一种基于增强图对比学习的欺诈检测方法

高熠辉[1]，李元庆[1]，张三峰[1,2]，杨望[1,2]
（1.东南大学网络空间安全学院，江苏 南京 211189;
2. 教育部计算机网络和信息集成重点实验室(东南大学)，江苏 南京 211189）

摘 要：图对比学习作为一种有效的预训练策略，能够有效解决基于图的欺诈检测方法中高质量标签数据匮乏的问题。然而，当前这类方法面临恶意行为特征在图神经网络聚合机制中被削弱或在数据增强过程中受损的挑战。为此，本文提出了一种结合图重构和动态数据增强技术的图对比学习优化方法，旨在提升欺诈检测的效果。该方法通过调整图的边权重，减少因邻居特征聚合而产生的冲突，从而提高检测准确性。同时，利用标签不变性和分布多样性指标动态调整数据增强过程，以确保增强数据既能保留关键的欺诈特征，又具备必要的多样性。在多个图欺诈检测数据集上的实验结果表明，该方法的有效性，相较于最先进的技术，检测性能提升了 2%至 5%。

关键词：欺诈检测；图对比学习；图数据增强；图重构
中图法分类号：TP389.1 **文献标志码：**A

A Fraud Detection Framework based on Enhanced Contrastive Graph Learning

GAO Yihui[1], Li Yuanqing[1], Zhang Sanfeng[1,2], Yang Wang[1,2]
(1. School of Cyber Science and Engineering, Southeast University, Nanjing 211189, China;
2.Key Laboratory of Computer Network and Information Integration (Southeast University), Ministry of Education, Nanjing 211189, China)

Abstract: Graph contrastive learning, as an effective pre-training strategy, addresses the issue of scarce high-quality labeled data in graph-based fraud detection methods. However, current approaches encounter challenges related to the weakening of malicious behavior features within graph neural network aggregation mechanisms and potential damage during data augmentation processes. To tackle these issues, this paper proposes an optimization method for graph contrastive learning that integrates graph reconstruction and dynamic data augmentation techniques, aimed at enhancing fraud detection effectiveness. This method reduces conflicts arising from neighbor feature aggregation by adjusting the graph's edge weights, thereby improving detection accuracy. Additionally, the data augmentation process is dynamically modified using indicators of label invariance and distribution diversity, ensuring that the augmented data retains essential fraudulent features while maintaining necessary diversity. Experimental results on multiple graph fraud detection datasets demonstrate the method's effectiveness, achieving a detection performance improvement of 2% to 5% over state-of-the-art techniques.

Key words： Fraud detection; Graph contrastive learning; Graph data augmentation; Graph reconstruction

收稿日期：20**-**-**; **修回日期:** 20**-**-**.
基金项目：国家重点研发计划（2022YFB3104601）;自然科学基金项目(No. 62172093)
通信作者：张三峰(sfzhang@seu.edu.cn)
通信地址：211189 江苏省南京市东南大学网络空间安全学院
Address： School of Cyber Science and Engineering, Southeast University, Nanjing 211189, Jiangsu, P.R.China

1 引言

随着数字经济的快速发展，欺诈行为成为金融、电子商务等领域的一大挑战。传统的欺诈检测方法，如基于规则的系统和传统机器学习技术[1]，专注各实体的个体特征，而忽略了实体间的关联关系，在面对复杂和动态变化的欺诈策略时效果不佳，因此基于图的欺诈检测方法开始成为新的研究热点[2]。

这些方法通过将实体建模为带属性的节点，并将相应的交互建模为边，从而在图层次上检测可疑实体[3]。例如，这些方法可以用于分类评论-用户-评论图中的欺骗性评论[4]、预测用户-IP-用户网络中的虚假账号[5]，或通过点击引发的跳转关系过滤恶意网站[6]。图神经网络（Graph Neural Network, GNN）能够充分利用数据中隐含的复杂关系和结构信息，将节点间的连接关系纳入分析。此外，GNN 具有良好的自适应性和扩展性，能够处理大型复杂网络，使其在处理非结构化数据和识别复杂欺诈行为模式方面优于传统方法[7]。然而，基于 GNN 的欺诈检测方法在实际应用中普遍面临标签数据稀缺的挑战[8]，因此，基于图的欺诈检测方法常常采用图对比学习进行预训练[9]，以降低对标签数据的依赖。

然而，如图 1 所示，基于图对比学习的欺诈检测方法仍面临两个主要挑战。

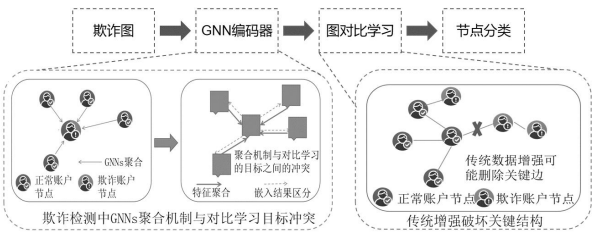

Figure 1. Challenges Faced by Fraud Detection Methods Based on Graph Contrastive Learning

图 1 基于图对比学习的欺诈检测方法所面临的挑战

首先，现有的 GNN 聚合机制容易忽视欺诈节点的恶意行为特征。GNN 通过聚合相邻节点的特征生成每个节点的嵌入表示。在欺诈检测场景中，部分欺诈用户可能与大量正常用户相连接，导致其嵌入表示被正常节点的特征平均化，从而削弱了其恶意特征。此外，这种聚合机制与对比学习的优化目标相悖，因为图对比学习旨在使不同节点的嵌入尽可能不同，以增强模型对节点特征的区分能力。

为此，本文针对图欺诈检测问题设计了一种基于异常值的图结构优化方法，其核心思想是计算所有节点的异常值，从中选出异常值较大的节点，并调整其邻接矩阵以适应对比学习的需求。考虑到目标节点的邻居同时出现在正负梯度贡献中，导致优化方向相反，从而使节点特征趋同。因此，在消息传递阶段根据异常值设置消息传递的权重可以克服图对比学习中的特征平滑效应。具体而言，本文提出在图编码器的聚合操作之前，使用异常值计算器来获取各个邻居的权重。结合对比学习损失，优化节点嵌入表示，使得正样本对更加相似，负样本对更加不同，从而提高模型的表示能力和分类性能。

其次，现有的图对比学习数据增强方法不适用于图欺诈检测场景。现有的增强方法通常通过随机删除节点或边、扰动节点特征等方式生成不同的视图[10]，旨在提升模型的鲁棒性。然而，在社交网络中，某些边（如虚假账户与其他账户之间的异常连接）可能是识别欺诈行为的关键。现有的图数据增强方法可能会意外删除这些关键边，从而破坏图中的重要结构信息，导致模型对虚假账户行为的误判，降低检测准确性。

为此，本文首次在图对比学习框架下提出了可学习的动态图数据增强策略。为了更好地将受控噪声引入数据集中进行增强，本文引入了基于标签不变性和嵌入多样性的动态数据增强机制。这些指标评估原始数据与其增强版本之间的关系：标签不变性强调保留相同标签，而分布多样性则强调它们在表示分布上的差异。

总体而言，本文贡献可总结如下：

(1) 基于异常值的图结构优化方法：针对图欺诈检测场景，分析 GNN 邻居节点正负梯度冲突问题，设计图结构优化算法，使其更适合节点级的图对比学习。

(2) 动态图数据增强策略：提出基于标签不变性和分布多样性的动态增强机制，确保增强后的图数据保留关键欺诈特征，同时引入必要的多样性，从而提升模型的鲁棒性和检测准确性。

(3) 增强的欺诈检测能力：在三种公开的图欺诈检测任务数据集上的实验结果表明，本文模型在节点分类基准上表现优异，超过了多个最先进的基线方法。

2 相关工作

2.1 数据增强

图数据增强技术在 GNN 和图对比学习（Graph Contrastive Learning, GCL）中发挥着重要作用。现有研究主要集中于如何生成高质量的增强图，以提升模型的鲁棒性和泛化能力。典型

的图数据增强方法可分为几类：节点特征增强，通过添加噪声、扰动或替换节点特征生成不同视图。例如，GraphAug[11]通过对节点特征添加高斯噪声来增强数据，从而提高模型的鲁棒性。拓扑结构增强，通过修改图的结构（如节点添加、节点删除、边添加和边删除）生成增强图。例如，DGI[12]通过随机删除部分边生成新的图视图，提升了模型的对比学习效果。混合增强方法，结合多种增强策略以提高效果。AugNet[13]提出了一种混合增强方法，结合节点特征扰动和拓扑结构扰动，显著提升了图对比学习的性能。

然而，这些图数据增强方法可能无意间删除与欺诈检测相关的关键边，从而破坏图中的重要结构信息。这种损失可能导致模型对欺诈行为的误判，降低整体检测的准确性。

2.2 嵌入传播冲突

目前，一些研究将重点放在节点级图对比学习上，如 BGRL[14]、CCA-SSG[15]和 GRADE[16]。这些方法通过不同的自我监督策略在节点级别进行对比学习，提升了节点表示的鲁棒性和泛化能力。与节点级 GCL 不同，GraphCL[17]则侧重于图级任务，通过多种图数据增强策略（如节点特征扰动和拓扑结构修改）来提升模型性能。此外，近年来的研究还探索了通过增强机制[18]和负采样技术[19]来改进 GCL 的方法。这些研究在不同任务上验证了其有效性，并为 GCL 提供了新的思路。

然而，图对比学习与 GNN 消息传递之间的冲突尚未得到深入研究。GNN 的聚合机制导致节点特征混杂，这不利于图对比学习进行有效的节点区分，从而可能导致下游检测过程中的误判。

3 方法

3.1 总体结构

如图 2 所示，本文提出的图欺诈检测框架包括两个主要部分：基于异常值的欺诈图结构优化和使用可学习的动态图数据增强的图对比学习。这两个部分协同工作，以提高欺诈检测的精度和鲁棒性。

在基于异常值的图结构优化过程中，首先利用一个 GNN 网络提取节点的嵌入表示。为了解决图对比学习中的正负样本冲突问题，本文设计了一种通过计算异常值调整邻接矩阵以优化图结构的方法。具体来说，调整邻接矩阵的边权重，以减少与高冲突负样本相关的影响，从而确保目标节点在信息聚合过程中更好地保留关键特征。此外，我们还引入高阶邻居以扩展潜在相邻候选池，从而替换原始图结构，进一步增强模型的表示能力。

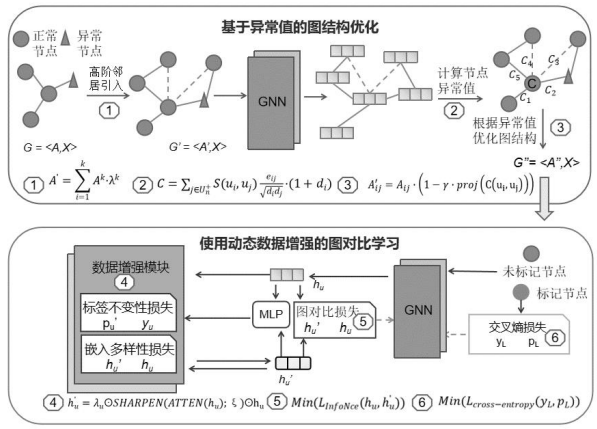

Figure 2 Overview of the Framework

图 2 整体架构

在动态图数据增强模块中，提出了一种可学习的增强策略，该策略基于标签不变性和嵌入多样性两个指标，动态调整数据增强过程。该模块依赖于可学习的掩码机制生成增强数据。具体而言，掩码机制使用注意力机制为输入特征向量的每个维度计算权重，然后通过锐化函数将这些权重二值化为 0 或 1，以保留对一致性和多样性失贡献较大的特征维度，同时掩盖其他维度。在增强过程中，通过保留特定特征维度，并基于标签不变性和嵌入多样性对网络的相关参数进行调整，从而生成高一致性和高多样性的特征表示。

最终，这两个部分共同作用，通过减少特征混杂和引入受控噪声，提升图对比学习的效果，使模型在欺诈检测任务中展现出更强的泛化能力和鲁棒性。

在欺诈图结构优化模块和图对比学习模块中，均采用相同结构的 GNN 网络。该网络由输入层、多层图卷积层、特征拼接层和输出投影层构成。首先，输入节点特征通过多层感知机（MLP）进行初步线性变换，主要目的是将输入特征维度映射到隐藏层维度，以便后续处理。模型包含多个图卷积层，这些层负责处理图结构数据。在每一层中，节点通过图卷积操作从其邻居处聚合信息，以更新自身的特征表示。每层的输出可选择进行批归一化，以进一步规范化节点特征。随后，通过激活函数进行非线性变换，并应用丢弃层以防止过拟合，增强模型的泛化能力。在特征拼接与输出投影部分，模型将每一层的输出特征拼接在一起，形成更长的特征向量。这些拼接的特征包含了每一层图卷积的输出信息，能够捕捉不同层次的特征表示。最后，将拼接后的特征通过另

一个 MLP 投影到最终输出维度，以生成适用于具体节点分类任务的分类或回归结果。

3.2 基于异常值的欺诈图结构优化

3.2.1 异常值计算分析

一个图可以表示为 $G=(V,E,X)$，其中 V 表示节点集合，E表示边集合，X表示节点的特征矩阵，设$A\in\{0,1\}^{N\times N}$表示邻接矩阵。对于目标节点u_i，其邻居节点集合为$N(i)$。在单层图卷积网络 GCN 中，目标节点u_i的嵌入表示为(1)。

$$u_i=\Phi\left(\sum_{j\in N(i)\cup\{i\}}\frac{e_{ji}}{\sqrt{d_i d_j}}x_j\right)\#(1)$$

其中，节点嵌入u_i通过图卷积计算得到，$\Phi=W^T\theta^T$表示 GCN 和投影函数的参数，$N(i)$表示节点i的邻居集合，e_{ji}是从节点j到i的边的权重，d_i和d_j分别表示节点i和节点j的度，W是权重矩阵。在对比学习中，本文使用 InfoNCE 损失函数来优化嵌入。

$$\begin{aligned}L_i&=-log\frac{f(u_i,v_i)}{f(u_i,v_i)+\sum_{k\neq i}f(u_i,v_k)+\sum_{k\neq i}f(u_i,u_k)}\\&=-log[f(u_i,v_i)]\\&+log\left[f(u_i,v_i)+\sum_{k\neq i}f(u_i,v_k)+\sum_{k\neq i}f(u_i,u_k)\right]\end{aligned}\#(2)$$

当使用 InfoNCE 损失函数(2)作为图对比损失函数时，对于目标节点的邻居节点而言，本文主要考虑损失函数中$\sum_{k\neq i}f(u_i,u_k)$相关部分对目标节点u_i的影响。设 $\hat{u}_i=\frac{u_i}{||u_i||}$ 是u_i的规范化形式，将损失函数关于 $\hat{u}_i$求导，得到(3)。

$$\begin{aligned}\frac{\partial L_i}{\partial\hat{u}_i}=\frac{1}{\tau}((S(u_i,v_i)-1)\cdot\hat{v}_i\\+\sum_{k\neq i}S(u_i,v_k)\cdot\hat{v}_k\\+\sum_{k\neq i}S(u_i,u_k)\cdot\hat{u}_k)\end{aligned}\#(3)$$

其中，S (i, j) = softmax(f(i, j)) ∈ [0、1]是i被识别为j的概率，将视图内负样本的部分提取出来，用 I(u_i,u_k)表示视图内所有负样本对的影响，并且用图卷积计算进行替换，其影响可以表示为(4)。

$$I(u_i,u_k)=\sum_{k\neq i}\sum_{j\in N(k)\cup\{k\}}S(u_i,u_k)\frac{e_{jk}}{\sqrt{d_k d_j}}x_j\#(4)$$

我们观察到，在既作为目标节点的邻居又作为图对比学习负样本的节点中，存在冲突。这是由于目标节点参与了邻居节点的消息传递，而这些邻居节点又出现在负样本中。为量化这种冲突的程度，本文计算每个节点的异常值。异常值的计算主要基于与目标节点相邻的节点，因为这些邻居节点对目标节点的嵌入表示产生影响，从而影响对比学习的效果。

设U_n为与目标节点 u_i相关的视图内负样本集合，其中U_n^+表示与目标节点u_i相邻的样本集。所以，当本文给定目标节点u_i，U_n^+内的冲突可以通过这些邻居节点关于x_i的影响来量化。同时，本文考虑到欺诈节点为了隐藏和其他欺诈节点的连接，通常会连接大量的正常节点导致节点度异常。为了更好的防止正常节点的特征混淆欺诈节点的特征，本文将节点度数也纳入异常值计算的考虑当中，最终节点的异常值由每一条边的异常值的和构成，计算公式如(5)。

$$C=\sum_{j\in U_n^+}S(u_i,u_j)\frac{e_{ij}}{\sqrt{d_i d_j}}\cdot(1+d_i)\#(5)$$

3.2.2 基于异常值的图重构

为了获得节点的异常值，本文首先使用$GNN(A,X;\theta)$作为异常值计算器来获得嵌入，然后将其输入到编码器中。并且，异常值计算器与图编码器$GNN(A,X;\theta)$具有相同的参数关联。最终，计算异常值过程如(6)。

$$C=F(GNN(A,X;\theta),A)\#(6)$$

其中，$F(\cdot)$表示公式(5)的处理过程。在处理视图内负样本冲突时，图重构方法通过调整邻接矩阵以减弱这些负样本对目标节点的负面影响。当节点的异常值高于设定阈值时，将对节点与其邻居节点之间的边权重进行调整。具体而言，通过调整图的邻接矩阵A来改变边的权重，有效降低与高冲突负样本相关的边的权重。对于一条边(u_i,u_j)，如果其量化指标$C(u_i,u_j)$超过设定的阈值，这表明该边在优化过程中引入了显著的冲突，可能会对目标节点的学习方向产生误导。在这种情况下，本文通过调整邻接矩阵来减少该边的权重：

$$A'_{ij}=A_{ij}\cdot\left(1-\gamma\cdot proj\left(\mathrm{C}(u_i,u_j)\right)\right)\#(7)$$

这里，$C(u_i,u_j)$表示通过归一化处理，边(i,j)相对于所有边的冲突强度，参数γ控制权重的调整幅度，确保仅对冲突性强的边施加足够的惩罚。

$proj$函数将冲突映射到 0 到 1 中。这样不仅可以降低高冲突负样本边的影响，也使得模型在聚合特征时能够更好地聚焦于重要的正样本特征。

另外，为了抵消梯度引导的图重构在消息传递过程中削弱信息量的情况，本文引入了高阶邻居来增强编码器。将高阶结构引入模型中，从而替换原始的图形结构。在这一过程中，本文还注意削弱高阶邻居所传递的信息。引入后的邻接矩阵可以用以(8)来表示。

$$A^{'} = \sum_{i=1}^{k} A^{k} \cdot \lambda^{k} \#(8)$$

其中，$\lambda \in [0,1]$，用来控制高阶邻居信息的削弱程度。因为对比学习中的负样本范围涵盖了整个数据集，这样可以通过吸纳潜在相邻候选者来获取更精确的邻接矩阵。

3.3 基于动态数据增强的图对比学习

3.3.1 高质量节点选择

首先，基于模型对未标记节点的预测置信度来选择高质量节点。设节点v的表示为h_v，我们可以利用预测器P进行预测。对于每个未标记节点，根据其预测概率计算置信度分数。在选择高质量节点时，考虑了图数据中的类别不平衡性以及度中心性指标，进行综合评估。如果评分超过预设的固定阈值，该节点将被视为高质量节点并赋予伪标签。鉴于类别不平衡问题，正常节点和异常节点的预测评分通常具有不同的置信水平。因此，为每个类别分配独立的固定阈值，以确保有效选择正常节点和异常节点。最终计算公式如(9)。

$$S(v) = \mu \cdot Conf(v) + (1-\mu) \cdot Central(v) \#(9)$$

通过这种方式，高质量节点的选择能够动态适应数据分布的变化，从而为后续训练提供更稳定的伪标签。

3.3.2 标签不变性和嵌入多样性

标签不变性指标用于衡量经过数据增强后，未标记节点的增强版本与原始版本在标签预测上的一致性。尽管数据增强可能引入一定的扰动，模型对增强数据的标签预测应尽量与原始数据保持一致。嵌入多样性指标则评估增强数据与原始数据在特征表示上的差异。我们可以通过计算增强数据与原始数据特征向量之间的距离（例如欧氏距离）来衡量增强的多样性。理想情况下，一个有效的数据增强方法应在保持标签一致性的同时，尽可能引入更多的数据多样性。

3.3.3 动态数据增强

在本文的图对比学习框架中，为了有效生成增强后的高质量节点表示，本文引入了一种动态的数据增强模块。其网络结构如图 3 所示。

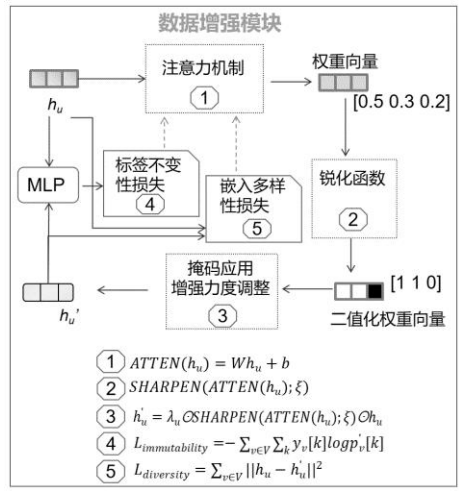

Figure 3. Dynamic Augmented Network

图 3.动态增强网络

对于筛选出的高质量节点，动态数据增强模块通过学习生成动态掩码，以在增强过程中选择性地保留特定特征维度，从而实现高一致性和高多样性的特征表示。具体来说，给定一个高质量节点u及其特征表示h_u，本文通过可学习的掩码机制生成其增强版本$h_u^{'}$，其表示为(10)。

$$h_u^{'} = f(h_u;\theta) \#(10)$$

其中，$f(\cdot;\theta)$是由参数 θ 控制的可学习掩码函数。为了有效选择适合增强的特征维度，本文利用注意力机制计算每个特征维度的重要性权重。具体实践中，注意力机制输出一个与输入向量h_v维度相同的向量，用于指示每个维度的重要性。

然后，本文应用锐化函数，将这些权重调整为 0 和 1，从而选择性地保留特征。在这个过程中，注意力函数通过计算输入特征的权重向量来确定增强所需保留的特征维度。权重的锐化由 SHARPEN 函数实现，该函数通过设定一定的比例ξ将权重中较小的部分截断为零，以保证特征维度的选择性保留。λ_u随着节点特征的变化而动态调整并作为模型的一部分通过反向传播进行优化。在每次反向传播时，λ_u的值会根据损失函数的梯度自动更新，从而有效地调整特征增强的力度。该过程表示如(11)。

$$h_u^{'} = \lambda_u \odot SHARPEN(ATTEN(h_u);\xi) \odot h_u \#(11)$$

其中，$ATTEN(\cdot)$表示注意力函数。将待增强的表示作为输入，输出一个相同维度的向量用来表示每个维度的重要程度。该注意力模块的简化形式可以表示为$ATTEN(h_u) = Wh_u + b$，其中$W \epsilon R^{d*d}$和$b \epsilon R^d$分别为权重矩阵和偏置向量。

此外，该锐化函数的实现不会截断梯度，从而支持模型的端到端训练，模型的各个部分紧密耦合，数据从输入到输出的路径是连续的，没有人为干预的中间步骤。这种方式使得模型能够在整体目标上达到最佳效果，而不需要人为干预或手动调整中间过程。增强后的节点表示用于计算标签不变性损失和嵌入多样性损失，这两者共同指导数据增强策略的优化。本文定义标签不变性损失为增强视图与其伪标签间的差异为(12)。

$$L_{immutability} = -\sum_{v \in V}\sum_{k} y_v[k] log p_v^{'}[k] \#(12)$$

其中，y_v是为节点v分配的伪标签，$p_v^{'}$是其增强后的预测概率。本文希望通过最小化该损失，使增强前后的预测保持一致。同时，多样性损失用于鼓励增强后的表示在特征空间中具备足够的变异性。本文通过原始和增强表示之间的距离(13)来表示多样性损失。

$$L_{diversity} = \sum_{v \in V} \left|\left| h_u - h_u^{'} \right|\right|^2 \#(13)$$

通过最小化该损失，模型能够在特征表示中发现更多潜在模式。最终，本文的总损失函数是这两者的加权和(14)。

$$L_{total} = \alpha L_{immutability} + \beta L_{diversity} \#(14)$$

其中，α和β是损失的加权系数，用于平衡一致性和多样性在优化过程中的影响。

通过可学习的数据增强模块，本文实现了增强样本的自适应生成，保证了模型在对比学习中的鲁棒性和泛化性。与传统的静态增强方法不同，本文的方法能够动态调整增强策略，以适应数据特性的变化，提供更有效的图对比学习性能。

3.3.4 图对比学习

为了有效利用大量的无标签节点，本文采用了图对比学习与有监督学习相结合的训练方法，旨在同时学习节点的特征和结构信息。

首先在无标签数据上应用图对比学习，通过生成伪标签来指导模型训练。模型先通过评估未标记节点的特征表示，生成弱预测结果。然后根据预设的阈值，将这些无标签节点分类为伪正样本和伪负样本，并为其生成伪标签。

然后，采用动态数据增强方法生成更加合适的增强图数据。在训练过程中，本文先冻结图对比学习模型的参数，仅对动态数据增强模块进行优化。通过反向传播，调整模型以最小化总损失函数。这个损失函数结合了标签不变性损失和嵌入多样性损失，并加入了$L2$正则化项，以控制模型的复杂性。

在对动态数据增强中的参数进行学习后，解冻图对比学习模型的参数，同时冻结动态数据增强模块的参数。接着，结合动态数据增强模块，对增强视图进行前向传播，以获得增强后的嵌入。通过对比增强后的嵌入和原始嵌入，模型能够学习到更加丰富且鲁棒的节点表示。欺诈行为通常表现为异常的节点活动或不寻常的关系模式，图对比学习能够有效捕捉这些隐蔽特征，通过最大化同一节点在不同视图下的相似性，提升模型对欺诈行为的识别能力。

此外，还在图对比学习中融入了有监督学习。模型在有标签数据上进行前向传播，可以获得模型对这些有标签节点的预测概率。通过计算这些概率与对应标签之间的交叉熵损失，模型得以学习不同类别节点的特征模式。将图对比损失与交叉熵损失结合，形成最终的损失函数(15)。

$$L = \mu_u L_{InfoNce} + \mu_L L_{cross-entropy} \#(15)$$

这种训练方法使得图对比学习与有监督学习在同一训练框架下协同工作，确保模型不仅能够利用无标签数据的潜在信息，还能从有标签数据中获取明确指导，从而提升模型在欺诈检测任务中的表现。

上述过程在整个训练数据集上反复迭代，每次迭代都经过前向传播、损失计算、反向传播和梯度更新，逐步优化模型。随着训练的进行，模型参数不断更新，最终使得模型在任务上的表现达到最优。

4 实验

4.1 实验数据集

本文使用三个常用的公开数据集和一个新的私有数据集进行实验。这三个公开数据集分别为亚马逊[20]，YeclpChi[21]和 T-Finance[22]。

Yelp Chi 数据集包含 Yelp 平台的酒店和餐厅评论数据，包括带标签的垃圾评论和合法评论。每个节点具有 32 个手工设计的特征。节点之间

有三种关系：(1) R-U-R 连接同一用户发布的评论；(2) R-S-R 连接同一产品下的相同星级评论；(3) R-T-R 连接同一产品在同一月份发布的评论。

亚马逊数据集由用户的产品评论组成，其中有用性投票超过 80%的用户被标记为良性用户，而低于 20%的用户则被标记为欺诈用户。该图同样包含三种关系：1) U-P-U 连接评论一个相同产品的用户；2) U-S-V 连接在一周内至少有一个相同星级评分的用户；3) U-V-U 连接用户与前 5%互评文本相似度较高的用户。

T-Finance 数据集旨在识别交易网络中的异常账户。每个节点具有与注册天数、日志活动和交互频率相关的 10 维特征，边表示有交易记录的账户之间的关系。如果节点被标记为欺诈、洗钱或在线赌博等类别，人类专家会将其注释为异常。

Table 1 Datasets

表 1 数据集

数据集	节点数	边数	特征数	欺诈节点比例
Amazon	11944	4398392	25	6.87%
Yelpchi	45954	3846979	32	14.53%
T-Finance	39357	3846979	32	14.53%
FDCompCN	5317	7407	57	10.51%

此外，本文还引入了一个新的舞弊检测数据集 FDCompCN[23]。FDCompCN 包含部分公司的财务报表舞弊信息，数据来源于股票市场与会计研究（CSMAR）数据库，样本涵盖 2020 年至 2023 年期间在上海等证券交易所交易的 5317 家上市公司。FDCompCN 共有三种关系：(1) C-I-C 连接有投资关系的公司；(2) C-P-C 连接公司与其披露的客户；(3) C-S-C 连接公司与其披露的供应商。每家公司包含基本信息和财务报表信息，其中基本信息包括注册资本、货币、经营状况、公司类型、行业、城市、人员规模和参保人数等；财务报表信息则包括长期应收账款、长期负债、总资产等。根据文献[24]，财务报表舞弊包括中国监管机构披露的 7 类违规行为，如虚增利润、虚增资产、虚假陈述、延迟披露、遗漏重大信息、欺诈性披露和一般会计违规。违规次数超过 3 次的公司被标记为欺诈样本，其余公司标记为良性样本。最终，数据集包含 559 个舞弊样本和 4758 个良性样本，舞弊样本占比 10.51%。表 1 汇总了四个数据集的信息。

4.2 实验结果评估

本文实验主要专注于有限监督设置。根据数据集大小，我们将 Amazon、Yelp Chi、T-Finance 和 FDCompCN 的训练比例设定为 1%，在所有场景中，将剩余数据进行测试，同时将所有数据作为无标签用于图对比学习训练。考虑到不平衡的设置，本文使用三个指标，即 AUROC，AUPRC 和 Macro F1，来综合评估模型的性能。这些指标的取值范围均为 0~1，分值越高，说明绩效越好。本文选择的超参数如表 2 所示。

Table 2 Hyperparameter Selection for Four Experimental Datasets

表 2. 4 个数据集上的超参数设置

	Amazon	Yelpchi	T-Finance	FDCompCN
learning rate	0.001	0.001	0.001	0.001
batch size	32	128	128	32
weight-decay	0.0001	0.0001	0.0001	0.0001
normal-th	5	7	5	6
fraud-th	85	88	88	87
hidden-dim	64	64	64	64

本文和每种对比方法报告 5 次独立运行的平均得分，结果如表 3 和表 4 所示。

Table 3 Comparison on Amazon and Yelpchi Datasets

表 3 Amazon 和 Yelpchi 数据集上的检测性能对比

	Amazon			Yelpchi		
	AUC	AP	F1	AUC	AP	F1
MLP	91.82	78.45	86.97	71.45	30.86	61.04
GCN	86.63	47.51	69.40	53.59	16.42	34.78
GAT	79.31	44.92	62.88	69.68	28.25	60.44
CARE-GNN	88.79	48.62	75.03	71.89	30.56	60.37
GDN	90.43	78.62	88.47	75.68	38.42	62.13
BWGNN	84.82	65.79	81.98	77.94	40.70	66.15
GHRN	84.59	61.23	80.24	76.28	37.66	65.51
Ours	**93.84**	**83.52**	**90.27**	**82.86**	**46.98**	**70.07**

Table 4 Comparison on T-Finance and FDCompCN Datasets

表 4 T-Finance 和 FDCompCN 数据集的检测性能

	T-Finance			FDCompCN		
	AUC	AP	F1	AUC	AP	F1
MLP	91.89	51.64	81.72	56.63	11.22	42.28
GCN	88.26	52.26	77.15	51.89	9.25	40.26
GAT	86.13	45.70	76.18	50.68	11.83	38.20
CARE-GNN	90.38	72.04	82.33	65.18	23.77	49.00
GDN	87.32	52.96	75.28	63.92	18.56	62.13
BWGNN	92.61	77.80	86.64	62.79	20.09	50.41

GHRN	90.79	64.62	80.07	64.28	19.55	48.75
Ours	**97.24**	**83.06**	**89.77**	**69.49**	**27.76**	**63.58**

4.3 消融实验

本节通过消融实验来证明图重构、动态数据增强两种策略的有效性。我们用 A、B、C、D 分别代表图对比学习、图重构、可学习数据增强和传统图数据增强。在保持模型其余部分不变的情况下，测试每种组合的效果。结果如表 5 所示。

Table 5 Ablation Experiment Results

表 5 消融实验结果

	AUC			
	Amazon	Yelpchi	T-finance	CompCN
A+D	80.89	72.64	76.72	52.63
A+B	85.26	76.26	88.15	56.89
A+C	88.13	78.70	92.19	60.68
A+B+D	90.32	79.73	94.79	66.46
A+B+C	**93.84**	**82.86**	**97.24**	**69.49**

可以看出，去除任一部分都会导致性能下降，这突出了它们对欺诈检测的协同作用。

另外，如表 6 所示为模型不同的标签数据比例下的表现情况，可以看出随着标签数据比例的增加，分类性能逐渐提升，但比例提升的边际效用不断降低。

Table 6 Results with Different Training Data Percentages

表 6.不同训练数据百分比结果展示

	AUC			
Training Ratio	Amazon	Yelpchi	T-finance	CompCN
1%	93.84	82.86	97.24	69.49
5%	94.59	86.83	97.58	71.93
10%	95.27	90.96	98.07	73.04
15%	95.32	91.27	97.95	73.08

Table 7 Ablation Study Results for the Dynamic Augmentation Component

表 7.动态增强部分消融实验结果展示

	AUC			
	Amazon	Yelpchi	T-finance	CompCN
No-immutability	89.57	78.63	90.18	66.83
No-Diversity	91.88	80.25	94.36	67.45
Our method	**93.84**	**82.86**	**97.24**	**69.49**

我们还分析了标签不变性和嵌入多样性的重要性，如表 7 所示。可以看出，两个损失互相协作，共同提升欺诈检测的效果，其中保持标签不变性更加重要。

总之，本文提出的模型在结果上表现优异，相较于其他先进的模型，效果提升了 2%-5%。这主要归因于基于异常值的图结构优化和使用动态数据增强的图对比学习部分。通过消融实验，本文也验证了各个部分的效果，去除任一部分都会导致性能下降，这突出了它们对欺诈检测的协同影响。

5 总结

本文提出了一种基于图对比学习的新方法，以应对欺诈检测中面临的两个主要挑战：一是图神经网络在聚合过程中恶意节点特征被淡化的问题，二是传统图数据增强方法可能破坏关键结构信息的问题。针对这些挑战，本文采用异常值计算优化图结构，从而缓解消息传递与图对比学习之间的冲突。此外，本文引入了一种动态数据增强策略，以确保增强过程不改变节点的关键特征，同时保证增强的强度，使图对比学习能够充分挖掘节点的有用特征。实验结果表明，本文提出的方法在多个数据集上显著优于现有的最先进方法，验证了该方法在提高图欺诈检测模型准确性和鲁棒性方面的有效性。

未来，将缓解冲突的手段与图数据增强结合，通过更有效的可学习增强实现图神经网络与图对比学习的有机统一，可能成为一个具有潜力的研究方向。

参考文献：

[1] Ali A, Abd Razak S, Othman SH, et al. Financial Fraud Detection Based on Machine Learning: A Systematic Literature Review[J]. Applied Sciences, 2022, 12(19): 9637.

[2] Duan M, Zheng T, Gao Y, et al. DGA-GNN: Dynamic Grouping Aggregation GNN for Fraud Detection[C]//Proc of the AAAI Conference on Artificial Intelligence, 2024: 11820-11828.

[3] Wang Y, Zhang J, Huang Z, et al. Label Information Enhanced Fraud Detection against Low Homophily in Graphs[C]//Proc of the ACM Web Conference 2023 (WWW'23), 2023: 1-11.

[4] Dou Y, Liu Z, Sun L, et al. Enhancing graph neural network-based fraud detectors against camouflaged fraudsters[C]//Proc of the 29th ACM International Conference on Information & Knowledge Management, 2020: 315-324.

[5] Yuan D, Miao Y, Gong N Z, et al. Detecting fake accounts in online social networks at the time of registrations[C]//Proc of the 2019 ACM SIGSAC Conference on Computer and Communications Security, 2019: 1423-1438.

[6] Zhang M, Meng W, Lee S, et al. All your clicks belong to me: investigating click interception on the web[C]//Proc of the 28th USENIX Security Symposium, 2019: 941-957.

[7] Veličković P. Everything is connected: Graph neural networks[J]. Current Opinion in Structural Biology, 2023, 79: 102538.

[8] Zhu X, Liu J, Wang G, et al. GraphFC: Customs Fraud Detection with Label Scarcity[J]. arXiv preprint arXiv:2305.11377, 2023.
[9] Liu Z, Dou Y, Yu P S, et al. Alleviating the inconsistency problem of applying graph neural network to fraud detection[C] // Proc of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020: 1251-1260.
[10] Feng W, Zhang J, Dong Y, et al. Graph random neural networks for semi-supervised learning on graphs[J]. Advances in Neural Information Processing Systems, 2020, 33: 22092-22103.
[11] Wang M, Zheng D, Ye Z, et al. Deep graph library: A graph-centric, highly-performant package for graph neural networks[J]. arXiv preprint arXiv:1909.01315, 2019.
[12] Velickovic P, Fedus W, Hamilton W L, et al. Deep Graph Infomax[C]//Proc of the International Conference on Learning Representations (ICLR), 2019.
[13] Chen M, Chang Z, Lu H, et al. Augnet: End-to-end unsupervised visual representation learning with image augmentation[J]. arXiv preprint arXiv:2106.06250, 2021.
[14] Thakoor S, Tallec C, Azar M G, et al. Bootstrapped representation learning on graphs[C]//Proc of the International Conference on Learning Representations 2021 Workshop on Geometrical and Topological Representation Learning, 2021.
[15] Zhang H, Wu Q, Yan J, et al. From canonical correlation analysis to self-supervised graph neural networks[J]. Advances in Neural Information Processing Systems, 2021, 34: 76-89.
[16] Xia J, Wu L, Wang G, et al. ProGCL: Rethinking Hard Negative Mining in Graph Contrastive Learning[C]//Proc of the International Conference on Machine Learning, 2022: 24332-24346.
[17] You Y, Chen T, Sui Y, et al. Graph contrastive learning with augmentations[C]//Proc of Advances in Neural Information Processing Systems, 2020: 5812-5823.
[18] Zhang Y, Zhu H, Song Z, et al. Spectral feature augmentation for graph contrastive learning and beyond[C]//Proc of the AAAI Conference on Artificial Intelligence, 2023, 37: 11289-11297.
[19] Liu Y, Jin M, Pan S, et al. Graph self-supervised learning: A survey[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 35(6): 5879-5900.
[20] McAuley J J, Leskovec J. From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews[C]//Proc of the 22nd International Conference on World Wide Web, 2013: 897-908.
[21] Rayana S, Akoglu L. Collective opinion spam detection: Bridging review networks and metadata[C]//Proc of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015: 985-994.
[22] Tang J, Li J, Gao Z, et al. Rethinking graph neural networks for anomaly detection[C]//Proc of the International Conference on Machine Learning, 2022: 21076-21089.
[23] Wu B, Yao X, Zhang B, et al. SplitGNN: Spectral Graph Neural Network for Fraud Detection against Heterophily[C]//Proc of the ACM International Conference on Knowledge Discovery and Data Mining, 2023: 2737-2746.
[24] Liao L, Chen G, Zheng D. Corporate Social Responsibility and Financial Fraud: Evidence from China[J]. Accounting & Finance, 2019, 59(5): 3133-3169.