软件学报 ISSN 1000-9825, CODEN RUXUEW
2026,37(4):1492−1510 [doi: 10.13328/j.cnki.jos.007523] [CSTR: 32375.14.jos.007523]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于强-弱互信息掩码学习的可解释动态不完整图异常检测*

骆祥峰, 顾峻铨, 余 航

(上海大学 计算机工程与科学学院, 上海 200444)

通信作者: 余航, E-mail: yuhang@shu.edu.cn

摘 要: 大多数图异常检测方法依赖图神经网络 (GNN) 在相对高质量的图数据上进行学习. 然而, 在现实应用中, 这种理想场景极为罕见, 大多数数据存在标签缺失、动态变化和结构不完整等问题, 这些问题统称为动态不完整图. 针对 GNN 在极端条件下性能下降的挑战, 提出一种可解释的动态不完整图异常检测方法 EXDIG (explainable dynamic incomplete graph anomaly detection), 其核心是一种结合强-弱互信息优化的图掩码自编码器框架. 该框架通过对图结构 (节点/边) 和节点特征进行掩码, 模拟现实中的动态不完整场景. 此外, 通过强-弱互信息损失, EXDIG 捕捉结构与特征之间的关系, 同时保持结构完整性, 降低过拟合风险, 并提升泛化能力. 此外, 该方法通过在节点、边及特征上引入掩码扰动, 提高动态不完整图异常检测的可解释性, 使其能够识别关键组成部分, 并为异常检测结果提供透明且可信的解释. 在 9 个真实世界图数据集上进行评估, 实验结果表明, EXDIG 在不同程度的动态不完整场景下, 在多种下游任务和表示学习评估 (包括有监督和无监督设定) 中均优于现有最先进方法. 其中, 在异常检测数据集 Amazon 上, EXDIG 的 NMI 和 ARI 指标分别提升了超过 13% 和 15%; 在动态不完整比率从 25% 到 99% 的设置下, 其 *F*1 分数波动被控制在 5% 以内. 此外, EXDIG 还实现了在动态不完整图中对各节点的可解释性分析.

关键词: 可解释的图学习; 掩码自编码器; 互信息; 图异常检测

中图法分类号: TP18

中文引用格式: 骆祥峰,顾峻铨,余航.基于强-弱互信息掩码学习的可解释动态不完整图异常检测.软件学报,2026,37(4):1492–1510. http://www.jos.org.cn/1000-9825/7523.htm

英文引用格式: Luo XF, Gu JQ, Yu H. Explainable Dynamic Incomplete Graph Anomaly Detection Based on Masked Learning with Strong-weak Mutual Information. Ruan Jian Xue Bao/Journal of Software, 2026, 37(4): 1492–1510 (in Chinese). http://www.jos.org.cn/1000-9825/7523.htm

Explainable Dynamic Incomplete Graph Anomaly Detection Based on Masked Learning with Strong-weak Mutual Information

LUO Xiang-Feng, GU Jun-Quan, YU Hang

(School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China)

Abstract: Most graph anomaly detection methods leverage graph neural network (GNN) to learn from relatively high-quality graph data. Unfortunately, such ideal scenarios are rare in real-world applications, where most data suffer from issues such as missing labels, dynamic changes, and structural incompleteness, collectively referred to as dynamic incomplete graph (DIG). To address the challenge of performance degradation of GNN under extreme conditions, this study proposes an explainable dynamic incomplete graph anomaly detection (EXDIG) method. The core is a graph masked autoencoder framework optimized with strong-weak mutual information. This framework simulates real-world DIG scenarios by masking graph structures (nodes/edges) and node features. In addition, through the

* 基金项目: 国家自然科学基金 (62302287)
本文由“高可信自适应机器学习”专题特约编辑王熙照教授、张敏灵教授、曹付元教授推荐.
收稿时间: 2025-04-16; 修改时间: 2025-06-30, 2025-08-15; 采用时间: 2025-08-20; jos 在线出版时间: 2025-09-02
CNKI 网络首发时间: 2026-01-02

strong-weak mutual information (SWMI) loss, it captures the relationship between structure and features while maintaining structural integrity, reducing overfitting, and improving generalization. Furthermore, EXDIG enhances the interpretability of anomaly detection in DIGs by incorporating masked perturbations on nodes, edges, and features, enabling the identification of key components and providing transparent, trustworthy explanations for anomaly detection results. This study evaluates EXDIG on nine real-world graph datasets, and the results demonstrate its superiority over state-of-the-art methods across different levels of DIG scenarios and across various downstream tasks and representation learning evaluations, both supervised and unsupervised. Specifically, on the Amazon anomaly detection dataset, EXDIG achieves improvements over 13% and 15% in NMI and ARI, respectively. It maintains *F*1-score fluctuations within 5% across dynamic incompleteness ratios from 25% to 99%. Notably, EXDIG enables node-level interpretability in dynamic incomplete graphs.

Key words: explainable graph learning; masked autoencoder; mutual information; graph anomaly detection

图异常检测[1−3]是一种用于识别图中异常节点的技术, 广泛应用于社交媒体垃圾检测、网络安全、欺诈检测、金融风险分析、医疗诊断、推荐系统等多个领域. 因此, 一些研究者将图异常检测转化为节点分类问题, 并利用图神经网络 (graph neural network, GNN)[4−9]来学习节点的有效表示, 通过聚合多跳邻居信息, 捕捉隐藏模式, 来增强对抗攻击的鲁棒性.

然而, 大多数 GNN 方法假设观测到的图数据足够理想, 能够提供充足的信息 (结构、特征、标签) 用于训练[10−15]. 事实上, 在真实世界的异常检测[16−18]任务中, 这一假设通常并不成立, 如图 1 所示. 图结构的动态变化和特征的不完整性阻碍了 GNN 在图异常检测任务上的训练[19,20]. 此外, 异常检测模型缺乏可解释性, 这严重影响其在实际场景中的应用. 大多数基于 GNN 的方法本质上是黑箱模型, 使得难以理解为何某个节点被分类为异常. 在动态和不完整图环境下, 结构与特征的持续变化进一步增加了决策的验证难度和可信度, 增加了偏差或误导性预测的风险. 现有的解释方法, 如注意力机制或特征归因方法[21,22], 通常依赖于相对完整的数据和稳定的结构, 因此在噪声环境下难以提供可靠的见解.

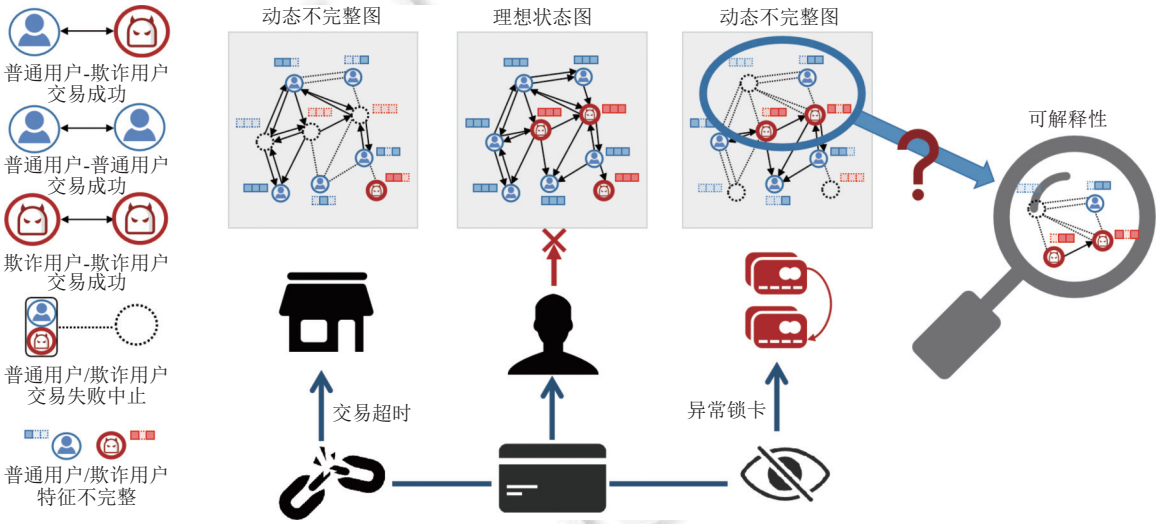


图 1 金融交易网络中的动态不完整图问题

具体而言, 存在 3 大难点: 1) 图结构的动态变化会破坏邻域聚合, 由于缺失或过时的边导致邻域信息不完整, 消息传递过程变得不准确, 从而削弱模型捕捉关键关系模式[23]的能力; 2) 节点特征的不完整性阻碍了有效的表示学习, 导致聚合信息的偏差并降低泛化能力, 同时增加过拟合风险. 尽管时序模型可以适应动态图的变化, 但它们难以处理结构信息的缺失. 而重构或特征补全方法 (如 GAN、GAE) 通常面临梯度冲突、次优优化和噪声引入的问题, 限制了它们在异常检测中的有效性; 3) 在动态不完整图 (dynamic incomplete graph, DIG) 环境下, 稳定结构和完整特征的缺失加剧了解释性挑战. 随着图的演化和特征的波动, 识别和解释异常节点变得愈发困难, 尤其是现有的解释方法往往针对更稳定、更完整的图进行设计. 这种不稳定性削弱了解释技术的可靠性, 使得验证异常检

测结果的合理性变得更加困难.

本文提出了一种可解释的动态不完整图异常检测 (explainable dynamic incomplete graph anomaly detection, EXDIG) 方法, 该方法通过掩码数据增强和互信息优化来处理动态不完整图. 我们通过对节点、边和特征进行掩码来模拟动态不完整图场景, 以提高模型的鲁棒性和泛化能力. 稀疏字典学习方法通过将数据表示为已知特征的线性组合, 从而有效恢复缺失特征. 对于动态图结构, 使用两个预训练模型分别对节点和边进行编码, 并采用强互信息 (strong mutual information, SMI) 损失来最大化二者之间的互信息, 从而确保结构完整性. 弱互信息 (weak mutual information, WMI) 损失用于捕捉局部关系, 平衡全局和局部特征, 从而降低过拟合并提高泛化能力. 此外, EXDIG 通过在训练过程中对节点、边和特征施加掩码扰动, 提高了动态不完整图异常检测的可解释性. 这些扰动有助于识别关键组成部分, 并训练局部线性模型来逼近 GNN 预测结果, 最终增强模型的可解释性, 使其能够提供更透明、可信的异常检测解释. 这些技术使 EXDIG 在无标签、动态图和不完整图环境下能够更好地检测异常. 因此, 本文主要贡献如下.

(1) 针对动态不完整图异常检测问题进行研究. 我们提出的模型确保 GNN 在极端条件下 (如动态图结构和缺失特征) 仍然能够保持有效性, 而不受标签可用性的影响.

(2) 提出一种基于强-弱互信息 (SWMI) 的优化机制, 其中强互信息损失用于最大化全局结构信息, 而弱互信息损失用于捕捉局部关系, 使图神经网络能够在不完整图数据上进行高效表示学习.

(3) 引入一种新的动态不完整图异常检测可解释性机制, 通过在节点、边和特征上施加掩码扰动, 使得局部线性模型能够逼近图神经网络预测, 从而提供更透明、可信的异常检测解释.

(4) 进行大规模实验评估, 包括有标签和无标签场景. 实验结果表明, 所提方法在动态不完整图场景下展现出更强的鲁棒性和抗噪能力, 并能显著提升异常检测精度.

1 相关工作

1.1 图异常检测

图异常检测在图神经网络 (GNN) 的推动下取得了显著进展, 图神经网络能够有效捕捉复杂的图结构. 利用图神经网络的方法, 如卷积模型和扩散模型, 通过建模局部邻域信息, 在异常检测任务中取得成功. 例如, DOMINANT[24] 提出基于自监督学习的动态图异常检测, 通过重构误差识别异常节点; CoLA[25]基于对比自监督学习, 提升全局与局部异常检测能力. 这些方法通常假设图结构较为完整, 依赖邻居聚合以捕捉潜在的异常模式. 此外, 扩散过程也被探索用于图异常检测, 例如 GRAND[26]通过随机扩散增强 GNN 表达能力. 然而, 在不完整图中, 扩散机制可能因缺失节点或边而传播不准确的信息, 导致潜在表示偏离真实结构, 降低对异常的检测敏感度. 然而, 这些方法往往难以应对动态和不完整的图数据, 而这正是现实世界中常见的情况. 例如, 基于局部信息的方法可能无法处理缺失的节点特征和边, 而扩散技术可能会引入噪声, 导致学习任务受误导[1,27,28].

一些研究结合反事实推理和多尺度动态学习来增强异常检测能力, 这些方法能够捕捉时序模式, 但通常假设特征数据是完整的, 难以有效处理结构的不完整性[29,30]. 此外, 针对层次结构建模的双曲神经网络方法也被提出, 但它们往往忽略了特征的不完整性[31]. 尽管这些方法取得了一定进展, 但在处理不完整和动态图数据方面仍然存在显著的挑战. 传统的图神经网络以及近年来基于图神经网络的方法通常假设图数据是理想且完全可观测的, 而在实际应用中, 由于隐私保护、数据错误以及动态变化等因素, 这一假设往往并不成立. 因此, 开发能够从不完整数据和动态数据中学习的鲁棒模型, 对于在现实条件下实现准确的异常检测至关重要[32−34].

另一类相关工作是掩码自编码器方法, 如 GraphMAE[23]通过随机掩码节点或边的特征, 利用自编码器重建缺失部分, 以增强图表示学习. 这类方法通常聚焦于静态且完整的图环境, 旨在提升特征或结构的重建精度. 然而, 它们并未专门设计用于动态不完整图场景, 尤其缺乏对特征-结构一致性问题的建模. 相比之下, 本文不仅采用多掩码视图机制来模拟不同缺失模式, 还引入了强-弱互信息机制, 从特征与结构两个层面共同增强异常检测能力, 构成了与现有掩码自编码器方法的核心差异.

1.2 不完整图学习

不完整图学习通过从不完整数据中学习鲁棒的图表示来应对现实应用中的数据缺失问题. 部分图神经网络能够有效处理缺失的节点特征和边, 即使数据存在不完整性, 仍能表现出较高的性能, 但在高度动态图环境下可能面临困难[35]. 另一种方法是引入噪声容忍机制, 以增强对噪声和缺失信息的鲁棒性, 然而, 依赖噪声估计可能导致在噪声水平不可预测的情况下表现欠佳[36].

基于共识的多视角聚类方法利用共识学习来提高聚类准确性, 但在动态图的时序变化方面可能存在不足[37,38]. 相关研究强调跨多视角整合信息的重要性, 但许多方法在特征严重缺失的情况下仍显不足[39]. 针对不完整多视角数据的创新聚类技术, 如低秩张量图学习和自补全策略, 提供了高效的解决方案, 但在大规模数据集上的可扩展性可能受到限制[40,41]. 非负表示学习通过基于图的方法来重构缺失信息, 在许多场景下表现良好, 但可能难以处理高度稀疏的数据集[42]. 总体而言, 尽管不完整图学习取得了一定进展, 但在处理动态和高度不完整的图数据方面仍然存在挑战, 凸显了对更鲁棒和自适应模型的需求. 针对不完整图的异常检测, 现有研究相对较少, 且多集中于静态场景. 总体而言, 针对动态不完整图的异常检测尚缺乏系统性研究. 本文提出的 EXDIG 模型正是面向动态不完整图, 通过多掩码视图与强-弱互信息机制, 试图解决此类问题.

1.3 图学习的可解释性

在图神经网络 (GNN) 中, 可解释性对于异常检测等应用至关重要, 因为理解模型决策的依据是关键. 然而, 图神经网络常被批评为黑箱模型, 尤其是在动态和不完整图环境下. 常见的方法如注意力机制能够突出对决策重要的邻居[20], 但它们依赖于稳定数据, 在动态图环境下表现不佳. 同样, 特征归因方法 (例如 LIME) 用于识别关键特征[21], 但在处理噪声或缺失数据时存在困难.

为了解决这一问题, 局部可解释性方法通过使用决策树等更简单、可解释的模型来逼近图神经网络预测[43]. 这些方法能够提供对局部邻域动态的洞察, 但通常难以捕捉更广泛的图结构背景, 特别是在不完整或不断变化的图中. 一些技术进一步扩展了这一方法, 利用替代模型或特征重要性排序来解释节点行为[44,45]. 然而, 在高度动态或稀疏的图数据中, 这些方法仍然面临挑战, 因为不断变化的图结构和缺失特征会阻碍其准确性. 近期研究尝试结合全局和局部可解释性方法, 以提高其鲁棒性, 但在现实应用中仍然受到数据不完整性和噪声的限制[46]. 尽管在可解释性方面取得了一定进展, 但在动态不完整图 (DIG) 中, 可解释性仍然是一个重大挑战, 这突显了对能够有效应对这些复杂性的更鲁棒技术的需求.

2 基础知识

本文所提方法主要基于变分自编码器和最大均值差异, 下面就相关概念和基本知识予以介绍.

2.1 动态不完整图 (DIG) 学习

在许多现实世界的应用场景中, 由于隐私限制、数据录入错误或动态变化, 图数据通常是不完整的. 这种不完整性通常表现为两种形式: 结构不完整性 (缺失节点或边) 和特征不完整性 (缺失节点特征).

形式化表示为, 设 $G=(X,A,E)$ 表示一个图, 其中 $X\in\mathbb{R}^{n\times d}$ 是节点特征矩阵, 包含 n 个节点和 d 个特征, $A\in\mathbb{R}^{n\times n}$ 是邻接矩阵, E 是边的集合. 对于不完整图, 特征矩阵和边集分别表示为 $\hat{X}$ 和 $\hat{E}$. 动态不完整图学习的目标是开发能够在不完整图数据上进行鲁棒学习的模型, 生成准确的表示, 以用于分类和异常检测等下游任务.

然而, 信息缺失会破坏图神经网络 (GNN) 中的基本消息传递机制. 在标准 GNN 结构中, 节点表示通过邻域聚合迭代更新:

$$H^{(l)}=\sigma\left(W^{(l)}\tilde{A}H^{(l-1)}\right) \tag{1}$$

其中, $H^{(l)}$ 表示第 l 层的节点嵌入, $W^{(l)}$ 是可训练的变换矩阵, σ 是激活函数, $\tilde{A}$ 是归一化后的邻接矩阵. 在边缺失的情况下, 聚合结构被改变, 导致信息传递不完整, 从而降低关系建模的有效性.

对于特征不完整性, 观测到的特征矩阵 $\hat{X}$ 可以用二进制掩码矩阵 $M\in\{0,1\}^{n\times d}$ 来表示缺失值:

$$\hat{X} = M \odot X \tag{2}$$

其中, $\odot$ 表示按元素相乘. 特征缺失会导致节点嵌入的扭曲, 使得消息聚合出现偏差, 从而降低模型的泛化能力. 传统的填充策略 (如零填充或均值填充) 会引入偏差, 而基于学习的重构方法可能面临次优优化和噪声放大等问题.

为了应对这些挑战, 我们提出的 EXDIG 框架并非直接修改 GNN 结构, 而是利用基于互信息优化的图掩码自编码器 (graph masked autoencoder, GMAE). 通过引入掩码扰动, 并通过特征和结构重构学习鲁棒表示, EXDIG 使 GNN 能够更好地适应动态不完整图场景, 从而有效缓解节点、边和特征缺失的影响.

2.2 图掩码自编码器 (GMAE)

图掩码自编码器 (GMAE) 旨在通过重构被掩码的节点特征, 从不完整图中学习有效表示. GMAE 由编码器和解码器两个主要组件组成. 编码器 F 处理被掩码的节点特征 $\hat{X}$, 这些特征由以下特征掩码函数生成:

$$\hat{X} = M \odot X, \quad M \sim Bernoulli(p) \tag{3}$$

其中, $M \in \{0,1\}^{n \times d}$ 是从伯努利分布中采样的随机二进制掩码矩阵, p 表示被掩码特征的比例. 该策略确保模型能够在不同程度的特征不完整性下学习鲁棒的表示. 编码器将被掩码的特征映射到潜在表示:

$$Z_s = F(\hat{X}) \tag{4}$$

即使在特征缺失的情况下, 仍能够捕捉关键信息. 解码器 Φ 通过对潜在表示进行解码来重构原始节点特征:

$$\hat{X} = \Phi(Z_s) \tag{5}$$

为了确保模型学习到有意义的表示, GMAE 通过最小化特征重构损失进行训练:

$$\mathcal{L}_r = |X - \hat{X}|_2^2 \tag{6}$$

其中, $\mathcal{L}_r$ 衡量了原始特征和重构特征之间的差异.

尽管 GMAE 能够从掩码输入中有效重构节点特征, 但它并未显式建模结构不完整性, 而结构不完整性是动态不完整图 (DIG) 中的关键因素. 为了突破这一局限性, 我们提出的 EXDIG 在 GMAE 的基础上进一步引入结构嵌入和互信息约束, 以确保更鲁棒地适应动态不完整图场景.

2.3 稀疏字典学习

稀疏字典学习将数据表示为少量字典原子的线性组合, 即使在特征不完整的情况下, 也能实现有效的特征重构. 该方法广泛应用于计算机视觉领域, 例如图像去噪和图像修复[47−49].

给定特征矩阵 $X \in \mathbb{R}^{n \times d}$, 我们初始化一个字典 $\mathcal{D} \in \mathbb{R}^{d \times k}$, 其中 k 是字典原子的数量 (通常 $k > d$). 在本研究中, 字典 $\mathcal{D}$ 初始化为服从正态分布 $\mathcal{N}(0,\sigma^2)$ 的随机矩阵, 并在稀疏编码过程中作为可学习参数, 通过梯度下降进行更新. 对于每个特征向量 x_i, 稀疏编码 s_i 通过求解以下优化问题获得:

$$s_i = \arg\min_{s_i} |x_i - \mathcal{D}s_i|_2^2 + \lambda |s_i|_1 \tag{7}$$

其中, λ 控制解的稀疏性. 重构的特征向量表示如下:

$$\hat{x}_i = \mathcal{D}s_i \tag{8}$$

通过将稀疏字典学习应用于节点特征, 我们能够利用高维图数据的冗余性, 有效缓解特征不完整问题. 与基于填充的方法不同, 这些方法依赖于统计假设, 而稀疏编码在保留节点嵌入固有结构的同时, 提高了对缺失信息的鲁棒性.

为了优化整个节点特征矩阵, 我们求解以下全局最小化问题:

$$S^* = \arg\min_{S} |X - \mathcal{D}S|_F^2 + \lambda |S|_1 \tag{9}$$

其中, $S \in \mathbb{R}^{k \times n}$ 为所有节点的稀疏编码矩阵, $|\cdot|_F$ 表示 Frobenius 范数, S^* 表示该问题的最优解.

通过在 EXDIG 框架中引入稀疏字典学习, 我们能够在动态不完整图 (DIG) 条件下实现更有效的表示学习. 与仅依赖观测特征的传统 GNN 方法不同, 我们的方法学习了一种结构化的潜在表示, 从而增强特征重构能力, 确保即使在高度不完整的环境下也能获得鲁棒的图嵌入. 本文在图学习任务中利用稀疏编码对节点特征进行建模, 以提升节点表示能力.

3 一种基于强-弱互信息的可解释掩码异常检测模型

3.1 模型概述

如图 2 所示, 底部为掩码后的多种视图, 包括结构动态和特征不完整的掩码增强, 这些增强视图输入到模型后, 在左侧的结构编码器部分生成结构嵌入 Z_s, 即节点嵌入 Z_n 与掩码边嵌入 Z_e; 右侧为特征编码器部分, 生成特征嵌入 Z_f. 我们的 EXDIG 框架不同于传统编码器处理静态且完整的图数据, 而是专为动态不完整图的异常检测任务设计. 此外, EXDIG 通过掩码机制和强-弱互信息优化, 不仅能够增强模型在缺失数据和动态图环境下的鲁棒性, 还能够提高异常检测的可解释性. 其掩码扰动机制可以识别关键节点、边和特征, 并结合局部线性模型来逼近 GNN 预测, 从而提供更透明、可信的异常检测结果.

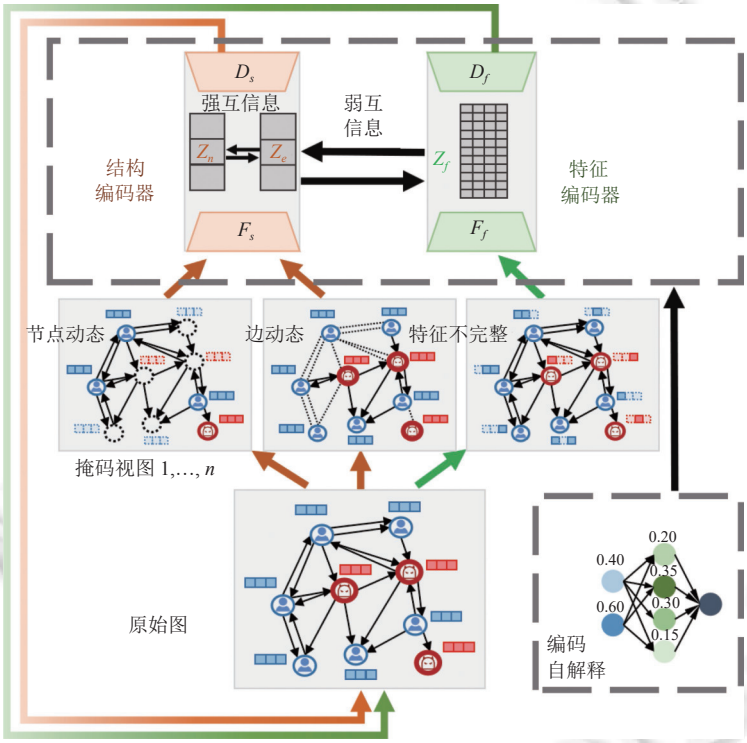


图 2 EXDIG 框架通过处理掩码视图来处理动态不完整图

3.1.1 特征编码器

为了模拟现实世界图异常检测任务中动态不完整性的特征, 我们构造了多个掩码特征视图. 定义: $X_n \in \mathbb{R}^{N\times d}$ 表示节点嵌入矩阵, 由预训练图编码器从图 G 提取的节点表示. $X_e \in \mathbb{R}^{E\times d}$ 表示边嵌入矩阵, 由另一预训练图编码器从图 G 提取的边表示. 掩码矩阵 $M \in \{0,1\}^{N\times F}$ 是一个二值矩阵, 用于随机掩蔽节点特征, 其中 $M_{i,j}=0$ 表示第 i 个节点的第 j 个特征被掩码. 每个掩码特征视图矩阵 $\hat{A}_i$ 通过选择性地移除每个节点的一部分特征生成, 而不是简单地将其设为零. 具体地, 我们令 $\hat{A}_i$ 在实现中表示为二值矩阵 $M \in \{0,1\}^{N\times F}$, 用于对原始节点特征矩阵 X 进行逐元素乘积, 生成不同的特征视图. 每个特征视图的节点特征为 $X^{(i)}=M^{(i)}\odot X$, 其中 $\odot$ 表示逐元素乘积. 这样可以确保每个视图具有独特的特征缺失模式, 从而模拟由于隐私保护、数据录入错误或其他因素导致的特征缺失. 通过从原始特征矩阵 A 中移除部分特征, 我们构造了一组掩码特征视图矩阵 $\hat{A}_1,\hat{A}_2,\ldots,\hat{A}_n$, 从而保留了特征不完整性的动态特性. 这种方法使得特征编码器能够学习对不同程度和不同模式的特征缺失具有鲁棒性的表示.

给定特征矩阵 $A \in \mathbb{R}^{n\times d}$, 我们初始化一个字典 $\mathcal{D} \in \mathbb{R}^{d\times k}$, 其中 $k>d$ 为字典原子数. 对于每个特征向量 a_i, 稀疏编码 s_i 通过求解以下优化问题获得:

$$s_i = \arg\min_{s_i} \|a_i - \mathcal{D}s_i\|_2^2 + \alpha\|s_i\|_1 \tag{10}$$

其中, α 是控制编码稀疏性的正则化参数. 该优化目标确保每个特征向量可以表示为字典原子的稀疏组合, 从而增强模型对缺失数据的鲁棒性. 特征编码器的重构损失定义如下:

$$\mathcal{L}_f = \|A - \hat{A}\|_2^2 = \|A - \mathcal{D}S\|_2^2 \tag{11}$$

采用稀疏字典学习的优势在于它能够有效利用数据的固有冗余性, 使特征表示在数据不完整的情况下仍然保持鲁棒性, 从而提升整体模型性能.

3.1.2 结构编码器

对于原始图 $G = (X, A, E)$, 我们使用两个预训练图模型分别提取节点和边的初始表示. 预训练模型对输入图进行处理, 并分别生成全面的节点和边表示 $P_n(X)$ 和 $P_e(X)$, 从而捕捉局部和全局的结构特征.

为了模拟图结构的动态特性, 我们构造了多个掩码视图 $\hat{G}_1, \hat{G}_2, \ldots, \hat{G}_n$, 其中每个视图通过随机掩码部分节点特征和边来生成. 这些掩码视图有助于模型在动态和不完整环境下学习鲁棒的结构表示.

结构编码器 F_s 处理来自掩码节点和掩码边视图的嵌入. 对于掩码节点, 预训练模型 P_n 生成节点嵌入:

$$E_n = P_n(\hat{X}_n) \tag{12}$$

对于掩码边, 预训练模型 P_e 生成边嵌入:

$$E_e = P_e(\hat{X}_e) \tag{13}$$

参考公式 (4), 结构编码器基于邻接矩阵和节点特征, 通过 GNN 进一步生成节点嵌入: $Z_s = GNN(A, X)$. 最终, 结构编码器整合这些嵌入, 以生成潜在表示 Z_s.

3.1.3 重构模块

特征编码器生成潜在表示 Z_f, 并通过解码器 Φ_f 重新构造为原始特征矩阵, 即: $\hat{A} = \Phi_f(Z_f)$. 通过最小化重构误差, 特征编码器和解码器能够有效捕捉并恢复节点特征, 从而增强对缺失特征的鲁棒性.

结构编码器的潜在表示 Z_s 被用于重构原始图结构, 包括邻接矩阵, 通过解码器 Φ_s 进行恢复: $\hat{E} = \Phi_s(Z_s)$. 通过最小化重构误差, 结构编码器和解码器能够学习在节点和边动态缺失的情况下准确重构图结构.

3.2 强-弱互信息损失

为了确保特征编码器和结构编码器能够学习共享信息并生成一致的表示, 我们设计了强互信息损失和弱互信息损失. 在实现中, 强互信息使用同一结构编码器生成的掩码节点与掩码边嵌入 (同一表征空间, 耦合更紧密), 弱互信息使用稀疏特征编码器生成的掩码特征嵌入与结构编码器生成的节点嵌入 (跨表征空间, 需要投影对齐). 因此前者为“强耦合”, 后者为“弱耦合”.

强互信息重构损失旨在最大化掩码节点嵌入 E_n (公式 (12)) 与掩码边嵌入 E_e (公式 (13)) 之间的互信息, 其定义如下:

$$\mathcal{L}_{\text{strong}} = L_{D1}(Z_s, E_n) + L_{D2}(Z_s, E_e) + L_{D3}(Z_s, E_n, E_e) \tag{14}$$

基于噪声对比估计 (noise contrastive estimation, NCE)[50], 我们定义损失项 L_{D1} 和 L_{D2} 如下:

$$L_{D1}(Z_s, E_n) = \sum_{i=1}^{N} \log \frac{D(Z_s^i, E_n^i)}{\sum_{j=1}^{N} D(Z_s^i, E_n^j)} \tag{15}$$

$$L_{D2}(Z_s, E_e) = \sum_{i=1}^{N} \log \frac{D(Z_s^i, E_e^i)}{\sum_{j=1}^{N} D(Z_s^i, E_e^j)} \tag{16}$$

其中, 判别函数 D 定义如下:

$$D(x, y) = \exp\left(\frac{\mathbf{f}_{\text{proj}}(x) \cdot y}{|\mathbf{f}_{\text{proj}}(x)| \cdot |y| \cdot \tau}\right) \tag{17}$$

其中, $\mathbf{f}_{\text{proj}}:\mathbb{R}^{|n|\times d_x}\to\mathbb{R}^{|n|\times d_y}$ 是投影函数, τ 是温度参数. 进一步推导后, L_{D3} 的详细展开形式如下:

$$L_{D3}(Z_s,E_n,E_e)=\exp\left(\frac{\mathbf{f}_{\text{proj}}(Z_s)\cdot[E_n;E_e]}{|\mathbf{f}_{\text{proj}}(Z_s)|\cdot|[E_n;E_e]|\cdot\tau}\right) \tag{18}$$

弱互信息损失旨在通过捕捉特征表示 Z_f 与结构表示 Z_s 之间的弱相关性, 增强特征编码器和结构编码器之间的一致性, 其定义如下:

$$\mathcal{L}_{\text{weak}}=-\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp(sim(Z_f^i,Z_s^i)/\tau)}{\sum_{j=1}^{N}\exp(sim(Z_f^i,Z_s^j)/\tau)} \tag{19}$$

其中, 相似度函数定义如下:

$$sim(Z_f^i,Z_s^j)=\frac{\mathbf{f}_{\text{proj}}(Z_f^i)\cdot\mathbf{f}_{\text{proj}}(Z_s^j)}{|\mathbf{f}_{\text{proj}}(Z_f^i)|\cdot|\mathbf{f}_{\text{proj}}(Z_s^j)|} \tag{20}$$

模型的整体损失函数结合了特征重构损失、强互信息重构损失以及弱互信息损失, 具体定义如下:

$$\mathcal{L}_{\text{total}}=\lambda_1\mathcal{L}_f+\lambda_2\mathcal{L}_{\text{strong}}+\lambda_3\mathcal{L}_{\text{weak}} \tag{21}$$

其中, λ_1、λ_2 和 λ_3 为超参数, 用于平衡各损失项的权重. 在本研究中, $\mathcal{L}_{\text{weak}}$ 是一种基于 InfoNCE 的相似性对齐损失, 用于度量由特征编码器生成的节点表示 Z_f 与由结构编码器生成的节点表示 Z_s 之间的一致性. 与结构内的强互信息损失 $\mathcal{L}_{\text{strong}}$ 不同, $\mathcal{L}_{\text{weak}}$ 主要用于跨特征域与结构域的对齐. 在动态不完整图场景下, 通过对不同掩码视图的对齐, $\mathcal{L}_{\text{weak}}$ 能有效减少由图不完整性引起的表征变化. 例如, 在反洗钱场景下, 一个节点的部分银行卡特征缺失时, 结构编码器仍可通过邻居关系捕捉该节点与异常行为的潜在联系, $\mathcal{L}_{\text{weak}}$ 则帮助两者共享信息, 提升检测鲁棒性. 由于直接估计互信息较为困难, 本文将互信息最大化作为建模目标, 并采用 InfoNCE[51]作为互信息下界的估计方法. 这种设计不同于通常在同一表征空间内进行对比学习的传统方法, 而是针对跨表征空间构建一致性约束, 以提升模型在异常检测任务中的鲁棒性.

通过优化 $\mathcal{L}_{\text{total}}$, EXDIG 框架能够从动态不完整图数据中学习鲁棒且准确的节点和结构表示, 从而提升分类和异常检测等下游任务的性能.

3.3 可解释性讨论

为了揭示 EXDIG 框架的决策过程, 我们设计了一个可解释性模块, 用于评估缺失特征、边和节点对异常检测的影响.

按照前述方法, 通过扰动原始图 G 的特定特征、边和节点, 生成掩码图 $\hat{G}_i$. 对于每个掩码图, 我们计算对应的异常分数 $\hat{y}_i$, 并测量其变化量:

$$\Delta\hat{y}_i=\hat{y}(G)-\hat{y}(\hat{G}_i) \tag{22}$$

这一变化量反映了被扰动组件对模型预测的影响, 变化幅度越大, 表明相应的特征或结构对模型决策的重要性越高.

特征重要性通过线性回归方法进行评估. 对于每个掩码特征矩阵 $\hat{A}_i$, 局部线性模型定义如下:

$$\hat{y}_i=\mathbf{w}_i^{\mathrm{T}}\hat{A}_i+b_i \tag{23}$$

其中, $\mathbf{w}_i$ 表示特征的重要性权重. 通过最小化重构损失来优化回归模型:

$$\mathcal{L}_{\text{reg}}=\sum_{i=1}^{m}\left(\hat{y}_i-\mathbf{w}_i^{\mathrm{T}}\hat{A}_i\right)^2 \tag{24}$$

这一过程能够量化各特征对异常检测决策的贡献, 权重值越大, 表明该特征的重要性越高.

最后, 我们基于异常分数的变化计算节点和边的相对重要性. 对于边, 其重要性通过评估异常分数对邻接矩阵 A 的扰动敏感性得到, 而节点的重要性则由每个节点的特征向量 X_i 或嵌入表示 Z_s^i 对模型输出的影响来确定. 具体的得分计算如下:

$$edge\_score_{ij}=\left|\frac{\partial \hat{y}}{\partial A_{ij}}\right|,\ node\_score_i=\left|\frac{\partial \hat{y}}{\partial X_i}\right| \tag{25}$$

这些得分揭示了哪些组件对异常检测决策具有最显著的影响.

4 实验分析

在本节中, 我们对所提出的方法 (EXDIG) 在不同的动态不完整图 (DIG) 场景下进行了广泛的实验评估. 本实验旨在解答以下研究问题.

RQ1: 在动态不完整图环境下, 我们的方法能否有效发挥作用?

RQ2: 我们的方法在不同程度的动态不完整图场景下是否仍然保持良好性能?

RQ3: 关键的设计选择和超参数对我们方法的性能有何影响?

RQ4: 我们的方法是否能够在动态不完整环境下直观地增强表示能力?

RQ5: 在动态不完整图场景中, 我们的方法是否能够保持可解释性, 并且在图结构不完整和动态变化的情况下, 能否有效解释模型的决策?

4.1 实验设置

● 数据集

我们在 9 个公开的真实世界图数据集上评估所提出的方法, 包括 Cora、CiteSeer、Amazon Photo、Amazon Computers、CoAuthor CS、CoAuthor Physics、EllipticBitcoin、YelpChi 和 DGraphFin. 其中, Amazon Photo、Amazon Computers、EllipticBitcoin、YelpChi 和 DGraphFin 是与异常检测及其子领域相关的数据集, 用于验证我们方法在这些场景下的有效性. 由于 DGraphFin 具有较大的数据规模, 因此仅用于验证 RQ4. 数据集的详细信息见表 1.

表 1 实验数据集

名称	本文简称	节点	边	特征	类别
Amazon Computers	Computers	13 752	491 722	767	10
Amazon Photo	Photo	7 650	238 162	745	8
CoAuthor CS	CS	18 333	163 788	6 805	15
CoAuthor Physics	Physics	34 493	495 924	8 415	5
Cora	Cora	2 708	10 556	1 433	7
CiteSeer	CiteSeer	3 327	9 104	3 703	6
EllipticBitcoin	Bitcoin	203 769	234 355	165	2
YelpChi	Yelp	45 954	3 846 979	32	2
DGraphFin	DGraphFin	3 700 550	4 300 999	17	2

● 动态不完整图场景实现

在本研究中, 我们通过对图数据施加随机扰动并限制训练节点数量, 模拟各种动态不完整图 (DIG) 场景. 具体而言, 我们随机删除原始图结构中 x% 的边和 y% 的节点, 以实现动态结构的不完整性. 与其他处理不完整图的任务不同, 为了更贴近现实世界中动态不完整异常图的特性, 我们不使用零填充缺失特征, 而是直接从特征矩阵中移除 z% 的元素. 这种方法更符合异常检测任务的实际情况, 即用户或实体在所有维度上通常无法获得完全对齐的数据, 缺失的数据字段在物理上表示消失, 而不仅是数值上的零填充. 因此, 从原始图中分别移除 x% 的边、y% 的节点和 z% 的特征后, 得到一个动态不完整的异常图. 我们将动态不完整比率 (dynamic incomplete graph ratio, DIG ratio) 定义为图中缺失组件 (边、节点和特征) 所占的比例, 以更准确地反映现实场景.

● 实验细节

所有实验均基于 PyTorch[52]和 PyG (PyTorch geometric)[53]进行实现. 我们在 5 次实验中报告了测试集上的平均准确率及标准差. 对于本文方法, 我们在验证集上执行网格搜索 (grid search) 以选择最优超参数. 由于图中的每个节点都经历了特征级不完整性掩码和全局结构级不完整性掩码, 无法使用基于节点级的小批量训练 (如 PyG 中的 NeighborSampler 或 ClusterLoader). 因此, 对于一些较大的数据集, 如 Bitcoin、DGraphFin、CS 和 Physics, 随机

采样部分数据构造子图进行实验. 所有实验均在配备 NVIDIA GeForce RTX 4080 的设备上运行.

4.2 性能比较 (RQ1)

针对 RQ1, 我们通过移除 75% 的节点、边和特征 (动态不完整比率 75%) 构造异常图, 以模拟现实动态不完整场景中的常见情况, 即中等程度的动态不完整. 在该场景下比较了 EXDIG 和无弱互信息损失版本 (EXDIG-SMI) 与 11 种基线方法的性能, 实验结果如表 2 所示, 后续表格中加粗数据为最优结果. 具体而言, 对于所有对比方法, 我们将节点/边/特征的动态补全比率设为 0.5, 神经网络的隐藏层维度设为 64 维, 输出层维度设为 512 维, 学习率统一设置为 0.001, 并对每种方法训练 300 轮. 我们的观察结果如下.

表 2 EXDIG 与现有方法的 NMI 指标性能比较 (%)

Method	Anomaly detection			General		
	Computers	Photo	Bitcoin	Cora	CiteSeer	CS
GCN	18.90±4.43	30.07±4.42	0.01±0.00	50.67±2.24	8.22±0.61	5.75±0.80
GraphSAGE	2.89±1.11	4.45±0.89	0.01±0.01	2.90±1.06	2.03±0.73	3.13±0.78
GAT	36.27±2.66	50.48±4.15	6.55±2.80	16.64±1.19	5.68±1.68	5.89±0.44
GAE	8.23±2.14	11.83±2.80	2.21±1.94	1.99±0.43	4.13±0.46	2.66±0.33
VGAE-non-neg	24.36±4.01	34.70±2.54	0.01±0.00	21.15±2.62	10.87±1.49	6.59±0.72
VGAE-rand-neg	26.31±4.59	31.99±4.49	0.01±0.01	21.71±3.26	11.20±2.69	6.72±0.50
GPS	14.36±4.19	18.12±3.70	0.03±0.04	14.36±4.19	27.57±4.26	59.24±3.39
ADGN	15.07±4.32	24.92±4.52	9.32±0.36	16.08±4.98	24.60±2.89	60.48±1.64
EGC	6.72±2.04	6.43±2.98	0.04±0.02	6.72±2.04	3.21±0.80	3.29±0.45
GIN	8.84±0.68	7.86±1.54	0.03±0.04	0.76±0.01	1.79±0.02	1.32±0.12
GREET	18.04±1.21	19.09±0.21	0.04±0.01	9.69±0.19	11.32±0.07	19.90±0.19
EXDIG-SMI	**49.47±1.20**	62.55±4.04	8.25±0.33	51.32±2.76	33.40±2.54	60.02±0.52
EXDIG	47.07±1.56	**64.94±2.02**	**9.89±0.41**	**55.65±2.45**	**35.28±3.29**	**62.08±1.44**

● NMI

表 2 中的 NMI 结果显示, EXDIG 在多个数据集上的表现优越, 尤其在异常检测任务中展现了卓越的性能. 在 Photo (64.94%±2.02%) 和 CiteSeer (35.28%±3.29%) 数据集上 EXDIG 取得了最高的 NMI 得分, 表明其在保持图结构和检测异常方面具有强大的能力. 然而, 在 Bitcoin (9.89%±0.41%) 和 Computers (47.07%±1.56%) 数据集上的表现虽具竞争力, 但相较之下不够突出, 表明该方法可能在结构更复杂或多样性更强的数据集上更具优势. 结果表明, 尽管 EXDIG 在小型、结构化数据集上表现出色, 但在处理较为简单或噪声较大的数据时, 其优势可能有所降低.

● ARI

表 3 中的 ARI 结果进一步验证了 EXDIG 在一般任务类别中的稳健性. 在 Photo (52.18%±2.81%) 和 CiteSeer (34.51%±3.80%) 数据集中, EXDIG 取得了最高的 ARI 得分, 表明即使在动态不完整的环境下, 该方法依然能够高效聚类节点. 在异常检测任务中, EXDIG 在 Bitcoin (17.01%±0.72%) 和 Computers (47.35%±1.14%) 数据集上的表现良好, 但在较大或更复杂的数据集上未必始终优于 ADGN. 值得注意的是, 在 CS 数据集 (46.99%±2.71%) 上的表现突出, 这表明 EXDIG 在超越异常检测任务的更广泛应用中具有潜力, 特别是在一般任务中的表现尤为优异. 因此, 尽管 EXDIG 整体上表现强劲, 但在处理大规模、复杂图结构时仍可能受益于进一步的改进.

此外, 为了进一步验证各方法在异常检测任务中的实际区分能力, 我们补充了基于下游二分类器的评测. 具体而言, 所有方法均在无监督条件下完成节点表征学习, 训练过程中不使用任何标签信息. 在此基础上, 我们将各模型生成的节点表征输入同一个下游二分类器 (逻辑回归), 并使用 Accuracy (ACC) 和 $F1$ 值作为二分类指标. 实验结果如图 3 所示. 整体来看, EXDIG 在最具代表性的 Bitcoin 和 Cora 数据集上均取得了 Accuracy 和 $F1$ 值的最高成绩, 表现出较强的竞争力. 在 Cora 上, EXDIG 相较其他方法具备更明显的优势, 而在 Bitcoin 上, 尽管 EXDIG 依旧取得最高分, 其优势相对不够明显, 部分对比方法 (如 ADGN、EXDIG-SMI) 表现接近, 显示出在不同图结构和噪声条件下, 模型仍存在进一步提升的空间. 总体而言, EXDIG 在无监督条件下所学习的表征展现出良好的异常检测能力, 验证了其鲁棒性与泛化性.

表 3 EXDIG 与现有方法的 ARI 指标性能比较 (%)

Method	Anomaly detection			General		
	Computers	Photo	Bitcoin	Cora	CiteSeer	CS
GCN	9.87±3.55	15.08±2.46	0.12±0.30	16.15±0.56	5.98±0.80	2.38±0.48
GraphSAGE	0.21±1.27	1.77±0.37	0.00±0.08	0.00±0.20	0.73±0.29	0.78±0.58
GAT	29.51±3.42	37.12±4.35	11.49±4.40	8.74±1.91	4.58±1.74	2.37±0.61
GAE	5.97±2.48	7.88±2.38	5.28±3.23	0.22±0.24	0.73±0.28	0.87±0.15
VGAE-non-neg	12.40±3.39	17.91±1.41	0.11±0.30	14.18±2.94	9.51±1.72	3.07±0.31
VGAE-rand-neg	14.24±3.78	15.96±2.90	0.11±0.31	14.16±2.67	9.88±2.65	2.66±0.90
GPS	4.89±3.76	8.18±1.91	0.07±0.35	4.89±3.76	24.83±4.17	42.01±7.54
ADGN	6.94±2.90	12.30±3.25	16.38±0.59	7.25±3.49	18.53±4.19	46.85±3.75
EGC	2.52±1.29	2.98±1.68	0.05±0.08	2.52±1.29	2.61±0.49	1.61±0.59
GIN	0.00±2.27	2.10±0.23	0.01±0.25	1.35±0.30	0.00±0.02	0.08±0.06
GREET	2.01±0.53	11.66±0.07	0.09±0.01	4.83±0.13	7.36±0.03	8.06±0.51
EXDIG-SMI	**50.10±1.24**	47.37±5.17	13.89±0.09	38.56±2.79	33.42±3.08	42.91±1.82
EXDIG	47.35±1.14	**52.18±2.81**	**17.01±0.72**	**39.88±1.42**	**34.51±3.80**	**46.99±2.71**

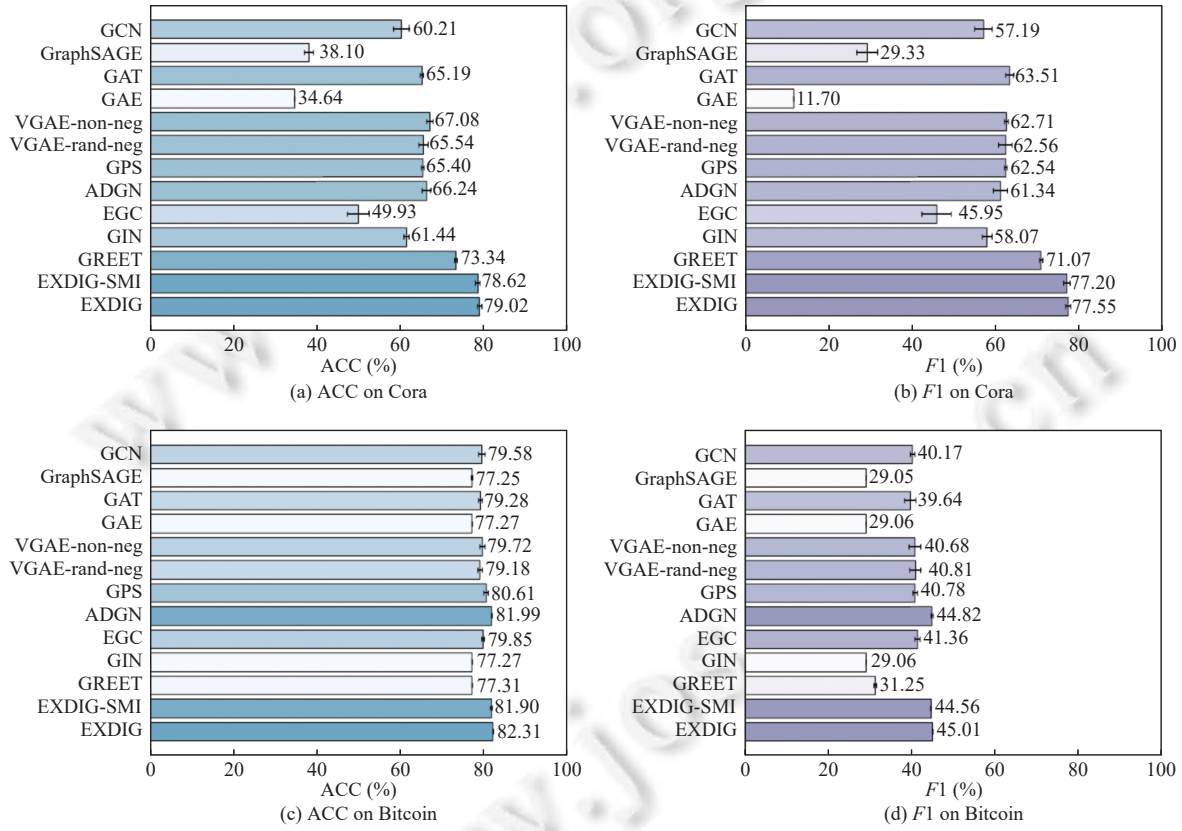


图 3 EXDIG 在 Cora 和 Bitcoin 数据集上的准确率 (ACC) 和 *F*1 值补充实验

4.3 定量评估 (RQ2)

在训练过程中, 模型学习能够捕捉结构和特征信息的节点表示, 并可应用于标签分类等下游任务. 一个简单的分类器 (如逻辑回归) 可用于评估这些表示的质量——更优的表示将带来更高的分类精度和准确率. 为了回答 RQ2, 我们从两个关键的图学习维度对 EXDIG 进行基准测试: 下游任务性能与动态不完整的严重程度.

为了评估不同动态不完整级别下的下游任务质量, 我们在不同数据集上调整动态不完整比率, 并使用多种评估指标. 首先对模型生成的嵌入进行标准化, 并划分为训练集和测试集. 然后, 使用优化后的超参数训练一个逻辑回归分类器, 并在测试集上计算评估指标, 所有实验均运行 5 次, 并计算其均值, 以衡量不同动态不完整比率下的表示质量, 详细分析如下.

图 4 展示了 EXDIG 在 3 个数据集上的性能, 随着动态不完整比率在节点、边和特征维度上均匀增加. 评估指标包括准确率 (ACC)、$F1$、精确率 (Precision) 和召回率 (Recall). 在 Photo 数据集上, 该模型在面对日益增加的不完整性时仍能保持较强的性能, 表现出较高的鲁棒性. 在 Computers 数据集中, EXDIG 依然表现良好, 但在 Precision 指标上略显敏感, 反映出其对结构复杂度的敏感性. 而在 CiteSeer 数据集中, 尽管动态不完整比率不断增加, 该模型的性能依然保持稳定, 甚至在部分指标上出现轻微提升. 这些结果表明, EXDIG 具有较强的鲁棒性, 尽管在极端情况下性能可能有所波动.

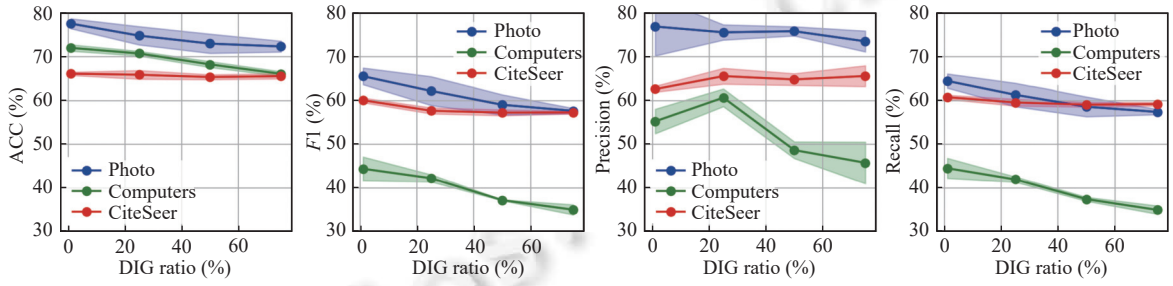


图 4 EXDIG 在不同动态不完整比率下对 Photo、Computers 和 CiteSeer 数据集的性能变化

图 5 提供了 Cora 和 Bitcoin 数据集上 AUC 指标的热力图分析. 在此实验中, 节点和边的完整性比率独立变化, 而特征补全比率固定为 1%. 该分析展示了模型在不同类型的图不完整性下的鲁棒性. 在 Cora 数据集中, 当节点和边的补全度较低时, AUC 评分显著下降, 强调了结构完整性对于该类图的重要性. 相比之下, 在 Bitcoin 数据集中, 即使节点完整性较低, AUC 依然在不同的边完整比率下保持稳定, 表明边信息对模型性能的影响较小, 这可能是由于该数据集具有较高的连接密度或关键结构节点的存在.

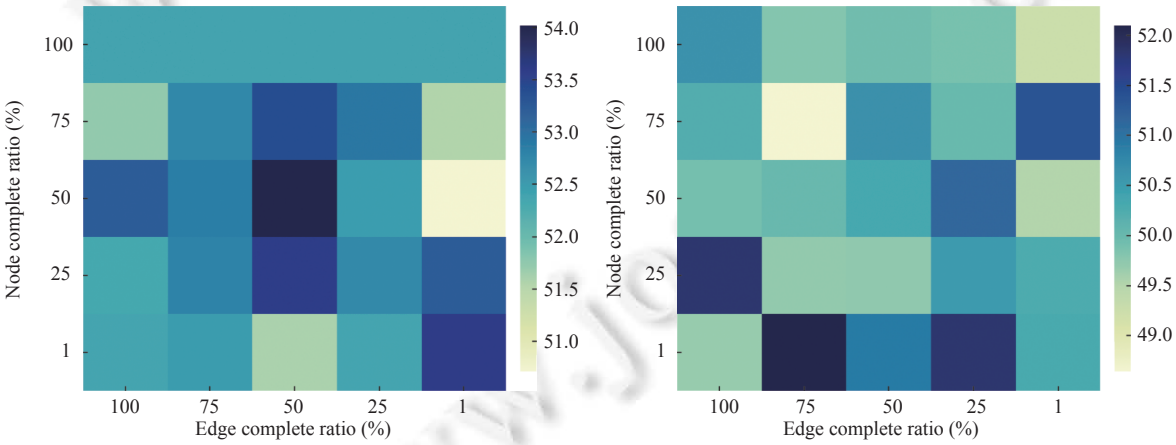


图 5 EXDIG 在下游任务中性能变化的热力图分析

总体而言, 图 4 和图 5 的实验结果表明, EXDIG 在不同级别和类型的动态不完整场景下仍然保持有效, 且性能依赖于具体数据集. 这些结果证实了该方法在处理不同程度的结构和特征不完整性方面的能力, 并在结构属性较为完备的数据集中展现出较强的泛化能力.

4.4 消融实验与超参分析 (RQ3)

为了解答 RQ3, 我们对提出的 EXDIG 模型进行了消融研究和超参数分析. 在“w/o SMI”配置下, 模型不计算公式 (14)–公式 (18) 定义的强互信息损失项; 在“w/o WMI”配置下, 模型不计算公式 (19)、公式 (20) 定义的弱互信息损失项. 其余训练流程保持一致.

● 消融研究

表 4 展示了在 50% DIG 比率下的消融实验结果. 带有完整强-弱互信息 (SWMI) 损失的模型在 Photo 和 Cora 数据集上的表现均优于其他配置, 凸显了强互信息损失 (SMI) 和弱互信息损失 (WMI) 组件的重要性. 具体而言, 在 Photo 数据集中, 当去除 WMI 组件时, 模型性能显著下降, 表明 WMI 在维持特征与结构的一致性方面起到了关键作用. 另一方面, 在 Cora 数据集中, 去除 SMI 组件导致更大幅度的性能下降, 这表明 SMI 在动态图环境中促进节点与边信息融合的关键作用. 表 5 展示了在 99% DIG 比率下的实验结果. 尽管 WMI 在 Photo 数据集的表现至关重要, 但其缺失对 Cora 数据集的影响较小. 相比之下, 去除 SMI 仍然会导致显著的性能下降, 尤其是在 Cora 数据集中, 这表明即使在低不完整性水平下, SMI 仍然是动态结构中信息流通的关键因素.

表 4 SWMI 方法在 50% DIG 比率的 Photo 和 Cora 数据集上的消融研究 (%)

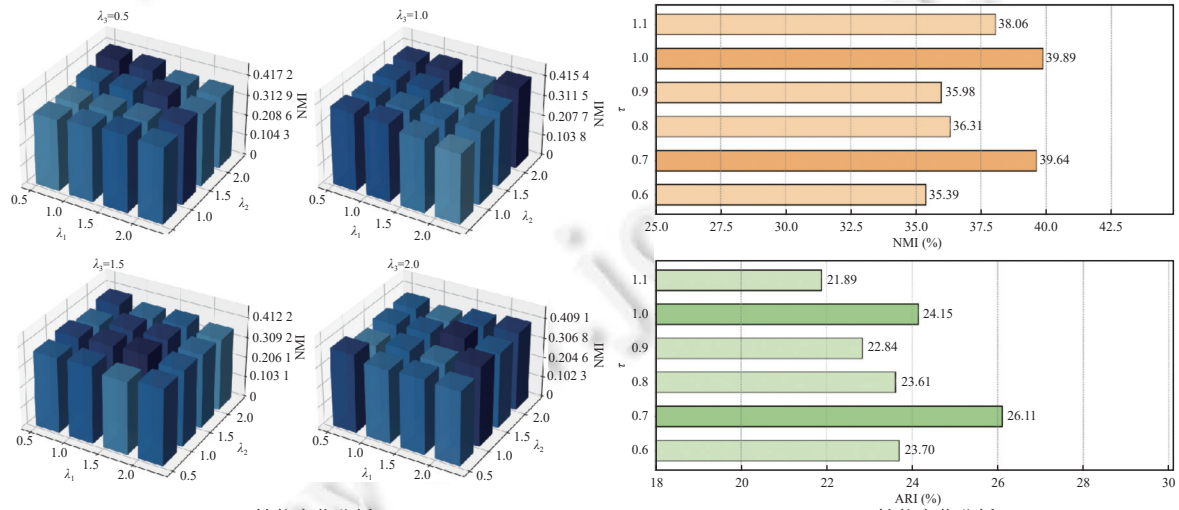
Method	Photo		Cora	
	NMI	ARI	NMI	ARI
SWMI	64.80±6.81	**54.67±5.86**	**54.80±2.89**	**42.07±1.80**
w/o WMI	**65.41±4.57**	54.31±3.21	54.19±3.88	41.01±2.31
w/o SMI	62.85±3.41	50.92±6.87	52.94±3.37	38.92±1.75
w/o SWMI	60.54±1.71	47.67±8.97	50.33±3.17	37.64±2.68

表 5 SWMI 方法在 99% DIG 比率的 Photo 和 Cora 数据集上的消融研究 (%)

Method	Photo		Cora	
	NMI	ARI	NMI	ARI
SWMI	**63.40±3.42**	**51.82±6.74**	**52.17±0.58**	**39.53±1.42**
w/o WMI	62.85±1.47	51.09±7.40	49.55±3.13	38.07±2.13
w/o SMI	61.94±2.35	49.50±7.92	51.70±1.94	38.10±2.23
w/o SWMI	59.79±1.50	45.98±5.80	48.16±3.76	36.57±1.66

● 超参数分析

图 6(a) 展示了模型在不同 λ_1、λ_2 和 λ_3 配置下的性能变化. 结果表明, $\lambda_3 = 1$ 的适中设置提供了最佳平衡, 在不同的 λ_1 和 λ_2 取值下保持了稳定的 NMI 分数. 然而, 进一步增大 λ_3 会降低模型性能, 这可能是由于过度强调特征编码器和结构编码器之间的耦合, 导致模型过拟合. 值得注意的是, 所提出的框架在不同超参数设置下均表现出稳健性, 这反映了其对各种不完整图场景的适应能力, 并进一步验证了强-弱互信息损失机制在增强模型泛化能力方面的有效性和鲁棒性.



(a) λ 性能变化分析 (b) τ 性能变化分析

图 6 λ_1、λ_2、λ_3 和 τ 的超参分析结果图

从图 6(b) 来看, NMI 和 ARI 分数在不同 τ 取值下表现出一定的变化趋势. 其中, NMI 在 $\tau = 1.0$ 时最高, 为 39.89%, 其次是 $\tau = 0.7$ (39.64%). ARI 分数则在 $\tau = 0.7$ 时达到最高 (26.11%), 相比其他取值更优. 这一趋势表明, τ 设定对模型性能影响显著. 过小或过大的 τ 可能导致信息利用失衡, 从而影响聚类效果. 当 $\tau = 0.7$ 时, NMI 和 ARI 均接近最优, 说明该值有助于平衡结构与特征信息, 增强模型稳健性. 这进一步验证了强-弱互信息损失在不同超参数下的鲁棒性, 并突出了合理调整 τ 以优化性能的重要性.

4.5 可视化分析 (RQ4)

在本节中, 我们使用 t-SNE 选择 3 个关键特征维度, 并为 Yelp 和 DGraphFin 数据集创建多维散点矩阵. 对角线上的图展示了每个类别的核密度估计, 而非对角线上的图则显示了不同特征维度之间的散点分布.

在图 7 中, 红色点表示异常类别节点, 蓝色点表示正常类别节点. 散点图所示的 3 个维度分别对应特征嵌入中的第 1–第 3 主成分, 用于展示节点间的分布差异. 如图 7 所示, YelpChi 数据集的可视化结果显示, 在 DIG 条件下, 两个类别之间存在明显的区分. 散点图突出显示了类别之间的分离, 尤其是在第 1 维度上, 类别差异最为显著. 核密度估计进一步强调了这一点, 在第 1 维度上, 红色和蓝色分布之间的重叠最小, 表明即使在 95% DIG 比率且仅保留 1/10 的节点和边的情况下, 模型仍能有效地学习类别表示. 而第 2 和第 3 维度的类别分布重叠相对较多.

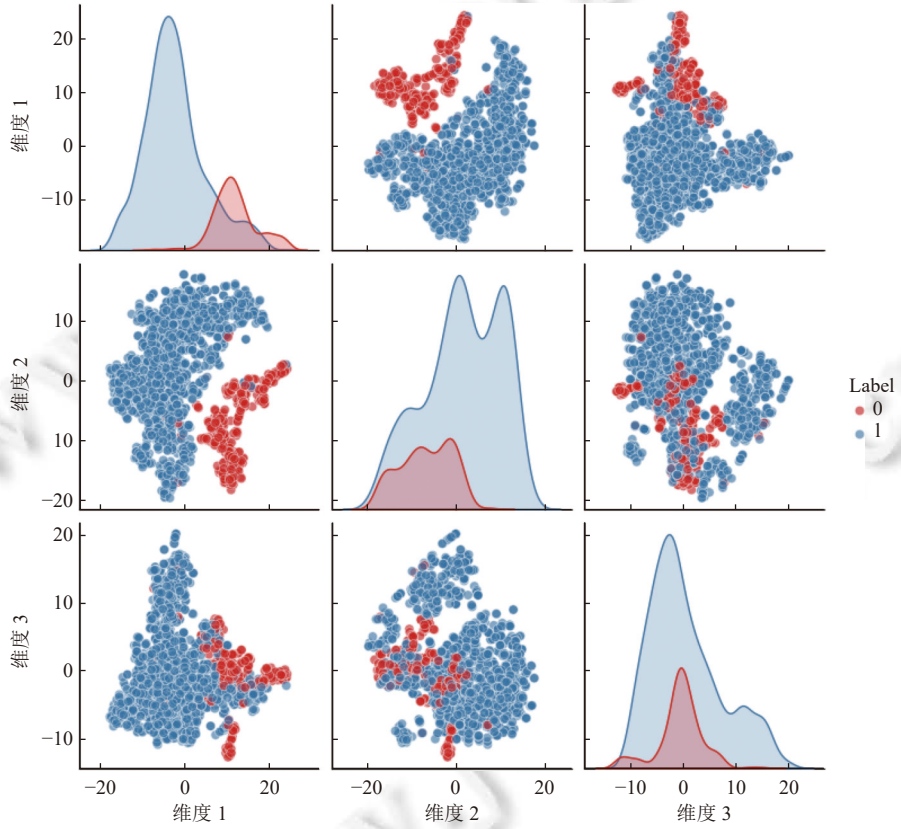


图 7 EXDIG 在 YelpChi 数据集中 95% DIG 比例下的可视化分析

图 8 中, 红色与蓝色点分别表示异常和正常节点类别. 所绘制的三维散点图对应节点嵌入向量的第 1–第 3 主成分. 通过该可视化结果可以观察到, 即使在 95% 的动态不完整条件下, 不同类别节点仍呈现出明显的分布差异. 相比之下, 图 8 展示了 DGraphFin 数据集的可视化结果. 尽管该数据集规模更大且复杂度更高, 但模型在 DIG 条件下仍能保持清晰的类别分离, 即使仅保留 1/300 的节点和边. 散点图显示了明显的聚类现象, 尤其是在第 3 维度上, 类别之间的分离最为明显. 而在第 1 和第 2 维度上, 类别之间的重叠较多. 核密度估计进一步证实了这一点, 尽

管某些维度上类别分离较清晰, 但仍存在一定的重叠, 这反映了数据集本身的复杂性.

总体而言, 这些可视化结果表明, 我们的方法在动态不完整图条件下能够有效增强数据表示, 在两个数据集中均展现了清晰的类别分离. 特别是在更复杂和更大规模的 DGraphFin 数据集中, 该方法仍能保持类别区分能力, 这进一步验证了其在 DIG 场景下的有效性.

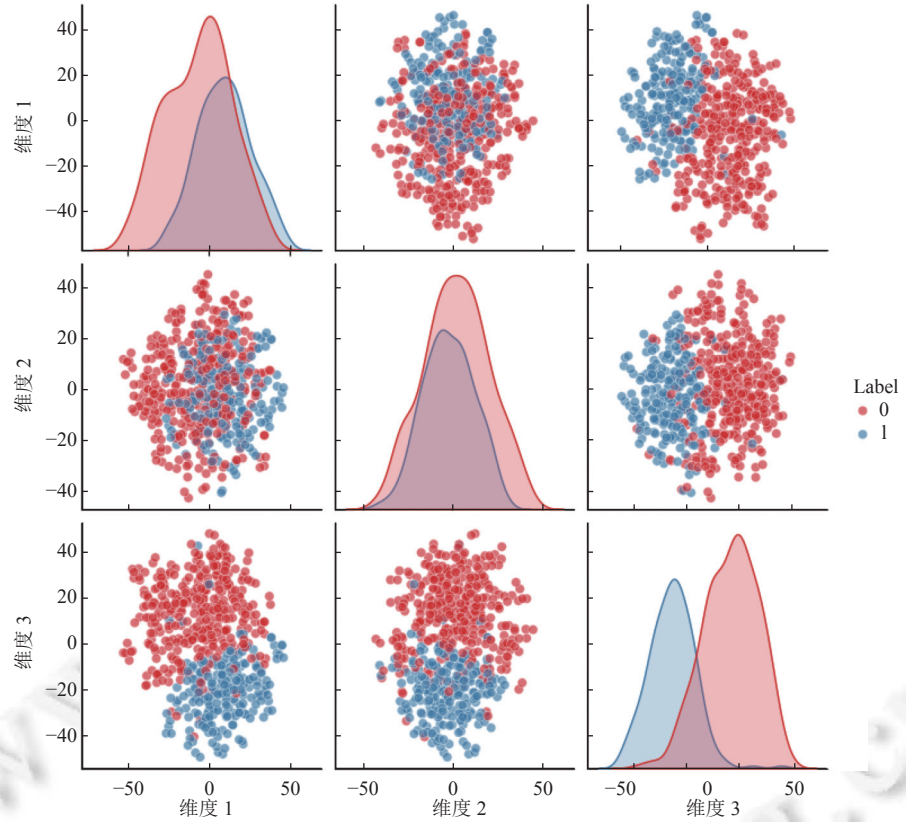


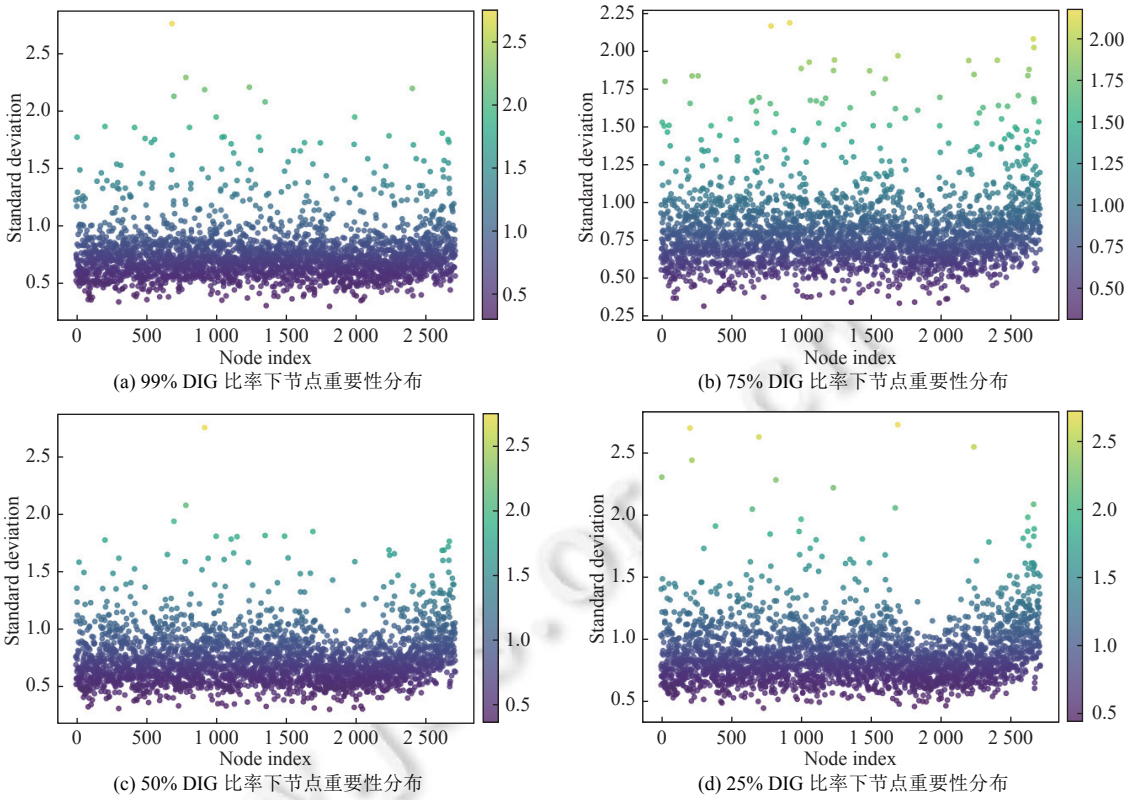
图 8 EXDIG 在 DGraphFin 数据集中 95% DIG 比例下的可视化分析

4.6 可解释性研究 (RQ5)

我们的可解释性模块实验结果表明, 即使在不同程度的动态不完整图 (DIG) 环境下, EXDIG 框架仍然能够保持可解释性. 实验结果通过可视化展示了模型如何在缺失节点、边和特征的情况下, 仍能有效地解释其决策过程, 并在不同的 DIG 比率下提供合理的解释.

● 节点重要性分布聚类: 如图 9 所示, 在第 1 组可视化实验中, 我们将节点重要性得分聚类为低、中、高这 3 个类别. 随着 DIG 比率的降低, 高重要性节点的分布保持稳定, 而低重要性节点对缺失组件表现出更强的敏感性. 在 99% DIG 比率下, 由于大量数据缺失, 重要性分布出现更大的变化. 然而, 即使在极端扰动条件下, 模型仍能识别出高重要性节点, 证明了该方法的鲁棒性.

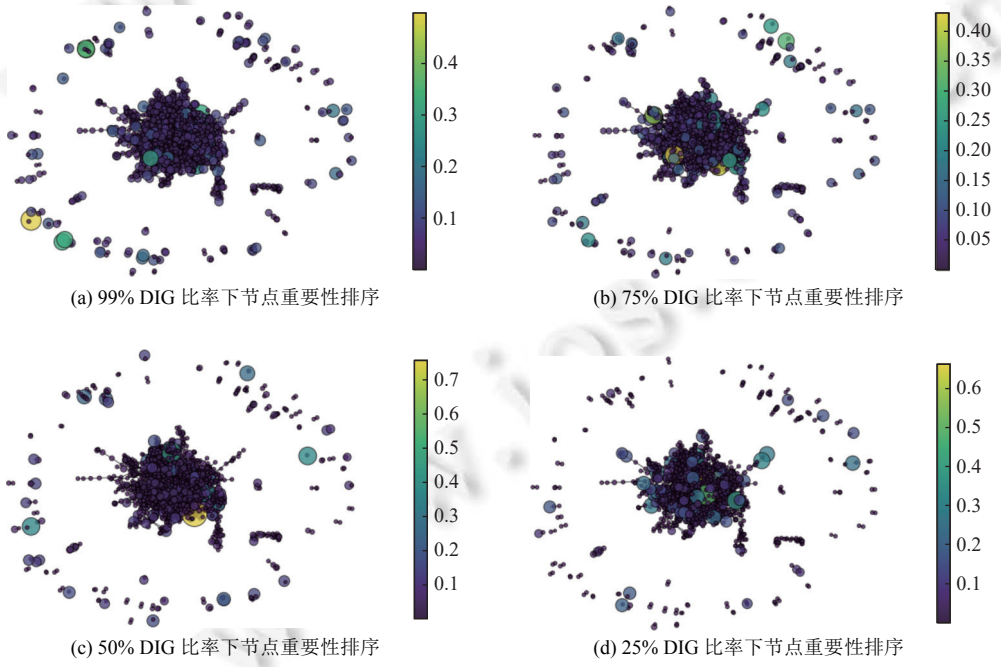
● 基于重要性得分的图可视化: 如图 10 所示, 在第 2 组可视化实验中, 我们对节点的重要性得分进行颜色编码和大小调整, 以突出哪些节点对模型决策影响最大. 结果清晰地展示了, 即使在 99% 的图数据缺失的情况下, 模型仍然能够突出关键节点. 当 DIG 比率从 99% 降至 25% 时, 高重要性节点和低重要性节点之间的区分度更加明显, 特别是在 25% DIG 比率下, 图结构更完整, 节点重要性更加清晰可见. 尽管数据缺失, EXDIG 仍能有效突出关键节点, 确保决策过程的透明度.



(a) 99% DIG 比率下节点重要性分布 (b) 75% DIG 比率下节点重要性分布

(c) 50% DIG 比率下节点重要性分布 (d) 25% DIG 比率下节点重要性分布

图 9 节点重要性分布的聚类结果: 对所有节点的异常得分进行聚类



(a) 99% DIG 比率下节点重要性排序 (b) 75% DIG 比率下节点重要性排序

(c) 50% DIG 比率下节点重要性排序 (d) 25% DIG 比率下节点重要性排序

图 10 基于重要性排序的节点可视化结果

总体而言, 在所有 DIG 比率下的实验结果均表明, EXDIG 在不同程度的缺失图组件情况下仍能保持可解释性. 模型在不同扰动级别下始终能够稳定地识别关键节点和边, 进一步验证了其在动态不完整图环境下的鲁棒性和可靠性.

5 总 结

我们提出了 EXDIG 框架, 以解决动态不完整图 (DIG) 中的异常检测问题, 通过应对特征和结构的不完整性来提升模型性能. 大量实验验证了该方法在不同动态不完整条件下的鲁棒性和有效性, 能够恢复 GNN 的表示能力, 并适用于多种下游任务. 此外, EXDIG 通过识别关键组成部分, 提高了异常检测结果的可解释性. 该框架为扩展到更复杂、更贴近现实的工业生产图场景提供了新的可能性. 未来的研究将重点优化 EXDIG 的计算效率和可扩展性, 使其能够应用于包含数百万个节点和边的大规模图数据. 具体而言, 一是探索针对演化图流的动态图数据流建模框架, 以实时捕捉动态不完整图中结构与特征的快速演化; 二是研究轻量化的结构增强掩码策略, 以降低模型的计算负担, 并提升在大规模动态图场景下的适应性. 为此, 我们将探索优化模型架构的方法, 并结合分布式或并行计算策略, 以提升其在真实大规模部署中的可行性.

References

[1] Ma XX, Wu J, Xue S, Yang J, Zhou C, Sheng QZ, Xiong H, Akoglu L. A comprehensive survey on graph anomaly detection with deep learning. IEEE Trans. on Knowledge and Data Engineering, 2023, 35(12): 12012–12038. [doi: 10.1109/tkde.2021.3118815]

[2] Kim H, Lee BS, Shin WY, Lim S. Graph anomaly detection with graph neural networks: Current status and challenges. IEEE Access, 2022, 10: 111820–111829. [doi: 10.1109/ACCESS.2022.3211306]

[3] Noble CC, Cook DJ. Graph-based anomaly detection. In: Proc. of the 9th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. Washington: ACM, 2003. 631–636. [doi: 10.1145/956750.956831]

[4] Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv:1609.02907, 2016.

[5] Veličković P, Cucurull G, Casanova A, Romero A, Liò P, Bengio Y. Graph attention networks. In: Proc. of the 6th Int'l Conf. on Learning Representations. Vancouver: OpenReview.net, 2018. [doi: 10.48550/arXiv.1710.10903]

[6] Hamilton WL, Ying R, Leskovec J. Inductive representation learning on large graphs. In: Proc. of the 31st Int'l Conf. on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2017. 1025–1035. [doi: 10.5555/3294771.3294869]

[7] Song LY, Ma ZY, Li ZH, Shang XQ. Review on temporal graph neural networks for financial risk prediction. Ruan Jian Xue Bao/Journal of Software, 2024, 35(8): 3897–3922 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/7087.htm [doi: 10.13328/j.cnki.jos.007087]

[8] Yu H, Liu ZY, Luo XF. Barely supervised learning for graph-based fraud detection. In: Proc. of the 38th AAAI Conf. on Artificial Int'l and 36th Conf. on Innovative Applications of Artificial Intelligence and 14th Symp. on Educational Advances in Artificial Intelligence. Menlo Park: AAAI Press, 2024. 16548–16557. [doi: 10.1609/aaai.v38i15.29593]

[9] Liu ZY, Yu H, Luo XF. Federated graph anomaly detection via disentangled representation learning. In: Proc. of the 2025 ACM on Web Conf. Sydney: ACM, 2025. 1216–1224. [doi: 10.1145/3696410.3714567]

[10] Wang RJ, Mou S, Wang X, Xiao WP, Ju Q, Shi C, Xie X. Graph structure estimation neural networks. In: Proc. of the 2021 Web Conf. Ljubljana: ACM, 2021. 342–353. [doi: 10.1145/3442381.3449952]

[11] Lai XC, Zhang Z, Chen H, Zhang LY, Li ZH, Lu W. Tracking-removed neural network with graph information for classification of incomplete data. Applied Intelligence, 2025, 55(4): 256. [doi: 10.1007/s10489-024-06031-7]

[12] Zhou BS, Ji JT, Gu ZB, Zhou ZH, Ding GY, Feng SH. One-step graph-based incomplete multi-view clustering. Multimedia Systems, 2024, 30(1): 32. [doi: 10.1007/s00530-023-01225-4]

[13] Zhang H, Chen XH, Zhang EH, Wang LP. Incomplete multi-view learning via consensus graph completion. Neural Processing Letters, 2023, 55(4): 3923–3952. [doi: 10.1007/s11063-022-10973-9]

[14] Ge Y, Chen SC. Graph convolutional network for recommender systems. Ruan Jian Xue Bao/Journal of Software, 2020, 31(4): 1101–1112 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5928.htm [doi: 10.13328/j.cnki.jos.005928]

[15] Liu J, Shang XQ, Song LY, Tan YC. Progress of graph neural networks on complex graph mining. Ruan Jian Xue Bao/Journal of Software, 2022, 33(10): 3582–3618 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6626.htm [doi: 10.13328/j.cnki.jos.006626]

[16] Ding KZ, Zhang CX, Tang J, Chawla N, Liu H. Toward graph minimally-supervised learning. In: Proc. of the 28th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. Washington: ACM, 2022. 4782–4783. [doi: 10.1145/3534678.3542602]

[17] Li QM, Wu XM, Liu H, Zhang XT, Guan ZC. Label efficient semi-supervised learning via graph filtering. In: Proc. of the 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Long Beach: IEEE, 2019. 9574–9583. [doi: 10.1109/CVPR.2019.00981]

[18] You JX, Ma XB, Ding DY, Kochenderfer M, Leskovec J. Handling missing data with graph representation learning. In: Proc. of the 34th Int'l Conf. Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2020. 19075–19087. [doi: 10.5555/3495724.3497325]

[19] Jin W, Ma Y, Liu XR, Tang XF, Wang SH, Tang JL. Graph structure learning for robust graph neural networks. In: Proc. of the 26th ACM SIGKDD Int'l Conf. on Knowledge Discovery & Data Mining. ACM, 2020. 66–74. [doi: 10.1145/3394486.3403049]

[20] Sun K, Lin ZC, Zhu ZX. Multi-stage self-supervised learning for graph convolutional networks on graphs with few labeled nodes. In: Proc. of the 34th AAAI Conf. on Artificial Intelligence. Palo Alto: AAAI Press, 2020. 5892–5899. [doi: 10.1609/aaai.v34i04.6048]

[21] Antwarg L, Miller RM, Shapira B, Rokach L. Explaining anomalies detected by autoencoders using Shapley additive explanations. Expert Systems with Applications, 2021, 186: 115736. [doi: 10.1016/j.eswa.2021.115736]

[22] Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": Explaining the predictions of any classifier. In: Proc. of the 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016. 1135–1144. [doi: 10.1145/2939672.2939778]

[23] Hou ZY, Liu X, Cen YK, Dong YX, Yang HX, Wang CJ, Tang J. GraphMAE: Self-supervised masked graph autoencoders. In: Proc. of the 28th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining. Washington: ACM, 2022. 594–604. [doi: 10.1145/3534678.3539321]

[24] Ding KZ, Li JD, Bhanushali R, Liu H. Deep anomaly detection on attributed networks. In: Proc. of the 2019 SIAM Int'l Conf. on Data Mining (SDM). Philadelphia: Society for Industrial and Applied Mathematics, 2019. 594–602. [doi: 10.1137/1.9781611975673.67]

[25] Liu YX, Li Z, Pan SR, Gong C, Zhou C, Karypis G. Anomaly detection on attributed networks via contrastive self-supervised learning. IEEE Trans. on Neural Networks and Learning Systems, 2022, 33(6): 2378–2392. [doi: 10.1109/TNNLS.2021.3068344]

[26] Chamberlain B, Rowbottom J, Gorinova MI, Bronstein M, Webb S, Rossi E. GRAND: Graph neural diffusion. In: Proc. of the 38th Int'l Conf. on Machine Learning, PMLR. 2021. 1407–1418.

[27] Atkinson O, Bhardwaj A, Englert C, Ngairangbam VS, Spannowsky M. Anomaly detection with convolutional graph neural networks. Journal of High Energy Physics, 2021, 2021(8): 80. [doi: 10.1007/JHEP08(2021)080]

[28] Pang SK, Xiao CJ, Tai WX, Cheng ZT, Zhou F. Graph anomaly detection with diffusion model-based graph enhancement (student abstract). In: Proc. of the 38th AAAI Conf. on Artificial Intelligence. Washington: AAAI Press, 2024. 23610–23612. [doi: 10.1609/aaai.v38i21.30494]

[29] Wei YT, Shu WZ, Cheng ZT, Tai WX, Xiao CJ, Zhong T. Counterfactual graph learning for anomaly detection with feature disentanglement and generation (student abstract). In: Proc. of the 38th AAAI Conf. on Artificial Intelligence. Washington: AAAI Press, 2024, 38: 23682–23683. [doi: 10.1609/aaai.v38i21.30524]

[30] Jin YX, Wei YT, Cheng ZT, Tai WX, Xiao CJ, Zhong T. Multi-scale dynamic graph learning for time series anomaly detection (student abstract). In: Proc. of the 38th AAAI Conf. on Artificial Intelligence. Washington: AAAI Press, 2024, 38: 23523–23524. [doi: 10.1609/aaai.v38i21.30456]

[31] Gu J, Zou DM. Three revisits to node-level graph anomaly detection: Outliers, message passing and hyperbolic neural networks. In: Proc. of the 2nd Learning on Graphs Conf. (LoG 2023). 2023. 27–30. [doi: 10.48550/arXiv.2403.04010]

[32] Tang JH, Li JJ, Gao ZQ, Li J. Rethinking graph neural networks for anomaly detection. In: Proc. of the 39th Int'l Conf. on Machine Learning. Baltimore, 2022. [doi: 10.48550/arXiv.2205.15508]

[33] Papadimitriou P, Dasdan A, Garcia-Molina H. Web graph similarity for anomaly detection (poster). In: Proc. of the 17th Int'l Conf. on World Wide Web. Beijing: ACM, 2008. 1167–1168. [doi: 10.1145/1367497.1367709]

[34] Mesgaran M, Hamza AB. Graph fairing convolutional networks for anomaly detection. Pattern Recognition, 2024, 145: 109960. [doi: 10.1016/j.patcog.2023.109960]

[35] Jiang B, Zhang ZY. Incomplete graph representation and learning via partial graph neural networks. arXiv:2003.10130, 2021.

[36] Berger P, Hannak G, Matz G. Efficient graph learning from noisy and incomplete data. IEEE Trans. on Signal and Information Processing over Networks, 2020, 6: 105–119. [doi: 10.1109/TSIPN.2020.2964249]

[37] Wen J, Liu CL, Xu GH, Wu ZH, Huang C, Fei LK, Xu Y. Highly confident local structure based consensus graph learning for incomplete multi-view clustering. In: Proc. of the 2023 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE,

2023. 15712–15721. [doi: 10.1109/CVPR52729.2023.01508]
[38] Zhou W, Wang H, Yang Y. Consensus graph learning for incomplete multi-view clustering. In: Proc. of the 23rd Pacific-Asia Conf. on Advances in Knowledge Discovery and Data Mining. Macao: Springer, 2019. 529–540. [doi: 10.1007/978-3-030-16148-4_41]
[39] Tang JJ, Yi QQ, Fu SJ, Tian YJ. Incomplete multi-view learning: Review, analysis, and prospects. Applied Soft Computing, 2024, 153: 111278. [doi: 10.1016/j.asoc.2024.111278]
[40] Ji GY, Lu GF. One-step incomplete multiview clustering with low-rank tensor graph learning. Information Sciences, 2022, 615: 209–225. [doi: 10.1016/j.ins.2022.10.026]
[41] Zhao XJ, Shen QQ, Chen YY, Liang YS, Chen JX, Zhou YC. Self-completed bipartite graph learning for fast incomplete multi-view clustering. IEEE Trans. on Circuits and Systems for Video Technology, 2024, 34(4): 2166–2178. [doi: 10.1109/TCSVT.2023.3302326]
[42] Sun SL, Zhang N. Incomplete multiview nonnegative representation learning with graph completion and adaptive neighbors. IEEE Trans. on Neural Networks and Learning Systems, 2024, 35(3): 4017–4031. [doi: 10.1109/TNNLS.2022.3201562]
[43] Liu NH, Feng QZ, Hu X. Interpretability in graph neural networks. In: Wu LF, Cui P, Pei J, *et al*, eds. Graph Neural Networks: Foundations, Frontiers, and Applications. Singapore: Springer, 2022. 121–147. [doi: 10.1007/978-981-16-6054-2_7]
[44] Yuan H, Yu HY, Gui SR, Ji SW. Explainability in graph neural networks: A taxonomic survey. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2023, 45(5): 5782–5799. [doi: 10.1109/TPAMI.2022.3204236]
[45] Wang T, Zheng XW, Zhang LF, Cui Z, Xu CY. A graph-based interpretability method for deep neural networks. Neurocomputing, 2023, 555: 126651. [doi: 10.1016/j.neucom.2023.126651]
[46] Ying R, Bourgeois D, You JX, Zitnik M, Leskovec J. GNNExplainer: Generating explanations for graph neural networks. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 9244–9255. [doi: 10.5555/3454287.3455116]
[47] Mairal J, Elad M, Sapiro G. Sparse representation for color image restoration. IEEE Trans. on Image Processing, 2008, 17(1): 53–69. [doi: 10.1109/TIP.2007.911828]
[48] Yang M, Dai DX, Shen LL, van Gool L. Latent dictionary learning for sparse representation based classification. In: Proc. of the 2014 IEEE Conf. on Computer Vision and Pattern Recognition. Columbus: IEEE, 2014. 4138–4145. [doi: 10.1109/CVPR.2014.527]
[49] Mairal J, Bach F, Ponce J, Sapiro G, Zisserman A. Supervised dictionary learning. In: Proc. of the 22nd Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2008. 1033–1040. [doi: 10.5555/2981780.2981909]
[50] Gutmann M, Hyvärinen A. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In: Proc. of the 13th Int'l Conf. on Artificial Intelligence and Statistics. Sardinia, 2010. 297–304.
[51] van den Oord A, Li YZ, Vinyals O. Representation learning with contrastive predictive coding. arXiv:1807.03748, 2019.
[52] Paszke A, Gross S, Massa F, *et al*. PyTorch: An imperative style, high-performance deep learning library. In: Proc. of the 33rd Int'l Conf. on Neural Information Processing Systems. Vancouver: Curran Associates Inc., 2019. 8026–8037. [doi: 10.5555/3454287.3455008]
[53] Fey M, Lenssen JE. Fast graph representation learning with PyTorch geometric. arXiv:1903.02428, 2019.

附中文参考文献

[7] 宋凌云, 马卓源, 李战怀, 尚学群. 面向金融风险预测的时序图神经网络综述. 软件学报, 2024, 35(8): 3897–3922. http://www.jos.org.cn/1000-9825/7087.htm [doi: 10.13328/j.cnki.jos.007087]
[14] 葛尧, 陈松灿. 面向推荐系统的图卷积网络. 软件学报, 2020, 31(4): 1101–1112. http://www.jos.org.cn/1000-9825/5928.htm [doi: 10.13328/j.cnki.jos.005928]
[15] 刘杰, 尚学群, 宋凌云, 谭亚聪. 图神经网络在复杂图挖掘上的研究进展. 软件学报, 2022, 33(10): 3582–3618. http://www.jos.org.cn/1000-9825/6626.htm [doi: 10.13328/j.cnki.jos.006626]

作者简介

骆祥峰, 博士, 教授, 博士生导师, 主要研究领域为机器学习, 知识图谱, 脑认知, 知识系统.
顾峻铨, 博士生, 主要研究领域为机器学习, 图学习, 欺诈检测.
余航, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为机器学习, 概念漂移.