面向表达能力强的图神经网络的不变分层传播

Asela Hevapathige
asela.hevapathige@unimelb.edu.au
机械工程系 墨尔本大学 墨尔本，澳大利亚

Ahad N. Zehmakan
ahadn.zehmakan@anu.edu.au
计算机学院 澳大利亚国立大学 堪培拉，澳大利亚

Asiri Wijesinghe
asiriwijesinghe.wijesinghe@data61.csiro.au
Data61 澳大利亚联邦科学与工业研究组织 堪培拉，澳大利亚

Saman Halgamuge
saman@unimelb.edu.au 机械工程系 墨尔本大学 墨尔本，澳大利亚

arXiv:2603.01388v1 [cs.LG] 2 Mar 2026

摘要

图神经网络（GNN）在表达能力和捕捉结构异质性方面存在根本局限。标准消息传递架构受限于一维 Weisfeiler-Leman（1-WL）测试，无法区分仅度序列不同以外的图，并且会对来自邻居的信息进行统一聚合，因而无法捕捉节点在高阶模式中所处的不同结构位置。尽管已有方法能够提升表达能力，但它们的计算成本高得难以承受，且缺乏能够灵活编码多样结构属性的统一框架。为解决这些局限，我们提出不变性分层传播（ISP），这是一个同时包含新型 WL 变体（ISP-WL）及其高效神经网络实现（ISP-GNN）的框架。ISP 根据图不变量对节点进行分层，并在层次化的层中处理节点，从而揭示 1-WL 无法识别的结构差异。通过分层结构异质性编码，ISP 量化节点在高阶模式中结构位置的差异，区分参与者所处角色迥异的交互与参与者角色一致的交互。我们通过形式化的理论分析证明，该方法具有超越 1-WL 的增强表达能力、收敛保证，并且天然能够抵抗过平滑。在图分类、节点分类和影响力估计任务上的大量实验表明，与标准架构和最先进的表达能力强的基线方法相比，该方法均能稳定提升性能。

CCS 概念

• **计算方法** → **机器学习**。

ACM 参考格式：
Asela Hevapathige、Ahad N. Zehmakan、Asiri Wijesinghe 和 Saman Hal- gamuge。2026年。《面向表达能力强的图神经网络的不变分层传播》网络。载于。 ACM，美国纽约州纽约市，26页。https://doi.org/10.1145/nnnnnnn.nnnnnnn

允许免费出于个人或课堂用途制作本作品全部或部分内容的数字或纸质副本，但不得为营利或商业利益制作或分发副本，且副本必须附有本声明及首页上的完整引文。对于本作品中由作者以外的其他方拥有版权的组成部分，必须尊重其版权。注明出处后可以摘录。若要以其他方式复制、重新出版、发布到服务器或分发给他人，必须事先获得特定许可和/或支付费用。请通过 permissions@acm.org 申请许可。*Conference'17*，美国华盛顿特区© 2026年版权所有者 / 作者所有。出版权已授权给 ACM。ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 引言

图神经网络（图神经网络）已成为一种强大的图结构化数据学习工具，应用涵盖分子性质预测 [57, 61], 社交网络分析 [6, 25], 和知识推理 [37, 74]。许多图神经网络建立在消息传递机制之上，节点通过该机制迭代地从邻居聚合信息，以优化其表示，从而对关系结构进行建模 [21]。然而，一些根本性限制制约了其表示复杂图结构的能力：

• **计算与表达能力权衡**。标准图神经网络从根本上受到一维Weisfeiler-Leman（1-WL）测试的限制 [60, 68], ，仅能根据邻域度序列区分图。因此，具有相同局部结构模式的节点会得到无法区分的表示，而无论其全局角色如何 [62, 70]。这形成了一个关键权衡：实现更高表达能力的方法，例如 k-WL变体 [2, 43, 77] 和子图枚举方法 [7, 14, 75], 会带来难以承受的计算成本，并且难以扩展到大规模图。虽然结构信息注入技术 [11, 24, 62, 67] 提供了计算效率，但它们通常纳入由架构设计预先确定的固定结构编码。这种刚性使模型无法根据任务特定需求和数据特征自适应地选择并组合多样的结构属性，例如几何特征、拓扑特征和高阶模式。

• **无法捕捉高阶模式中的结构异质性**。 加剧上述表达能力挑战的是，标准图神经网络会对所有邻居的信息进行均匀聚合，将高阶结构一概而论，而不考虑参与节点的结构位置 [49, 59]。它们无法区分节点处于截然不同结构位置的高阶交互。在各种应用中，高阶模式中的这种结构异质性至关重要：从网络分析到分子性质预测，其中实体在局部模式中的相对

结构位置会显著影响其属性和行为。

- **静态、任务无关的结构编码**。当前图神经网络方法主要通过彼此孤立的机制来编码结构信息。例如，位置编码使用谱特征或随机游走特征 [16, 54],子图方法枚举预先设定的模式 [7, 75], 而重连技术则借助曲率等几何工具修改拓扑结构 [29, 55]。这些方法缺乏统一框架，无法灵活结合多个图不变量、将全局结构排序与局部高阶交互联系起来，也无法通过学习使结构偏置适应任务需求。这种碎片化使模型无法发现，对于特定预测任务，哪些结构属性及其相互作用最具信息量。

本文提出不变性分层传播（ISP），这是一个统一框架，包含一种新颖的 WL 变体（ISP-WL）及其高效的神经网络实现（ISP-GNN），系统地解决上述局限。我们工作的核心洞见是由图不变量诱导的排序，我们根据不变量值的序数排名将其称为"分层"。它自然揭示了高阶结构中的结构异质性模式，而这些模式在统一聚合方案中不可见。ISP-WL 按照这些由不变量诱导的排序处理节点，从而能够区分在标准消息传递下表现相同、但结构上各异的节点。ISP-GNN 通过一种可微机制实现这些原理，并学习组合多种结构属性。我们的主要贡献如下:

- **不变分层传播：** 我们提出一种框架，通过引入基于不变量的归纳偏置来增强标准消息传递，从而克服度序列障碍。我们依据图不变量对节点进行分层，并采用能够揭示结构差异的方式处理这些节点，这些差异可能会被统一聚合忽略。该方法在保持计算效率的同时，实现了更强的表达能力。

- **结构异质性编码：** 我们提出一种编码层次化结构差异的方法，以捕捉节点在高阶模式中的不同作用。该方法依据节点的结构位置区分交互，与忽略这些差异的标准统一聚合形成对比。

- **结构编码的统一框架：** 我们的框架整合了多种结构信息，既支持预定义不变量，也支持可学习不变量。它统一了现有的零散方法，使得能够在单一框架中灵活组合多个图不变量。

除核心方法论贡献之外，我们还提供了形式化理论分析，确立了 ISP 框架超越 1-WL 的增强表达能力、收敛保证、与稀疏图上标准消息传递相匹配的计算效率，以及通过深度不变结构嵌入对过平滑的固有抵抗能力。在涵盖图分类、节点分类和动态影响估计任务等多样化图学习场景中的大量实验表明，我们的方法相较于标准消息传递架构和表达能力强的基线方法均取得了持续且显著的改进。

2 相关工作

2.1 超越 1-WL 表达能力的图神经网络

标准消息传递图神经网络从根本上受限于一维 Weisfeiler-Leman 测试，其区分图的能力无法超越度序列 [60, 68]。高阶 WL 方法 [3, 43, 77] 扩展到节点元组以获得更强的表达能力，但会带来难以承受的计算成本，且难以扩展到大规模图。基于子图的方法 [7, 11, 39] 通过子图枚举与聚合增强表达能力，但会产生巨大的计算开销，并且缺乏根据任务需求对结构模式进行加权的自适应机制。结构编码方法 [24, 62, 67, 70] 注入异构节点标识符或结构特征，以区分原本相同的节点。然而，这些方法通常使用固定的、预定义的结构属性，这些属性可能与任务特定需求不一致，并且缺乏灵活组合多种图不变量的统一框架。

ISP-WL 通过基于不变量的分层，在稀疏图上保持计算效率的同时，实现了超越标准消息传递的表达能力，从而缓解了计算与表达能力权衡。我们的框架既提供高效的基于三角形的聚合，也提供一种统一机制，可通过端到端训练自适应地组合多个结构不变量。

2.2 图学习中的排序与层次结构

近期研究探索了多种方法，试图在统一聚合之外组织图神经网络的计算。层次化方法 [8, 19, 64, 69, 72] 采用粗化和池化操作，学习多尺度图表示，但主要用于图级任务，在消息传递过程中并未提供节点级的结构分层。位置感知图神经网络 [18, 44, 71] 引入基于谱和随机游走的位置编码，以区分节点位置；而 DirGNN [50] 等方向感知方法，以及 OSAN [49] 和 GOAT [12] 等有序聚合方法，则对边方向性和邻居排序进行显式建模。然而，这些方法将结构属性视为静态输入特征，或采用局部排序启发式方法，却没有基于全局图结构确定排序的原则性框架，也没有显式建模节点在高阶模式中所处的不同结构位置。

ISP-GNN 利用不变量诱导的分层，为节点建立全局一致的排序，并根据多种图不变量衡量节点的结构重要性，以此处理节点。这种层次化方法凸显了结构异质性，因为处于不同不变量层级的节点具有不同的结构角色。通过基于不变量的间隙度量，明确量化三角形内部的差异，模型能够区分结构重要性各异的节点。将结构层次融入传播机制后，模型便能自适应地学习哪些层次模式与特定任务相关。

3 ISP 框架

我们考虑一个无向图 $G = (V, E)$，其节点集为 V，边集为 E。对于节点 $v \in V$，其邻域为 $N(v) = \{u \in V : (v, u) \in E\}$，其度为 $d_v = |N(v)|$。与 v 相 incident 的三角形

由 $T(v) = \{\{u, w\} : u, w \in N(v), (u, w) \in E\}$ 给出。我们用 $\{\{\cdot\}\}$ 表示多重集，并使用 Hash$(\cdot)$ 表示单射哈希函数。图不变量是一个函数 $\mu : V \to \mathbb{R}$，在图同构下保持不变，即对于任意同构 $\pi : V \to V$，都有 $\mu(\pi(v)) = \mu(v)$。有关与我们研究相关的详细背景概念，请参阅附录第 7 节。

3.1 ISP-WL算法

对于不变量，我们采用**全局排序**：$\phi(v = \mu u u \in \bigcup_i V_i \mu v 1 2 \ldots L)$ rank$(\{\{(\,) : \}(\,)\})$，将取值映射为连续整数，其中 $L = |\{\mu(v) : v \in \bigcup_i V_i\}|$ 表示比较集中所有图里不同不变量值的数量。这确保了表示的一致性，并定义了 L 个不同的层。

3.1.1 分层结构异质性编码。由 ϕ 确定的层次化处理顺序形成了自然的分层，揭示了高阶结构内部的结构异质性。

定义 3.1（层次化结构异质性编码）。对于高阶交互 (v, u, w)，其中 v是中心节点，且 $u, w \in N(v)$ 与 $(u, w) \in E$：

$$\delta(v, u, w) = (\delta_1, \delta_2, \delta_3) \tag{1}$$

其中：

$$\delta_1 = \max(\phi(v) - \phi(u), \phi(v) - \phi(w)) \quad \text{(largest gap)} \tag{2}$$

$$\delta_2 = \min(\phi(v) - \phi(u), \phi(v) - \phi(w)) \quad \text{(smallest gap)} \tag{3}$$

$$\delta_3 = |\phi(u) - \phi(w)| \quad \text{(inter-neighbor gap)} \tag{4}$$

这三个值完整地、以与顺序无关的方式刻画了三角形中的间隙结构，是结构异质性的最小充分统计量。这种编码的设计依据，以及将其自然扩展到三角形以外高阶结构的方法，见附录第 9 节。

以下命题形式化地说明了该编码对邻居对进行对称处理，这对正确性至关重要，并确保无论邻居以何种顺序枚举，图的表示都保持一致。

命题3.2（置换不变性）。$\delta(v, u, w) = \delta(v, w, u)$ 对所有 $u, w \in N(v)$。

3.1.2 层次颜色细化。该算法迭代细化两条颜色流，直至收敛。WL 流遵循标准 1-WL 细化，而 ISP 流则处理位于对应于不同不变量值的层中的节点。在迭代 $\ell \in \{1, 2, \ldots, L\}$ 时，我们处理所有具有 $\phi(v) = \ell$ 的节点 v，并构造其 ISP 颜色。在处理具有 $\phi(v) = \ell$ 的节点 v 时，我们构造：

$$M_v^{\text{struct}} = \{(c_{\text{ISP}}(u), c_{\text{ISP}}(w), \delta(v, u, w)) : (u, w) \in T(v)\} \tag{5}$$

其中 $c_{\text{ISP}}(u)\ c_{\text{ISP}}(w) \in \mathbb{N} \cup \{\perp\}$ 是 ISP 颜色（$\perp$ 表示未着色的节点）。ISP 颜色的赋值方式如下：

$$c_{\text{ISP}}(v) = \begin{cases} \text{Hash}(\phi(v), M_v^{\text{struct}}) & \text{if } M_v^{\text{struct}} \neq \{\} \\ \phi(v) & \text{otherwise} \end{cases} \tag{6}$$

ISP-WL 的一个关键特性是，每个节点的 ISP 颜色只计算一次，此后便保持不变，从而消除冗余

算法 1ISP-WL算法

输入： 图 $G = (V, E)$，全局排序不变量 $\phi : V \to \{1, \ldots, L\}$
输出： 着色 $c : V \to \mathbb{N} \times \mathbb{N}$
初始化：$c_{WL}(v) \leftarrow 0$，$c_{ISP}(v) \leftarrow \perp$ 对所有 $v \in V$
$\ell \leftarrow 1$
重复
  // WL细化
  对于 $v \in V$ **执行**
    $m_v \leftarrow \{(c_{WL}(u)\ c_{ISP}(u)) : u \in N(v)\}$
    $c'_{WL}(v) \leftarrow$哈希$((c_{WL}(v)\ c_{ISP}(v))\ m_v)$
  结束循环
  $c_{WL}^{\text{prev}} \leftarrow c_{WL}$
  $c_{WL} \leftarrow c'_{WL}$
  // ISP 分配
  if $\ell \leq L$ **then**
    $V_\ell \leftarrow \{v \in V : \phi(v) = \ell\}$
    对于 $v \in V_\ell$ **执行**
      $M_v^{\text{struct}} \leftarrow \{(c_{ISP}(u)\ c_{ISP}(w)\ \delta(v, u, w)) : (u, w) \in T(v)\}$
      如果 $M_v^{\text{struct}} \neq \{\}$ **则**
        $c_{ISP}(v) \leftarrow$哈希$(\phi(v)\ M_v^{\text{struct}})$
      else
        $c_{ISP}(v) \leftarrow \phi(v)$
      结束条件判断
    结束循环
  结束条件判断
  $\ell \leftarrow \ell + 1$
直到 $(c_{WL} = c_{WL}^{\text{prev}})$ 且 $(\ell > L)$ *// 两个流均已收敛*
返回 $c(v) =$哈希$(c\ _{WL}(v)\ c_{ISP}(v))$ 对所有 $v \in V$

计算，并确保 ISP流恰好经过 L 次迭代后完成。下文将对此进行形式化说明。

命题 3.3 (单次赋值性质)。对于任意节点 $v \in V$，$c_{ISP}(v)$ 在迭代 $\ell^* = \phi(v)$ 中恰好被赋值一次，此后保持不变。

当两条颜色流都稳定时，算法终止；这一过程至多需要 $\max(K_{\text{WL}}, L)$ 次迭代，其中 K_{WL} 是标准 1-WL 的收敛时间。ISP-WL 的伪代码见算法 1。

3.2 ISP-GNN架构

我们通过学习得到的注意力机制和软门控，将离散的 ISP-WL算法转化为连续、可微的神经网络。ISP-GNN采用双流设计，并使用学习得到的层次注意力，根据结构异质性对高阶交互进行加权。在第 k 层，节点维护 $h^{(k)}{}_{,v}{}^{\text{WL}}$(WL特征) 以及 $h^{(k)}{}_{,v}{}^{\text{ISP}}$(ISP特征)。

3.2.1 神经层次结构异质性编码。对于每个高阶交互（v，u，w），我们计算层次化异质性特征，作为 δ的连续类比：

$$d_{vuw} = \begin{bmatrix} \max(\phi(v) - \phi(u), \phi(v) - \phi(w)) \\ \min(\phi(v) - \phi(u), \phi(v) - \phi(w)) \\ |\phi(u) - \phi(w)| \end{bmatrix} \in \mathbb{R}^3 \tag{7}$$

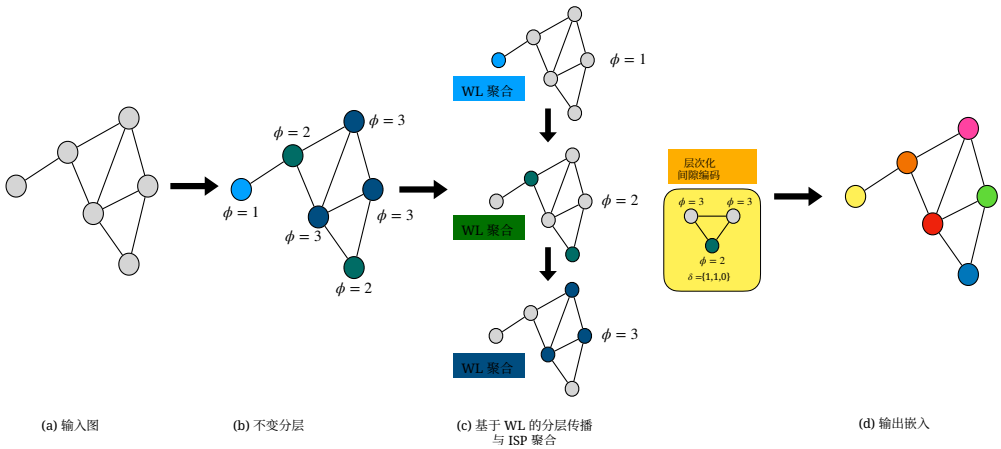


图1：ISP-GNN架构概览。(a) 输入图。(b) 基于不变量值的ISP着色。(c) 每层采用WL与ISP聚合的分层传播。三角形细节展示了具有更高层邻居的节点的间隙编码。(d) 用于下游任务的输出嵌入。

这些特征的计算方式与 $\delta(v,\ u,\ w)$相同，但作用于连续值学习不变量 ϕ，从而支持基于梯度的优化。随后，我们学习结构感知注意力权重：

$$\alpha_{vuw} = \sigma(\mathrm{MLP}_{\mathrm{struct}}(d_{vuw})) \tag{8}$$

其中 σ 将输出限制在 $[0,1]$。结构感知消息传递通过置换不变聚合来组合邻居嵌入：

$$m_{vuw}^{\mathrm{struct}} = \alpha_{vuw} \cdot \mathrm{MLP}_{\mathrm{tri}}(\mathrm{AGG}(h_u^{(k-1),\mathrm{ISP}}, h_w^{(k-1),\mathrm{ISP}})) \tag{9}$$

3.2.2 架构组件。

初始化。 WL特征以输入属性初始化；ISP特征从零开始：

$$h_v^{(0),\mathrm{WL}} = x_v, \quad h_v^{(0),\mathrm{ISP}} = 0 \quad \forall v \in V \tag{10}$$

*WL*流。 在每一层，WL特征使用组合表示从邻居进行聚合：

$$h_v^{(k),\mathrm{WL}} = \mathrm{UPD}(h_v^{(k-1),\mathrm{comb}}, \mathrm{AGG}_{u\in N(v)} h_u^{(k-1),\mathrm{comb}}) \tag{11}$$

其中$h_v^{\mathrm{comb}} = [h_v^{\mathrm{WL}} \| h_v^{\mathrm{ISP}}]$ 结合了两个流。

门控。 这确保每个节点的 ISP特征在其不变量层中恰好分配一次。对于预定义不变式 $\phi\colon V \to \{1,\ldots,L\}$:

$$\gamma_v^{(k)} = \mathbb{I}[\phi(v) = k] \cdot \mathbb{I}[\|h_v^{(k-1),\mathrm{ISP}}\|_2 = 0] \tag{12}$$

其中，第二个指示函数可防止覆盖先前分配的特征。

*ISP*流。 对于门控节点，ISP特征聚合结构加权消息：

$$h_v^{(k),\mathrm{TRI}} = \begin{cases} \mathrm{AGG}_{(u,w)\in T(v)} m_{vuw}^{\mathrm{struct}} & \text{if } |T(v)| > 0 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

$$h_v^{(k),\mathrm{ISP}} = \begin{cases} h_v^{(k),\mathrm{TRI}} & \text{if } \gamma_v^{(k)} = 1, |T(v)| > 0 \\ \mathrm{Embed}(k) & \text{if } \gamma_v^{(k)} = 1, |T(v)| = 0 \\ h_v^{(k-1),\mathrm{ISP}} & \text{otherwise} \end{cases} \tag{14}$$

输出。 最终表示结合两个流：

$$h_v^{(K)} = \mathrm{MLP}_{\mathrm{out}}([h_v^{(K),\mathrm{WL}} \| h_v^{(K),\mathrm{ISP}}]) \tag{15}$$

3.2.3 可学习不变量分层。 预定义不变量提供了固定层次结构，而我们采用基础不变量 $\{\phi_1,\ldots,\phi_m\}$ 引入了一种可学习机制，以实现任务自适应分层。

定义 3.4 （可学习不变式）。 设 $\phi_1,\ldots,\phi_m\colon V \to \mathbb{R}$ 为基础结构不变量。可学习不变式定义为：

$$\phi_{\mathrm{learn}}(v) = (S-1) \cdot \sigma(\mathrm{MLP}_\phi([\phi_1(v),\ldots,\phi_m(v)])) + 1 \tag{16}$$

其中，多层感知机$\phi\colon \mathbb{R}^m \to \mathbb{R}$是学习得到的，$\sigma\colon \mathbb{R} \to (0,1)$ 是有界激活函数（例如 sigmoid），而 $S \in \mathbb{N}$ 控制最大层级数。

处理不可微性。 直接离散化 $\phi_{\mathrm{discrete}}(v) = \lfloor \phi_{\mathrm{learn}}(v) \rfloor$ 会破坏梯度流。因此，我们采用结合温度退火的软到硬训练。由于 $\sigma(\cdot) \in (0,1)$，因此有 $\phi_{\mathrm{learn}}(v) \in (1,S)$，这自然涵盖了所有层。在训练过程中，我们为每一层 $k \in \{1,\ldots,S\}$计算软成员权重：

$$w_k(v) = \frac{\exp(-\beta \cdot |\phi_{\mathrm{learn}}(v) - k|)}{\sum_{k'=1}^{S} \exp(-\beta \cdot |\phi_{\mathrm{learn}}(v) - k'|)} \tag{17}$$

其中，$\beta > 0$ 是一个在训练过程中逐渐增大的温度参数。ISP特征通过可微的加权聚合进行更新：

$$h_v^{(k),\mathrm{ISP}} = \sum_{k'=1}^{k} w_{k'}(v) \cdot h_v^{(k'),\mathrm{TRI}} \tag{18}$$

在推理阶段，我们使用完全离散的分层 $\phi_{\mathrm{discrete}}(v) = \lceil \phi_{\mathrm{learn}}(v) \rceil$（向下取整），并采用公式 12 中的硬门控，将每个节点恰好分配到 {1，……， S} 中的一个层级。一个关键性质是，学习得到的不变量继承了其基础组件的图不变性，从而确保学习得到的分层遵循图结构，并对结构等价节点一视同仁。

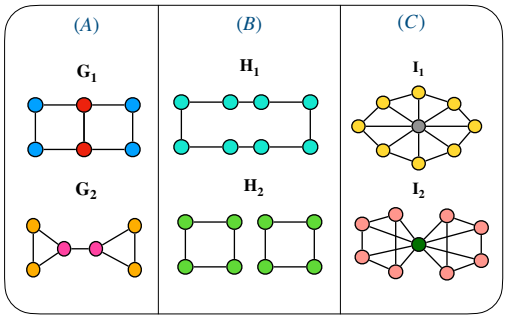


图 2：ISP-WL 区分 1-WL 无法区分的图的机制：(A) 三角形聚合：G_1 和 G_2 的度序列相同，但三角形结构不同。(B) 不变分层：无三角形图 H_1 和 H_2的结构属性不同，这些属性可由不变量捕获（例如局部围长）。(C) 层次化间隙编码：I_1 和 I_2 的三角形模式相同，但不变量层次不同（例如介数中心性）。

命题 3.5（结构一致性）。如果每个基础不变量 ϕ_i 都是图不变量，那么 ϕ_{learn} 也同样是图不变量。

ISP-WL 与 ISP-GNN 之间的对应关系见附录第 8 节。ISP-GNN 的整体架构如图1所示。

4 理论分析

我们证明了其表达能力、收敛性、复杂度和抗过平滑性。证明见附录。

4.1 表达能力

ISP-WL 通过多种互补机制增强 1-WL：基于不变量的分层和高阶聚合（见图 2）。这些机制使 ISP-WL 能够区分超越度序列障碍的图。

定理 4.1（超越 1-WL 的增强表达能力）。*ISP-WL* 的表达能力严格强于 *1-WL*。对于任意图 G_1, G_2：如果 *1-WL* 能区分 G_1 和 G_2，那么 *ISP-WL* 也能区分它们。此外，存在非同构图 G_1 和 G_2，*1-WL* 无法区分它们，但 *ISP-WL* 可以。

此外，与标准 1-WL 算法相比，ISP-WL 具有更强的节点可区分性。

定理 4.2（节点可区分性）。对于任意节点 u, v：如果$c_{1\text{-}WL}(u) \neq c_{1\text{-}WL}(v)$，那么 $c_{ISP\text{-}WL}(u) \neq c_{ISP\text{-}WL}(v)$；此外，即使 $c_{1\text{-}WL}(u) = c_{1\text{-}WL}(v)$，*ISP-WL* 仍能区分节点。

接下来，我们考察它与高阶WL变体的比较。不同于表达能力由各自细化规则固定决定的1-WL和3-WL [28], , ISP-WL的区分能力取决于不变量 ϕ。

定理4.3（依赖不变量的表达能力）。如果着色A能够区分图 G_1和 G_2，则它会产生不同的颜色多重集。令 $A \preceq B$ 表示任何被 A 区分的图对也会被 B区分，令 $A \prec B$ 表示严格不等。*ISP-WL* 的表达能力由 ϕ决定：

(1) 如果 $\phi \preceq$ *1-WL*，则：*1-WL* $\prec$ *ISP-WL*(ϕ) $\prec$ *3-WL (2)* 对任意 $k > 1$，如果ϕ和 *k-WL* 不可比较，则 *ISP-WL*(ϕ)与 *k-WL* 不可比较 *(3)* 如果对任意 $k \geq 1$，*k-WL* $\preceq \phi$，则：*k-WL* $\preceq$ *ISP-WL*(ϕ)

4.2 收敛性与复杂度

第3.1.2节确立的单次赋值性质保证了正确性和效率，因为它确保每个节点的结构信息恰好计算一次。

定理4.4（收敛）。*ISP-WL*至多经过 K =最大值（K_{WL}, L）$\leq |V|$ 次迭代即可终止，其中 L 为不同不变量值的数量，K_{WL} 为标准 *1-WL*的收敛迭代次数。

以下定理证明了ISP-WL在稀疏图上的计算效率，表明它保持了与标准 1-WL相同的渐近复杂度。

定理 4.5（时间复杂度）。*ISP-WL*的时间复杂度为O（K（$|V| + |E|$）$+ T$），其中$T = \sum_v |T(v)|$为三角形的总数。对于退化度为 d的有界退化度稀疏图，我们有$T = O(d|E|)$，当 d为常数时，总体复杂度为 O（K（$|V| + |E|$）），与标准 *1-WL*相匹配。

该结果广泛适用于现实世界网络，这些网络通常具有有界退化度 [1], ，例如社交网络、分子图和引文网络。

4.3 过平滑抵抗

深层 GNN 通常会面临过平滑问题，即随着深度增加，节点表示变得难以区分 [73]。ISP-GNN 通过固定 ISP 流维护结构信息，将其作为结构锚点，从而解决这一问题。

定理 4.6（特征坍缩抵抗性）。令 $\bar{h}_v^{ISP} := h^{(\phi(v)),\ v}$ 表示第 ϕ 层的固定结构嵌入（v）。对于具有不同结构嵌入 $\bar{h}_u^{ISP} \neq \bar{h}_v^{ISP}$ 的节点，最终的组合表示满足：

$$[h_u^{(K),WL} \| \bar{h}_u^{ISP}] \neq [h_v^{(K),WL} \| \bar{h}_v^{ISP}] \quad \forall K$$

即使 *WL*特征发生过平滑，如果 MLP_{out} 是单射的，最终节点输出在所有深度下仍保持不同。

因此，深度不变结构信息 $\bar{h}_v^{ISP}$能够防止特征坍缩。

5 实验

我们对 ISP-GNN 在多个图学习领域进行了全面评估。首先，我们通过标准基准评估其表达能力：图分类用于区分图级结构，节点分类用于捕捉局部邻域模式。其次，我们通过影响力估计任务评估模型捕捉时间动态的能力[38, 66], ；在这些任务中，信息级联以层次化传播模式在网络中传播，而这正是我们的不变量分层框架旨在建模的内容。

在主要实验中，我们主要采用度、k-核[41], 和洋葱分解 [23] 作为不变量，其计算效率介于 $O(|V|)$和 $O(|E|)$之间，同时提供

表1：不同数据集和方法的图分类性能比较。结果报告为准确率±标准差%。最佳、次佳和第三佳结果已突出显示。OOM 表示内存溢出。

方法	MUTAG	PTC_MR	BZR	DHFR	COX2	PROTEINS	D&D	IMDB-BINARY	IMDB-MULTI
GIN（2018）	92.8±5.9	65.6±6.5	91.1±3.4	84.9±4.0	88.9±2.3	78.8±4.1	82.2±3.7	78.1±3.5	54.6±3.0
ID-GNN（2021）	97.8±3.7	74.4±4.2	93.3±4.8	86.1±3.8	87.8±3.1	78.1±3.9	81.9±3.8	79.3±2.9	55.4±2.8
DropGIN (2021)	92.5±4.9	67.1±8.7	86.6±5.0	67.6±4.8	82.9±3.7	77.4±2.9	79.2±4.3	78.2±3.4	54.3±2.9
GSN（2022）	93.1±4.1	70.6±5.3	93.1±4.3	85.7±4.3	88.1±3.7	78.2±2.8	83.9±2.3	79.6±3.2	54.6±2.5
GraphSNN (2022)	94.7±1.9	70.6±3.1	91.1±3.0	85.7±3.8	86.3±3.3	78.4±2.7	81.8±3.7	78.5±2.3	55.3±3.0
GIN-AK+ (2022)	95.0±6.1	74.1±5.9	92.8±4.6	85.3±4.2	88.2±3.6	78.9±5.4	OOM	79.4±3.3	55.2±2.8
KP-GIN（2022）	95.6±4.4	76.2±4.5	93.5±4.1	86.8±3.9	88.8±3.2	79.5±4.4	84.0±3.4	79.7±3.0	55.5±2.6
NC-GNN (2024)	92.8±5.0	71.8±6.2	92.6±4.3	80.7±3.4	88.4±3.3	78.4±3.1	OOM	78.4±4.0	55.0±2.7
UnionGIN（2024）	93.6±4.7	74.8±6.3	92.4±4.2	84.2±4.5	87.6±3.4	78.9±2.5	82.3±3.5	79.1±3.1	55.0±2.7
AC-GIN (2025)	96.8±3.5	75.6±7.1	92.9±4.4	85.5±3.3	88.6±2.7	79.5±2.4	82.7±3.4	79.0±3.4	55.1±2.6
BEC-GIN (2025)	96.1±3.6	72.9±5.7	92.4±3.6	87.5±3.0	89.3±3.1	79.1±3.7	81.8±3.2	80.8±3.3	54.9±3.2
ISP-GNN^{Degree}	97.9±2.6	75.8±5.9	94.1±4.9	89.5±3.8	89.7±4.1	79.8±2.4	83.5±2.8	80.0±4.1	55.9±3.1
ISP-GNN$_{\text{Core}}$	96.7±5.3	75.5±6.0	94.3±5.3	86.5±4.8	88.9±4.1	79.6±2.6	83.4±2.4	79.8±3.8	55.7±2.9
ISP-GNN$_{\text{Onion}}$	96.8±4.8	75.9±6.1	94.6±3.9	85.9±4.5	88.7±2.9	79.4±2.8	83.8±3.1	79.6±3.6	55.6±3.0
ISP-GNN†	97.5±2.8	76.4±5.7	94.5±4.5	90.1±3.5	90.3±3.8	80.2±2.3	83.5±3.4	81.2±3.8	56.3±2.8

多样的结构视角。我们在消融研究中使用了其他不变量，以展示我们方法的有效性。关于这些不变量的更多详情，请参阅附录 10。

5.1 实验设置

数据集。我们使用 TU [42]和 OGB [27] 数据集基准评估图分类任务，其中涵盖分子、生物信息学和社交网络领域的数据集。对于节点分类，我们评估具有同配和异配标签模式的数据集，这些数据集来自 CitationFull [9],Amazon [53], WebKB [47], 和 Heterophilic [48] 基准数据集。对于影响力估计，我们采用四个真实世界社交网络 [9, 51]。

基线。对于图分类，我们将 ISP-GNN 与传统图神经网络进行比较，这类网络的表达能力受限于 1-WL（GIN [68]），以及旨在纳入高阶结构以超越 1-WL 的图神经网络（ID-GNN [62], DropGNN [46], GSN [11], GraphSNN[62], GIN-AK+ [76]、KP-GIN [17], UnionSNN [67], NC-GNN [39], AC-GNN [24], BEC-GIN [26]）。这些基线分为四类：（i）结构注入（GSN、GraphSNN、UnionSNN）；（ii）子图枚举（GIN-AK+、KP-GIN）；（iii）结构增强（ID-GNN、DropGNN、NC-GNN）；（iv）使用曲率的几何方法（AC-GNN、BEC-GIN）。对于节点分类，我们采用传统的图神经网络架构：GCN [32], GAT [56], GraphSAGE[22],，以及现代基线：方向感知模型 Dir-GNN [50], 和基于曲率的模型 BEC-GCN [26],。对于影响力估计，我们将其与传统图神经网络（GIN [68], GCN[32], GAT [56], GraphSAGE [22]）、方向感知的 DirGNN 架构 [50],、特定任务的影响力估计方法（SGNN[34],DeepIM [38],DeepIS [66],GLIE [45]），以及采用与影响力估计类似的观点传播归纳偏置设计的扩散导向模型（APPNP [20], ODNet [80], UniGO [35]）进行比较。

评估设置。对于图分类，我们采用两种设置。对于 TU 数据集，我们遵循 Feng等人 [17]，进行 10 折交叉验证，并报告最佳准确率的平均值和标准差。对于 OGB数据集，我们采用 Hu等人 [27] 的设置，重点关注 10 个随机种子下 ROC-AUC 分数的平均值和标准差。在节点分类中，我们按照 Zheng 等人 [79] 的方法采用 60/20/20 划分，并报告 10 次初始化的平均准确率。对于影响力估计，我们评估三种扩散模型：线性阈值（LT）、独立级联（IC）和易感者–感染者–易感者（SIS），并将初始激活节点设为节点总数的 10%。我们遵循 Ling 等人 [38] 的设置，报告 10 折上的平均绝对误差（MAE）均值和标准差。可用时，基线结果取自先前工作；否则，我们使用指定的超参数复现结果。

有关下游任务、基线、数据集统计、模型超参数及实现细节的更多信息见附录第 12 节。

5.2 主要结果

在我们的实验中，我们使用记号 ISP-GNN$_\phi$ 表示采用特定预定义不变量 ϕ实例化的模型，并用 ISP- GNN† 表示具有可学习不变量分层的变体。

5.2.1 图分类。表1展示了在九个 TU 基准数据集上的图分类性能。ISP-GNN 变体始终优于传统图神经网络架构和表达能力强的图神经网络基线，或与之相比具有竞争力。值得注意的是，可学习的 ISP-GNN† 变体在大多数数据集上明显优于预定义变体，并在五个基准数据集上取得了最佳性能。我们认为，这表明了学习任务特定结构表示的优势。不同的预定义变体在不同图类型上的表现各不相同，表明结构不变量的选择与特定图特征相关。

5.2.2 影响力估计。表2评估了 ISP-GNN 在四个社交网络数据集上的影响力估计性能。结果表明，ISP-GNN 的各个变体始终表现出色，跻身前三，证明了层次化结构编码在建模级联动态方面的有效性。可学习的 ISP-GNN† 变体在各种扩散模型中均表现强劲，这表明学习得到的结构分层能够有效捕捉随网络拓扑和

表2：不同数据集和方法的影响力估计性能比较。结果以平均绝对误差±标准差报告。数值越低，性能越好。最佳，次佳和第三佳结果以高亮显示。

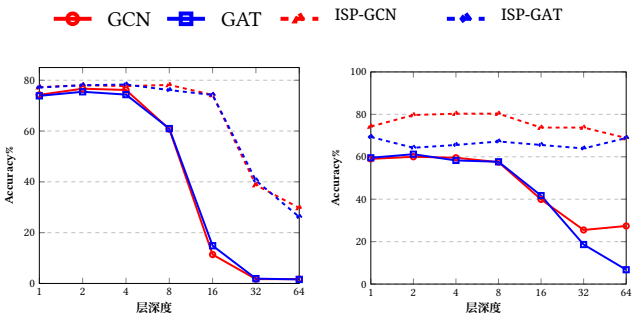
方法	Jazz			Cora-ML			网络科学			电网		
	IC	LT	SIS	IC	LT	SIS	IC	LT	SIS	IC	LT	SIS
GCN（2017）	$0.233_{\pm.010}$	$0.199_{\pm.006}$	$0.344_{\pm.023}$	$0.277_{\pm.007}$	$0.255_{\pm.008}$	$0.365_{\pm.065}$	$0.270_{\pm.019}$	$0.190_{\pm.012}$	$0.180_{\pm.007}$	$0.313_{\pm.024}$	$0.335_{\pm.023}$	$0.207_{\pm.001}$
GraphSAGE（2017）	$0.201_{\pm.028}$	$0.120_{\pm.004}$	$0.301_{\pm.018}$	$0.255_{\pm.010}$	$0.203_{\pm.019}$	$0.222_{\pm.051}$	$0.241_{\pm.010}$	$0.112_{\pm.005}$	$0.102_{\pm.005}$	$0.313_{\pm.024}$	$0.341_{\pm.018}$	$0.133_{\pm.001}$
GIN（2018）	$0.148_{\pm.012}$	$0.174_{\pm.029}$	$0.344_{\pm.013}$	$0.247_{\pm.001}$	$0.305_{\pm.008}$	$0.225_{\pm.019}$	$0.277_{\pm.005}$	$0.265_{\pm.006}$	$0.140_{\pm.007}$	$0.276_{\pm.001}$	$0.471_{\pm.002}$	$0.184_{\pm.029}$
GAT (2018)	$0.342_{\pm.005}$	$0.156_{\pm.100}$	$0.288_{\pm.017}$	$0.352_{\pm.004}$	$0.192_{\pm.010}$	$0.208_{\pm.008}$	$0.274_{\pm.002}$	$0.114_{\pm.008}$	$0.123_{\pm.013}$	$0.331_{\pm.002}$	$0.280_{\pm.015}$	$0.133_{\pm.001}$
APPNP (2018)	$0.200_{\pm.006}$	$0.124_{\pm.003}$	$0.357_{\pm.059}$	$0.265_{\pm.022}$	$0.220_{\pm.031}$	$0.321_{\pm.042}$	$0.248_{\pm.006}$	$0.084_{\pm.006}$	$0.100_{\pm.012}$	$0.290_{\pm.006}$	$0.189_{\pm.011}$	$0.132_{\pm.001}$
DeepIS（2021）	$0.151_{\pm.003}$	$0.219_{\pm.002}$	$0.434_{\pm.003}$	$0.203_{\pm.001}$	$0.301_{\pm.005}$	$0.304_{\pm.001}$	$0.223_{\pm.001}$	$0.306_{\pm.001}$	$0.256_{\pm.001}$	$0.206_{\pm.001}$	$0.374_{\pm.001}$	$0.251_{\pm.001}$
SGNN（2022）	$0.183_{\pm.004}$	$0.164_{\pm.014}$	$0.330_{\pm.007}$	$0.210_{\pm.003}$	$0.134_{\pm.004}$	$0.211_{\pm.006}$	$0.213_{\pm.003}$	$0.049_{\pm.004}$	$0.127_{\pm.000}$	$0.257_{\pm.002}$	$0.192_{\pm.001}$	$0.175_{\pm.004}$
DeepIM（2023）	$0.178_{\pm.002}$	$0.134_{\pm.014}$	$0.383_{\pm.010}$	$0.210_{\pm.002}$	$0.271_{\pm.010}$	$0.270_{\pm.006}$	$0.216_{\pm.003}$	$0.118_{\pm.003}$	$0.135_{\pm.004}$	$0.258_{\pm.003}$	$0.331_{\pm.002}$	$0.205_{\pm.005}$
GLIE（2024）	$0.136_{\pm.003}$	$0.055_{\pm.028}$	$0.454_{\pm.062}$	$0.199_{\pm.041}$	$0.286_{\pm.016}$	$0.205_{\pm.029}$	$0.163_{\pm.031}$	$0.160_{\pm.063}$	$0.103_{\pm.023}$	$0.183_{\pm.009}$	$0.384_{\pm.020}$	$0.132_{\pm.026}$
DirGNN（2024）	$0.205_{\pm.016}$	$0.124_{\pm.002}$	$0.379_{\pm.032}$	$0.255_{\pm.006}$	$0.193_{\pm.012}$	$0.229_{\pm.025}$	$0.236_{\pm.011}$	$0.084_{\pm.017}$	$0.142_{\pm.019}$	$0.287_{\pm.010}$	$0.257_{\pm.062}$	$0.132_{\pm.001}$
ODNet (2024)	$0.180_{\pm.003}$	$0.053_{\pm.003}$	$0.322_{\pm.006}$	$0.232_{\pm.001}$	$0.210_{\pm.004}$	$0.196_{\pm.002}$	$0.255_{\pm.002}$	$0.104_{\pm.015}$	$0.106_{\pm.005}$	$0.274_{\pm.001}$	$0.296_{\pm.002}$	$0.145_{\pm.001}$
UniGO（2025）	$0.192_{\pm.013}$	$0.159_{\pm.020}$	$0.335_{\pm.002}$	$0.255_{\pm.001}$	$0.155_{\pm.019}$	$0.212_{\pm.001}$	$0.201_{\pm.001}$	$0.110_{\pm.049}$	$0.108_{\pm.023}$	$0.231_{\pm.002}$	$0.235_{\pm.005}$	$0.206_{\pm.023}$
ISP-GNN^{Degree}	$0.132_{\pm.001}$	$0.137_{\pm.028}$	$0.315_{\pm.013}$	$0.162_{\pm.001}$	$0.106_{\pm.007}$	$0.191_{\pm.012}$	$0.205_{\pm.001}$	$0.045_{\pm.008}$	$0.100_{\pm.005}$	$0.245_{\pm.004}$	$0.195_{\pm.004}$	$0.126_{\pm.006}$
ISP-GNN$_{\text{Core}}$	$0.147_{\pm.011}$	$0.162_{\pm.009}$	$0.314_{\pm.012}$	$0.179_{\pm.002}$	$0.152_{\pm.007}$	$0.184_{\pm.004}$	$0.210_{\pm.004}$	$0.051_{\pm.009}$	$0.094_{\pm.001}$	$0.240_{\pm.002}$	$0.181_{\pm.004}$	$0.127_{\pm.002}$
ISP-GNN$_{\text{Onion}}$	$0.156_{\pm.009}$	$0.155_{\pm.003}$	$0.294_{\pm.014}$	$0.176_{\pm.002}$	$0.134_{\pm.006}$	$0.187_{\pm.004}$	$0.160_{\pm.003}$	$0.060_{\pm.002}$	$0.095_{\pm.001}$	$0.263_{\pm.001}$	$0.222_{\pm.005}$	$0.120_{\pm.002}$
ISP-GNN†	$0.135_{\pm.003}$	$0.058_{\pm.005}$	$0.285_{\pm.011}$	$0.168_{\pm.003}$	$0.112_{\pm.009}$	$0.189_{\pm.007}$	$0.168_{\pm.005}$	$0.048_{\pm.006}$	$0.096_{\pm.003}$	$0.180_{\pm.003}$	$0.186_{\pm.006}$	$0.124_{\pm.004}$

传播机制而变化的扩散相关模式。不同的预定义变体在互补场景中各有所长：基于度的分层能够捕捉直接影响力传播模式，而基于核心和洋葱的分层则更擅长刻画社区结构对级联的影响。总体而言，ISP-GNN 的表现可与针对特定任务的影响力估计方法相媲美，同时能够泛化到多种扩散模型和图结构，表明我们的方法不仅适用于静态图属性预测，还能带来额外收益。

5.2.3 节点分类。 表3评估了ISP-GNN在八个基准数据集上的节点分类性能。我们通过将可学习的结构颜色嵌入与基线嵌入相结合来增强基线架构，从而生成ISP变体。ISP-GNN† 持续提升性能，证明了层次化结构编码对节点级预测任务的有效性。我们的方法在异配数据集上取得了显著改进，而传统消息传递在这些数据集上表现不佳，这表明当局部同配假设失效时，ISP 颜色嵌入能够捕获与基于特征的邻域聚合互补的结构模式。在同配网络上，ISP-GNN† 带来了适度但稳定的改进，表明即使节点特征与标签高度相关，结构分层仍能提供有价值的位置信息。

5.3 消融研究

5.3.1 过平滑分析。图 3 考察了随着网络深度增加，ISP-GNN 对过平滑的鲁棒性。在同配数据集（Citeseer）和异配数据集（Texas）上，基线 GCN 和 GAT 模型在超过 8 层后均出现严重的性能下降，准确率因过平滑而大幅降低。相比之下，ISP-GNN 变体在更深的架构中仍能保持显著更高的性能，这表明结构颜色嵌入能够缓解过平滑问题，即使特征表示趋于收敛，也能保留节点的可区分性。



(a) Citeseer **(b) Texas**

图3：过平滑对比

5.3.2 组件分析。 为评估各机制的贡献，我们对 ISP-GNN 的核心组件进行了消融，包括基于不变量的分层、三角形聚合以及层次化结构异质性编码，并在四个图分类基准上加以评估。表4中的结果表明，三角形聚合是最关键的要素，移除它会导致最大的性能下降，在社交网络中尤为明显。不变分层的重要性位居其次，而层次间隙则带来稳定而适度的提升。总体而言，三种机制的组合所带来的性能优于任何单一组件。

表4：图分类的组件分析。结果以平均准确率±标准差（%）报告。最佳结果以粗体标出。

变体	MUTAG	PTC-MR	PROTEINS	IMDB-BINARY
仅WL	92.8 ± 5.9	65.6 ± 6.5	78.8 ± 4.1	78.1 ± 3.5
无分层	95.2 ± 3.4	73.5 ± 6.2	79.2 ± 2.8	79.4 ± 4.2
无三角形聚合	93.5 ± 4.1	71.5 ± 6.8	79.0 ± 3.1	78.8 ± 4.5
无层次间隙	94.2 ± 3.6	72.8 ± 6.4	79.3 ± 2.9	79.2 ± 4.3
ISP-GNN†	**97.5 ± 2.8**	**76.4 ± 5.7**	**80.2 ± 2.3**	**80.5 ± 3.9**

表3：不同数据集和方法的节点分类性能对比。结果以准确率 ±标准差（%）表示。 Δ 表示 ISP 变体相较于各自基线的最大百分比提升。

方法	Cora ML	Citeseer	Pubmed	DBLP	Film	康奈尔	威斯康星	Texas
GraphSAGE (2017)	86.52 ± 1.32	76.04 ± 1.30	88.45 ± 0.50	86.16 ± 0.50	34.23 ± 0.99	75.95 ± 5.01	81.18 ± 5.56	82.43 ± 6.14
ISP-GraphSAGE†	87.23 ± 1.78	77.91 ± 0.92	88.51 ± 0.58	83.93 ± 0.60	35.89 ± 1.18	78.51 ± 5.98	87.38 ± 5.05	83.61 ± 6.96
Δ (%)	+0.98	+2.89	+0.07	-2.10	+6.31	+6.45	+7.64	+7.39
GCN (2018)	87.07 ± 1.21	76.68 ± 1.64	86.74 ± 0.47	83.93 ± 0.34	30.26 ± 0.79	55.14 ± 8.46	61.60 ± 7.00	60.00 ± 6.45
ISP-GCN†	87.29 ± 1.20	78.05 ± 1.70	89.01 ± 0.42	84.13 ± 0.53	35.10 ± 1.03	64.47 ± 8.94	66.75 ± 6.35	79.67 ± 3.53
Δ (%)	+0.25	+1.89	+2.62	+0.82	+15.99	+22.71	+8.36	+32.78
GAT (2018)	84.12 ± 0.55	75.46 ± 1.72	87.24 ± 0.55	80.61 ± 1.21	26.28 ± 1.73	53.64 ± 11.1	60.00 ± 11.0	61.21 ± 8.17
ISP-GAT†	87.14 ± 1.18	78.02 ± 1.32	88.05 ± 0.91	84.25 ± 0.63	32.67 ± 3.68	71.70 ± 4.47	62.75 ± 12.55	64.26 ± 2.44
Δ (%)	+3.83	+4.04	+1.09	+4.52	+26.33	+33.67	+6.25	+4.98
DIRGNN (2024)	85.66 ± 0.31	77.71 ± 0.78	86.94 ± 0.55	81.22 ± 0.54	35.76 ± 1.68	76.51 ± 6.14	80.50 ± 5.50	76.25 ± 6.31
ISP-DIRGNN†	86.38 ± 1.09	77.86 ± 1.73	89.16 ± 0.78	83.53 ± 0.54	36.27 ± 1.74	81.28 ± 4.34	88.63 ± 4.05	84.26 ± 3.13
Δ (%)	+1.68	+1.12	+2.55	+2.91	+1.85	+6.23	+11.49	+13.30
BEC-GCN（2025）	88.65 ± 1.32	78.14 ± 1.31	87.00 ± 0.28	85.20 ± 0.48	34.34 ± 2.08	65.96 ± 9.03	66.25 ± 4.55	68.85 ± 11.02
ISP-BEC-GCN†	89.11 ± 1.70	78.58 ± 1.68	88.21 ± 0.92	86.43 ± 0.87	35.09 ± 1.88	68.09 ± 9.06	73.75 ± 4.65	70.49 ± 10.89
Δ (%)	+0.51	+0.56	+1.39	+1.44	+2.18	+3.23	+11.32	+2.38

5.3.3 可扩展性分析。 为评估 ISP-GNN 在大型数据集上的效率，我们在 ogbg-molpcba 上评估其运行时间与性能；该基准包含 437,929 个图，用于分子性质预测。为公平比较，所有模型均使用 5 层，嵌入维度为 300。结果如图4所示。

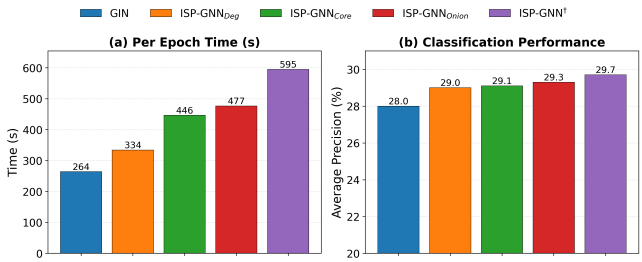


图4：ogbg-molpcba数据集上的运行时间和性能。(a) 每个轮次的耗时（秒）。(b) 平均精确率。结果展示了采用不同结构不变量的 ISP-GNN。

ISP-GNN 在大型分子图上展现出良好的实际可扩展性。基于度的分层得益于高效的不变量计算，仅使基线 GIN 的开销增加了 26%。核分解由于需要通过迭代剥离识别 k-核，成本更高；而洋葱分解通过构建粒度更细的层次层，进一步增加了运行时间，因为这需要额外的处理迭代。可学习变体通过对多个不变量进行优化，以最高的计算成本换取最佳性能。这一系列方案使从业者能够根据应用需求平衡效率与表达能力，同时所有变体在大型数据集上仍具有良好的可处理性。

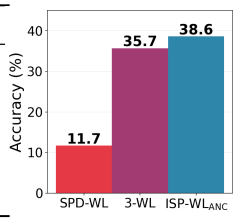
5.3.4 表达能力比较。 我们使用BREC基准 [58], 评估ISP-WL的表达能力，同时报告每个图对的平均处理时间与准确率。结果（表5）表明，ISP-WL始终优于1-WL和SPD-WL [58]等标准WL变体。性能随不变量复杂度提升：度仅带来最小增益，介数中心性（BC）提供中等改进，而复杂不变量（如平均邻域聚类（ANC））取得最强性能。值得注意的是，ISP-WL变体保持了实用效率。即使ISP-WL$_{ANC}$ 处理图的速度也比3-WL快七倍，同时达到其三分之二的表达能力。这种依赖于不变量的表达能力直接验证了定理4.3。关键在于，当不变量与3-WL不可比时，ISP-WL能够区分3-WL无法区分的图，同时保持更低的复杂度，这一点已在正则图上得到证明。尽管在标准BREC上，使用通用不变量时ISP-WL仍以上界受限于3-WL，这些结果证实，策略性的不变量选择使ISP-WL能够高效地补充高阶方法。

表5：BREC基准上的表达能力评估

(a) 整体性能

方法	准确率(%)	时间 (ms)
1-WL	0.0	2.58
SPD-WL	20.8	7.08
ISP-WL^{Deg}	23.3	4.35
ISP-WL$_{BC}$	30.5	7.59
ISP-WL$_{ANC}$	44.5	15.25
3-WL	67.5	101.42

(b) 正则图性能



5.3.5 补充实验。在附录中，我们展示以下实验：(1) 不同图不变量的性能比较（第 13.1 节），(2) 层次化结构间隙编码的消融研究，包括组件分析和向高阶模体的扩展（第 9 节），(3) ISP-GNN† 的任务自适应结构发现和超参数敏感性（第 13.3 节），(4) 具有多种不变量的 ISP-WL 可扩展性分析（第 13.4 节），以及 (5) 关于跨不变量着色细化、不变量正交性和收敛动态的可视化分析（第 13.5 节）。

6 结论、局限性与未来展望 Work

我们提出了不变分层传播，这是一种基于图不变量对节点进行层次化分层、突破 1-WL 表达能力壁垒的框架。通过编码结构

异质性，该方法能够捕捉节点在局部模式中占据不同结构位置的差异。该框架包含一种新颖的 WL 变体，具有形式化的表达能力保证；还包含其采用可学习不变量的高效神经网络实现。大量实验表明，在各种下游任务中，与标准消息传递架构和最先进的表达能力强的基线方法相比，该框架均取得了稳定的性能提升。

我们在附录第14节中深入讨论了本工作的局限性及未来可能的研究方向。此外，对于我们可学习的 ISP-GNN 变体，层级层数和温度退火参数被设为超参数。虽然我们提供了敏感性分析，以指导从业者选择这些值，但我们计划在未来的工作中以数据驱动方式学习这些参数。

参考文献

[1] Peter Allen、Julia Böttcher 和 Anita Liebenau。2023。有界退化度图的普适性。*arXiv*预印本 *arXiv:2309.05468* (2023)。[2] Waiss Azizian 等。2021。不变与等变图神经网络的表达能力。发表于国际学习表征会议。[3] Waïss Azizian 和 Marc Lelarge。2021。不变与等变图神经网络的表达能力。发表于国际学习表征会议。[4] Albert-László Barabási 和 Réka Albert。1999。随机网络中的标度涌现。《科学》286，5439 (1999)，509–512。[5] Marc Barthelemy。2004。大型复杂网络中的介数中心性。《欧洲物理学杂志 *B*辑》38，2 (2004)，163–168。[6] Nesrine Ben Yahia。2024。利用基于堆叠式图神经网络的社区检测增强社会与协作学习。《社会网络分析与挖掘》14，1 (2024)，205。[7] Beatrice Bevilacqua、Fabrizio Frasca、Derek Lim、Balasubramaniam Srinivasan、Chen Cai、Gopinath Balamurugan、Michael M Bronstein 和 Haggai Maron。2022。等变子图聚合网络。国际学习表征会议(2022)。[8] Filippo Maria Bianchi、Daniele Grattarola 和 Cesare Alippi。2020。利用图神经网络进行图池化的谱聚类。发表于国际机器学习会议。PMLR，874–883。[9] Aleksandar Bojchevski 和 Stephan Günnemann。2018。图的深度高斯嵌入：基于排序的无监督归纳学习。发表于国际学习表征会议。[10] Phillip Bonacich。1987。权力与中心性：一类度量。《美国社会学杂志》92，5 (1987)，1170–1182。[11] Giorgos Bouritsas、Fabrizio Frasca、Stefanos P Zafeiriou 和 Michael Bronstein。2022。通过子图同构计数提升图神经网络的表达能力。*IEEE*模式分析与机器智能汇刊 (2022)。[12] Michail Chatzianastasis、Johannes Lutzeyer、George Dasoulas 和 Michalis Vazirgiannis。2023。图排序注意力网络。发表于 *AAAI*人工智能会议论文集，第37卷。7006–7014。[13] Rongqin Chen、Yan Li、Dan Wu、Fan Mo、Shenghui Zhang、Pak Lon Ip、Hoi Cheong Iam、Ye Li 和 Leong Hou U。2025。通过局部连通性、谱编码和距离编码增强2-FWL图神经网络中的子图学习。发表于第*31*届*ACM SIGKDD*知识发现与数据挖掘会议论文集，第*2*卷。227–238。[14] Leonardo Cotta、Christopher Morris 和 Bruno Ribeiro。2021。强大图表示的重构。神经信息处理系统进展 (2021)。[15] Erik D Demaine、Felix Reidl、Peter Rossmanith、Fernando Sánchez Villaamil、Somnath Sikdar 和 Blair D Sullivan。2019。复杂网络的结构稀疏性：随机模型和现实世界图中的有界扩张。《计算机系统科学杂志》105 (2019)，199–241。[16] Vijay Prakash Dwivedi 和 Xavier Bresson。2020。将Transformer网络推广到图。*arXiv*预印本 *arXiv:2012.09699* (2020)。[17] Jiarui Feng、Yixin Chen、Fuhai Li、Anindya Sarkar 和 Muhan Zhang。2022。k跳消息传递图神经网络的表达能力有多强？神经信息处理系统进展35 (2022)，4776–4790。[18] Yaniv Galron、Fabrizio Frasca、Haggai Maron、Eran Treister 和 Moshe Eliasof。2025。理解并改进时序图神经网络的拉普拉斯位置编码。发表于欧洲机器学习与数据库知识发现联合会议。Springer，420–437。

[19] Hongyang Gao 和 Shuiwang Ji。2019。Graph U-Net。在国际机器学习会议上。PMLR，2083–2092。[20] Johannes Gasteiger、Aleksandar Bojchevski 和 Stephan Günnemann。2018。先预测，再传播：图神经网络与个性化PageRank相遇。在国际学习表征会议上。[21] Justin Gilmer、Samuel S Schoenholz、Patrick F Riley、Oriol Vinyals 和 George E Dahl。2017。用于量子化学的神经消息传递。在国际机器学习会议上。PMLR，1263–1272。[22] Will Hamilton、Zhitao Ying 和 Jure Leskovec。2017。在大规模图上进行归纳表示学习。神经信息处理系统进展30（2017）。[23] Laurent Hébert-Dufresne、Joshua A Grochow 和 Antoine Allard。2016。通过一种新的网络统计量——洋葱分解，研究多尺度结构和拓扑异常检测。《科学报告》6，1（2016），31708。[24] Asela Hevapathige。2025。用于分子表示学习的曲率增强图神经网络。Physica Scripta 100，9（2025），096010。[25] Asela Hevapathige、Qing Wang 和 Ahad N Zehmakan。2025。DeepSN：用于影响力最大化的层束神经网络框架。在AAAI人工智能会议论文集，第39卷。17177–17185。[26] Asela Hevapathige、Ahad N Zehmakan 和 Qing Wang。2025。通过可学习的Bakry-Émery曲率实现深度自适应图神经网络。在第31届ACM SIGKDD知识发现与数据挖掘会议论文集V. 2。944–955。[27] Weihua Hu、Matthias Fey、Marinka Zitnik、Yuxiao Dong、Hongyu Ren、Bowen Liu、Michele Catasta 和 Jure Leskovec。2020。开放图基准：用于图上机器学习的数据集。神经信息处理系统进展33（2020），22118–22133。[28] Ningyuan Teresa Huang 和 Soledad Villar。2021。Weisfeiler-Lehman检验及其变体简明教程。在ICASSP 2021—2021 IEEE国际声学、语音与信号处理会议（ICASSP）上。IEEE，8533–8537。[29] Kedar Karhadkar、Pradeep Kr Banerjee 和 Guido Montufar。2023。FoSR：用于解决图神经网络中过挤压问题的一阶谱重连。在第十一届国际学习表征会议上。[30] Sung Jin Kim 和 Sang Ho Lee。2002。改进PageRank算法的计算方法。在欧洲信息检索会议上。Springer，73–85。[31] DP Kingma。2015。Adam：一种随机优化方法。在国际学习表征会议上。[32] Thomas N Kipf 和 Max Welling。2017。基于图卷积网络的半监督分类。在国际学习表征会议上。[33] Johannes Kobler、Uwe Schöning 和 Jacobo Torán。2012。图同构问题：其结构复杂度。Springer Science & Business Media。[34] Sanjay Kumar、Abhishek Mallik、Anavi Khetarpal 和 Bhawani Sankar Panda。2022。利用图嵌入和图神经网络实现社交网络中的影响力最大化。《信息科学》607（2022），1617–1636。[35] Hao Li、Hao Jiang、Yuke Zheng、Hao Sun 和 Wenying Gong。2025。UniGO：用于建模图上观点动力学的统一图神经网络。在2025年ACM网络会议论文集上。530–540。[36] Yusheng Li、Yilun Shang 和 Yiting Yang。2017。大型网络的聚类系数。《信息科学》382（2017），350–358。[37] Ke Liang、Lingyuan Meng、Meng Liu、Yue Liu、Wenxuan Tu、Siwei Wang、Sihang Zhou、Xinwang Liu、Fuchun Sun 和 Kunlun He。2024。图类型上的知识图谱推理综述：静态、动态和多模态。《IEEE模式分析与机器智能汇刊》46，12（2024），9456–9478。[38] Chen Ling、Junji Jiang、Junxiang Wang、My T Thai、Renhao Xue、James Song、Meikang Qiu 和 Liang Zhao。2023。用于影响力最大化的深度图表示学习与优化。在国际机器学习会议上。PMLR，21350–21361。[39] Meng Liu、Haiyang Yu 和 Shuiwang Ji。2024。通过边感知Weisfeiler-Leman算法增强图神经网络。机器学习研究汇刊（2024）。[40] Laurens van der Maaten 和 Geoffrey Hinton。2008。使用t-SNE可视化数据。《机器学习研究期刊》9，11月（2008），2579–2605。[41] Fragkiskos D Malliaros、Christos Giatsidis、Apostolos N Papadopoulos 和 Michalis Vazirgiannis。2020。网络的核分解：理论、算法与应用。VLDB期刊29，1（2020），61–92。[42] Christopher Morris、Nils Morten Kriege、Franka Bause、Kristian Kersting、Petra Mutzel 和 Marion Neumann。2020。TUDataset：用于图学习的基准数据集集合。在ICML 2020图表示学习及其他主题研讨会（GRL+ *2020*）。[43] Christopher Morris、Martin Ritzert、Matthias Fey、William L Hamilton、Jan Eric Lenssen、Gaurav Rattan 和 Martin Grohe。2019。Weisfeiler 和 Leman 走向神经网络：高阶图神经网络。在AAAI人工智能会议论文集，第33卷。4602–4609。[44] Sunil Nishad、Shubhangi Agarwal、Arnab Bhattacharya 和 Sayan Ranu。2021。GraphReach：利用可达性估计的具位置感知能力的图神经网络。第三十届国际人工智能联合会议（2021）。

[45] George Panagopoulos、Nikolaos Tziortziotis、Michalis Vazirgiannis、Jun Pang 和 Fragkiskos D Malliaros。2024。学习用于影响力最大化的图表示。《社会网络分析与挖掘》14，1（2024），203。[46] Pál András Papp、Karolis Martinkus、Lukas Faber 和 Roger Wattenhofer。2021。DropGNN：随机丢弃可提升图神经网络的表达能力。神经信息处理系统进展34（2021），21997–22009。[47] Hongbin Pei、Bingzhe Wei、Kevin Chen Chuan Chang、Yu Lei 和 Bo Yang。2020。GEOM-GCN：几何图卷积网络。在第八届国际学习表征会议上。[48] Oleg Platonov、Denis Kuznedelev、Michael Diskin、Artem Babenko 和 Liudmila Prokhorenkova。2023。审视异配条件下图神经网络的评估：我们真的在取得进展吗？在第十一届国际学习表征会议上。[49] Chendi Qian、Gaurav Rattan、Floris Geerts、Mathias Niepert 和 Christopher Morris。2022。有序子图聚合网络。神经信息处理系统进展35（2022），21030–21045。[50] Emanuele Rossi、Bertrand Charpentier、Francesco Di Giovanni、Fabrizio Frasca、Stephan Günnemann 和 Michael M Bronstein。2024。边方向性可提升异配图上的学习效果。在图学习会议上。PMLR，25–1。[51] Ryan Rossi 和 Nesreen Ahmed。2015。包含交互式图分析与可视化功能的网络数据仓库。在AAAI人工智能会议论文集，第29卷。[52] Ketan Rajshekhar Shahapure 和 Charles Nicholas。2020。使用轮廓系数进行聚类质量分析。在2020年IEEE第七届数据科学与高级分析国际会议（DSAA）上。IEEE，747–748。[53] Oleksandr Shchur、Maximilian Mumme、Aleksandar Bojchevski 和 Stephan Günnemann。2018。图神经网络评估中的陷阱。arXiv预印本arXiv:1811.05868（2018）。[54] Balasubramaniam Srinivasan 和 Bruno Ribeiro。2020。位置节点嵌入与结构图表示的等价性。在国际学习表征会议上。[55] Jake Topping、Francesco Di Giovanni、Benjamin Paul Chamberlain、Xiaowen Dong 和 Michael M Bronstein。2022。通过曲率理解图上的过度挤压与瓶颈。在国际学习表征会议上。[56] Petar Veličković、Guillem Cucurull、Arantxa Casanova、Adriana Romero、Pietro Liò 和 Yoshua Bengio。2018。图注意力网络。在国际学习表征会议上。[57] Yuyang Wang、Zijie Li 和 Amir Barati Farimani。2023。分子图神经网络。在分子科学中的机器学习。Springer，21–66。[58] Yanbo Wang 和 Muhan Zhang。2024。对已实现的图神经网络表达能力进行实证研究。在第41届国际机器学习会议论文集上。52134–52155。[59] Yasida I nsika Wanigatunga 和 Asela Hevapathige。2025。利用富人俱乐部引导的表征学习揭示分子中的结构层级。在2025年第15届高级计算机信息技术国际会议（ACIT）上。IEEE，987–991。[60] Boris Weisfeiler 和 Andrei Leman。1968。将图约化为规范形式及其中出现的代数。nti，系列2，9（1968），12–16。[61] Oliver Wieder、Stefan Kohlbacher、Mélaine Kuenemann、Arthur Garon、Pierre Ducrot、Thomas Seidel 和 Thierry Langer。2020。图神经网络分子性质预测简明综述。今日药物发现：技术37（2020），1–12。[62] Asiri Wijesinghe 和 Qing Wang。2022。关于图神经网络如何超越Weisfeiler-Lehman检验的新视角？在国际学习表征会议上。[63] Ron Wilson 和 Alexander Din。2018。计算不同尺度下地点之间的聚类情况。《城市景观》20，1（2018），215–232。[64] Dengsheng Wu、Huidong Wu 和 Jianping Li。2024。层级感知自适应图神经网络。《IEEE知识与数据工程汇刊》（2024）。[65] Jian Wu、Alison Goshulak、Venkatesh Srinivasan 和 Alex Thomo。2018。在单台消费级计算机上对大型网络进行K-嵌套分解。在2018年IEEE/ACM社会网络分析与挖掘进展国际会议（ASONAM）上。IEEE，873–880。[66] Wenwen Xia、Yuchen Li、Jun Wu 和 Shenghong Li。2021。DeepIS：社交网络上的易感性估计。在第14届ACM国际网络搜索与数据挖掘会议论文集上。761–769。[67] Jiaxing Xu、Aihu Zhang、Qingtian Bian、Vijay Prakash Dwivedi 和 Yiping Ke。2024。并集子图神经网络。在AAAI人工智能会议论文集，第38卷。16173–16183。[68] Keyulu Xu、Weihua Hu、Jure Leskovec 和 Stefanie Jegelka。2018。图神经网络的能力有多强？在国际学习表征会议上。[69] Zhitao Ying、Jiaxuan You、Christopher Morris、Xiang Ren、Will Hamilton 和 Jure Leskovec。2018。利用可微池化进行层次化图表征学习。神经信息处理系统进展31（2018）。

[70] Jiaxuan You、Jonathan M Gomes-Selman、Rex Ying和Jure Leskovec. 2021. 身份感知图神经网络. 载于*AAAI*人工智能会议论文集. [71] Jiaxuan You、Rex Ying和Jure Leskovec. 2019. 位置感知图神经网络. 载于国际机器学习会议. PMLR, 7134–7143.[72] Shengzhong Zhang、Yimin Zhang、Bisheng Li、Wenjie Yang、Min Zhou和Zengfeng Huang. 2025. 面向可扩展图神经网络的图批次粗化框架. 神经网络 183 (2025), 106931. [73] Xu Zhang、Yonghui Xu、Wei He、Wei Guo和Lizhen Cui. 2023. 图神经网络中过平滑问题的全面综述. 载于*CCF*计算机支持的协同工作与社会计算会议. Springer, 451–465. [74] Yongqi Zhang和Quanming Yao. 2022. 基于关系有向图的知识图谱推理. 载于*2022*年*ACM Web*会议论文集. 912–924. [75] Lingxiao Zhao、Wei Jin、Leman Akoglu和Neil Shah. 2022. 从星形结构到子图：利用局部结构感知提升任意图神经网络. 国际学习表征会议 (2022). [76] Lingxiao Zhao、Wei Jin、Leman Akoglu和Neil Shah. 2022. 从星形结构到子图：利用局部结构感知提升任意图神经网络. 载于国际学习表征会议. [77] Lingxiao Zhao、Neil Shah和Leman Akoglu. 2022. 一种实用的渐进式表达图神经网络. 神经信息处理系统进展 35 (2022), 34106–34120.[78] Wanying Zhao、Jin Mao和Kun Lu. 2018. 共词网络上的主题排序：探索不同指标之间的关系. 信息处理与管理 54, 2 (2018), 203–218. [79] Yizhen Zheng、He Zhang、Vincent Lee、Yu Zheng、Xiao Wang和Shirui Pan. 2023. 寻找缺失半边：面向同配性图与异配性图的图互补学习. 载于国际机器学习会议. PMLR, 42492–42505. [80] Bingxin Zhou、Outongyi Lv、Jing Wang、Xiang Xiao和Weishu Zhao. 2024. ODNet：面向图与超图的受意见动力学启发的神经消息传递. 机器学习研究汇刊 (2024).

附录

7 背景

7.1 图神经网络

图神经网络通过迭代邻域聚合学习节点表示。在第 k层，每个节点 v通过将其先前状态与聚合后的邻居信息相结合来更新其嵌入 $h_v^{(k)}$:

$$h_v^{(k)} = \mathrm{UPD}^{(k)}\left(h_v^{(k-1)}, \mathrm{AGG}^{(k)}\left(\{h_u^{(k-1)} : u \in N(v)\}\right)\right) \quad (19)$$

其中，AGG(·) 是一种置换不变函数，用于从邻居收集信息；UPD(·) 是一种可学习的变换，用于生成新的节点嵌入 [68]。这一消息传递框架保持置换等变性，并可扩展到不同规模的图。

7.2 图同构

定义 7.1 （图同构）。如果存在一个双射 $\pi: V_1 \to V_2$，使得：图 $G_1 = (V_1, E_1)$ 和 $G_2 = (V_2, E_2)$ 称为同构，记作 $G_1 \cong G_2$:

$$(u, v) \in E_1 \iff (\pi(u), \pi(v)) \in E_2 \quad \forall u, v \in V_1 \quad (20)$$

双射 π 称为同构。

直观而言，在节点重新标记 [33]后，同构的图具有相同的结构。

7.3 Weisfeiler-Leman 检验

1-WL 检验 [60] 通过迭代细化节点颜色，为图同构提供一种启发式方法。首先，对所有 $v \in V$进行统一着色$c^{(0)}(v) = c_0$，然后在每次迭代中，根据邻居颜色多重集更新颜色:

$$c^{(k)}(v) = \mathrm{Hash}\left(c^{(k-1)}(v), \{\!\{c^{(k-1)}(u) : u \in N(v)\}\!\}\right) \quad (21)$$

若两个图的最终颜色多重集不同，则可将它们区分开。1-WL 检验为标准消息传递神经网络的表达能力设定了上限：在单射聚合和统一初始化的条件下，图神经网络无法区分 1-WL 无法区分的图 [43, 68]。

7.4 图不变量

定义 7.2 （图不变量）。 图不变量是指在图同构下保持不变的性质。对于节点级不变量，这意味着存在一个函数 $\mu: V \to \mathbb{R}$，使得对于任意图同构 $\pi: V \to V$，对于所有 $v \in V$，均有 $\mu(\pi(v)) = \mu(\{$v13)。

这确保不变量反映图的内在结构：具有相同结构角色的节点会获得相同的值，与节点标记无关。此类标准不变量包括节点度、核数、洋葱数、聚类系数、PageRank 和介数中心性 [23, 41, 78]。

8 ISP-WL 与 ISP-GNN 之间的对应关系

我们证明，ISP-GNN 忠实地实现了 ISP-WL 的组合原理。该神经网络架构直接映射离散算法的双流结构，其中 ISP-WL 的每项操作在 ISP-GNN 中都有对应的连续可微形式：

我们通过以下表达能力结果对这种对应关系进行形式化：

定理 8.1（ISP-WL $\subseteq$ ISP-GNN 表达能力）。如果 *ISP-GNN* 使用单射聚合和更新函数，且嵌入维度足够，则 *ISP-WL* 的区分结果得以保留：

$$\textit{ISP-WL}_K(G_1) \neq \textit{ISP-WL}_K(G_2)$$

这意味着存在参数 θ^*，使得

$$\textit{ISP-GNN}_K(G_1; \theta^*) \neq \textit{ISP-GNN}_K(G_2; \theta^*)$$

证明。我们通过归纳法构造一个从 ISP-WL 颜色到 ISP-GNN 嵌入的单射映射 $\Psi^{(k)}: \mathsf{C}^{(k)} \to \mathbb{R}^d$。在第零次迭代中，我们将初始颜色映射到输入特征。对于归纳步骤，假设 $\Psi^{(k-1)}$ 是单射。由于 ISP-WL 通过对当前颜色、邻居颜色和结构信息进行哈希来更新颜色，而 ISP-GNN 则通过对这些相同组成部分进行单射聚合与变换来更新嵌入，因此不同的颜色会产生不同的嵌入。根据归纳假设，不同的邻居颜色对应不同的邻居嵌入，而这些嵌入会经过单射聚合。在容量足够时，更新函数能够在有限颜色空间上实现单射映射，从而确保 $\Psi^{(k)}$ 是单射。在图级别上，在 $\Psi^{(K)}$下，不同的颜色多重集会映射为不同的嵌入多重集，并可通过单射图级聚合加以区分。□

这一结果表明，ISP-GNN 的表达能力至少与 ISP-WL 相当，验证了该神经网络架构是离散算法的一种有理论依据的实现，同时支持端到端学习。

表 6：ISP-WL算法与 ISP-GNN 神经网络实现之间的对应关系。

ISP-WL 组件	ISP-GNN 组件
基于哈希的颜色细化	神经聚合 + 多层感知机
多重集 $\{\{c(u) : u \in N(v)\}\}$	聚合$\{h_u : u \in N(v)\}$
间隙编码 $\delta(v, u, w)$	间隙特征 d_{vuw}
$\phi(v)$处的层分配	门控 $\gamma(k)_v = \mathbb{I}[\phi(v) = k]$
颜色空间 $\mathsf{C}^{(k)}$	嵌入空间 $\mathbb{R}^d$

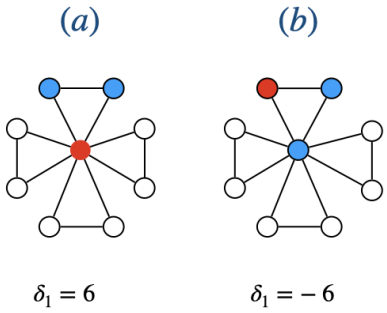
9 层次间隙的设计依据与扩展

本节将讨论我们设计层次间隙编码机制的依据，以及其在三角形之外的潜在扩展。

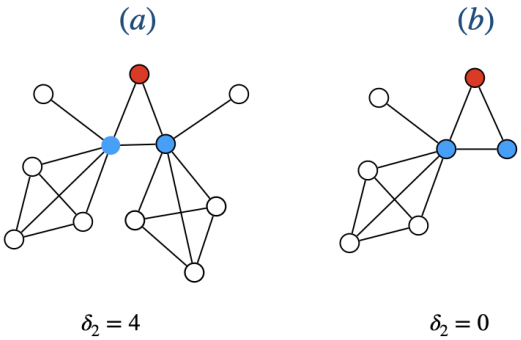
9.1 层次间隙编码属性

三分量间隙编码 $\delta(v, u, w) = (\delta_1, \delta_2, \delta_3)$ 旨在从三个互补方面捕捉三角形完整的结构异质性：

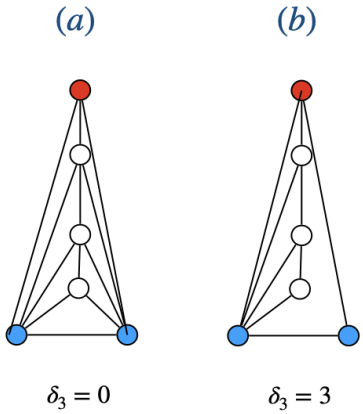
层次化支配（δ_1）：最大间距 $\delta_1 = \max(\phi(v) - \phi(u)\ \phi(v) - \phi(w))$衡量中心节点在结构排序中领先其最远邻居的程度。较大的正值表示其扮演枢纽型角色，而负值则表示其处于外围位置。



(a) 层次化支配（δ_1）



(b) 层次广度（δ_2）



(c) 邻居同质性（δ_3）

图 5：间隙编码的各组成部分以度为不变量，捕捉互补的结构属性。(a) δ_1 区分枢纽节点与外围角色。(b) δ_2 捕捉与最近邻居之间的结构邻近性。(c) δ_3 衡量邻居相似性。

层次广度（δ_2）：最小间距 $\delta_2 = \min(\phi(v)$ _ $\phi\ u\ \ \phi\ v\ -\phi\ w$ ( ) ( ) ( ))衡量与最近邻居之间的间距。这表明中心节点是与两个邻居都相距较远（$|\delta_2$较大），还是至少与一个邻居相近（$|\delta_2$较小）。

邻居同质性 (δ_3)：邻居间隙 $\delta_3 = |\phi(u) - \phi(w)|$ 衡量邻居之间的结构相似性，不受中心节点位置的影响。数值较大表示中心节点连接了结构上不相似的区域，而数值较小则表明存在局部社区结构。

这三个分量构成一个最小充分统计量：给定$\phi(v)$和 δ，即可还原多重集$\{\phi(v)\ \phi(u)\ \phi(w)\}$，且没有任何分量是冗余的。此外，根据命题3.2，该编码满足 $\delta(v,\ u,\ w) = \delta(v,\ w,\ u)$，从而能够在 ISP-WL 中高效地进行基于哈希的着色，并实现一致的神经处理。

9.2 扩展到高阶模体

间隙编码可以从三角形推广到任意高阶模式。对于任意一个指定了中心节点v和参与邻居$\{u_1, \ldots\ldots, u_m\}$的模体，我们计算：

$$\delta^{\text{center}}(v) = \text{sort}([\phi(v) - \phi(u_i)]_{i=1}^{m}) \in \mathbb{R}^m \tag{22}$$

用于捕捉排序后的中心到邻居的间隙，并且

$$\delta^{\text{inter}} = \text{sort}([|\phi(u_i) - \phi(u_j)|]_{(i,j)\in E_M}) \in \mathbb{R}^{|E_M|} \tag{23}$$

用于捕捉模体中相连邻居之间的间隙，其中 E_M 表示模体中邻居间的边。排序可确保置换不变性。

对于三角形，这会简化为我们的原始编码，其中包含 $\delta^{\text{center}} = (\delta_1,\ \delta_2)$ 和 $\delta^{\text{inter}} = (\delta_3)$。对于 k-团，其中所有邻居都相连，$|E_M| = \binom{k-1}{2}$，我们得到 $_{(k-1)_2}{}^{k}$ 个总值。对于其他模体，例如路径或星形结构，邻居间隙的数量取决于模体的边结构。

虽然该框架可扩展至任意模式，但出于计算效率考虑，我们聚焦于三角形。

10 图不变量

ISP-WL 支持满足置换等变性的任何图不变量。表7总结了我们实验中使用的不变量及其计算复杂度。

11 证明

本节将对论文正文中提出的所有理论结果给出详细证明。

11.1 方法论部分

命题3.2（置换不变性）。 $\delta(v,\ u,\ w) = \delta(v,\ w,\ u)$ 对于所有 $u,\ w \in N(v)$。

证明。根据定义，$\delta(v, u, w) = (\delta_1, \delta_2, \delta_3)$，其中：

$$\delta_1 = \max(\phi(v) - \phi(u), \phi(v) - \phi(w))$$
$$\delta_2 = \min(\phi(v) - \phi(u), \phi(v) - \phi(w))$$
$$\delta_3 = |\phi(u) - \phi(w)|$$

表7：实验中使用的图不变量及其在有界退化度稀疏图上的计算复杂度。

不变量	复杂度	描述
度	$O(\|V\|)$	邻居数
K-核 [41]	$O(\|V\| + \|E\|)$	最大核归属
洋葱分解 [23]	$O(\|V\| + \|E\|)$	分层 k 核层
聚类系数 [36]	$O(d\|E\|)$	局部三角形密度比
平均邻域聚类 [63]	$O(d^2\|E\|)$	邻居的平均聚类系数
K-桁架 [65]	$O(\|E\|^{1.5})$	三角形支持分解
PageRank [30]	$O(k(\|V\| + \|E\|))$	平稳分布（k 次迭代）
特征向量 [10]	$O(k(\|V\| + \|E\|))$	主特征向量（k 次迭代）
介数 [5]	$O(\|V\|\|E\|)$	最短路径中心性

由于最大值和最小值是对称函数，且绝对值也是对称的，因此：

$$\begin{aligned}\delta_1(v,u,w) &= \max(\phi(v)-\phi(u),\phi(v)-\phi(w)) \\ &= \max(\phi(v)-\phi(w),\phi(v)-\phi(u)) \\ &= \delta_1(v,w,u) \\ \delta_2(v,u,w) &= \min(\phi(v)-\phi(u),\phi(v)-\phi(w)) \\ &= \min(\phi(v)-\phi(w),\phi(v)-\phi(u)) \\ &= \delta_2(v,w,u) \\ \delta_3(v,u,w) &= |\phi(u)-\phi(w)| \\ &= |\phi(w)-\phi(u)| = \delta_3(v,w,u)\end{aligned}$$

因此，$\delta(v,\ u,\ w) = \delta(v,\ w,\ u)$。 □

命题 3.3 (单次赋值性质**)。** 对于任意节点 $v \in V$，c_{ISP}（v） 在迭代 $\ell^* = \phi$（v） 中恰好被赋值一次，此后保持不变。

证明。在迭代 ℓ $V_\ell = v \in$ 中，算法按照 {: $\phi(\,)$ } 处理节点，其中 $v = \ell$ $\ell \in$ {, , , ⋯}。L 对于不变值为 $\phi(v) \in \{1, \ldots, L\}$ 的节点，当且仅当 $v \in V_\ell$ 时：

$$\phi(v) = \ell$$

由于 $\phi(v)$ 是固定的，且恰好取自 $\{1,\ldots,L\}$ 中的一个值，因此恰好存在一次迭代 $\ell^* = \phi(v)$，使得 $v \in V_{\ell^*}$。

在迭代 ℓ^*时，赋予 ISP 颜色：

$$c_{\text{ISP}}(v) \leftarrow \begin{cases}\text{Hash}(\phi(v), M_v^{\text{struct}}) & \text{if } M_v^{\text{struct}} \neq \{\!\{\}\!\} \\ \phi(v) & \text{otherwise}\end{cases}$$

对于所有 $\ell \neq \ell^*$，都有 $\phi(v) \neq \ell$，因此 $v \notin V_\ell$，且 $c_{\text{ISP}}(v)$ 不会被修改。□

命题 3.5 （结构一致性）。 如果每个基础不变量 ϕ_i 都是图不变量，那么 ϕ_{learn} 也是图不变量。

证明。对于任意自同构 $\pi: V \to V$，它将结构等价的节点映射到彼此：

$$\begin{aligned}\phi_{\text{learn}}(\pi(v)) &= L \cdot \sigma(\text{MLP}_\phi([\phi_1(\pi(v)),\ldots,\phi_m(\pi(v))])) \\ &= L \cdot \sigma(\text{MLP}_\phi([\phi_1(v),\ldots,\phi_m(v)])) \\ &= \phi_{\text{learn}}(v)\end{aligned}$$

其中第二个等式成立，是因为每个 ϕ_i 都是图不变量。由于向下取整运算保持相等，$\phi_{\text{discrete}}(\phi_{\text{discrete}}\ v$ □

( )) () 也成立。

11.2 表达能力

定理 (超越 1-WL 的增强表达能力）。*ISP-WL* 的表达能力严格强于 *1-WL*。对于任意图 G_1、G_2：若 *1-WL* 能区分 G_1 与 G_2，则 *ISP-WL* 也能区分它们。此外，存在非同构图 G_1 与 G_2，*1-WL* 无法区分，但 *ISP-WL* 可以区分。

证明。首先，我们证明：若 1-WL 能区分任意图对 G_1 与 G_2，则 ISP-WL 也能区分它们。

回顾可知，ISP-WL 维护一条 WL 流，并在每次迭代中执行标准 1-WL 细化。具体而言，对于每个节点v，其 WL 颜色按如下方式更新：

$$c'_{WL}(v) \leftarrow \text{Hash}((c_{WL}(v), c_{ISP}(v)), m_v)$$

其中 $m_v = \{(c_{WL}(u)\ c_{ISP}(u)) : u \in N(v)\}$ 是邻居颜色的多重集。当 ISP流被禁用时（即对所有 v 有 $c_{ISP}(v) = \perp$），WL细化退化为标准 1-WL。由于 ISP-WL 包含 WL 流，且仅通过 ISP流增加额外信息，若 $c_{1\text{-}WL}(G_1) \neq c_{1\text{-}WL}(G_2)$，则 ISP-WL 中的 WL 流将在某次迭代 k 区分它们。最终的 ISP-WL 着色为 $c(v)$ =哈希$(c_{WL}(v)\ c_{ISP}(v))$，其中纳入了 $c_{WL}(v)$。因此，若 WL 颜色不同，则最终的 ISP-WL 颜色必定不同。由此，ISP-WL 的表达能力至少与 1-WL 相当。

要证明表达能力严格更强，只需展示至少一对 1-WL 无法区分、但 ISP-WL 可以区分的非同构图对。图 2 给出了三个这样的例子。因此，ISP-WL 的表达能力严格强于 1-WL。

□

定理 4.2（节点可区分性）。对于任意节点 u, v：如果$c_{1\text{-}WL}(u) \neq c_{1\text{-}WL}(v)$，那么 $c_{ISP\text{-}WL}(u) \neq c_{ISP\text{-}WL}(v)$；此外，即使在 $c_{1\text{-}WL}(u) = c_{1\text{-}WL}(v)$ 的情况下，*ISP-WL* 也能区分节点。

证明。我们首先证明，对于任意节点 $u\ v$ $c_{1\text{-WL}}\ u \neq c_{1\text{-WL}}\ v$ $c_{\text{ISP-WL}}\ u \neq c_{\text{ISP-WL}}\ v$,, 若 ( ) ( )，则 ( ) ( )。ISP-WL 维护一个 WL 流，在每次迭代中执行标准 1-WL 细化：

$$c_{\text{WL}}^{(k)}(v) = \text{Hash}\big((c_{\text{WL}}^{(k-1)}(v), c_{\text{ISP}}^{(k-1)}(v)), \{\!\{(c_{\text{WL}}^{(k-1)}(u), c_{\text{ISP}}^{(k-1)}(u)) : u \in N(v)\}\!\}\big)$$

忽略 ISP 颜色时（全部 $c_{\text{ISP}} = \perp$），此更新规则会退化为标准 1-WL 细化。由于 ISP-WL 仅通过 ISP 流添加信息，并使用单射哈希，因此标准 1-WL 作出的任何区分都会保留在 ISP-WL 的 WL 流中。更准确地说，若标准 1-WL 在迭代 k 时区分了节点 u 和 v，即 $c_1^{(k)}$-WL$(u) \neq c^{(k)}$1-WL(v)，那么它们在 ISP-WL 中的 WL 颜色也必然不同：$c^{(k)}\text{WL}(u) \neq c^{(k)}\text{WL}(v)$。最终的 ISP-WL 着色通过 $c_{\text{ISP-WL}}(v) = \text{Hash}(c_{\text{WL}}(v)\ c_{\text{ISP}}(v))$ 合并两个流。根据哈希函数的单射性，如果 WL 组件不同，最终的着色也必然不同。因此，我们得出结论：$c_{1\text{-WL}}(u) \neq c_{1\text{-WL}}(v)$ 蕴含 $c_{\text{ISP-WL}}(u) \neq c_{\text{ISP-WL}}(v)$。

为了证明 ISP-WL 即使在 $c_{1\text{-WL}}(u) = c_{1\text{-WL}}(v)$时也能区分节点，只需提供一个包含这两个节点的图对。考虑图6中的图 G_1 和 G_2，它们具有相同的度序列。令 $u \in G_1$ 和 $v \in G_2$ 表示任意节点。两者的度均为 2，且具有相同的邻域度序列，因此 $c_{1\text{-WL}}(u) = c_{1\text{-WL}}(v)$。

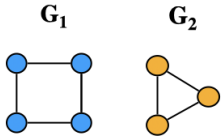


图6：图 G_1 和 G_2 具有相同的度序列，但三角形结构不同。

然而，$|T(u)| = 0 \neq 1 = |T(v)|$。ISP-WL 的计算如下：

$$M_u^{\text{struct}} = \{\!\{(c_{\text{ISP}}(u'), c_{\text{ISP}}(u''), \delta(u,u',u'')) : (u',u'') \in T(u)\}\!\}$$

由于 $|M_u^{\text{struct}}| \neq |M_v^{\text{struct}}|$，根据单射哈希：

$$c_{\text{ISP}}(u) \neq c_{\text{ISP}}(v)$$

因此，$c_{\text{ISP-WL}}(u) \neq c_{\text{ISP-WL}}(v)$，尽管 $c_{1\text{-WL}}(u) = c_{1\text{-WL}}(v)$。 □

定理4.3（依赖不变量的表达能力）。一种着色A区分图 G_1 与 G_2，当它产生不同的颜色多重集时。令 $A \preceq B$ 表示任何被 A 区分的图对也被 B区分，并令 $A \prec B$ 表示严格不等。*ISP-WL*的表达能力由 ϕ决定：

(1) 若 $\phi \preceq$ *1-WL*，则：*1-WL* $\prec$ *ISP-WL*$(\phi) \prec$ *3-WL (2)* 对任意 $k > 1$，若 ϕ与 *k-WL*不可比，则*ISP-WL*(ϕ)与 *k-WL*不可比 *(3)* 若对任意 $k \geq 1$有 *k-WL* $\preceq \phi$，则：*k-WL* $\preceq$ *ISP-WL*(ϕ)

证明。我们首先证明陈述 1。定理 4.1 表明，ISP-WL 的表达能力严格强于 1-WL，因此对于任意 ϕ，都有 1-WL $\prec$ ISP-WL(ϕ)。对于上界，若为 $\phi \preceq$ 1-WL，则可通过 1-WL 着色计算得出。ISP-WL 使用 1-WL 颜色和三角形聚合的信息。由于 3-WL 根据所有三元组 [13], 区分图，而 ISP-WL 处理的是局部三元组（邻域），而非全局三元组，因此其表达能力受 3-WL 限制。

接下来，我们证明结论 2。根据不可比性的定义，ϕ能够区分某些 k-WL无法区分的图对，反之亦然。我们证明 ISP-WL(ϕ)兼具这两种能力。对于第一个方向，

如果 ϕ能够区分图 G_1和 G_2，但 k-WL无法区分，那么它们的 ϕ-值多重集不同：$\{\!\{\phi(u) : u \in V(G_1)\}\!\} \neq \{\!\{\phi(v) : v \in V(G_2)\}\!\}$。由于 ISP-WL 通过分层纳入 ϕ 值，且对于不含三角形的节点，ISP流会为其赋予 $c_{\text{ISP}}(u) = \phi(u)$，因此最终的颜色多重集不同：$\{\!\{c_{\text{ISP-WL}}(u) : u \in V(G_1)\}\!\} \neq \{\!\{c_{\text{ISP-WL}}(v) : v \in V(G_2)\}\!\}$。对于第二个方向，如果 k-WL能够区分图H_1和 H_2，但 ϕ无法区分，那么它们的 ϕ-值多重集相同。ISP-WL 的表达能力受其 WL 流（受 1-WL 限制；对于 $k > 1$，1-WL 弱于 k-WL）和 ISP 流（由 ϕ和三角形聚合得到）的约束。由于 ϕ无法区分H_1和 H_2，分层和间隙编码完全相同；仅靠 WL 流无法弥补k-WL 通过元组级着色捕获的高阶模式。因此，ISP-WL(ϕ)与 k-WL不可比。

对于陈述 3，假设 k-WL $\preceq \phi$，这意味着任何由 k-WL 区分的图也能被 ϕ 区分。我们将证明 ISP-WL(ϕ) 保留 ϕ 所作的所有区分，因而也保留 k-WL 所作的区分。考虑任意一对非同构图 $G_1 \not\cong G_2$，它们可被 k-WL 区分。根据假设，ϕ 也能区分它们：$\{\!\{\phi(u \in V\ G_1 \neq \phi\ v \quad v \in V\ G_2) : ( )\}\!\}$ $\{\!\{ ( ) : ( )\}\!\}$。ISP流通过两种机制直接纳入 ϕ 值：(1) 根据 ϕ 值的排序分层为 $\ell \in \{1, \ldots, L\}$；(2) 对于没有三角形的节点，基础 ISP 颜色满足 $c_{\text{ISP}}(u) = \phi(u)$。对于有三角形的节点，ISP 颜色计算为 $c_{\text{ISP}}(u) = \text{哈希}(\phi(u)\ M_u^{\text{struct}})$，其中仍直接包含 $\phi(u)$。根据哈希函数的单射性，若 $\phi(u) \neq \phi(v)$，则无论结构消息如何，$c_{\text{ISP}}(u) \neq c_{\text{ISP}}(v)$。因此，不同的 ϕ 多重集会产生不同的 ISP 颜色多重集，并在最终的 ISP-WL 着色中得到保留：$\{\!\{ c_{\text{ISP-WL}}\ u \quad u \in V\ G_1 = c_{\text{WL}}\ u\ c_{\text{ISP}}\ u \quad u \in ( ) : ( )\}\!\}$ $V\ G_1 \neq c_{\text{WL}}\ v\ c_{\text{ISP}}\ v \quad v \in V\ G_2$ $\{\!\{\text{Hash}(( ) ( )) : ( )\}\!\}$ $\{\!\{\text{Hash}(( ) ( )) : ( )\}\!\}$。因此，任何一对由 k-WL 区分的图也能被 ϕ 区分，而后者又能被 ISP-WL(ϕ) 区分，由此得到 k-WL $\preceq$ ISP-WL(ϕ)。证明完毕。

□

11.3 收敛与复杂度

定理4.4（收敛）。*ISP-WL*最多在 $K = \max(K_{WL}, L) \leq |V|$ 次迭代内终止，其中 L 是不同不变量值的数量，K_{WL} 是标准 *1-WL* 的收敛迭代次数。

证明。该算法维护两条流：WL 流和ISP流。WL细化在由 $|V|$界定的有限颜色空间上运行。该细化是单调的：不同的颜色保持不同。根据标准 1-WL 的收敛结果，它最多在 $K_{\text{WL}} \leq |V|$次迭代内达到不动点。在ISP流中，由于对所有 $v \in V$均有$\phi(v) \in \{1, 2, \ldots, L\}$，因此恰好存在 L 个不同的不变量层。在迭代 $\ell \in \{1, \ldots, L\}$时，算法处理 $V_\ell = \{v \in V : \phi(v) = \ell\}$ 中的节点。根据单次赋值性质，每个节点在其层 $\ell = \phi(v)$被到达时恰好被处理一次。ISP流在处理完所有 L 层后完成。

因此，该算法在 $K = \max(K_{\text{WL}}, L) \leq |V|$ 时终止，因为 $L \leq |V|$。

□

定理 4.5（时间复杂度）。*ISP-WL* 的时间复杂度为$O(K(|V| + |E|) + T)$，其中 $T = \sum_v |T(v)|$是三角形的总数。对于退化度有界的稀疏图 d，有$T = O(d|E|)$，因此总体复杂度为 $O(K(|V| + |E|))$（当 d为常数时），与标准 *1-WL* 相匹配。

证明。对所有节点计算 $\phi(v)$，对于度这类简单不变量需要 $O(|V|)$，而对于 K-核和洋葱分解这类不变量则需要 $O(|E|$。

对于每个节点 v，收集邻居多重集需要 $O(d_v)$ 时间。对所有节点求和：$\sum_{v\in V} O(d_v) = O(|E|)$。每个节点计算哈希需要 $O(1)$，因此总计需要 $O(|V|)$。因此，每次迭代的 WL 细化需要 $O(|V| + |E|)$，经过 K次迭代后需要 $O(K(|V| + |E|))$。

根据单次赋值性质，每个节点只处理一次。对于节点 v，枚举三角形 $T(v)$ 需要检查所有邻居对：$O(d_v^2)$。计算 $|T(v)|$个三角形的层次化间隙编码并进行哈希，需要 $O(|T(v)|)$。每个节点的总开销：$O(d_v^2 + |T(v)|) = O(d_v^2)$，因为 $|T(v)| \leq \binom{d_v}{2}$。对所有节点求和，三角形枚举的总开销为 $O(T)$，因为每个三角形会被枚举三次（每个节点各一次）。

总复杂度为 $O(|E|\log|V|) + O(K(|V| + |E|)) + O(T) = O(K(|V| + |E|) + T)$，其中预处理的主要成本来自 $K \geq 1$ 的迭代成本。

对于退化度为 d 的图，我们有 $T \leq d \cdot |E|$。当 $d = O(1)$ $O(K(|V| + |E|) + T) = O(K(|V| + |E|) + O(|E|) = O(K(|V| + |E|)$ 时，我们可以得出结论：$(\ (\ )\ )\ (\ (\ )\ )\ (\ (\ ))$. □

11.4 过平滑抵抗

定理 4（特征坍塌抵抗性）。令 $\bar{h}_v^{ISP} := h_v^{(\phi(v)),ISP}$ 表示第 $\phi(v)$ 层固定的结构嵌入。对于结构嵌入各不相同的节点 $\bar{h}_u^{ISP} \neq \bar{h}_v^{ISP}$，最终的组合表示满足：

$$[h_u^{(K),WL} \| \bar{h}_u^{ISP}] \neq [h_v^{(K),WL} \| \bar{h}_v^{ISP}] \quad \forall K$$

即使 WL特征 发生过度平滑。若多层感知机$_{out}$ 为单射，最终节点输出在所有深度上仍保持互异。

证明。持久性 $h_v^{(k),ISP} = \bar{h}_v^{ISP}$ 对于 $k \geq \phi(v)$ 由门控机制得出：

$$\gamma_v^{(k)} = \mathbb{I}[\phi(v) = k] \cdot \mathbb{I}[\|h_v^{(k-1),ISP}\|_2 = 0]$$

对于 $k = \phi(v)$，两个指示函数均为真，因此 $\gamma_v^{(k)} = 1$ 且 $h_v^{(k),ISP}$ 被赋值。对于 $k > \phi(v)$，第一个指示函数为假，从而得到 $\gamma_v^{(k)} = 0$ 且 $h_v^{(k),ISP} = h_v^{(k-1),ISP} = \bar{h}_v^{ISP}$，依据归纳法。

由于 $\bar{h}_u^{ISP}$与 $\bar{h}_v^{ISP}$与深度无关且彼此不同，拼接后的输入互不相同：

$$[h_u^{(K),WL} \| \bar{h}_u^{ISP}] - [h_v^{(K),WL} \| \bar{h}_v^{ISP}] = [h_u^{(K),WL} - h_v^{(K),WL} \| \bar{h}_u^{ISP} - \bar{h}_v^{ISP}]$$

ISP 分量 $\bar{h}_u^{ISP} - \bar{h}_v^{ISP} \neq 0$ 确保无论 WL 流行为如何，输入都保持互异。若多层感知机$_{out}$ 为单射，不同输入产生不同输出。 □

深度不变的结构信息 $\bar{h}_v^{ISP}$充当锚点，防止深度网络中的特征坍缩，同时 WL 流继续细化高阶模式。

12 实验设计

12.1 下游任务

我们在三项基本图学习任务上评估 ISP-GNN，这些任务从不同方面考察模型的表达能力和泛化能力：

12.1.1 节点分类。 该任务旨在利用节点特征和图结构预测各个节点的标签。

12.1.2 影响力估计。 该任务被表述为节点级回归，旨在预测信息扩散模型（独立级联、线性阈值和 SIS）下每个节点的激活概率。

12.1.3 图分类。目标是根据整个图的结构与节点特征预测单一标签。

12.2 数据集统计

我们在 26 个基准数据集上对 ISP-GNN 进行了全面评估，涵盖三种图学习任务：节点分类、影响力估计和图分类。

12.2.1 节点分类。我们采用了表 8 所示的 10 个基准数据集，这些数据集涵盖同配和异配网络结构。同配数据集包括 Cora ML、Citeseer、Pubmed、DBLP 和 Amazon Photo，其同配性较强，连接的节点通常具有相似的标签。相比之下，异配数据集包括康奈尔、威斯康星、Texas、Film 和 Amazon-Ratings，其具有异配混合特征，相邻节点往往属于不同类别。

12.2.2 影响力估计。我们使用了表 9 所列的 4 个真实世界社交网络，它们呈现出多样的拓扑特性，从小世界特征到无标度特征不等。这些网络包括爵士乐、Cora-ML、网络科学和电网。

12.2.3 图分类。我们评估了来自 TU 基准和开放图基准的 12 个数据集，涵盖不同领域和规模。基准数据集包括：(1) 分子图（MUTAG、PTC-MR、BZR、DHFR、COX2、ogbg-moltox21、ogbg-molpcba），其中节点为原子；(2) 生物信息学（PROTEINS、D&D），包含蛋白质相互作用网络；以及 (3) 社交网络（IMDB-BINARY、IMDB-MULTI、Reddit-Binary），用于研究协作关系和社群。

表 8：节点分类数据集统计。

数据集	# 节点	# 边	# 特征	# 类别
Cora ML	2,995	8,416	2,879	7
Citeseer	3,327	4,732	3,703	6
Pubmed	19,717	44,338	500	3
DBLP	17,716	105,734	1,639	4
Film	7,600	33,544	932	5
Amazon Photo	7,650	119,081	745	8
Amazon-Ratings	24,492	93,050	300	5
康奈尔	183	295	1,703	5
威斯康星	251	499	1,703	5
Texas	183	309	1,703	5

表9：影响力估计的数据集统计。

数据集	# 节点	# 边
Jazz	198	2,742
Cora-ML	2,810	7,981
网络科学	1,565	13,532
电网	4,941	6,594

表10：图分类的数据集统计。

数据集	领域	图数	平均节点数	平均边数	类别数
MUTAG	分子	188	17.93	19.79	2
PTC_MR	分子	344	14.29	14.69	2
BZR	分子	405	35.75	38.36	2
DHFR	分子	467	42.43	44.54	2
COX2	分子	467	41.22	43.45	2
PROTEINS	生物信息学	1,113	39.06	72.82	2
D&D	生物信息学	1,178	284.32	715.66	2
IMDB-BINARY	社交	1,000	19.77	96.53	2
IMDB-MULTI	社交	1,500	13.00	65.94	3
Reddit-Binary	社交	2,000	429.63	497.75	2
ogbg-moltox21	分子	7,831	18.51	19.26	12
ogbg-molpcba	分子	437,929	26.0	28.1	128

表 11：采用不同结构不变量的节点分类和图分类性能。数值表示准确率（%）及标准差。最佳结果以粗体标出。

方法	节点分类		图分类	
	Amazon Photo	Amazon-Ratings	ogbg-moltox21	Reddit-Binary
基线	89.30 ± 0.82	37.99 ± 0.61	74.9 ± 0.5	92.2 ± 2.3
度	93.12 ± 0.46	47.5 ± 1.03	75.1 ± 0.6	92.8 ± 2.2
K-核	93.97 ± 0.46	**48.43 ± 1.21**	75.1 ± 0.8	92.5 ± 2.1
洋葱分解	93.93 ± 0.35	44.95 ± 0.54	**75.6 ± 0.7**	92.4 ± 2.7
聚类系数	93.23 ± 0.31	46.27 ± 0.78	74.8 ± 0.6	93.0 ± 2.7
平均邻域聚类	93.11 ± 0.67	46.42 ± 1.00	74.9 ± 1.0	92.1 ± 3.0
K-桁架	**94.01 ± 0.65**	47.83 ± 1.59	75.0 ± 0.5	**93.4 ± 2.6**

表 13：软件环境规格

软件包	版本
Python	3.11.5
PyTorch	2.3.1
PyTorch Geometric	2.7.0
torch-cluster	1.6.3
torch-scatter	2.0.9
torch-sparse	0.6.18

12.3 模型超参数

为确保模型在不同数据集上达到最佳性能，我们进行了超参数搜索。所有模型均使用 Adam优化器 [31]进行训练。对于OGB数据集，我们采用学习率调度器，每 10 个训练轮次将学习率乘以衰减因子 0.5。表 12 总结了实验中使用的完整超参数搜索空间。

表 12：ISP-GNN 变体的超参数搜索空间

超参数	搜索空间
层数	{2,3,4,5, 6}
温度 (β)	{0.5, 1.0, 1.5, 2.0}
分层数（S）	{4, 6, 8}
丢弃率	{0。0，0。1，0。2，0。5，0。6，0。9}
学习率	{0.005, 0.009, 0.01, 0.1}
批次大小	{32, 64, 100}
隐藏维度	{32，64，128，256，300}
权重衰减	$\{5\times10^{-4}, 9\times10^{-3}, 1\times10^{-2}, 1\times10^{-1}\}$
训练轮数	{100, 200, 1000}

12.4 计算资源

所有实验均在一台配备 Intel Xeon W-2175 处理器（2.50GHz，28 核）、NVIDIA RTX A6000 GPU 和 512GB 内存的 Linux 服务器上完成。软件环境规格列于表 13。

13 补充实验

13.1 不同不变量的性能比较

表 11 展示了 ISP-GNN 在使用不同结构特征时，在节点分类和图分类任务上的分类性能。图 7 展示了各数据集对应的结构特征。

结果表明，ISP-WL 通过融入强结构偏置，显著提升了基线性能，从而大幅提高节点分类性能。对于 Amazon-Ratings，该图呈现异配特征，即相连节点的标签并不相似。基于核心/洋葱分解的不变量性能更优，因为它们能够捕捉结构位置和层级角色，而基于聚类的特征则无法

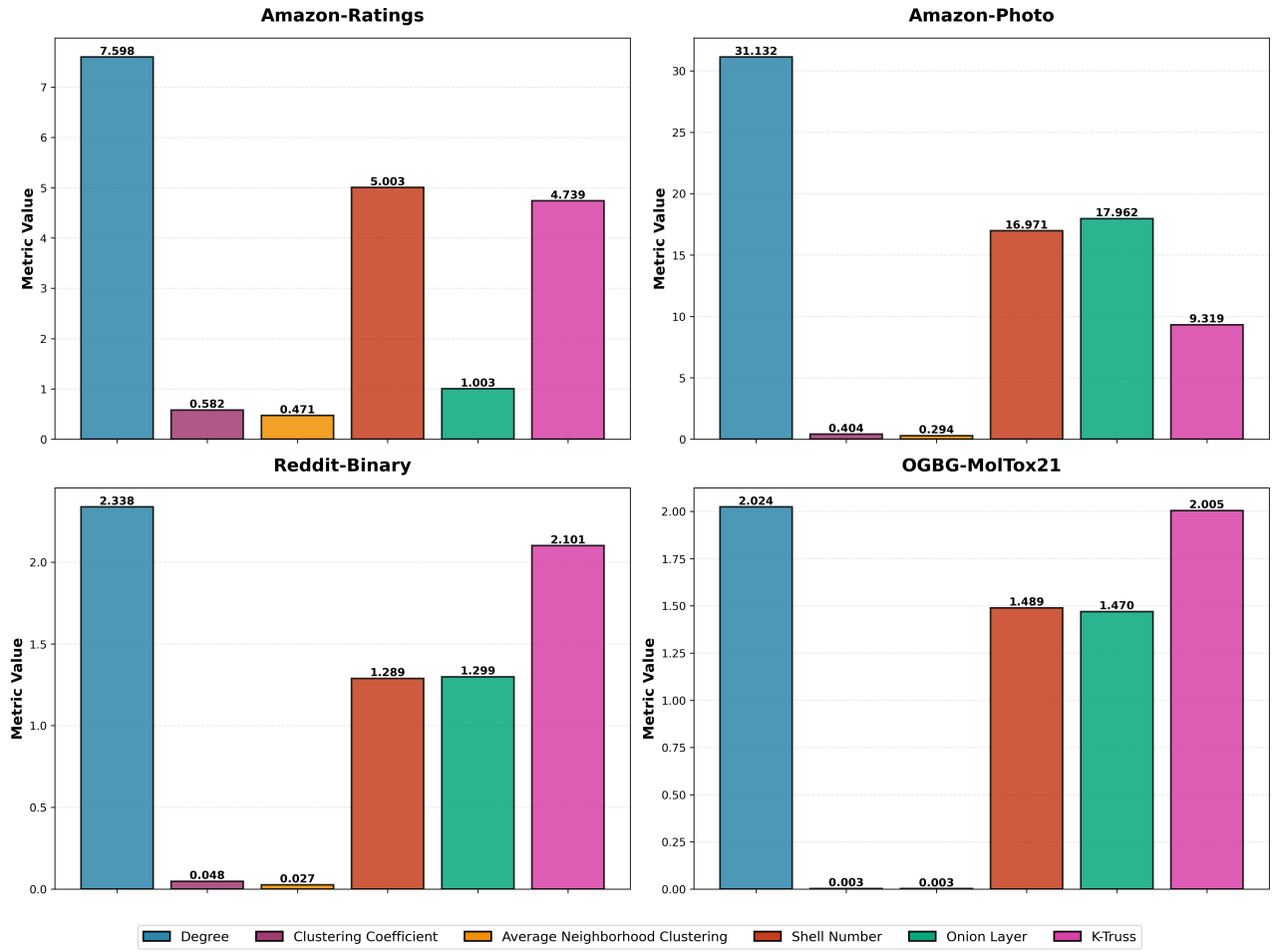


图 7：不同数据集的结构指标比较。

区分这些角色。Amazon-Photo 具有稠密结构和深层次的层次化组织，同样受益于结构不变量，其中 K-桁架和 k-核带来了显著的性能提升。在图分类任务中，分子基准 ogbg-moltox21 在所有不变量上的提升都较小，因为分子性质取决于特定领域的化学语义，而非图论拓扑。作为一个具有稀疏三角结构的社交网络基准，Reddit-Binary 从 K-桁架中获益最大，因为它捕捉的是层次化社区模式和凝聚子结构，而非局部三角模体。总体而言，这些发现表明，ISP-GNN 通过选择与目标图的底层标签同质性、结构特征和领域语义相契合的不变量，实现了最佳性能。

13.2 层次化结构间隙编码

在本小节中，我们通过实验验证与层次化结构间隙编码机制相关的设计选择。

表 14：模体类型的消融研究。结果以图分类的平均准确率 ± 标准差（%）表示。最佳结果以粗体标出。请注意，度用作不变量。

模体类型	MUTAG	PTC MR_	IMDB-MULTI
GIN（基线）	92.8 ± 5.9	65.6 ± 6.5	54.6 ± 3.0
4-团	95.2 ± 4.4	72.7 ± 5.6	55.2 ± 2.8
4-环	94.7 ± 4.0	71.0 ± 7.5	55.0 ± 3.2
菱形	94.6 ± 4.1	73.4 ± 7.7	55.4 ± 2.9
5-团	95.3 ± 5.5	72.2 ± 7.7	55.3 ± 3.0
6-环	94.7 ± 5.0	74.3 ± 6.6	54.9 ± 3.1
三角形	**97.9 ± 2.6**	**75.8 ± 5.9**	**55.9 ± 3.1**

13.2.1 模体类型的影响。表 14 通过对五种模体类型进行消融研究，验证了我们选择三角形的合理性。在所有数据集上，基于三角形的聚合效果始终优于高阶替代方案（4-团、5-团、4-环、菱形、6-环）。这种优势源于三个相辅相成的因素。首先是模体丰度：三角形比其他

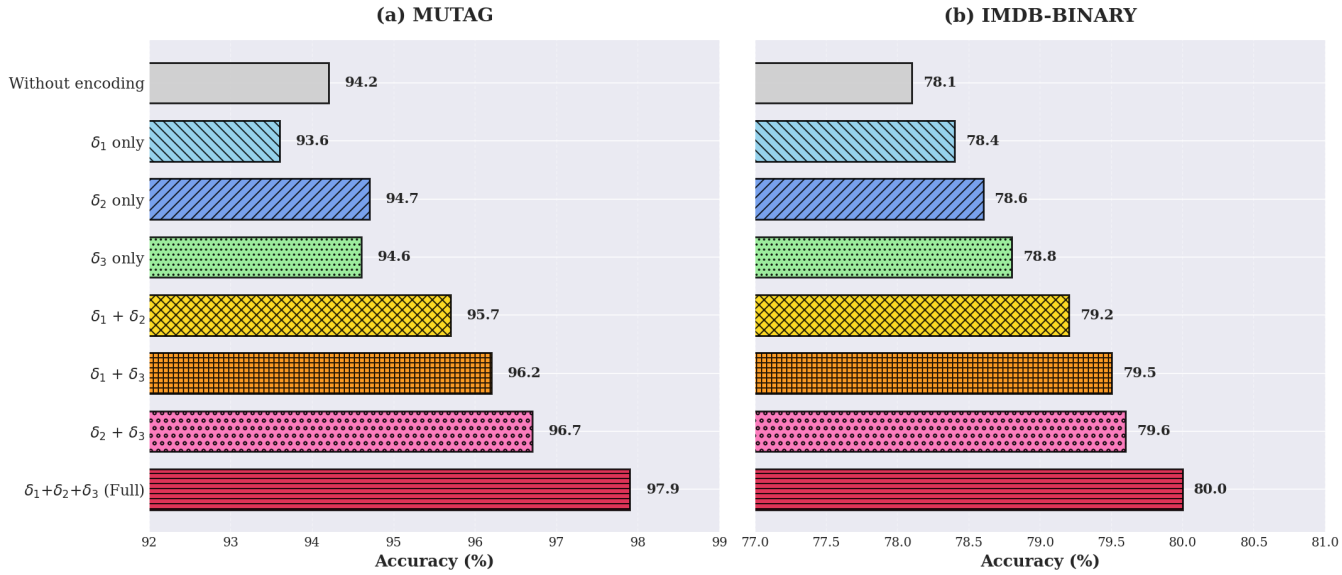
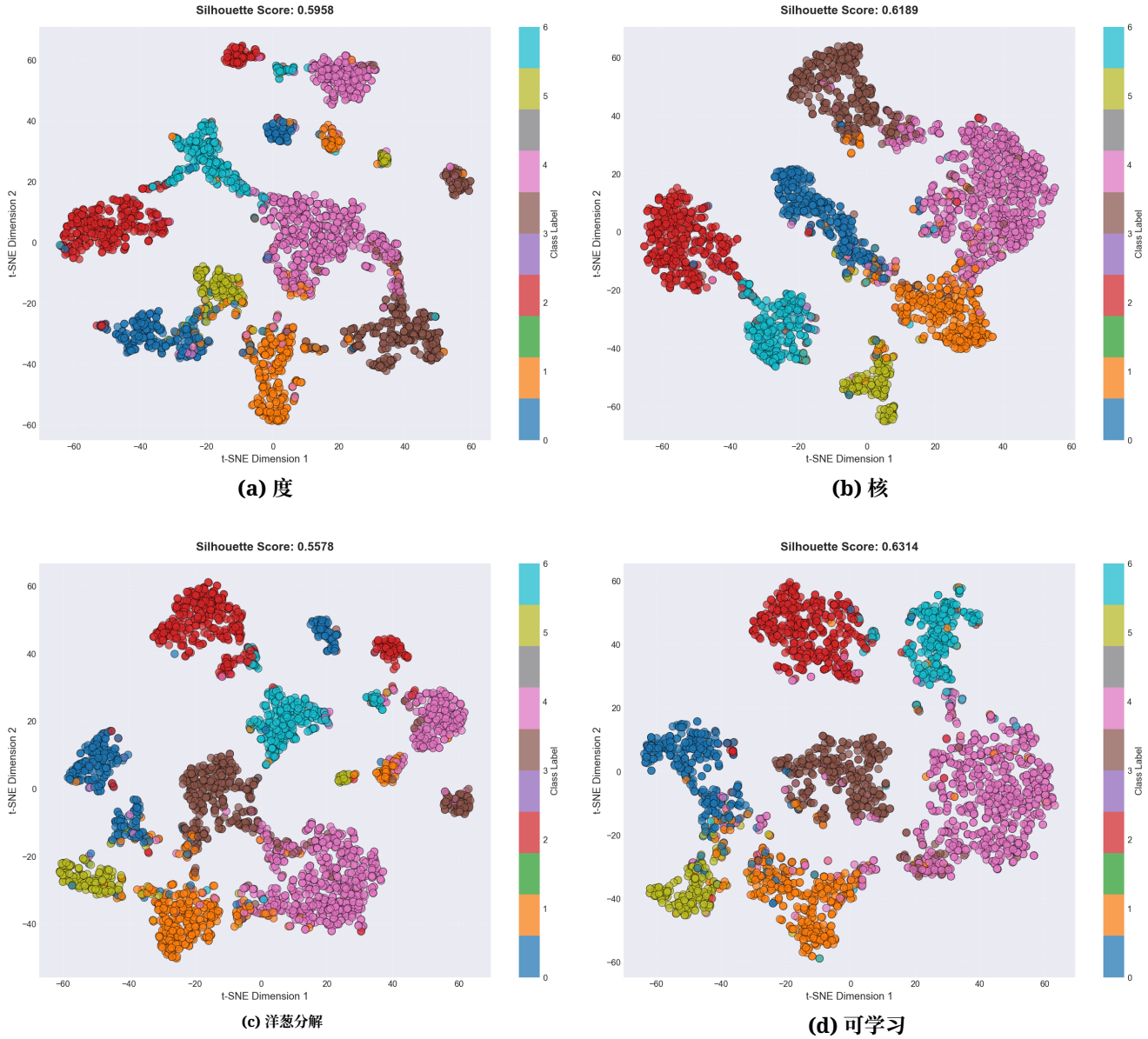


图10：图分类中层次化间隙编码组件的消融研究。结果表明，δ_1、δ_2 和 δ_3 具有互补性。从单个组件到两两组合，再到完整编码，性能逐步提升，验证了这三种间隙特征的必要性。

高阶模体更为常见，因而能够在整个图中提供稳健的学习信号。其次是信息完整性：三角形是高阶结构的基本构成单元，因为较大的团可以分解为相互重叠的三角形，因此，基于三角形的聚合无需显式枚举罕见模体，即可捕获关键结构信息。第三是参数更少：三角形所需参数更少，却能提供数量显著更多的训练实例，从而降低过拟合风险。总而言之，对于 ISP 框架而言，三角形在表达能力、覆盖度和效率之间实现了最佳平衡。

13.2.2 层次化间隙编码中不同组件的影响。为验证层次化间隙编码中各组件的作用，我们通过系统地移除单个间隙特征及其组合开展消融研究。图10展示了MUTAG和IMDB-BINARY数据集上的结果。

结果揭示了三个关键发现。首先，与不进行编码的基线相比，所有单独的间隙组件都能提升性能；在两个数据集上，δ_2 和 δ_3 的贡献均大于 δ_1，这表明最小间隙和邻居同质性比最大间隙单独具有更强的判别能力。其次，两两组合始终优于单个组件，表明间隙特征之间具有互补性。第三，完整编码 $\delta_1+\delta_2+\delta_3$ 达到了最佳性能，证实结合全部三个组件能够提供独特且协同的结构信息。



(a) 度 **(b) 核**

(c) 洋葱分解 **(d) 可学习**

图11：Cora-ML 上 ISP-GNN 嵌入的 t-SNE 可视化。可学习不变式形成的簇最紧凑、分离度最高（轮廓系数：0.631），优于预定义不变式（轮廓系数：0.558–0.619）。这表明，可学习分层能够自动发现与语义类别相契合、适应任务的结构模式。

13.3 可学习不变量分析

本节将分析 ISP-GNN† 变体的能力及其对超参数的敏感性。

13.3.1 任务自适应结构发现。 为验证 ISP-GNN 的可学习不变式能否捕捉与任务相关的结构模式，我们在 Cora-ML数据集上针对节点分类任务，通过 t-SNE投影 [40]和轮廓系数 [52] 分析嵌入质量。图11显示，采用可学习不变式的 ISP-GNN 能形成最紧凑、分离度最高的簇，且轮廓系数最高，优于所有预定义变体。

固定不变量只能捕捉特定的结构属性（度：局部连通性；核：凝聚性；洋葱分解：层级），而可学习变体会调整其分层方式，使之与真实标签中的语义类别相契合，从而为每项任务学习结构信号的最优组合。

13.3.2 超参数敏感性分析。我们分析了 ISP-GNN† 对两个关键超参数的敏感性：分层数 S 和软分配温度 β。图12展示了图分类任务上的结果。

ISP-GNN† 在 $S \in [6, 8]$ 和 $\beta \in [0.5, 2.0]$ 下表现稳健，为超参数选择提供了实用指导。虽然最优值依赖于数据集，但在这些范围内，性能保持稳定，波动较小。值得注意的是，$S \in [6, 8]$ 表明，较少的层数就足以获得最佳性能，这有利于提高计算效率并降低内存占用。

13.4 可扩展性分析

我们通过在节点数从 10K 到 1M、平均度为 10 的合成 Barabási-Albert（BA）图上进行实验，实证验证了 ISP-WL 的理论复杂度界（定理 4.4 和 4.5）。BA 图是无标度网络，能够代表包括社交网络、引文网络和生物网络在内的真实世界拓扑结构 [4]。这些图具有有界退化度，这一性质在真实世界的稀疏图中十分常见 [15]。

在本实验中，我们测量了六种图规模（10K、50K、100K、250K、500K、1M 个节点）下 ISP-WL变体（度、K-核、洋葱分解、K-桁架）的每次迭代时间。所有实验均采用 5 次 WL 次迭代，并将预处理开销分摊到各次迭代中，以便单独测量每次迭代开销。

13.4.1 定理*4.5*的验证。图13(a)展示了与 $O(|V| + |E|)$ 复杂度相符的线性单次迭代规模伸缩：度、K-核和洋葱分解变体在三个数量级范围内均保持近线性增长。K-桁架由于其 $O(|E|^{1.5})$ 预处理而呈现略微超线性的规模伸缩，但即使在拥有 1M 个节点时仍然切实可行。

图13(b)中的开销比揭示了一种重要的规模伸缩行为：在中等规模图（50K–100K 个节点）上，开销最初会激增，这是因为预处理开销超过了较小的基线单次迭代成本；但随后，随着预处理开销得到更充分的摊销，这一开销会下降，并在大规模（500K–1M 个节点）时趋于稳定。这验证了 ISP-WL 的单次迭代复杂度在渐近意义上与标准 1-WL 相匹配。

13.4.2 预处理开销。图14(a)单独展示了预处理复杂度：度计算几乎瞬间完成，K-核和洋葱分解随边数线性增长，而 K-桁架对每条边进行三角形枚举，因此代价呈超线性增长。不过，按迭代次数摊销后（图14(b)），即使是 K-桁架的预处理开销也变得可以承受。

13.4.3 三角形枚举。 图15验证了稀疏图的复杂度 $T = O(d|E|)$ 假设。三角形计数呈次线性增长（面板a），枚举时间呈线性扩展（面板b），面板（c）展示了高度数下的二次爆炸，这使得有界退化度对于实际效率至关重要。

13.4.4 定理*4.4*的验证。 在所有图规模下，所有ISP-WL变体均在5次迭代内收敛，远低于 $K \leq |V|$ 上界。规模扩展下稳定的开销比证实，收敛时间仍与图规模无关。

总体而言，这些结果证实，ISP-WL在稀疏无标度图上保持了高效的复杂度，其开销稳定地取决于不变量的选择。度的开销最低，K-核/洋葱分解在效率与表达能力之间取得平衡，而K-桁架（5.5×）在预处理开销分摊到多次迭代后，以适中的成本提供最大的结构可区分性。这些结果可推广到具有类似有界退化度性质的真实世界稀疏网络。

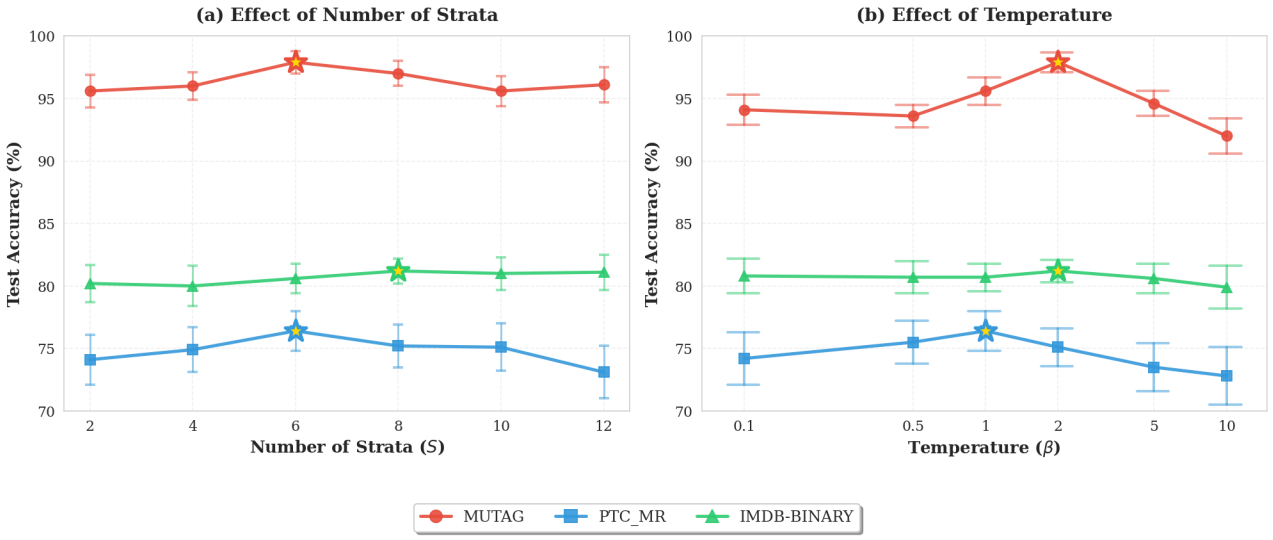


图12：超参数敏感性分析。(a) 分层数的影响 S (b) 温度的影响 β。

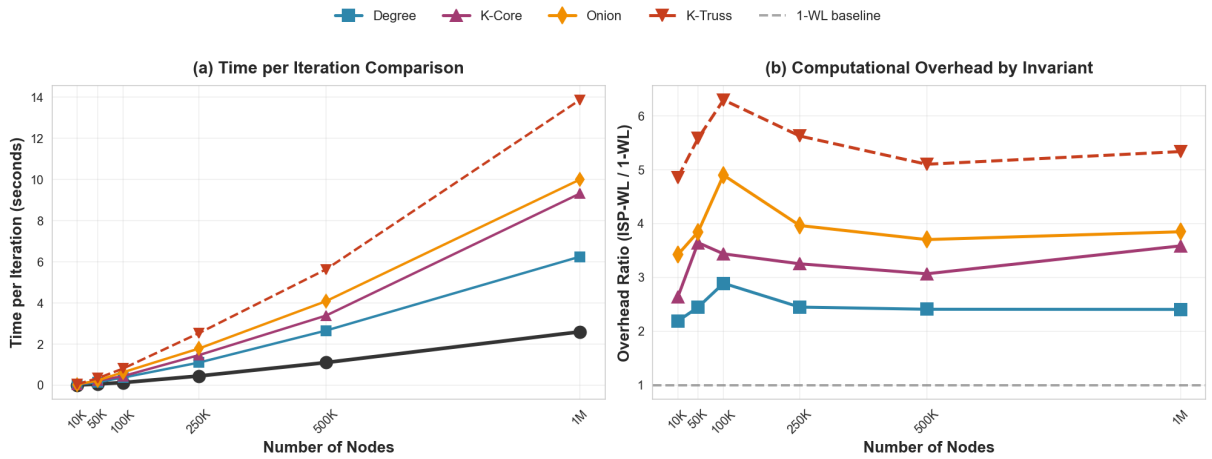


图13：ISP-WL 在合成 BA 图上的计算开销。(a) 所有变体的单次迭代时间均随图规模线性增长。(b) 开销比在大规模图上趋于稳定，预处理开销也随着规模扩大而得到更充分的摊销。

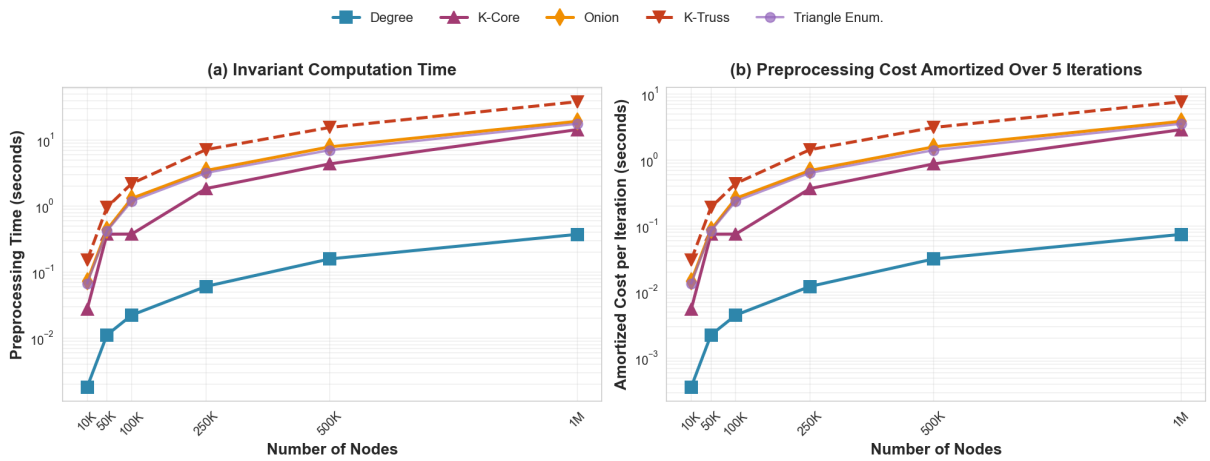


图14：预处理成本分析（对数刻度）。(a) 不变量计算时间显示，度（$O(|V|)$）几乎瞬时完成，K-核/洋葱分解（$O(|E|)$）仍然高效，而K-桁架（$O(|E|^{1.5})$）开销较大，但在规模增大时摊销效果良好。(b) 每次迭代的摊销成本显示，随着图规模增大，预处理的贡献相对于迭代成本逐渐降低。

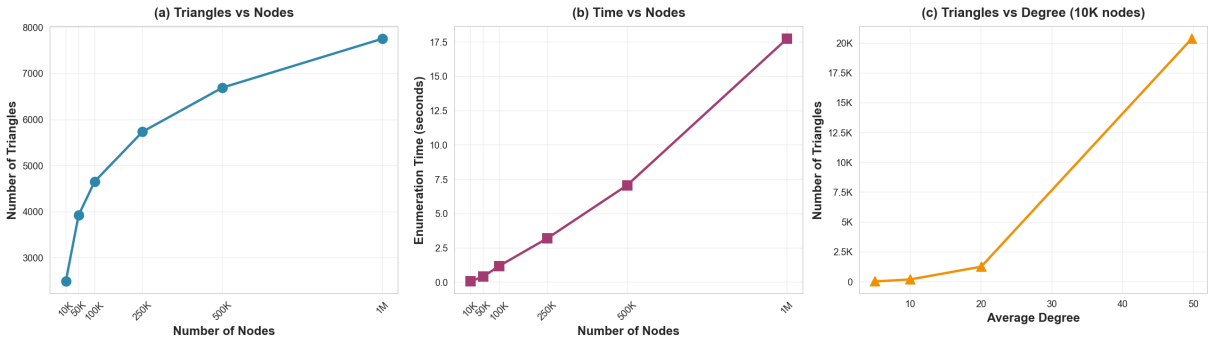


图15：BA图上的三角形枚举可扩展性。(a) 三角形数量的次线性增长证实了现实世界网络通常具有有界退化度。(b) 线性枚举时间验证了复杂度 $T = O(d|E|)$。(c) 高节点度下的二次爆炸表明有界退化度为何至关重要。

13.5 可视化分析

本节展示 ISP-WL 的多种可视化结果。我们采用《悲惨世界》共现网络，该网络由 77 个节点和 254 条边组成。

13.5.1 跨不变量着色细化。 图16展示了 ISP-WL 在七种图不变量上的判别能力。标准 1-WL 可实现 52 种不同的着色，而所有 ISP-WL 变体的表现均优于这一基线，提升幅度因所选不变量而异，从中等到显著不等。这种提升增强了节点的可区分性，验证了定理 4.2。强大的不变量（K-桁架、聚类系数）实现了最强的细化效果，表明与 ISP-WL 理论洞见相契合的度量（定理4.3）。基于分解的度量（K-核）紧随其后，能够有效捕捉网络的核心-边缘结构。全局中心性度量（PageRank、特征向量）通过识别 1-WL 无法区分的影响力节点，带来中等程度的提升。较简单的局部度量（度、洋葱分解）提升最弱，但仍稳定有效，因为其结构信息与 1-WL 的邻域聚合存在部分重叠。

从视觉上看，ISP-WL变体中颜色多样性的增加揭示了更细粒度的结构差异，尤其是在连接密集的中心区域：此处的字符具有相似的局部邻域，但高阶连接模式各不相同。不同不变量之间的性能差异印证了我们框架的核心原则：表达能力的提升在很大程度上取决于是否选择了能够捕捉与标准消息传递正交的结构属性的不变量。

13.5.2 不变量正交性分析。图17通过比较成对着色差异，量化了不变量之间的结构互补性。大多数不变量对都表现出高度差异，表明它们各自捕捉的结构属性大体不同。这种正交性印证了 ISP-WL 对不变量选择的敏感性：不同的结构视角会产生截然不同的节点区分结果。

一些显著的模式显现出来：K-桁架与聚类系数表现出最低的分歧，揭示了它们对三角形内聚性的共同侧重。度与所有其他不变量之间的分歧最为一致，证实它仅捕捉一阶连通性，而不包含任何高阶信息。以红色为主的矩阵证实，精心选择的不变量提供了正交的结构视角，而 ISP-WL 成功地利用这些视角来提升表达能力。

13.5.3 收敛动态。 图18揭示了各不变量之间截然不同的收敛模式。标准 1-WL 在第 3 次迭代时迅速收敛，确立了 52 种颜色的基线。ISP-WL 变体表现出差异化的收敛速度：K-桁架在第 3 次迭代时即实现早期稳定，共 66 种颜色，而聚类系数、K-核、PageRank 和特征向量则需要更多次迭代才能收敛。

K-桁架的快速收敛源于其粗粒度分层——该图包含的团结构有限，因而能够快速逐层处理。相比之下，聚类和 PageRank 等更细粒度的不变量会形成更丰富的初始分层，并在额外的迭代中传播结构差异。

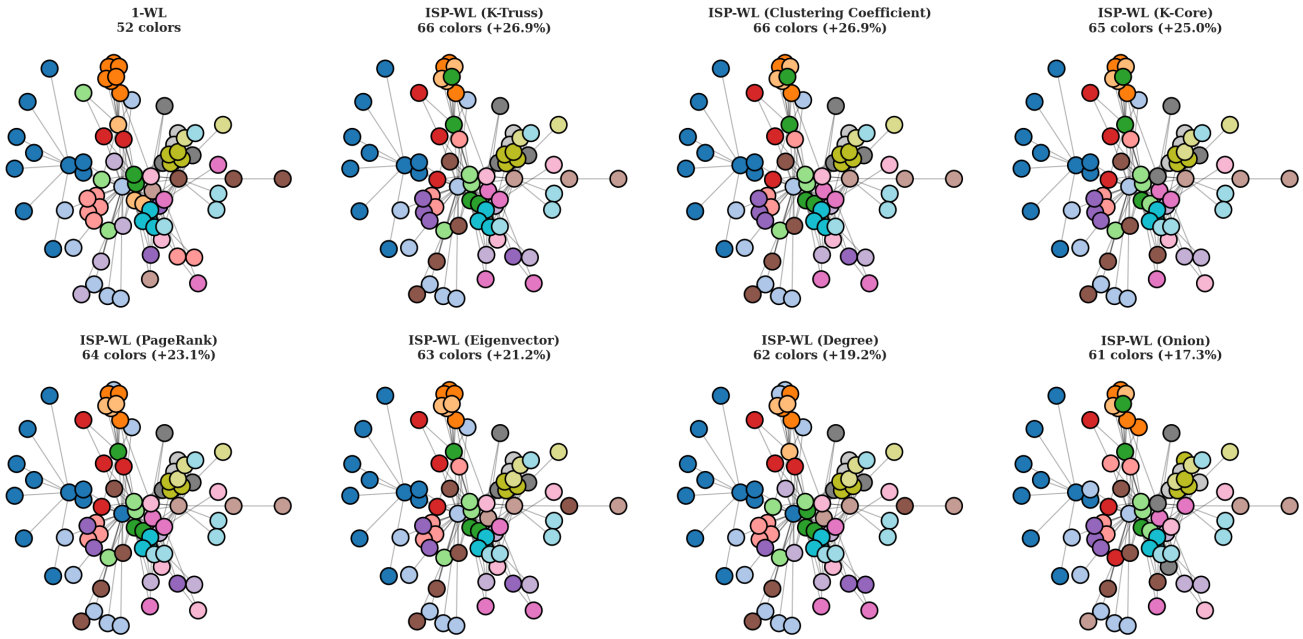


图16：1-WL与ISP-WL的着色细化比较。采用不同不变量的ISP-WL比标准1-WL多获得17-27%的颜色，从而提供更高的判别能力。

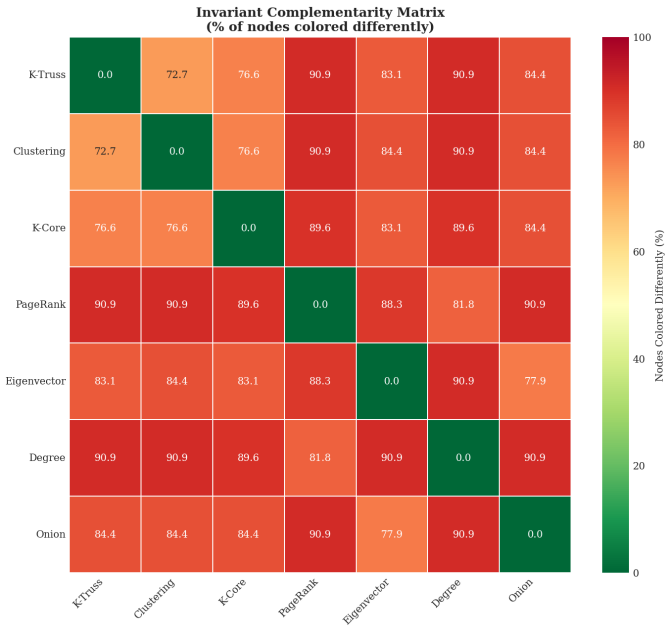


图17：不变量之间的成对互补性，以获得不同最终颜色的节点百分比衡量。高值（红色）表示正交结构信息，验证了不同不变量捕捉到不同的图性质。

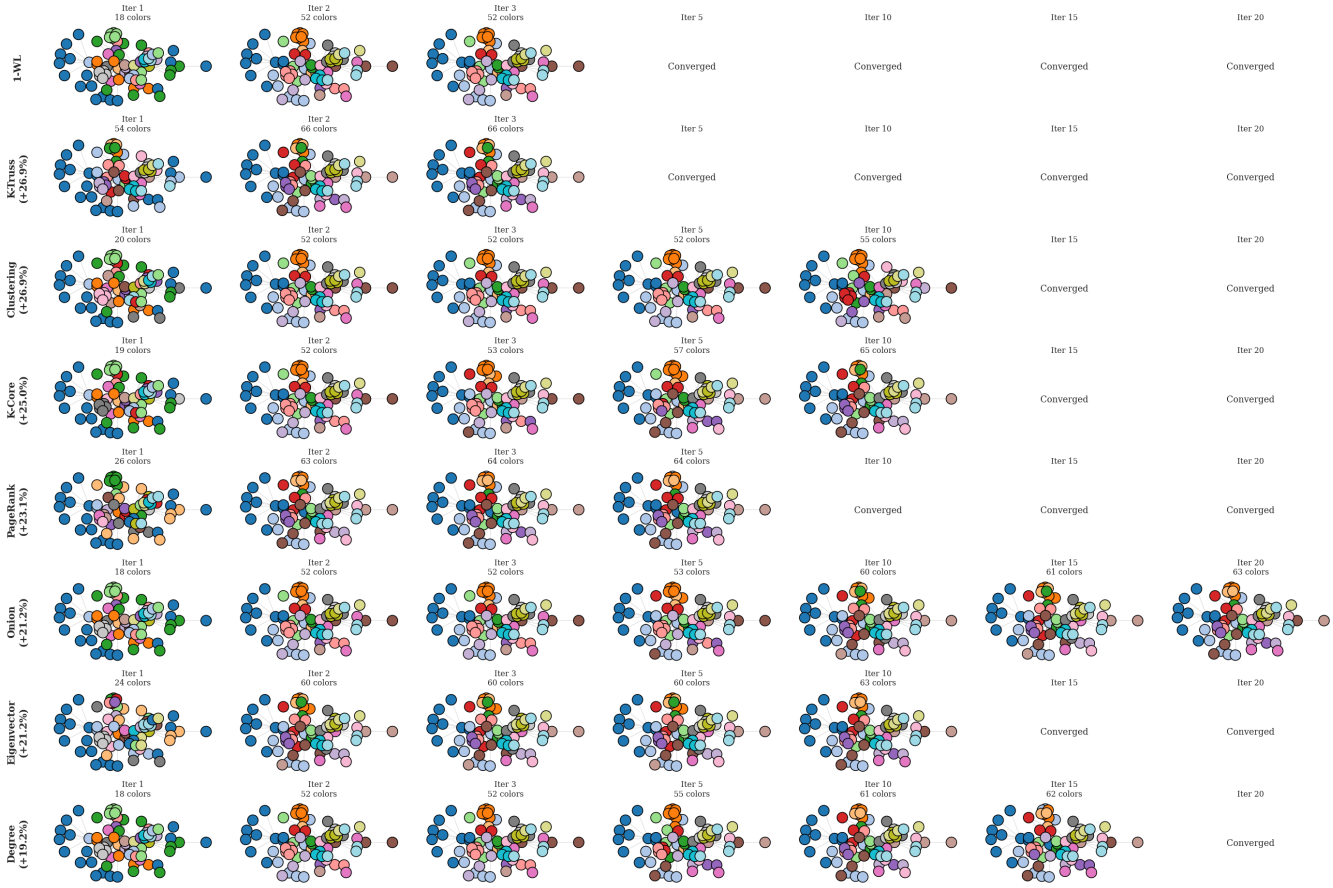


图18：不同变体之间的迭代颜色细化。K-桁架收敛最快，而其他更细粒度的不变量（如度与洋葱层）则需要更多迭代。

14 局限性与未来工作

虽然我们的 ISP 框架提供了一种有原则的方法，通过不变分层传播增强图神经网络的表达能力，但我们的工作仍存在若干局限，也为未来研究提供了机会。

14.1 稠密图的计算复杂度

我们的理论复杂度分析在定理4的 ISP-WL 的计算时 $\mathrm{O}(K(|V|+|E|))$ 复杂度。然而，这依赖于三角形数量为 $T=\mathrm{O}(d|E|)$ 且 d 为常数这一假设。虽然这一假设适用于许多现实世界网络，包括社交网络、分子图和引文网络，但对于稠密图或聚类系数较高的图，可能并不成立。如图15(c)所示，对于度数极高的节点，三角形枚举可能会变得难以承受。未来的研究可以探索近似三角形计数方法或基于采样的方法，在保持 ISP 所捕获的关键结构信息的同时，维持在更稠密图结构上的效率。

14.2 基于三角形的聚合

当前实现主要侧重于基于三角形的高阶聚合，因为这种方法计算效率高，且三角形在现实世界图中十分常见。虽然第9.2节从理论上将其扩展至任意 k-模体，且表 12 对多种模体类型进行了实证评估，但我们的实验表明，在所测试的基准数据集上，三角形始终优于这些替代方案。不过，这种优势未必适用于所有图类型。具体而言，三角形结构有限的图（例如二部网络、树状结构或稀疏生物网络）可能更适合采用其他模体模式。未来的研究应探讨自适应模体选择机制，以便自动识别并利用针对特定图族最具信息量的高阶模式，并将初步的比较分析扩展至更广泛的图结构和应用领域。

14.3 层次化间隙编码设计

定义 3.1 中提出的三分量间隙编码 $\delta(v, u, w)=(\delta_1, \delta_2, \delta_3)$，其设计基于对结构异质性的直觉。尽管消融研究（图10）表明，这三个分量都提供了互补信息，但选择最大值/最小值/邻居间隙并未被形式化地证明为最小或最优。对于特定领域，其他间隙函数或学习得到的间隙编码或许能更有效地捕捉结构异质性。将我们的框架扩展为支持可学习间隙函数，使其能够适应任务特定模式，将增强灵活性，并有望提升性能。

14.4 表达能力刻画

定理34表明 ISP-WL 的表达能力依赖于不变量，但这一刻画仍然有些不完整。虽然我们证明了当 $\phi \preceq 1$-WL 时，1-WL $\prec$ ISP-WL(ϕ) $\prec$ 3-WL，但 ISP-WL 在 Weisfeiler-Leman 层级中的确切位置取决于所选的不变量。更精确地刻画 ISP-WL 相对于高阶 WL 方法能够区分与不能区分哪些图对，将强化其理论基础。未来工作将识别在 $k>1$ 时 ISP-WL 达到或超过 k-WL 的特定图族，并刻画仍超出 ISP-WL 区分能力的结构模式。

14.5 不变量选择

虽然我们的框架支持多种图不变量，并表明不同不变量在不同任务中表现更优（如表 11 所示），但目前不变量的选择仍需人工干预，并依赖于领域知识。在未来工作中，我们旨在开发一种更系统的数据驱动方法，以选择最合适的不变量。

14.6 大规模图的可扩展性

尽管我们的可扩展性实验表明，在合成BA图上可线性扩展至多达100万个节点（图13），但现实世界的图学习越来越多地涉及拥有数千万乃至数亿个节点的图。在这样的规模下，即便是$\mathrm{O}(d|E|)$的三角形枚举，尽管在理论上具有效率，也可能变得具有挑战性。分布式计算方法、近似全局不变量排序的小批量训练策略，或基于采样的 ISP 机制，都可将我们的框架扩展到大规模图。此外，研究近似三角形计数与表达能力之间的权衡，将使实践者能够在计算约束与模型性能之间取得平衡。