ± 1.15	91.31 ± 0.6	<b>3.89</b>

表2：实验结果：10个真实世界基准数据集的平均测试精度±标准差。最好的结果被突出显示。结果 "\*"来自[7, 24]，结果 "+"来自[33]。NA表示报告的结果不可用，OOM表示内存不足。

## 7 未来的工作

本文定义的相似性矩阵和新的度量标准主要是捕捉聚集的节点特征的线性关系。但当特征向量中的非线性信息对分类很重要时，这有时可能是不够的。在未来，可以提出能够捕捉节点特征之间的非线性关系的相似性矩阵来定义新的同质性度量。

正如第4.3节中提出的限制，filterbank方法不能正确处理所有有害的异质性情况。在未来，我们需要用GNN探索和解决更多具有挑战性的异性图。

## 参考文献

- [1] S.Abu-El-Haija, B. Perozzi, A. Kapoor, N. Alipourfard, K. Lerman, H. Harutyunyan, G. Ver Steeg, and A. Galstyan. Mixhop: 通过稀疏化邻域混合的高阶图卷积架构。在 *国际机器学习会议上*, 第21-29页。PMLR, 2019年。
- [2] D. Bahdanau, K. Cho, and Y. Bengio. 通过联合学习对齐和翻译的神经机器翻译。 *arXiv 预印本 arXiv:1409.0473*, 2014。
- [3] P.W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, et al. Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv: 1806.01261*, 2018.
- [4] D. Bo, X. Wang, C. Shi, and H. Shen. 超越图卷积网络的低频信息. *arXiv 预印本 arXiv:2101.00797*, 2021.
- [5] M.M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst. Geometric deep learning: going beyond euclidean data. *arXiv*, abs/1611.08097, 2016.
- [6] M. Chen, Z. Wei, Z. Huang, B. Ding, and Y. Li. 简单和深入的图卷积网络。在 *国际机器学习会议上*, 第1725-1735页。PMLR, 2020.
- [7] E. Chien, J. Peng, P. Li, and O. Milenkovic. Adaptive universal generalized pagerank graph neural network. In *International Conference on Learning Representations*. <https://openreview.net/forum>, 2021.
- [8] F.R. Chung and F. C. Graham. *谱图理论*. 第92号。American Mathematical Soc., 1997.
- [9] M. Defferrard, X. Bresson, and P. Vandergheynst. 图上的卷积神经网络与快速局部频谱过滤。 *arXiv*, abs/1606.09375, 2016。
- [10] V.N. Ekambaram. *通过傅里叶透镜看图结构化数据*. 加州大学伯克利分校, 2014年。
- [11] M. Fey and J. E. Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019.
- [12] F. Gao, G. Wolf, and M. Hirn. 图形数据分析的几何散射。在 *国际机器学习会议上*, 第2122-2131页。PMLR, 2019.
- [13] A. Graves, A.-r. Mohamed, and G. Hinton. 用深度递归神经网络进行语音识别。在 *2013年IEEE声学、语音和信号处理国际会议上*, 第6645-6649页。Ieee, 2013.
- [14] W.L. Hamilton. 图表示学习. *Synthesis Lectures on Artificial Intelligence and Machine Learning*, 14(3):1-159, 2020.
- [15] W.L. Hamilton, R. Ying, and J. Leskovec. 大图上的归纳表征学习. *arXiv*, abs/1706.02216, 2017。
- [16] Y. Hou, J. Zhang, J. Cheng, K. Ma, R. T. Ma, H. Chen, and M.-C. Yang. Yang. 测量和改进图神经网络中的图信息使用。In *International Conference on Learning Representations*, 2019.
- [17] D.P. Kingma 和 J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [18] T.N. Kipf and M. Welling. 图卷积网络的半监督性分类. *arXiv*, abs/1609.02907, 2016。
- [19] J. Klicpera, A. Bojchevski, and S. Günnemann. 预测然后传播: 图形神经网络满足个性化的pagerank. *arXiv 预印本 arXiv:1810.05997*, 2018.
- [20] A. Krizhevsky, I. Sutskever, and G. E. Hinton. 用深度卷积神经网络进行图像网分类。In *Advances in neural information processing systems*, pages 1097-1105, 2012.

- [21] Y.LeCun, Y. Bengio, and G. Hinton.深度学习。《自然》，521（7553）：436，2015。
- [22] Y.LeCun, L. Bottou, Y. Bengio, P. Haffner, et al. 基于梯度的学习应用于文档识别。《IEEE 论文集》，86（11）：2278-2324，1998。



- [23] Q.Li, Z. Han, and X. Wu.对半监督学习的图卷积网络的深入了解。 *arXiv*, abs/1801.07606, 2018.
- [24] D.Lim, X. Li, F. Hohne, and S.-N. Lim.Lim.*arXiv preprint arXiv:2104.01404*, 2021.
- [25] V.Lingam, R. Ragesh, A. Iyer, and S. Sellamanickam.基于简单截断的svd模型的异质图上的节点分类. *arXiv预印本arXiv:2106.12807*, 2021.
- [26] M.Liu, Z. Wang, and S. Ji.非局部图神经网络。 *arXiv预印本arXiv:2005.14612*, 2020。
- [27] I.Loshchilov and F. Hutter.解耦权重衰减正则化.*arXiv预印本arXiv:1711.05101*, 2017.
- [28] S.Luan, M. Zhao, X.-W.Chang, and D. Precup.突破天花板： Stronger multi-scale deep graph convolutional networks. *arXiv preprint arXiv:1906.02174*, 2019.
- [29] S.Luan, M. Zhao, C. Hua, X.-W. Chang, and D. Precup.Chang, and D. Precup.完成缺失的一半：  
：*arXiv preprint arXiv:2008.08844*, 2020.
- [30] T.Maehara.Revisiting graph neural networks： All we have is low-pass filters. *arXiv preprint arXiv:1905.09550*, 2019.
- [31] M.McPherson, L. Smith-Lovin, and J. M. Cook.Birds of a feather： 社会网络中的同质性。  
*社会学年度评论*, 27 (1)： 415-444, 2001。
- [32] Y.Min, F. Wenkel, and G. Wolf.Scattering gcn： *arXiv preprint arXiv:2003.08414*, 2020.
- [33] H.Pei, B. Wei, K. C. -C.Chang, Y. Lei, and B. Yang.Geom-gcn： geometric graph convolutional networks. *arXiv preprint arXiv:2002.05287*, 2020.
- [34] F.Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini.图形神经网络模型。  
*IEEE transactions on neural networks*, 20(1):61-80, 2008.
- [35] P.Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio.Graph attention networks. *arXiv*, abs/1710.10903, 2017.
- [36] F.Wu, T. Zhang, A. H. d. Souza Jr, C. Fifty, T. Yu, and K. Q. Weinberger.Simplifying graph convolutional networks. *arXiv preprint arXiv:1902.07153*, 2019.
- [37] K.Xu, C. Li, Y. Tian, T. Sonobe, K.-i.Kawarabayashi, and S. Jegelka.图上的表征学习与跳跃的知识网络。 In J. Dy and A. Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5453-5462.PMLR, 2018年7月10-15日。
- [38] Y.Yan, M. Hashemi, K. Swersky, Y. Yang, and D. Koutra.同一枚硬币的两面： 图卷积神经网络中的异质性和过度平滑。 *arXiv预印本arXiv:2102.06462*, 2021。
- [39] J.Zhu, R. A. Rossi, A. Rao, T. Mai, N. Lipka, N. K. Ahmed, and D. Koutra.*arXiv preprint arXiv:2009.13566*, 2020.
- [40] J.Zhu, Y. Yan, L. Zhao, M. Heimann, L. Akoglu, and D. Koutra.超越图形神经网络中的同质性： 目前的限制和有效的的设计。 *Advances in Neural Information Processing Systems*, 33, 2020.

## 检查清单

1. 对于所有的作者...
  - (a) 摘要和引言中的主要主张是否准确反映了论文的贡献和范围? [是]。
  - (b) 你是否描述了你工作的局限性? [是]在附录G中, 我们讨论了高通滤波器无法解决的情况。
  - (c) 你是否讨论了你的工作可能产生的负面社会影响? [没有]这是在第8节, 我们没有想出重大的社会负面影响。
  - (d) 你是否已经阅读了伦理审查指南, 并确保你的论文符合指南的要求? [是]。
2. 如果你包括理论上的结果...
  - (a) 你是否说明了所有理论结果的全套假设? [是]在第3和第4节中, 我们主要定义了一个新的同质性度量, 后面有两个定理。
  - (b) 你是否包括所有理论结果的完整证明? [是]在附录B&D&E中, 我们证明了新的度量衡和两个定理。
3. 如果你做实验...
  - (a) 你是否包括重现主要经验-心理结果所需的代码、数据和说明(在补充材料中或作为URL)? [是]在细节上提供了设置, 在补充材料中提交了源代码。
  - (b) 你是否指定了所有的训练细节(例如, 数据分割, 超参数, 它们是如何选择的)? [是]在第6节中, 我们指定了模型的细节。
  - (c) 你是否报告了误差条(例如, 在多次运行实验后与随机种子的关系)? [是]我们包括运行次数的平均测试精度与标准偏差。
  - (d) 你是否包括计算总量和使用的资源类型(例如, GPU的类型, 内部集群, 或云供应商)? [是]我们在附录中包括了硬件细节, 这在计算上并不昂贵。
4. 如果你正在使用现有的资产(如代码、数据、模型)或策划/发布新的资产...
  - (a) 如果你的工作使用了现有的资产, 你是否引用了创造者的名字? [是]在第6节中, 我们在脚注中注明了数据集及其数据分割来源。
  - (b) 你有没有提到资产的许可证? [没有]
  - (c) 你是否在补充材料中或作为一个URL包括任何新的资产? [没有]
  - (d) 你是否讨论过是否以及如何从你正在使用/收集数据的人那里获得同意? [没有]
  - (e) 你是否讨论过你正在使用/收集的数据是否包含个人身份信息或攻击性内容? [没有] 不包括。
5. 如果你使用众包或进行以人为对象的研究...
  - (a) 你是否包括给参与者的说明全文和截图(如果适用)? [没有]没有包括。
  - (b) 你是否描述了任何潜在的参与者风险, 如果适用, 请附上机构审查委员会(IRB) 批准的链接? [没有] 不包括。
  - (c) 你是否包括支付给参与者的估计小时工资和用于参与者报酬的总金额? [没有] 没有包括。

## A 超参数和实验细节

### A.1 合成图上的GNN的超参数搜索范围

合成实验的超参数搜索范围				
模型和超参数	lr	重量_衰变	辍学者	隐藏的
MLP-1	0.05	{5-5, 1-4, 5-4}	-	-
SGC-1	0.05	{5-5, 1-4, 5-4}	-	-
ACM-SGC-1	0.05	{5-5, 1-4, 5-4}	{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}	-
MLP-2	0.05	{5-5, 1-4, 5-4}	{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}	64
GCN	0.05	{5-5, 1-4, 5-4}	{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}	64
ACM-GCN	0.05	{5-5, 1-4, 5-4}	{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}	64

表3：合成实验的超参数搜索范围

### A.2 消融研究中的GNN的超参数搜索范围

消融研究的超参数搜索范围				
模型和超参数	lr	weight_decay	辍学	隐藏
SGC-LP+HP	{0.01, 0.05, 0.1}	{0, 5-6, 1-5, 5-5, 1-4, 5-4, 1-3, 5-3, 1-2}	-	-
SGC-LP+Identity	{0.01, 0.05, 0.1}	{0, 5-6, 1-5, 5-5, 1-4, 5-4, 1-3, 5-3, 1-2}	-	-
ACM-SGC-没有自适应混合	{0.01, 0.05, 0.1}	{0, 5-6, 1-5, 5-5, 1-4, 5-4, 1-3, 5-3, 1-2}	{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}	-
GCN-LP+HP	{0.01, 0.05, 0.1}	{0, 5-6, 1-5, 5-5, 1-4, 5-4, 1-3, 5-3, 1-2}	{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}	64
GCN-LP+Identity	{0.01, 0.05, 0.1}	{0, 5-6, 1-5, 5-5, 1-4, 5-4, 1-3, 5-3, 1-2}	{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}	64
ACM-GCN-没有自适应混合	{0.01, 0.05, 0.1}	{0, 5-6, 1-5, 5-5, 1-4, 5-4, 1-3, 5-3, 1-2}	{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}	64

表4：消融研究的超参数搜索范围

### A.3 真实世界数据集上GNN的超参数搜索范围

见表5。

### A.4 基线和ACM(II)-GNN在真实世界任务中的最佳超参数

见表6和7。

### A.5 ACM和ACMII框架的实施细节

在ACM(II)框架中，我们首先在输入数据上使用dropout操作。ACM(II)-GCN和ACM(II)-snowball的实现很简单，但SGC-1、SGC-2、GCNII和GCNII\*却不能在ACM(II)框架下应用，我们将做如下解释。

- SGC-1和SGC-2: SGC不包含非线性, 所以步骤1中的选项1和选项2对于ACM-SGC和ACMII-SGC是一样的。因此, 我们只实施ACM-SGC。
- GCNII and GCNII\*:

$$\text{GCCII: } \mathbf{H}^{(A+1)} = \sigma \left( (1 - \alpha_A) \mathbf{A} \mathbf{H}^{(A)} + \alpha H_A^{(0)} (1 - \beta_A) \mathbf{I}_n + \beta W_A^{(A)} \right)$$

$$\text{GCCII*}: \mathbf{H}_A^{(A+1)} = \sigma \left( (1 - \alpha_A) \mathbf{A} \mathbf{H}^{(A)} (1 - \beta_A) \mathbf{I}_n + \beta W_A^{(A)} + \alpha H_A^{(0)} (1 - \beta_A) \mathbf{I}_n + \beta^{(A)} \right) \quad 2$$

如果不做重大修改, GCNII和GCNII\*很难被放入ACMII框架中。在ACMII框架中, 在应用

算子 $\hat{A}$ 之前, 我们首先实现一个非线性特征操作

<sup>A</sup>但在GCNII和GCNII\*中, 在乘以 $W_A^{(A)}$  (或 $W_1^A, W_2^A$ ) 以提取特征之前, 我们

需要再增加一个包括 $H^{(0)}$ 的项, 这些项没有被 $\hat{A}$ 过滤。这与以下情况不相容  
ACMII框架, 因此, 我们没有在ACMII框架中实现GCNII和GCNII\*。

开源代码将很快发布。

模型和超参数	lr	重量_衰变	辍学者	隐藏的	lambda	$\hat{\alpha}\hat{\alpha}\hat{\alpha}$	头	层数	JK型
H2GCN	0.01	0.001	{0, 0.5}	{8, 16, 32, 64}	-	-	-	{1, 2}	-
搅拌机	0.01	0.001	0.5	{8, 16, 32}	-	-	-	{2, 3}	-
GCN+JK	{0.1, 0.01, 0.001}	0.001	0.5	{4, 8, 16, 32, 64}	-	-	-	2	{pos(192,252)}最大的, 猫
GAT+JK	{0.1, 0.01, 0.001}	0.001	0.5	{4, 8, 12, 32}	-	-	{2,4,8}	2	{pos(192,252)}最大的, 猫
GCNII, GCNII*	0.01	{0, 5-6, 1-5, 5-5, 1-4, 5-4, 1-3}为Deezer-欧洲和{0, 5-6, 1-5, 5-5, 1-4, 5-4, 1-3, 5e-3, 1e-2}代表其他	0.5	64	{0.5, 1, 1.5}	{0.1,0.2,0.3,0.4,0.5}		{4, 8, 16, 32}为Deezer-欧洲和{4, 8, 16, 32, 64}为其他	-
基线: {SGC-1, SGC-2, GCN, 雪球-2, 雪球-3, fagcn}; acm- {sgc-1, SGC-2, GCN, Snowball-2, Snowball-3}; ACMII- {SGP-1, SGC-2, GCN, 图谱}	{0.002, 0.01, 0.05}, Deezer-5欧洲5e-3和{0.01, 0.05, 0.1}为其他人{0.01,0.05, 0.1}	{0, 5-6, 1-5, 5-5, 1-4, 5-4, 1-3}为Deezer-欧洲和{0, 5-6, 1-5, 5-5, 1e-4, 5e-4, 1e-3, 1e-2}为其他0	{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6,0.7, 0.8, 0.9}	64	-	-	-	-	-
acm- {gcni, gcni*}	0.01	{0, 5-6, 1-5, 5-5, 1-4, 5-4, 1-3}为Deezer-欧洲和{0, 5-6, 1-5, 5-5, 1-4, 5-4, 1-3, 5e-3, 1e-2}代表其他	{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}	64	-	-	-	{1,2,3,4}	-

表5：现实世界数据集的超参数搜索范围

## A.6 计算资源

在合成数据集和真实世界数据集的所有实验中，我们使用NVidia V100 GPU，16/32GB GPU内存，8核CPU，16G内存。软件的实现是基于PyTorch和PyTorch Geometric[11]。

基线GNN的超参数														
数据集	模型和超参数	lr	重量_衰变	锯齿学者	隐藏的_#层数	盖特头	JK型	lambda	AAA	结果	标准	平均纪元时间/平均总时间		
康奈尔大学	SGC-1	0.05	1.00E-02	0	64	-	-	-	-	70.98	8.39	2.53ms/0.51s		
	SGC-2	0.05	1.00E-03	0	64	-	-	-	-	72.62	9.92	2.46ms/0.53s		
	GCN	0.1	5.00E-03	0.5	64	2	-	-	-	82.46	3.11	3.67ms/0.74s		
	雪球-2	0.01	5.00E-03	0.4	64	2	-	-	-	82.62	2.34	4.24ms/0.87s		
	雪球-3	0.01	5.00E-03	0.4	64	3	-	-	-	82.95	2.1	6.66ms/1.36s		
	GCNII	0.01	1.00E-03	0.5	64	16	-	-	0.5	89.18	3.96	25.41ms/8.11s		
	GCNII*	0.01	1.00E-03	0.5	64	8	-	-	0.5	90.49	4.45	15.35ms/4.05s		
	FAGCN	0.01	1.00E-04	0.7	32	2	-	-	-	88.03	5.6	8.1ms/3.8858s		
	H2GCN	0.01	0.001	0.5	32	1	-	-	-	60.33	28.53	10.379ms/2.105s		
	搅拌站	0.01	0.001	0.5	64	1	-	-	-	86.23	4.71	4.381ms/1.123s		
	GCN+JK	0.1	0.001	0.5	64	2	-	猫	-	66.56	13.82	5.589ms/1.227s		
	GAT+JK	0.1	0.001	0.5	32	2	8	最大	-	74.43	10.24	10.725ms/2.478s		
威斯康星州	SGC-1	0.05	5.00E-03	0	64	-	-	-	-	70.38	2.85	2.83ms/0.57s		
	SGC-2	0.1	1.00E-03	0	64	-	-	-	-	74.75	2.89	2.14ms/0.43s		
	GCN	0.1	1.00E-03	0.7	64	2	-	-	-	75.5	2.92	3.74ms/0.76s		
	雪球-2	0.1	1.00E-03	0.5	64	2	-	-	-	74.88	3.42	3.73ms/0.76s		
	雪球-3	0.05	5.00E-04	0.8	64	3	-	-	-	69.5	5.01	5.46ms/1.12s		
	GCNII	0.01	1.00E-03	0.5	64	8	-	-	0.5	83.25	2.69			
	GCNII*	0.01	1.00E-03	0.5	64	4	-	-	1.5	89.12	3.06	9.26ms/1.96s		
	FAGCN	0.05	1.00E-04	0	32	2	-	-	-	89.75	6.37	12.9ms/4.6359s		
	搅拌站	0.01	0.001	0.5	16	2	-	-	-	77.25	7.80	10.281ms/2.095s		
	H2GCN	0.01	0.001	0.5	32	1	-	-	-	87.5	1.77	4.324ms/1.134s		
	GCN+JK	0.1	0.001	0.5	32	2	-	猫	-	62.5	15.75	5.117ms/1.049s		
	GAT+JK	0.1	0.001	0.5	4	2	8	最大	-	69.5	3.12	10.762ms/2.25s		
APNP	0.05	0.001	0.5	64	2	-	-	-	92	3.59	10.303ms/2.104s			
辽宁沈阳	0.05	0.001	0.5	256	2	-	-	-	93.75	2.37	11.856ms/2.415s			
德州	SGC-1	0.05	1.00E-03	0	64	-	-	-	-	83.28	5.43	2.55ms/0.54s		
	SGC-2	0.01	1.00E-03	0	64	-	-	-	-	81.31	3.3	2.61ms/2.53s		
	GCN	0.05	1.00E-02	0.9	64	2	-	-	-	83.11	3.2	3.59ms/0.73s		
	雪球-2	0.05	1.00E-02	0.9	64	2	-	-	-	83.11	3.2	3.98ms/0.82s		
	雪球-3	0.05	1.00E-02	0.9	64	3	-	-	-	83.11	3.2	5.66ms/1.12s		
	GCNII	0.01	1.00E-04	0.5	64	4	-	-	1.5	82.46	4.58			
	GCNII*	0.01	1.00E-04	0.5	64	8	-	-	0.5	88.52	3.02	15.64ms/3.47s		
	FAGCN	0.01	5.00E-04	0	32	2	-	-	-	88.85	4.39			
	搅拌站	0.01	0.001	0.5	32	2	-	-	-	76.39	7.66	11.099ms/2.329s		
	H2GCN	0.01	0.001	0.5	64	1	-	-	-	85.90	3.53	4.197ms/0.95s		
	GCN+JK	0.1	0.001	0.5	32	2	-	猫	-	80.66	1.91	5.28ms/1.085s		
	GAT+JK	0.1	0.001	0.5	8	2	2	猫	-	75.41	7.18	10.937ms/2.402s		
电影	SGC-1	0.01	5.00E-06	0	64	-	-	-	-	25.26	1.18	3.18ms/0.70s		
	SGC-2	0.01	5.00E-06	0	64	-	-	-	-	28.81	1.11	2.13ms/0.43s		
	GCN	0.1	5.00E-04	0	64	2	-	-	-	35.51	0.99	4.86ms/0.99s		
	雪球-2	0.1	5.00E-04	0	64	2	-	-	-	35.97	0.66	5.59ms/1.14s		
	雪球-3	0.1	5.00E-04	0.2	64	3	-	-	-	36	1.36	7.89ms/1.60s		
	GCNII	0.01	1.00E-04	0.5	64	8	-	-	1.5	40.82	1.79	15.85ms/3.22s		
	GCNII*	0.01	1.00E-06	0.5	64	4	-	-	1	41.54	0.99			
	FAGCN	0.01	5.00E-05	0.6	32	2	-	-	-	31.59	1.37	45.4ms/11.107s		
	搅拌站	0.01	0.001	0.5	8	3	8	最大	-	33.13	2.40	17.651ms/3.566s		
	H2GCN	0.01	0.001	0	64	1	8	最大	-	38.85	1.17	8.101ms/1.695s		
	GCN+JK	0.1	0.001	0.5	64	2	8	猫	-	32.72	2.62	8.946ms/1.807s		
	GAT+JK	0.001	0.001	0.5	32	2	4	猫	-	35.41	0.97	20.726ms/4.187s		
变色龙	SGC-1	0.1	5.00E-06	0	64	-	-	-	-	64.86	1.81	3.48ms/2.96s		
	SGC-2	0.1	0.00E+00	0	64	-	-	-	-	62.67	2.41	4.43ms/1.12s		
	GCN	0.01	1.00E-05	0.9	64	2	-	-	-	64.18	2.62	4.96ms/1.18s		
	雪球-2	1.00E-01	1.00E-05	0.9	64	2	-	-	-	64.99	2.39	4.96ms/1.00s		
	雪球-3	0.1	5.00E-06	0.9	64	3	-	-	-	65.49	1.64	7.44ms/1.50s		
	GCNII	0.01	5.00E-06	0.5	64	4	-	-	0.5	60.35	2.7	9.76ms/2.26s		
	GCNII*	0.01	5.00E-04	0.5	64	4	-	-	1.5	62.8	2.87	10.40ms/2.17s		
	FAGCN	0.002	1.00E-04	0	32	2	-	-	-	49.47	2.84	8.4ms/13.8696s		
	搅拌站	0.01	0.001	0.5	16	2	8	最大	-	36.28	10.2	11.372ms/2.297s		
	H2GCN	0.01	0.001	0	32	1	8	最大	-	52.3	0.48	4.059ms/0.82s		
	GCN+JK	0.001	0.001	0.5	32	2	8	猫	-	64.68	2.85	5.211ms/1.053s		
	GAT+JK	0.001	0.001	0.5	4	2	8	最大	-	68.14	1.18	13.772ms/2.788s		
松鼠	SGC-1	0.05	0.00E+00	0	64	-	-	-	-	47.62	1.27	4.65ms/1.44s		
	SGC-2	0.1	0.00E+00	0.9	64	-	-	-	-	41.25	1.4	35.06ms/7.81s		
	GCN	0.01	5.00E-05	0.7	64	2	-	-	-	44.76	1.39	8.41ms/2.50s		
	雪球-2	0.1	0.00E+00	0.9	64	2	-	-	-	47.88	1.23	8.96ms/1.92s		
	雪球-3	0.1	0.00E+00	0.8	64	3	-	-	-	48.25	0.94	14.00ms/2.90s		
	GCNII	0.01	1.00E-04	0.5	64	4	-	-	1.5	38.81	1.97	13.35ms/2.70s		
	GCNII*	0.01	5.00E-04	0.5	64	4	-	-	1.5	38.31	1.3	13.81ms/2.78s		
	FAGCN	0.05	1.00E-04	0	32	2	-	-	-	42.24	1.2	16ms/6.7961s		
	搅拌站	0.01	0.001	0.5	32	2	-	-	-	24.55	2.6	17.634ms/3.562s		
	H2GCN	0.01	0.001	0	16	1	-	-	-	30.39	1.22	9.315ms/1.882s		
	GCN+JK	0.001	0.001	0.5	32	2	-	最大	-	53.4	1.9	14.321ms/2.905s		
	GAT+JK	0.001	0.001	0.5	8	2	4	最大	-	52.28	3.61	29.097ms/5.878s		
Cora	SGC-1	0.1	5.00E-06	0	64	-	-	-	-	85.12	1.64	3.47ms/1.55s		
	SGC-2	0.1	1.00E-05	0	64	-	-	-	-	85.48	1.48	2.91ms/6.85s		
	GCN	0.1	5.00E-04	0.2	64	2	-	-	-	87.78	0.96	4.24ms/0.86s		
	雪球-2	0.1	5.00E-04	0.1	64	2	-	-	-	88.64	1.15	4.65ms/0.94s		
	雪球-3	0.05	1.00E-03	0.6	64	3	-	-	-	89.33	1.3	6.41ms/1.32s		
	GCNII	0.01	1.00E-04	0.5	64	16	-	-	0.5	88.98	1.33			
	GCNII*	0.01	5.00E-04	0.5	64	4	-	-	0.5	88.93	1.37	10.16ms/2.24s		
	FAGCN	0.05	5.00E-04	0	32	2	-	-	-	88.85	1.36	8.4ms/3.3183s		
	搅拌站	0.01	0.001	0.5	16	2	-	-	-	65.65	11.31	11.177ms/2.278s		
	H2GCN	0.01	0.001	0	32	1	-	-	-	87.52	0.61	4.335ms/1.209s		
	GCN+JK	0.001	0.001	0.5	64	2	-	猫	-	86.90	1.51	6.656ms/1.346s		
	GAT+JK	0.001	0.001	0.5	32	2	2	猫	-	89.52	0.43	12.91ms/2.608s		
CiteSeer	SGC-1	0.1	5.00E-04	0	64	-	-	-	-	79.66	0.75	3.43ms/7.30s		
	SGC-2	0.01	5.00E-04	0.9	64	-	-	-	-	80.75	1.15	5.33ms/4.40s		
	GCN	0.1	1.00E-03	0.9	64	2	-	-	-	81.39	1.23	4.18ms/0.86s		
	雪球-2	0.1	1.00E-03	0.8	64	2	-	-	-	81.53	1.71	5.19ms/1.11s		
	雪球-3	0.1	1.00E-03	0.8	64	3	-	-	-	80.93	1.32	7.64ms/1.69s		
	GCNII	0.01	1.00E-03	0.5	64	16	-	-	0.5	81.58	1.3			
	GCNII*	0.01	1.00E-03	0.5	64	16	-	-	0.5	81.83	1.78	32.50ms/10.29s		
	FAGCN	0.05	5.00E-04	0	32	2	-	-	-	82.37	1.46	9.4ms/4.7648s		
	搅拌站	0.01	0.001	0.5	16	2	-	-	-	49.52	13.35	13.793ms/2.786s		
	H2GCN	0.01	0.001	0	8	1	-	-	-	79.97	0.69	5.794ms/3.049s		
	GCN+JK	0.001	0.001	0.5	32	2	-	最大	-	73.77	1.85	5.264ms/1.063s		
	GAT+JK	0.001	0.001	0.5	8	2	4	最大	-	74.49	2.76	12.326ms/2.49s		
医学博士	SGC-1	0.05	5.00E-06	0.3	64	-	-	-	-	87.75	0.88	6.04ms/2.61s		
	SGC-2	0.05	5.00E-05	0.1	64	-	-	-	-	88.79	0.5	8.62ms/3.18s		
	GCN	0.1	5.00E-05	0.6	64	2	-	-	-	88.9	0.32	5.08ms/1.03s		
	雪球-2	0.1	5.00E-04	0	64	2	-	-	-	89.04	0.49	5.68ms/1.19s		
	雪球-3	0.1	5.00E-06	0	64	3	-	-	-	88.8	0.82	8.54ms/1.75s		
	GCNII	0.01	1.00E-03	0.5	64	16	-	-	0.5	89.8	0.3	10.98ms/3.21s		
	GCNII*	0.01	1.00E-06	0.5	64	4	-	-	0.5	81.83	1.78	11.47ms/3.24s		
	FAGCN	0.05	5.00E-04	0	32	2	-	-	-	89.98	0.54	14.5ms/6.411s		
	搅拌站	0.01	0.001	0.5	16	2	-	-	-	87.04	4.10	17.459ms/3.527s		
	H2GCN	0.01	0.001	0	64	1	-	-	-	87.78	0.28	8.039ms/2.28s		
	GCN+JK	0.01	0.001	0.5	32	2	-	猫	-	90.09	0.68	12.001ms/2.424s		
	GAT+JK	0.1	0.001	0.5	8	2	4							

表6：基线模型的超参数

ACM-GNNs和ACMII-GNNs的超参数														
数据集	模型和超参数	lr	重量	衰变	辍学者	隐藏的	#层数	盖特头	JK型	lambda	AAA	结果	标准	平均纪元时间/平均总时间
康奈尔大学	ACM-SGC-1	0.01	5.00E-03	0.6	64	-	-	-	-	-	-	93.77	1.91	5.53ms/2.31s
	ACM-SGC-2	0.01	5.00E-03	0.6	64	-	-	-	-	-	-	93.77	2.17	4.73ms/1.87s
	ACM-GCN	0.05	1.00E-02	0.2	64	2	-	-	-	-	-	94.75	3.8	8.25ms/1.69s
	ACMII-GCN	0.1	1.00E-02	0.5	64	2	-	-	-	-	-	95.25	2.79	8.43ms/1.71s
	ACM-GCNII	0.01	1.00E-03	0.5	64	1	-	-	-	0.5	0.4	92.62	3.13	6.81ms/1.43s
	ACM-GCNII*	0.01	5.00E-04	0.5	64	1	-	-	-	0.5	0.1	93.44	2.74	6.76ms/1.39s
	ACM-雪球-2	0.05	1.00E-02	0.2	64	2	-	-	-	-	-	95.08	3.11	9.15ms/1.86s
	ACM-雪球-3	0.1	1.00E-02	0.4	64	3	-	-	-	-	-	94.26	2.57	13.20ms/2.68s
	ACMII-雪球-2	0.05	1.00E-02	0.6	64	2	-	-	-	-	-	95.25	1.55	8.23ms/1.72s
	ACMII-雪球-3	0.05	1.00E-02	0.7	64	3	-	-	-	-	-	93.61	2.79	11.70ms/2.37s
威斯康星州	ACM-SGC-1	0.05	5.00E-03	0.7	64	-	-	-	-	-	-	93.25	2.92	5.96ms/1.34s
	ACM-SGC-2	0.1	5.00E-03	0.2	64	-	-	-	-	-	-	94	2.61	4.60ms/0.95s
	ACM-GCN	0.1	5.00E-03	0	64	2	-	-	-	-	-	95.75	2.03	8.11ms/1.64s
	ACMII-GCN	0.1	1.00E-02	0.2	64	2	-	-	-	-	-	96.62	2.44	8.28ms/1.68s
	ACM-GCNII	0.01	5.00E-03	0.5	64	1	-	-	-	1	0.1	94.63	2.96	9.31ms/2.19s
	ACM-GCNII*	0.01	1.00E-03	0.5	64	1	-	-	-	1.5	0.4	94.37	2.81	7.11ms/1.45s
	ACM-雪球-2	0.1	5.00E-03	0.1	64	2	-	-	-	-	-	96.38	2.59	8.63ms/1.74s
	ACM-雪球-3	0.05	1.00E-02	0.3	64	3	-	-	-	-	-	96.62	1.86	12.79ms/2.58s
	ACMII-雪球-2	0.1	1.00E-02	0.1	64	2	-	-	-	-	-	96.63	2.24	8.11ms/1.65s
	ACMII-雪球-3	0.1	5.00E-03	0.1	64	3	-	-	-	-	-	97	2.63	12.38ms/2.51s
德州	ACM-SGC-1	0.01	5.00E-03	0.6	64	-	-	-	-	-	-	93.61	1.55	5.43ms/2.18s
	ACM-SGC-2	0.05	5.00E-03	0.4	64	-	-	-	-	-	-	93.44	2.54	4.59ms/1.01s
	ACM-GCN	0.05	1.00E-02	0.6	64	2	-	-	-	-	-	94.92	2.88	8.33ms/1.70s
	ACMII-GCN	0.1	5.00E-03	0.4	64	2	-	-	-	-	-	95.08	2.54	8.49ms/1.72s
	ACM-GCNII	0.01	1.00E-03	0.5	64	1	-	-	-	0.5	0.4	92.46	1.97	6.47ms/1.36s
	ACM-GCNII*	0.01	1.00E-03	0.5	64	1	-	-	-	0.5	0.4	93.28	2.79	7.03ms/1.45s
	ACM-雪球-2	0.05	1.00E-02	0.1	64	2	-	-	-	-	-	95.74	2.22	8.35ms/1.71s
	ACM-雪球-3	0.01	5.00E-03	0.6	64	3	-	-	-	-	-	94.75	2.41	12.56ms/2.63s
	ACMII-雪球-2	0.1	1.00E-02	0.4	64	2	-	-	-	-	-	95.25	1.55	9.74ms/1.97s
	ACMII-雪球-3	0.05	1.00E-02	0.6	64	3	-	-	-	-	-	94.75	3.09	11.91ms/2.42s
电影	ACM-SGC-1	0.05	5.00E-05	0.7	64	-	-	-	-	-	-	39.33	1.25	5.21ms/2.33s
	ACM-SGC-2	0.1	5.00E-05	0.7	64	-	-	-	-	-	-	40.13	1.21	12.41ms/4.87s
	ACM-GCN	0.1	5.00E-04	0.5	64	2	-	-	-	-	-	41.62	1.15	10.72ms/2.66s
	ACMII-GCN	0.1	5.00E-04	0.5	64	2	-	-	-	-	-	41.24	1.16	10.51ms/2.44s
	ACM-GCNII	0.01	0.00E+00	0.5	64	3	-	-	-	1.5	0.2	41.37	1.37	13.65ms/2.74s
	ACM-GCNII*	0.01	1.00E-05	0.5	64	3	-	-	-	1.5	0.1	41.27	1.24	14.98ms/3.01s
	ACM-雪球-2	0.1	5.00E-03	0	64	2	-	-	-	-	-	41.4	1.23	10.30ms/2.08s
	ACM-雪球-3	0.05	1.00E-02	0	64	3	-	-	-	-	-	41.27	0.8	16.43ms/3.52s
	ACMII-雪球-2	0.1	5.00E-03	0	64	2	-	-	-	-	-	41.1	0.75	10.74ms/2.19s
	ACMII-雪球-3	0.05	5.00E-03	0.2	64	3	-	-	-	-	-	40.31	1.6	16.31ms/3.29s
变色龙	ACM-SGC-1	0.1	5.00E-06	0.9	64	-	-	-	-	-	-	63.68	1.62	5.41ms/1.21s
	ACM-SGC-2	0.1	5.00E-06	0.9	64	-	-	-	-	-	-	60.48	1.55	7.86ms/1.81s
	ACM-GCN	0.01	5.00E-05	0.8	64	2	-	-	-	-	-	68.18	1.67	10.55ms/3.12s
	ACMII-GCN	0.05	5.00E-05	0.7	64	2	-	-	-	-	-	68.38	1.36	10.90ms/2.39s
	ACM-GCNII	0.01	5.00E-06	0.5	64	4	-	-	-	0.5	0.1	58.73	2.52	18.31ms/3.68s
	ACM-GCNII*	0.01	1.00E-03	0.5	64	1	-	-	-	1	0.1	61.66	2.29	6.68ms/1.40s
	ACM-雪球-2	0.05	5.00E-05	0.7	64	2	-	-	-	-	-	68.51	1.7	9.92ms/2.06s
	ACM-雪球-3	0.01	1.00E-04	0.7	64	3	-	-	-	-	-	68.4	2.05	14.49ms/3.15s
	ACMII-雪球-2	0.1	5.00E-05	0.6	64	2	-	-	-	-	-	67.83	2.63	9.99ms/2.10s
	ACMII-雪球-3	0.05	1.00E-04	0.7	64	3	-	-	-	-	-	67.53	2.83	15.03ms/3.29s
松鼠	ACM-SGC-1	0.05	0.00E+00	0.9	64	-	-	-	-	-	-	46.4	1.13	6.96ms/2.16s
	ACM-SGC-2	0.05	0.00E+00	0.9	64	-	-	-	-	-	-	40.91	1.39	35.20ms/10.66s
	ACM-GCN	0.05	5.00E-06	0.6	64	2	-	-	-	-	-	58.02	1.86	14.35ms/2.98s
	ACMII-GCN	0.05	0.00E+00	0.7	64	2	-	-	-	-	-	53.76	1.63	14.08ms/3.39s
	ACM-GCNII	0.01	1.00E-05	0.5	64	4	-	-	-	0.5	0.1	40.9	1.58	20.72ms/4.17s
	ACM-GCNII*	0.01	1.00E-03	0.5	64	4	-	-	-	0.5	0.3	38.32	1.5	21.78ms/4.38s
	ACM-雪球-2	0.05	5.00E-06	0.6	64	2	-	-	-	-	-	55.97	2.03	15.38ms/3.15s
	ACM-雪球-3	0.01	1.00E-04	0.6	64	3	-	-	-	-	-	55.73	2.39	26.15ms/5.94s
	ACMII-雪球-2	0.1	5.00E-06	0.6	64	2	-	-	-	-	-	53.48	0.6	15.54ms/3.19s
	ACMII-雪球-3	0.05	5.00E-05	0.5	64	3	-	-	-	-	-	52.31	1.57	26.24ms/5.30s
Cora	ACM-SGC-1	0.01	5.00E-06	0.9	64	-	-	-	-	-	-	86.63	1.13	6.00ms/7.40s
	ACM-SGC-2	0.1	5.00E-05	0.6	64	-	-	-	-	-	-	87.64	0.99	4.85ms/1.17s
	ACM-GCN	0.1	5.00E-03	0.5	64	2	-	-	-	-	-	88.62	1.22	8.84ms/1.81s
	ACMII-GCN	0.1	5.00E-03	0.4	64	2	-	-	-	-	-	89	0.72	8.93ms/1.83s
	ACM-GCNII	0.01	1.00E-03	0.5	64	3	-	-	-	1	0.2	89.1	1.61	14.07ms/3.04s
	ACM-GCNII*	0.01	1.00E-02	0.5	64	4	-	-	-	1	0.2	89	1.35	11.36ms/2.48s
	ACM-雪球-2	0.05	1.00E-03	0.6	64	2	-	-	-	-	-	88.83	1.49	9.34ms/1.92s
	ACM-雪球-3	0.1	1.00E-02	0.3	64	3	-	-	-	-	-	89.59	1.58	13.33ms/2.75s
	ACMII-雪球-2	0.1	5.00E-03	0.5	64	2	-	-	-	-	-	88.95	1.04	9.29ms/1.90s
	ACMII-雪球-3	0.1	5.00E-03	0.5	64	3	-	-	-	-	-	89.36	1.26	14.18ms/2.89s
CiteSeer	ACM-SGC-1	0.01	5.00E-04	0.9	64	-	-	-	-	-	-	80.96	0.93	5.90ms/4.31s
	ACM-SGC-2	0.05	5.00E-04	0.9	64	-	-	-	-	-	-	80.93	1.16	5.01ms/1.42s
	ACM-GCN	0.05	5.00E-03	0.7	64	2	-	-	-	-	-	81.68	0.97	11.35ms/2.57s
	ACMII-GCN	0.05	5.00E-05	0.7	64	2	-	-	-	-	-	81.58	1.77	9.55ms/1.94s
	ACM-GCNII	0.01	1.00E-02	0.5	64	3	-	-	-	0.5	0.3	82.28	1.12	15.61ms/3.56s
	ACM-GCNII*	0.01	1.00E-02	0.5	64	3	-	-	-	0.5	0.5	81.69	1.25	15.56ms/3.61s
	ACM-雪球-2	0.05	5.00E-03	0.7	64	2	-	-	-	-	-	81.58	1.23	11.14ms/2.50s
	ACM-雪球-3	0.01	5.00E-03	0.9	64	3	-	-	-	-	-	81.32	0.97	15.91ms/5.36s
	ACMII-雪球-2	0.05	5.00E-03	0.7	64	2	-	-	-	-	-	82.07	1.04	10.97ms/2.55s
	ACMII-雪球-3	0.05	1.00E-04	0.6	64	3	-	-	-	-	-	81.56	1.15	14.95ms/3.03s
医学博士	ACM-SGC-1	0.05	5.00E-06	0.3	64	-	-	-	-	-	-	87.75	0.88	6.04ms/2.61s
	ACM-SGC-2	0.05	5.00E-05	0.1	64	-	-	-	-	-	-	88.79	0.5	8.62ms/3.18s
	ACM-GCN	0.1	5.00E-04	0.2	64	2	-	-	-	-	-	90.54	0.63	10.20ms/2.08s
	ACMII-GCN	0.1	5.00E-04	0.2	64	2	-	-	-	-	-	90.74	0.5	10.20ms/2.07s
	ACM-GCNII	0.01	1.00E-04	0.5	64	3	-	-	-	1.5	0.5	90.12	0.4	15.07ms/3.35s
	ACM-GCNII*	0.01	1.00E-04	0.5	64	3	-	-	-	1.5	0.5	90.18	0.51	16.62ms/3.72s
	ACM-雪球-2	0.1	1.00E-04	0.3	64	2	-	-	-	-	-	90.81	0.52	11.52ms/2.36s
	ACM-雪球-3	0.05	1.00E-03	0.2	64	3	-	-	-	-	-	91.44	0.59	18.06ms/3.69s
	ACMII-雪球-2	0.1	1.00E-04	0.3	64	2	-	-	-	-	-	90.56	0.39	11.74ms/2.39s
	ACMII-雪球-3	0.1	5.00E-04											



---

表7:  
ACM-  
GNNs和  
ACMII-  
GNNs的  
超参数

## B (5)中梯度计算的细节

### B.1 矩阵形式的推导

在输出层，我们有

$$\mathbf{1}^T \odot \exp(Y) > 0$$

$$Y = \text{softmax}(XW) \equiv \text{softmax}(Y) = \exp(Y) \mathbf{1}_C$$

$$L = -\text{trace}(Z^T \log Y)$$

其中， $\mathbf{1}_C \in \mathbb{R}^{C \times 1}$ ， $(\cdot)^{-1}$  是点-明智反函数， $Y$  的每个元素都是正数。那么

$$dL = -\text{trace } Z^T ((Y)^{-1} \odot dY) = -\text{trace } Z^T (\text{softmax}(Y))^{-1} \odot d \text{softmax}(Y)$$

请注意

$$\begin{aligned} d \text{softmax}(Y) &= -\exp(Y) \mathbf{1}_C^T \odot [(\exp(Y) \odot dY) \mathbf{1}_C^T] \odot \exp(Y) \\ &\quad + \exp(Y) \mathbf{1}_C^T \odot \exp(Y) \mathbf{1}_C^T \odot (dY) \\ &= -\text{softmax}(Y) \odot \exp(Y) \mathbf{1}_C^T \odot [(\exp(Y) \odot dY) \mathbf{1}_C^T] \\ &\quad + \text{softmax}(Y) \odot dY \\ &= \text{softmax}(Y) \odot -\exp(Y) \mathbf{1}_C^T \odot (\exp(Y) \odot dY) \mathbf{1}_C^T + dY \end{aligned}$$

然后

、

$$\begin{aligned} dL &= -\text{跟踪 } Z^T (\text{softmax}(Y))^{-1} \odot \text{softmax}(Y) \odot -\exp(Y) \mathbf{1}_C^T \odot (\exp(Y) \odot dY) \mathbf{1}_C^T + dY \\ &= -\text{trace } Z^T - \exp(Y) \mathbf{1}_C^T \odot (\exp(Y) \odot dY) \mathbf{1}_C^T + dY \\ &= \text{跟踪 } Z \odot \exp(Y) \mathbf{1}_C^T \odot \mathbf{1}_C^T \odot [\exp(Y) \odot dY] - Z^T dY \\ &= \text{trace } \exp(Y) \odot Z \odot \exp(Y) \mathbf{1}_C^T \odot \mathbf{1}_C^T dY - Z^T dY \\ &= \text{trace } \exp(Y) \odot \exp(Y) \mathbf{1}_C^T \odot \mathbf{1}_C^T dY - Z^T dY \\ &= \text{trace } (\text{softmax}(Y) - Z)^T dY \end{aligned}$$

其中第4个方程由于 $Z \odot \exp(Y) \mathbf{1}_C^T$ 而成立。 $\mathbf{1}_C^T \odot \mathbf{1}_C^T = \exp(Y) \mathbf{1}_C^T$ 。因此、  
我们有

$$\frac{dL}{dY} = \text{softmax}(Y) - Z = Y - Z$$

对于 $Y$ 和 $W$ ，我们有

$$dY = \hat{A} X dW, \quad dL = \text{跟踪 } \frac{dL}{dY}^T dY = \text{追 } \frac{dL}{dY}^T \hat{A} X dW = \text{追 } \frac{dL}{dW}^T dW$$

为了获得  $dW$  我们有

$$\frac{dL}{dW} = X^T A^T \frac{dL}{dY} = x^T a^t (y - z) \quad (14)$$

## B.2 从成分上推导

记为  $\tilde{X} = XW$ 。我们将L重写如下：

$$\begin{aligned}
 L &= -\text{trace } Z^T \log (\exp(Y') \mathbf{1}_c \mathbf{1}^T) \mathbf{1}_c^{-1} \odot \exp(Y') \\
 &= -\text{trace } Z^T - \log(\exp(Y') \mathbf{1}_c \mathbf{1}^T) + Y' \\
 &= -\text{trace } Z^T Y' + \text{trace } Z^T \log \exp(Y') \mathbf{1}_c \mathbf{1}_c^T \\
 &= -\text{trace } Z^T XW + \text{trace } Z^T \log \exp(Y') \mathbf{1}_c \mathbf{1}_c^T \\
 &= -\text{trace } Z^T XW + \text{trace } \mathbf{1}^T \log (\exp(Y') \mathbf{1}_c) \mathbf{1}_c \quad \square \\
 &= - \sum_{i=1}^N \sum_{j \in N_i} \hat{A}_{i,j} Z_{i,:} X_{j,:}^T + \sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} X_{j,c} \quad \square \\
 &= - \sum_{i=1}^N \log \exp \sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} Z_{i,c} X_{j,c} + \sum_{i=1}^N \sum_{c=1}^C \exp \sum_{j \in N_i} \hat{A}_{i,j} X_{j,c} \quad \square \\
 &= - \sum_{i=1}^N \log \frac{\sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} Z_{i,c} X_{j,c}}{\sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} X_{j,c}} \quad \square
 \end{aligned}$$

请注意  $\sum_{c=1}^C Z_{j,c} = 1$  for any  $j$ . 考虑L在X上的派生  $\tilde{X}_{j,c}'$ :

$$\begin{aligned}
 \frac{dL}{d\tilde{X}_{j,c}'} &= - \sum_{i=1}^N \frac{\sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} X_{j,c}}{\exp \sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} X_{j,c}} \\
 &\quad \times \frac{\hat{A}_{i,j'} Z_{i,c'} \sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} Z_{i,c} X_{j,c} - \sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} X_{j,c} \sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} X_{j,c}}{\left( \sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} X_{j,c} \right)^2} \\
 &= - \sum_{i=1}^N \frac{\hat{A}_{i,j'} \sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} Z_{i,c} X_{j,c} - \sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} X_{j,c} \sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} X_{j,c}}{\left( \sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} X_{j,c} \right)^2} \\
 &= - \sum_{i=1}^N \frac{\hat{A}_{i,j'} \sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} Z_{i,c} X_{j,c} - \sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} X_{j,c} \sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} X_{j,c}}{\left( \sum_{c=1}^C \sum_{j \in N_i} \hat{A}_{i,j} X_{j,c} \right)^2}
 \end{aligned}$$

$$j \in \mathbb{N}$$

$$A_j$$

!

$$x_{j_c})$$



$$\begin{aligned}
&= - \sum_{i=1}^N \sum_{j \in N_i} \hat{A}_{i,j}' \frac{\sum_{c=1, c \neq c'}^C (Z_{i,c}') \sum_{j \in N_i} A_{i,j} X_{j,c} + (Z_{i,c}' - 1) \sum_{j \in N_i} A_{i,j} X_{j,c'}}{\sum_{c=1}^C \sum_{j \in N_i} A_{i,j} X_{j,c}} \\
&= - \sum_{i=1}^N \hat{A}_{i,j}' Z_{i,c}' P^{\wedge}(Y_i \neq c') + (Z_{i,c}' - 1) P^{\wedge}(Y_i = c') \\
&= - \sum_{i=1}^N \hat{A}_{i,j}' Z_{i,c}' - P^{\wedge}(Y_i = c')
\end{aligned}$$

将上述内容写成矩阵形式，我们有

$$\frac{dL}{dX} = A^{\wedge}(Z-Y), \frac{dL}{dW} = x^T a^{\wedge} t(z-y), \Delta y^T x^T a^{\wedge} t(z-y) \quad (15)$$

## C 合成实验的细节

在我们的合成实验中，我们生成了具有边缘同质性水平  $h \in [0.05, 0.95] : 0.05$  和  $h \in [0.005, 0.05] : 0.005$ 。我们用更精细的尺度 0.005 来探索区间  $[0.05, 0.005]$ ，因为我们根据经验发现 GNN 的性能在更多的这个区域是敏感的。对于一个给定的  $h$ ，我们从 `numpy.random.multinomial(2, numpy.ones(399)/399, size=1)[0]` 生成类内边缘（不包括自循环）和类间边缘，来自 `numpy.random.multinomial(int(2/h - 2), numpy.ones(1600)/1600, size=1)[0]`。

## D 定理1的证明

*证明。* 根据给定的假设，对于节点  $v$ ，我们有  $A_{v,k} = 1$ ，预期的类内边数为  $dh$ （这里根据边的同质性定义和数据生成过程， $A^{\wedge}$  引入的自环边不计算在内），类间边数为  $(1-h)d$ 。假设有  $C \geq 2$  个类。考虑矩阵  $A$ 、

然后，我们有  $E(A)_{i_{v,c}} = E \sum_{k \in V} A_{k,1} \{Z_{k,:} = e_c^T\} = \sum_{k \in V} \frac{E h 1_{\{Z_{k,:} = e_c^T\}}}{d+1}$ ，其中  $1$  是指的功能。

当  $v$  属于  $c$  类时，我们有  $\sum_{k \in V} \frac{E h 1_{\{Z_{k,:} = e_c^T\}}}{d+1} = \frac{hd+1}{d+1} (hd+1 = hd \text{ 类内边缘} + 1 \text{ 自环})$ 。

当  $v$  不属于  $c$  类时，我们有  $\sum_{k \in V} \frac{E h 1_{\{Z_{k,:} = e_c^T\}}}{d+1} = \frac{(1-h)d}{(C-1)(d+1)} ((1-h)d \text{ 类间边缘})$  均匀地分布在其他  $C-1$  类中)。

对于节点  $v, u$ ，我们有  $(A)_{v,:}, (A)_{u,:} \in R^C$ ，由于  $A_{v,k}$  和  $A_{u,k'}$  中的元素对所有  $k, k' \in V$  是独立生成的，所以我们有

$$E(A)_{v,c} A_{u,c} = E \left( \sum_{k \in V} A_{k,1} \{Z_{k,:} = e_c^T\} \right) \left( \sum_{k' \in V} A_{k',1} \{Z_{k',:} = e_c^T\} \right)$$

$$= \mathbb{E} \left( \sum_{k \in V} A_{k,1} \mathbb{1}_{\{Z_{k,:} = e^T\}} \right) = \mathbb{E} \left( \sum_{k' \in V} A_{k',1}^c \mathbb{1}_{\{Z_{k',:}^c = e^T\}} \right)$$

因此

、

$$E h S(\hat{A}; Z)_{v,u} = E \left[ \sum_{k \in P} \sum_{k' \in P} \mathbf{1}_{\{Z_{k,:} = e_T\}} \mathbf{1}_{\{Z_{k',:} = e_T\}} \right]$$

$\mathbf{1}_{\{Z_{k,:} = e_T\}}$

$$= \frac{hd+1}{d+1} + \frac{(1-h)d}{(C-1)(d+1)^2}, \quad u, v \text{ 是在同一个班级}$$

$$= \frac{2(hd+1)(1-h)d}{(C-1)(d+1)^2} + \frac{(C-2)(1-h)^2 d^2}{(C-1)^2(d+1)^2}, \quad u, v \text{ 属于不同的类}$$

对于节点  $u_1, u_2$  和  $v$ , 其中  $Z_{u_1,:} = Z_{v,:}$  和  $Z_{u_2,:} \neq Z_{v,:}$

$$g(h) \equiv E S(\hat{A}; Z)_{v,u} - E S(\hat{A}; Z)_{2v,u} \quad (16)$$

$$= \frac{(C-1)^2(hd+1)^2 + (C-1)[(1-h)d]^2 - (C-1)(2(hd+1)(1-h)d) - (C-2)[(1-h)d]^2}{(C-1)^2(d+1)^2}$$

$$= \frac{(C-1)(hd+1) - (1-h)d}{(C-1)(d+1)}$$

设定  $g(h)=0$ , 我们得到最优的  $h$ :

$$h = \frac{d+1-C}{Cd} \quad (17)$$

对于合成实验中的数据生成过程, 我们固定  $d_{\text{intra}}$ , 那么  $d = d_{\text{intra}}/h$ , 它是  $h$  的函数。我们将 (17) 中的  $d$  改为  $d_{\text{intra}}/h$ , 从而得出

$$h = \frac{d_{\text{intra}}/h + 1 - C}{Cd_{\text{intra}}/h} \quad (18)$$

很容易观察到, 当  $g(h)$  中的  $d$  被  $d_{\text{intra}}/h$  取代时, 满足 (18) 的  $h$  仍然使  $g(h)=0$ 。从 (18) 中我们得到了以  $d$  为单位的最佳  $h_{\text{intra}}$ :

$$h = \frac{d_{\text{intra}}}{Cd_{\text{intra}} + C - 1}$$

□

## D.1 定理1的延伸

$$\text{萨格 } S(\hat{A}; Z) = \frac{nv \text{ 平均}_u \{S(\hat{A}; Z)_{v,u} | Z_{u,:} = Z_{v,:}\} \geq \text{平均}_u \{S(\hat{A}; Z)_{v,u} | Z_{u,:} \neq Z_{v,:}\}}{\sum_{v \in P} \frac{|\mathcal{M}|}{\text{InMean}_u \{S(\hat{A}; Z)_{v,u} | Z_{u,:} = Z_{v,:}\} \geq \text{Mean}_u \{S(\hat{A}; Z)_{v,u} | Z_{u,:} \neq Z_{v,:}\}}}$$

$$\text{然后 } \frac{\sum_{v \in P} \frac{|\mathcal{M}|}{\text{InMean}_u \{S(\hat{A}; Z)_{v,u} | Z_{u,:} = Z_{v,:}\} \geq \text{Mean}_u \{S(\hat{A}; Z)_{v,u} | Z_{u,:} \neq Z_{v,:}\}}}{|\mathcal{M}|}$$

$$E \text{ 萨格 } S(\hat{A}; Z) = E \frac{|\mathcal{M}|}{|\mathcal{M}|} \quad \square$$

$$= \frac{\sum_{v \in P} P \text{ 平均}_u \{S(\hat{A}; Z)_{v,u} | Z_{u,:} = Z_{v,:}\} \geq \text{平均}_u \{S(\hat{A}; Z)_{v,u} | Z_{u,:} \neq Z_{v,:}\}}{|\mathcal{M}|}$$

$$= P \text{ 平均数}_u \{S(\hat{A}; Z)_{v,u} | Z_{u,:} = Z_{v,:}\} - \text{平均值}_u \{S(\hat{A}; Z)_{v,u} | Z_{u,:} \neq Z_{v,:}\} \geq 0$$



Consider the random variable

$$RV = \text{平均}_u \{S(\hat{\mathbf{A}}; \mathbf{Z})_{v,u} / Z_{u,:} = Z_{v,:}\} - \text{平均值}_u \{S(\hat{\mathbf{A}}; \mathbf{Z})_{v,u} / Z_{u,:} \neq Z_{v,:}\}$$

由于 $RV$ 是对称分布的，在定理1的条件下，其期望值为 $E[RV]=g(h)$ ，如（16）所示。由于 $g(h)$ 的最小值是0，并且 $RV$ 是对称分布的，我们有 $P(RV \geq 0) \geq 0.5$ ，这可以解释为什么在许多现实世界的任务中 $H_{\text{agg}}(G)$ 总是大于0.5。

## E 定理2的证明

证明。定义  $W_v^c = (A^c)_{v,c}$ 。那么、

$$W_v^c = \sum_{k \in P} \hat{A}_{v,k} \mathbf{1}_{\{Z_{k,:} = e_c^T\}} \in [0, 1], \quad \sum_{c=1}^C W_v^c = 1$$

请注意

$$s(i - \hat{a}, z) = (i - \hat{a}) Z Z^T (i - \hat{a})^T = Z Z^T + \alpha \hat{Z} \hat{a}^T - \hat{a} Z^T - Z Z^T \hat{a}^T \quad (19)$$

对于任何节点  $v$ ，让  $v$  所属的类用  $c$  表示  $v$ 。对于两个节点  $v, u$ ，如果  $Z_{v,:} \neq Z_{u,:}$ ，我们有

$$\begin{aligned} (Z Z^T)_{v,u} &= 0 \\ (\hat{a} Z^T \hat{a}^T)_{v,u} &= \sum_{c=1}^C W_v^c W_u^c \\ (\hat{A} Z^T)_{v,u} &= W_v^c \\ (Z Z^T \hat{A}^T)_{v,u} &= (\hat{A} Z^T)_{u,v} = W_u^c \end{aligned}$$

然后，从 (19) 中可以看出

$$(s(I - \hat{A}, Z))_{v,u} = \sum_{c=1}^C W_v^c W_u^c - W_v^c W_u^c - W_v^c W_u^c$$

当  $C=2$  时、

$$s(I - \hat{A}, Z)_{v,u} = W_v^c W_u^c (W_u^c - 1) + W_u^c W_v^c (W_v^c - 1) \leq 0$$

如果  $Z_{v,:} = Z_{u,:}$ ，即  $c_v = c_u$ ，我们有

$$\begin{aligned} (Z Z^T)_{v,u} &= 1 \\ (\hat{a} Z^T \hat{a}^T)_{v,u} &= \sum_{c=1}^C W_v^c W_u^c \\ (\hat{A} Z^T)_{v,u} &= W_v^c \\ (Z Z^T \hat{A}^T)_{v,u} &= (\hat{A} Z^T)_{u,v} = W_u^c \end{aligned}$$

然后，从 (19) 中可以看出

$$\begin{aligned} s(I - \hat{A}, Z)_{v,u} &= 1 + \sum_{c=1}^C W_v^c W_u^c - W_v^c W_u^c - W_v^c W_u^c \\ &= \sum_{c=1, c \neq c_v}^C W_v^c W_u^c + 1 + W_v^c W_u^c - W_v^c W_u^c - W_v^c W_u^c \\ &= \sum_{c=1, c \neq c_v}^C W_v^c W_u^c + (1 - W_v^c)(1 - W_u^c) \geq 0 \end{aligned}$$

因此，如果  $C = 2$ ，对于任何  $v \in V$ ，如果  $Z_{u,:} \neq Z_{v,:}$ ，我们有  $s(I - \hat{A}, Z)_{v,u} \leq 0$ ；如果  $Z_{u,:} = Z_{v,:}$ ，我们有  $s(I - \hat{A}, Z)_{v,u} \geq 0$ 。显然，(10) 的两个条件得到了满足。因此， $v$  是可分化的， $\text{DDA}_{\hat{X}}(G) = 1$ 。该定理得到了证明。□

## F 合成图上的模型比较

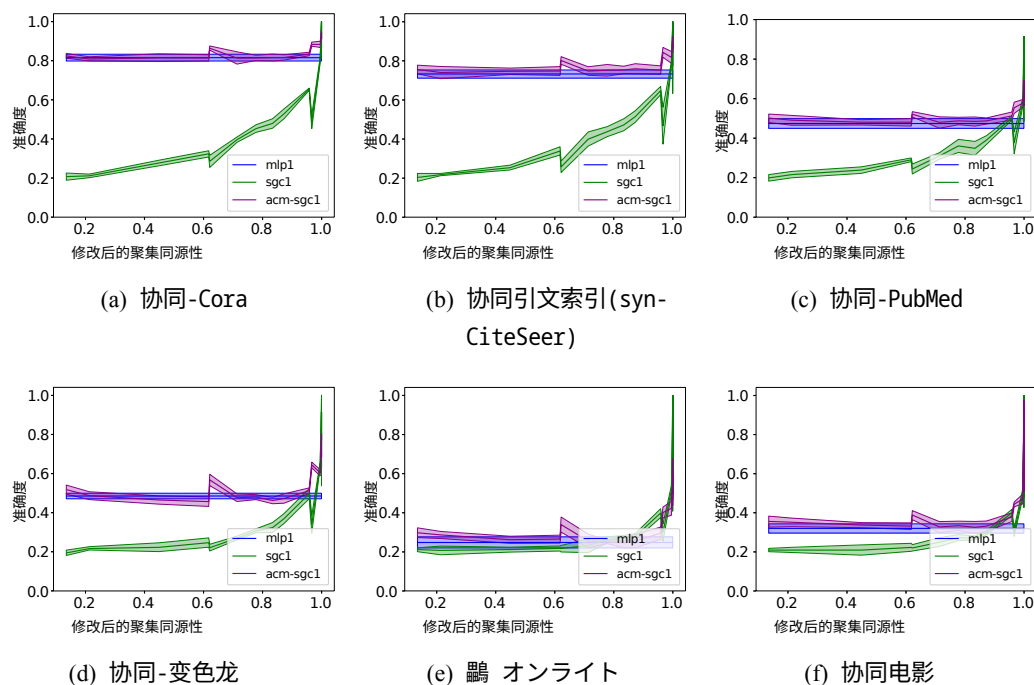


图4: MLP-1、SGC-1和ACM-SGC-1在合成数据集上的测试精度（平均值±标准差）的比较

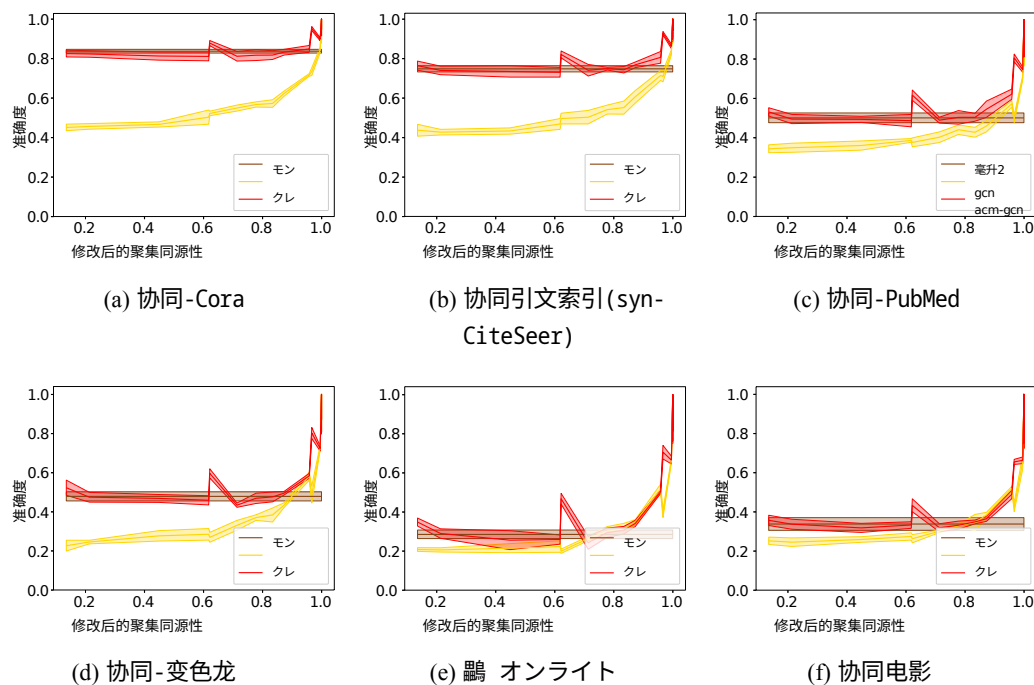


图5: MLP-2、GCN和ACM-GCN在合成数据集上的测试精度（平均值±标准差）的比较

为了分离非线性和图结构的影响，我们将1跳的sgc (sgc-1) 与MLP-1 (线性模型) 进行比较。对于包括非线性的GCN，我们使用MLP-2作为图无关的基线模型。我们训练上述GNN模型、图无关的基线模型和

ACM-GNN模型在所有合成数据集上的测试结果，并绘制了每个数据集上的平均测试精度与标准偏差。从图4和图5可以看出，在每个 $H^M(G)$ 水平上，ACM-GNNs的表现不会低于GNNs和图-agnostic模型。但当 $H^M(G)$ 较小时，GNNs的表现将大大优于图诊断模型。这证明了ACM框架的优势。

## G 多元化经营的局限性讨论

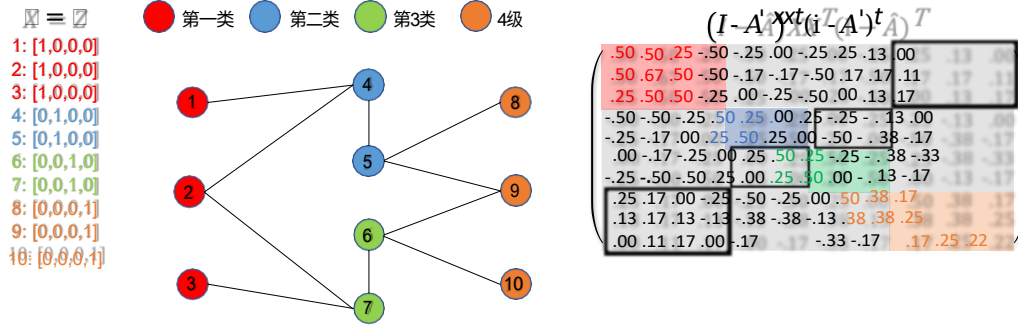


图6：HP过滤器对有害饮食习惯的情况（黑框内的区域）的例子

从图6的例子中 $s(I-A^{\wedge}, x)$ 的黑框区域，我们可以看到，第1类和第4类的节点相互分配非负的权重；第2类和第3类的节点也是相互分配非负的权重。这是因为1类的周围差异与4类相似，2类和3类也是如此。在现实世界的应用中，当几个小集群的节点连接到一个大的簇，小簇中的节点的周围差异将变得相似。在这种情况下，HP过滤器无法区分不同小簇中的节点。

## H 相似性、同质性和 $DDA^{\wedge}_{\mathcal{X}}(G)$ 度量及其估计

**附加指标** 有三个关键因素影响GNN在现实世界任务中的表现：标签、特征和图结构。经过修改的) 聚合同质性试图研究在标签和特征固定的情况下，图结构将如何影响性能。并通过合成实验验证了它们之间的关联性。

Besides graph-label consistency, we need to consider feature-label consistency and aggregated-feature-label consistency as well. With aggregation similarity score of the features  $S_{agg}(S(I, X))$  and aggregated features  $S_{agg}(S(\hat{A}, X))$  listed in appendix H, our methods open up a new perspective about analyzing and comparing graph diagnosis models and graph perception models in real-world tasks. Here are 2 examples.

**Example 1:** GCN (graph-aware model) underperforms MLP-2 (graph-agnostic model) on *Cornell, Wisconsin, Texas, Film*. Based on the aggregation homophily, the graph structure is not the main cause of the performance degradation. And from Table 6 we can see that the  $S_{agg}(S(\hat{A}, X))$  for the above 4 datasets are lower than their corresponding  $S_{agg}(S(I, X))$ , which implies that it is leading to performance degradation by the inconsistency of aggregated feature labels.

**Example 2:** According to the widely used analysis based on node or edge homophily, the graph structure of *Chameleon*, and *Squirrel* are heterophilous and bad for GNNs. But in practice, GCN outperforms MLP-2 on those 2 datasets which means the additional graph information is helpful for node classification instead of being harmful. Traditional homophily metrics fail to explain such phenomenon but our method can give an explanation from different angles: For *Chameleon*, its modified aggregation homophily is not low and its  $S_{agg}(S(\hat{A}, X))$  is higher than its  $S_{agg}(S(I, X))$ .

which means its graph-label consistency and aggregated-feature-label consistency help the graph-aware model obtain the performance gain; for Squirrel, its modified aggregation homophily is low but its  $S_{\text{agg}}(S(\hat{A}, X))$  is higher than its  $S_{\text{agg}}(S(l, X))$  which means although its graph-label consistency is bad but the aggregated-feature-label consistency is the key factor to help the graph-aware model perform better.

我们还需要指出的是，（修改后的）聚集相似度得分， $S_{\text{agg}}(S(\hat{A}, X))$ 和 $S_{\text{agg}}(S(l, X))$ 不是决定性的或阈值，因为它们没有考虑特征中的非线性结构。在实践中，一个低分并不能告诉我们GNN模型一定会表现不好。

	康奈尔大学	威斯康星州	德州	电影	变色龙	松鼠	Cora	CiteSeer	医学博士
哈格( $G$ )	0.9016	0.8884	0.847	0.8411	0.805	0.6783	0.9952	0.9913	0.9716
$S_{\text{agg}}(S(\hat{A}, X))$	0.8251	0.7769	0.6557	0.5118	0.8292	0.7216	0.9439	0.9393	0.8623
$S_{\text{agg}}(S(l, X))$	0.9672	0.8287	0.9672	0.5405	0.7931	0.701	0.9103	0.9315	0.8823
$DDA_{A,X}(G)$	0.3497	0.6096	0.459	0.3279	0.3109	0.2711	0.2681	0.4124	0.1889
$H_{\text{agg}}(G)$	$0.9046 \pm 0.0282$ $0.9147 \pm 0.0260$ $0.8596 \pm 0.0299$ $0.8451 \pm 0.0041$ $0.8041 \pm 0.0078$ $0.6788 \pm 0.0077$ $0.9959 \pm 0.0011$ $0.9907 \pm 0.0015$ $0.9724 \pm 0.0015$								
$\hat{S}_{\text{agg}}(S(\hat{A}, X))$	$0.8266 \pm 0.0526$ $0.8280 \pm 0.0351$ $0.6835 \pm 0.0498$ $0.5345 \pm 0.0421$ $0.8433 \pm 0.0070$ $0.7352 \pm 0.0132$ $0.9487 \pm 0.0023$ $0.9451 \pm 0.0038$ $0.8626 \pm 0.0021$								
$\hat{S}_{\text{agg}}(S(l, X))$	$0.9752 \pm 0.0174$ $0.8680 \pm 0.0270$ $0.9661 \pm 0.0336$ $0.5438 \pm 0.0184$ $0.8257 \pm 0.0050$ $0.7472 \pm 0.0089$ $0.9204 \pm 0.0044$ $0.9441 \pm 0.0036$ $0.8835 \pm 0.0019$								
$\hat{D}_{DA,X}(G)$	$0.3936 \pm 0.0663$ $0.6073 \pm 0.0436$ $0.4817 \pm 0.0762$ $0.3300 \pm 0.0136$ $0.3329 \pm 0.0151$ $0.3021 \pm 0.0101$ $0.3198 \pm 0.0225$ $0.4424 \pm 0.0136$ $0.1919 \pm 0.0046$								

表8：只用训练标签的其他指标及其估计值（平均值±标准差）

在大多数现实世界的应用中，并非所有的标签都可以用来计算数据集的统计数据。在本节中，我们将数据随机分成60%/20%/20%，用于训练/验证/测试，并且只使用训练标签来估计统计数据。我们将每个估计值重复10次，并报告平均值和标准差。结果显示在表8中。

**Estimation** The statistics we estimate are  $H_{\text{agg}}(G)$ ,  $S_{\text{agg}}(S(\hat{A}, X))$ ,  $S_{\text{agg}}(S(l, X))$  and  $DDA_{A,X}(G)$  and are denoted as  $\hat{H}_{\text{agg}}(G)$ ,  $\hat{S}_{\text{agg}}(S(\hat{A}, X))$ ,  $\hat{S}_{\text{agg}}(S(l, X))$  and  $\hat{D}_{DA,X}(G)$ . The two similarity scores  $S_{\text{agg}}(S(\hat{A}, X))$  and  $S_{\text{agg}}(S(l, X))$  measures the proportion of nodes, according to aggregated features and nodes features respectively, that will put larger weights on nodes in the same class than in other classes. The higher values of  $S_{\text{agg}}(S(\hat{A}, X))$  and  $S_{\text{agg}}(S(l, X))$  indicates the better quality of aggregated features and nodes features.

**分析** 从报告的结果中我们可以看出，估计是准确的，误差在可接受的范围内，这意味着所提出的指标和相似性分数可以用标签的子集来准确估计，这对现实世界的应用很重要。此外，我们注意到一些有趣的结果，例如，GNNs和MLP在Squirrel和Film上的表现很差，根据聚合同质性值，Film的图结构与其他数据集相比并不十分有害，但其特征和聚合特征比其他数据集差很多；Squirrel的特征和聚合特征并不太坏，但其图拓扑结构比其他数据集更有害。将本文定义的指标结合在一起，可以帮助我们区分图结构和特征中的不同因素，并确定什么可能导致GNN的性能下降。

## I 由[33]提供的关于固定分割的实验

结果见表9，最佳搜索的超参数见表10。

	康奈尔大学	威斯康星州	德克萨斯州	电影	变色龙	松鼠	科拉	ÄÄÄ	医学博士	级别
辽宁沈阳	78.11 ± 6.55	82.55 ± 6.23	81.35 ± 5.32	35.16 ± 0.9	62.59 ± 2.04	46.31 ± 2.46	87.95 ± 1.18	77.13 ± 1.67	87.54 ± 0.38	8.22
H2GCN	82.70 ± 5.28	87.65 ± 4.98	84.86 ± 7.23	35.70 ± 1.00	60.11 ± 2.15	36.48 ± 1.86	87.87 ± 1.20	77.11 ± 1.57	89.49 ± 0.38	6.78
FAGCN	76.76 ± 5.87	79.61 ± 1.58	76.49 ± 2.87	34.82 ± 1.35	46.07 ± 2.11	30.83 ± 0.69	<b>88.05 ± 1.57</b>	77.07 ± 2.05	88.09 ± 1.38	9.56
Geom-GCN*	60.54 ± 3.67	64.51 ± 3.66	66.76 ± 2.72	31.59 ± 1.15	60.00 ± 2.81	38.15 ± 0.92	85.35 ± 1.57	<b>78.02 ± 1.15</b>	89.95 ± 0.47	9.22
ACM-SGC-1	82.43 ± 5.44	86.47 ± 3.77	81.89 ± 4.53	35.49 ± 1.06	63.99 ± 1.66	45.00 ± 1.4	86.9 ± 1.38	76.73 ± 1.59	88.49 ± 0.51	8.44
ACM-SGC-2	82.43 ± 5.44	86.47 ± 3.77	81.89 ± 4.53	36.04 ± 0.83	59.21 ± 2.22	40.02 ± 0.96	87.69 ± 1.07	76.59 ± 1.69	89.01 ± 0.6	8.22
ACM-GCN	85.14 ± 6.07	<b>88.43 ± 3.22</b>	<b>87.84 ± 4.4</b>	36.28 ± 1.09	66.93 ± 1.85	<b>54.4 ± 1.88</b>	87.91 ± 0.95	77.32 ± 1.7	<b>90.00 ± 0.5</b>	<b>2.3</b>
ACM-雪球-2	85.41 ± 5.43	87.06 ± 2.87	57 ± 4.86	<b>36.89 ± 1.18</b>	<b>67.08 ± 2.04</b>	52.5 ± 1.49	87.42 ± 1.09	76.41 ± 1.38	89.89 ± 0.57	4.11
ACM-雪球-3	83.24 ± 5.38	86.67 ± 4.37	<b>87.84 ± 3.87</b>	36.82 ± 0.94	66.91 ± 1.73	53.31 ± 1.88	87.1 ± 0.93	75.91 ± 1.57	89.81 ± 0.43	5.22
ACMII-GCN	<b>85.95 ± 5.64</b>	87.45 ± 3.74	86.76 ± 4.75	36.16 ± 1.11	66.91 ± 2.55	51.85 ± 1.38	88.01 ± 1.08	77.15 ± 1.45	89.89 ± 0.43	3.22
ACMII-雪球-2	85.68 ± 5.93	87.45 ± 2.8	86.76 ± 4.43	36.55 ± 1.24	66.49 ± 1.75	50.15 ± 1.4	87.57 ± 0.86	76.92 ± 1.45	89.84 ± 0.48	4.67
ACMII-雪球-3	82.7 ± 4.86	85.29 ± 4.23	85.41 ± 6.42	36.49 ± 1.41	66.86 ± 1.74	48.87 ± 1.23	87.16 ± 1.01	76.18 ± 1.55	89.73 ± 0.52	7.00

表9：由[33]提供的关于固定分割的实验结果：在9个真实世界的基准数据集上的平均测试精度±标准偏差。最好的结果被突出显示。Geom- GCN、H<sub>2</sub> GCN和GPRGNN的结果来自[33, 40, 25]；其余模型的结果由我们自己运行，超参数搜索范围与表5相同。

## J 对ACM(II)-GNN和几个SOTA模型之间的差异的详细解释

- 与GPRGNN[7]的区别：解释通道混合机制的区别

和GPRGNN的学习机制，我们首先将GPRGNN重写为 $\mathbf{Z} = \sum_{k=0}^{\infty} \gamma_k \mathbf{H}^{(k)}$

$$\sum_{k=0}^{\infty} \gamma_k \mathbf{H}^{(k)} = \sum_{k=0}^{\infty} \text{diag}(\gamma_k, \gamma_k, \dots, \gamma_k) \mathbf{H}^{(k)} \quad \text{节点式通道混合机制}$$

在GPRGNN形式下是 $\sum_{k=0}^{\infty} \text{diag}(\gamma_k, \gamma_k, \dots, \gamma_k) \mathbf{H}^{(k)}$ ，其中 $N$ 是节点的数量

和 $\gamma_i, i = 1, \dots, N$ 是可学习的参数化混合权重。

- 与FAGCN[4]的区别：FAGCN不是使用一个固定的 $\hat{\mathbf{A}}$ ，而是学习一个新的过滤器 $\hat{\mathbf{A}}'$ ，基于在 $\hat{\mathbf{A}}$ 上。而 $\hat{\mathbf{A}}$ 可以被分解为 $\hat{\mathbf{A}} = \hat{\mathbf{A}}_1 \hat{\mathbf{A}}_2$ ，其中 $\hat{\mathbf{A}}_1$ 和 $\hat{\mathbf{A}}_2$ 代表正的

和负的边缘（传播）信息分别。在我们的论文中，我们并不讨论使用学习的滤波器 $\hat{\mathbf{A}}'$ 比固定的滤波器 $\hat{\mathbf{A}}$ 的优势，我们是在比较有和没有通道混合机制的模型。我们认为表2和表9中的实际任务的经验结果是比较带有固定滤波器和节点通道混合的模型和带有学习滤波器但没有通道混合的模型的最好方法。



数据集	模型和超参数	lr	重量_衰变	辍学者	隐藏的	结果	标准	平均纪元时间/平均总时间
康奈尔大学	ACM-SGC-1	0.01	5.00E-06	0	64	82.43	5.44	5.37ms/23.05s
	ACM-SGC-2	0.01	5.00E-06	0	64	82.43	5.44	5.93ms/25.66s
	ACM-GCN	0.05	5.00E-04	0.5	64	85.14	6.07	8.04ms/1.67s
	ACMII-GCN	0.1	1.00E-04	0	64	85.95	5.64	7.83ms/2.66s
	FAGCN	0.01	1.00E-04	0.6	64	76.76	5.87	8.80ms/7.67s
	ACM-雪球-2	0.05	5.00E-03	0.3	64	85.41	5.43	11.50ms/2.35s
	ACM-雪球-3	0.05	5.00E-03	0.2	64	83.24	5.38	15.06ms/3.12s
	ACMII-雪球-2	0.1	5.00E-03	0.2	64	85.68	5.93	12.63ms/2.58s
	ACMII-雪球-3	0.05	5.00E-03	0.2	64	82.7	4.86	14.59ms/3.06s
威斯康星州	ACM-SGC-1	0.1	5.00E-06	0	64	86.47	3.77	5.07ms/14.07s
	ACM-SGC-2	0.1	5.00E-06	0	64	86.47	3.77	5.30ms/16.05s
	ACM-GCN	0.05	1.00E-03	0.4	64	88.43	3.22	8.04ms/1.66s
	ACMII-GCN	0.01	5.00E-05	0.1	64	87.45	3.74	8.40ms/2.19s
	FAGCN	0.01	5.00E-05	0.5	64	79.61	1.59	8.61ms/5.84s
	ACM-雪球-2	0.01	1.00E-03	0.4	64	87.06	2	12.51ms/2.60s
	ACM-雪球-3	0.01	1.00E-02	0.1	64	86.67	4.37	14.92ms/3.15s
	ACMII-雪球-2	0.01	5.00E-04	0.1	64	87.45	2.8	11.96ms/2.63s
	ACMII-雪球-3	0.01	5.00E-03	0.5	64	85.29	4.23	14.87ms/3.10s
德州	ACM-SGC-1	0.01	1.00E-05	0	64	81.89	4.53	5.34ms/19.00s
	ACM-SGC-2	0.05	1.00E-05	0	64	81.89	4.53	5.50ms/9.26s
	ACM-GCN	0.05	5.00E-04	0.5	64	87.84	4.4	9.62ms/1.99s
	ACMII-GCN	0.01	1.00E-03	0.1	64	86.76	4.75	9.98ms/2.22s
	FAGCN	0.01	1.00E-05	0	64	76.49	2.87	10.45ms/5.70s
	ACM-雪球-2	0.01	5.00E-03	0.2	64	87.57	4.86	11.56ms/2.45s
	ACM-雪球-3	0.01	5.00E-03	0.2	64	87.84	3.87	15.17ms/3.15s
	ACMII-雪球-2	0.01	1.00E-03	0.2	64	86.76	4.43	11.36ms/2.30
	ACMII-雪球-3	0.01	5.00E-03	0.6	64	85.41	6.42	15.84ms/3.48s
电影	ACM-SGC-1	0.05	5.00E-04	0	64	35.49	1.06	5.39ms/1.17s
	ACM-SGC-2	0.05	5.00E-04	0.1	64	36.04	0.83	13.22ms/3.31s
	ACM-GCN	0.01	5.00E-03	0	64	36.28	1.09	8.96ms/1.82s
	ACMII-GCN	0.01	5.00E-03	0	64	36.16	1.11	9.06ms/1.83s
	FAGCN	0.01	5.00E-05	0.4	64	34.82	1.35	15.60ms/2.51s
	ACM-雪球-2	0.01	1.00E-02	0	64	36.89	1.18	14.77ms/3.01s
	ACM-雪球-3	0.01	1.00E-02	0.2	64	36.82	0.94	16.57ms/3.36s
	ACMII-雪球-2	0.01	5.00E-03	0.1	64	36.55	1.24	12.76ms/2.57s
	ACMII-雪球-3	0.05	5.00E-03	0.3	64	36.49	1.41	16.51ms/3.49s
变色龙	ACM-SGC-1	0.1	5.00E-06	0.9	64	63.99	1.66	5.92ms/1.74s
	ACM-SGC-2	0.1	0.00E+00	0.9	64	59.21	2.22	8.84ms/1.78s
	ACM-GCN	0.05	5.00E-05	0.7	64	66.93	1.85	8.40ms/1.71s
	ACMII-GCN	0.05	5.00E-06	0.8	64	66.91	2.55	8.90ms/2.10s
	FAGCN	0.01	5.00E-05	0	64	46.07	2.11	16.90ms/7.94s
	ACM-雪球-2	0.01	1.00E-04	0.7	64	67.08	2.04	12.50ms/2.69s
	ACM-雪球-3	0.01	1.00E-05	0.8	64	66.91	1.73	16.12ms/4.91s
	ACMII-雪球-2	0.01	5.00E-05	0.8	64	66.49	1.75	12.65ms/3.42s
	ACMII-雪球-3	0.05	5.00E-05	0.7	64	66.86	1.74	17.60ms/4.06s
松鼠	ACM-SGC-1	0.05	5.00E-06	0.9	64	45	1.4	6.10ms/2.18s
	ACM-SGC-2	0.05	0.00E+00	0.9	64	40.02	0.96	35.75ms/9.62s
	ACM-GCN	0.05	5.00E-06	0.7	64	54.4	1.88	10.48ms/2.68s
	ACMII-GCN	0.05	5.00E-06	0.7	64	51.85	1.38	11.69ms/2.91s
	FAGCN	0	5.00E-03	0	64	30.86	0.69	10.90ms/13.91s
	ACM-雪球-2	0.01	1.00E-04	0.7	64	52.5	1.49	17.89ms/5.78s
	ACM-雪球-3	0.01	5.00E-05	0.7	64	53.31	1.88	22.60ms/7.53s
	ACMII-雪球-2	0.05	5.00E-05	0.6	64	50.15	1.4	16.95ms/3.45s
	ACMII-雪球-3	0.01	5.00E-04	0.6	64	48.87	1.23	23.52ms/4.94s
Cora	ACM-SGC-1	0.05	5.00E-05	0.7	64	86.9	1.38	4.99ms/2.40s
	ACM-SGC-2	0.1	0	0.8	64	87.69	1.07	5.16ms/1.16s
	ACM-GCN	0.01	5.00E-05	0.6	64	87.91	0.95	8.41ms/1.84s
	ACMII-GCN	0.01	1.00E-04	0.6	64	88.01	1.08	8.59ms/1.96s
	FAGCN	0.02	1.00E-04	0.5	64	88.05	1.57	9.30ms/10.64s
	ACM-雪球-2	0.01	1.00E-03	0.5	64	87.42	1.09	12.54ms/2.72s
	ACM-雪球-3	0.01	5.00E-06	0.9	64	87.1	0.93	15.83ms/11.33s
	ACMII-雪球-2	0.01	1.00E-03	0.6	64	87.57	0.86	12.06ms/2.64s
	ACMII-雪球-3	0.01	5.00E-03	0.5	64	87.16	1.01	16.29ms/3.62s
CiteSeer	ACM-SGC-1	0.05	0.00E+00	0.7	64	76.73	1.59	5.24ms/1.14s
	ACM-SGC-2	0.1	0.00E+00	0.8	64	76.59	1.69	5.14ms/1.03s
	ACM-GCN	0.01	5.00E-06	0.3	64	77.32	1.7	8.89ms/1.79s
	ACMII-GCN	0.01	5.00E-05	0.5	64	77.15	1.45	8.95ms/1.80s
	FAGCN	0.02	5.00E-05	0.4	64	77.07	2.05	10.05ms/5.69s
	ACM-雪球-2	0.01	5.00E-05	0	64	76.41	1.38	12.87ms/2.59s
	ACM-雪球-3	0.01	5.00E-06	0.9	64	75.91	1.57	17.40ms/11.92s
	ACMII-雪球-2	0.01	5.00E-03	0.5	64	76.92	1.45	13.10ms/2.94s
	ACMII-雪球-3	0.1	5.00E-05	0.9	64	76.18	1.55	17.47ms/5.88s
医学博	Acm-SGC-1	0.05	5.00E-06	0.4	64	88.49	0.51	5.77ms/3.65s
	Acm-SGC-2	0.05	5.00E-06	0.3	64	89.01	0.6	8.50ms/5.18s
	Acm-GCN	0.01	5.00E-05	0.4	64	90	0.52	8.99ms/2.51s
	Acmii-GCN	0.01	1.00E-04	0.3	64	89.89	0.43	9.70ms/2.57s
	FAGCN	0.01	1.00E-04	0	64	88.09	1.38	10.30ms/8.75s
	ACM-雪球-2	0.01	1.00E-03	0.3	64	89.89	0.57	15.05ms/3.11s
	ACM-雪球-3	0.01	5.00E-03	0.1	64	89.81	0.43	20.51ms/4.63s
	ACM-雪球-2	0.01	5.00E-04	0.4	64	89.84	0.48	15.10ms/3.2s
	ACMII-雪球-2	0.01	1.00E-03	0.4	64	89.73	0.52	20.46ms/4.32s

表10: FAGCN和ACM-GNNs在固定分片上的超参数