文章编号:1671-4598(2025)07-0234-09 DOI:10.16526/j.cnki.11-4762/tp.2025.07.029 中图分类号:TP3 文献标识码:A

基于网络增强联合稀疏典型相关分析的异常节点检测模型设计

高 燕[1],崔鸿雁[2],刘 诚[2],符佳慧[2]
(1.中国电子科技集团公司 第三十研究所,成都 610000;
2.郑州信大先进技术研究院,郑州 450000)

摘要:文章对属性网络中的异常节点检测问题进行了研究,通过提出一种基于网络增强联合稀疏典型相关分析的异常节点检测模型(NEAnomSCCA),进行了网络结构增强、特征学习、重构误差与典型向量计算以及异常节点检测的综合分析;该模型首先利用k近邻算法生成每个节点的KNN图,以增强网络结构并补全离散节点;随后,采用图神经网络学习节点的属性特征和结构特征;接着,计算重构误差和典型稀疏向量,综合评估节点的异常程度;最终,根据重构误差和典型稀疏向量的相关性计算异常得分,检测出异常节点;实验结果表明,在BlogCatalog、Cora、Citeseer、Pubmed四个现实世界的数据集上,NEAnomSCCA模型与JAANE、ARISE、PROPOSED等模型进行对比,AUC值分别提高了1.13%、1.35%、1.11%、3.33%,显示出该模型在异常节点检测任务上的优越性能;得出了NEAnomSCCA模型通过结合网络增强和稀疏典型相关分析,能够有效提升属性网络异常节点检测的准确性和鲁棒性的结论。

关键词:属性网络;异常节点检测;稀疏典型向量分析;网络增强;图神经网络

Design of Anomaly Node Detection Model Based on Network-Enhanced Joint Sparse Canonical Correlation Analysis

GAO Yan[1], CUI Hongyan[2], LIU Cheng[2], FU Jiahui[2]
(1. The 30th Research Institute of China Electronics Technology Group Corporation, Chengdu 610000, China;
2. Zhengzhou Xinda Advanced Technology Research Institute, Zhengzhou 450000, China)

Abstract: Study on the anomaly node detection in attribute networks was conducted. An anomaly node detection model based on network-enhanced joint sparse canonical correlation analysis (NEAnomSCCA) was proposed, and a comprehensive analysis of the network structure enhancement, feature learning, reconstruction error and canonical vector calculation, as well as anomaly node detection was made. Firstly, this model uses the k-nearest neighbor algorithm to generate k-nearest neighbor (kNN) graphs for each node, and to enhance the network structure and complete the discrete nodes; Then, graph neural networks are used to learn the attribute and structural features of nodes; Next, the reconstruction error and typical sparse vectors are calculated to comprehensively evaluate the degree of node anomalies; Finally, the correlation between the reconstruction error and typical sparse vectors are reconstructed to calculate anomaly scores and detect abnormal nodes. Experimental results show that compared with models such as JAANE, ARISE, and PROPOSED on four real-world datasets of BlogCatalog, Cora, Citeseer, and Pubmed, the NEAnomSCCA model has improved the AUC by 1.13%, 1.35%, 1.11%, and 3.33%, respectively , with the superior performance of this model in detecting abnormal nodes. The results show that the NEAnomSCCA model can effectively improve the accuracy and robustness of attribute network anomaly node detection by combining the network enhancement and sparse canonical correlation analysis.

Keywords: attribute network; anomalous node detection; sparse canonical vector analysis; network enhancement; graph neural network (GNN)

收稿日期:2025-06-23; **修回日期**:2025-06-30。

作者简介:高 燕(1981-),女,大学本科,高级工程师。

引用格式:高 燕,崔鸿雁,刘 诚,等.基于网络增强联合稀疏典型相关分析的异常节点检测模型设计[J].计算机测量与控制,2025,33(7):234-242.

0　引言

属性网络是指网络中的节点含有属性信息的网络，随着互联网和社交媒体的迅猛增长，现实世界的属性网络如社交媒体网络、交通网络、金融交易网络和通信网络等，节点数量愈加庞大，关系更为复杂，节点本身携带的属性信息也更加丰富；同时这些现实中的属性网络也面临多种潜在威胁和风险，例如，社交媒体上的假信息传播、交通网络拥堵的异常现象、金融欺诈、通信中的电信诈骗等。属性网络异常节点检测技术，旨在检测出与大多数正常节点显著偏离的异常节点，能够有效应对现实世界属性网络面临的威胁。

传统的异常节点检测方法通过根据专业人士的经验制定具体的检测规则、聚类、机器学习等方式检测属性网络中的异常节点。然而，传统方法存在特征提取能力不足、未能充分利用网络的拓扑结构等问题，造成检测效果不佳。

近年来图神经网络（GNN，graph neural network）的发展，为解决属性网络异常节点检测问题提供了新思路。根据 Ding[1] 等人的研究，采用 GNN 提取特征，检测异常节点，能够充分利用属性网络的结构和属性信息，提升检测性能。

为了解决上述问题，本文提出基于网络增强联合稀疏典型相关分析的异常节点检测模型（NEAnomSCCA，anomaly node detection model based on network-enhanced joint sparse canonical correlation analysis）简称 NEAnomSCCA。该模型首先基于特征计算每个节点 k 近邻（KNN，k-nearest neighbor）图。kNN 图的特性可以保证每个节点都至少有 k 个邻居，从而保障了节点之间的连通性，增强属性网络的结构特征，对于离散点而言，这种连通性能有效促进节点间的信息传播，极大改善模型性能；同时添加随机特征，提升模型的泛化能力。然后，通过图神经网络提取属性网络的结构特征和属性特征，并计算重构误差和稀疏典型向量。重构误差能够确保模型检测出网络中大多数的异常，稀疏典型向量能够处理高维的稀疏数据，检测出特定的异常。最后，综合每个节点的重构误差和典型稀疏向量的相关性，计算异常得分，检测异常节点。

网络增强通过 KNN 图补全离散结构，通过添加随机特征增强泛化能力，为后续分析奠定良好基础。这使得稀疏典型相关分析能在更准确的特征数据上进行，其计算的典型向量可精准捕捉节点属性和结构相关性，弥补重构误差检测局限，从多维度提升检测效果。

1　异常节点检测方法

1.1　传统异常节点检测方法

在传统的网络异常节点检测中，常用诸如决策树算法[2] 检查异常节点，根据专业人士的经验制定具体的检测规则，利用这些规则检测异常节点，但这些规则面对多变和伪装的异常模式时常显不足。因此，研究者提出了基于统计学[3] 和聚类[4] 的方法。前者如基于直方图的离群值评分（HBOS，histogram-based outlier score）[5]，通过拟合数据分布来建立统计模型，从而区分正常与异常数据。后者，如基于聚类的局部离群因子算法（CBLOF，cluster-based local outlier factor）[6]，则通过数据聚类和计算节点到聚类中心的距离来评估异常度。此外，随着机器学习的发展，出现了基于机器学习的方法[7]，这些方法通过特征提取和数据维度降低[8]，进而使用分类器进行有效分类，以识别属性网络中的异常节点。

传统方法在低维和简单网络结构的异常节点检测方法虽然取得了一定的成效。然而，随着网络数据维度的增加，这些方法在特征提取方面的能力显得不足，无法有效识别高维数据中的复杂模式；其次，早期方法未能充分考虑网络的拓扑结构，忽视了节点间交互引起的异常模式，导致检测性能不佳。

1.2　图神经网络节点检测方法

图神经网络天然适用于处理属性网络数据，具备强大的网络特征提取和学习能力，从而显著提高了异常节点检测的准确性和效率。Ding 等人[1] 提出 DOMINANT 模型，首次采用 GNN 对异常节点进行检测。该方法通过编码器压缩网络信息，并使用属性与结构解码器重构网络信息，借此计算重构误差以识别异常节点。Zhu 等[9] 提出了 DeepAE 嵌入模型，该模型在嵌入过程中引入拉普拉斯锐化技术，放大正常节点与异常节点间的差异，使得编码器尽量忽略异常特征。Luo 等[10] 中提出了 ComGA 模型，针对基于 GNN 的异常检测方法存在的挑战，设计了一个关注社区感知的定制深度图卷积网络，并提出了一个新颖的社区感知属性图异常检测框架。Pei 等[11] 提出的 ResGCN 模型，通过图卷积网络（GCN，graph convolutional networks）对属性网络进行建模，有效地捕捉网络的稀疏性和非线性特征。FAN 等[12] 提出的 AnomalyDAE 模型，通过将图注意力网络（GAT，graph attention networks）应用于属性网络的嵌入过程中，学习不同邻居节点间的重要性。Fan 等[13] 提出 JAANE 模型，聚合过程中同时考虑目标节点及其邻居信息，融合二者学习节点的特征向量。Duan 等[14] 提出 ARISE 检测网络中的密集模式作为异常节点存在的可疑区域，接着计算节点对之间的相似度，根据平均相似度检测异常节点。Wasim 等[15] 提出 PROPOSED 基于稀疏典型分析和随机特征填充检测异常节点。

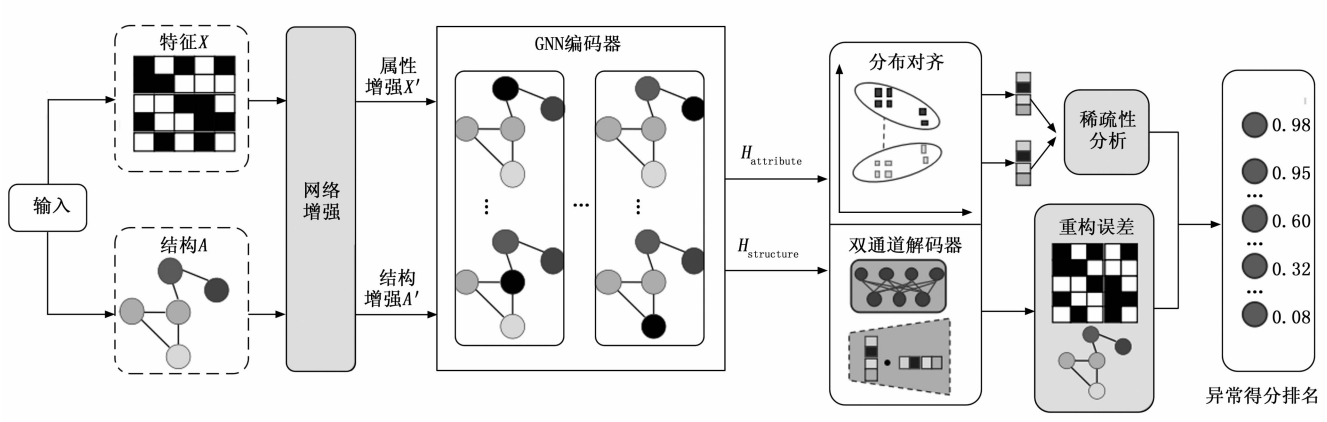

图 1 基于网络增强联合稀疏典型相关分析的异常节点检测模型

图神经网络的信息传播机制依赖于节点之间的连接关系来传递和更新特征信息。在现实世界的数据采集过程中，网络数据往往存在缺陷，离散的节点和子图较为常见。当网络中存在离散节点和子图时，信息传播无法有效地对这些不完整部分进行信息补全。例如，在一个社交网络中，如果存在一些孤立的用户节点（即离散节点），它们与其他节点没有连接关系，那么在图神经网络的信息传播过程中，这些孤立节点就无法获取到其他节点的信息来更新自身的特征表示，从而导致学习到的结构和属性特征不准确。这种不准确性会进一步影响到后续基于这些特征进行的异常节点检测任务，使得模型难以准确区分正常节点和异常节点。

此外，基于图神经网络的异常检测方法通常侧重于构建复杂的模型架构来捕获网络的整体信息。这些模型在训练过程中会尽可能地拟合训练数据中的各种模式，包括正常节点和异常节点的特征以及它们之间的关系。但是，这种设计方式并没有针对异常节点检测这一特定任务进行充分的优化。在实际应用中，异常节点在整个网络中所占比例通常较小，属于少数类。当模型过于追求对训练数据的拟合时，很容易过度学习到训练集中正常节点的特征和模式，而对异常节点的特征学习不够充分。例如，在一个金融交易网络中，正常的交易行为占绝大多数，异常的欺诈交易行为相对较少。基于图神经网络的模型可能会因为过度拟合正常交易的模式，而在面对新的、具有不同特征的欺诈交易（即异常节点）时，无法准确识别，导致在新的数据集上识别异常节点的效果不佳，表现出过拟合的风险。

2 基于网络增强联合稀疏典型相关分析的异常节点检测模型

本文提出一种基于网络结构增强联合稀疏典型相关分析的异常节点检测模型。如图 1 所示。模型主要包含四个部分：网络结构增强、网络特征学习、计算重构误差和典型向量、异常节点检测。首先，计算每个节点的 KNN 图，增强网络的结构特征；然后，使用图神经网络模型学习节点的属性特征和结构特征；接着根据学习到的特征计算重构误差和典型稀疏向量；最后结合重构误差和典型稀疏向量的相关性计算异常得分，检测异常节点。

步骤 1：网络增强。

首先，基于每个节点的特征值 X_i，采用 K 临近算法，计算出和节点 v_i 最相似的 k 个节点，得到节点 v_i 的 knn 图。然后，根据每个节点的 knn 图补全网络的结构信息。同时，采用高斯分布为节点添加随机特征，得到最终的增强网络 $G'=(A', X')$。

步骤 2：网络特征学习。

将结构增强后的增强视图 G'输入到图神经网络模型中进行训练，学习属性网络的结构特征和属性特征，得到学习后的结构特征 $H_{\text{structure}}$和属性特征 $H_{\text{attribute}}$。

步骤 3：重构误差和典型特征向量的计算。

根据步骤 2 学习得到的结构特征 $H_{\text{structure}}$、属性特征 $H_{\text{attribute}}$以及网络 G 的结构特征 A、属性特征 X，计算出网络的结构重构误差 $\hat{A}$ 和属性重构误差 $\hat{X}$；同时基于学习到的结构特征 $H_{\text{structure}}$、属性特征 $H_{\text{attribute}}$，使用稀疏典型相关分析方法计算获得结构典型向量 $\boldsymbol{U}$ 和属性典型向量 $\boldsymbol{V}$。

步骤 4：异常节点检测。

首先根据步骤 3 获得的结构典型向量 $\boldsymbol{U}$ 和属性典型向量 $\boldsymbol{V}$，计算出每个节点的典型向量的偏差 e_i；接着根据每个节点 v_i 的属性重构误差 $\hat{X}_i$、结构重构误差 $\hat{A}_i$ 以及典型向量偏差 e_i 计算 v_i 异常值得分 s_i；最后根据每个节点 v_i 的异常值得分 s_i 的排名检测异常节点。

2.1 网络增强

离散节点阻碍了属性网络的信息传播，为了解决这个问题，一种比较有效的做法是引入一个新的图结构，

确保所有节点都互相连通。然而，直接用新的图结构替换原有的图结构会不可避免地造成信息丢失。因此，本文提出一个在原始网络结构的基础上，采用 k 邻近的方法生成一个新的增强图结构，简称为 knn 结构增强，增强后的图称为 knn 图。模型在训练期间利用新的网络结构进行信息传播。具体过程如下：

首先，基于特征矩阵 $\boldsymbol{X}$，计算和每个节点 v_i 最相似的 k 个节点，这样就确保每个节点至少有 k 个邻居节点，从而使得整个图结构不存在离散的节点。接着，对比每个节点 v_i 在原图中的邻居节点和新生成的 k 个邻居节点，若原图的邻居节点大于等于 k 个，则不添加新的节点；若小于 k 个则给目标节点 v_i 添加新的邻居节点为 k 个。具体的见公式（1）：

$$A',\text{where},A'_{ij}=\begin{cases}1, s(\boldsymbol{X}_i,\boldsymbol{X}_j)\geqslant top_k \text{ and } |v_i|<k\\ 0, \text{otherwise}\end{cases} \tag{1}$$

其中：$s(X_i, X_j)$ 表示节点 v_i 和任意节点 v_j 之间的相似度的大小，top_k 表示目标节点 v_i 和节点 v_j 之间的相似度小为前 k 名，$|v_i|$ 表示节点 v_i 的邻居数量；整个公式的含义为，若目标节点 v_i 和节点 v_j 的相似度大小为前 top_k 且 v_i 的节点数量小于 k 个，那么就在让 A'_{ij} 之间添加一条变，也即使得 $A'_{ij}=1$。

同时，加入随机噪声以减少过拟合带来的影响。随机噪声从高斯分布中抽取。对于网络中的每个 v_i，其原始特征向量 $\boldsymbol{x}_i=\{x_{i1}, x_{i2}\cdots x_{in}\}$，从高斯分布中生成一个和 x_i 同维度的向量 $\boldsymbol{r}_i=\{r_{i1}, r_{i2}\cdots xr_{in}\}$，然后将两者对应元素相加，得到添加随机特征后节点特征向量，$\boldsymbol{x}_i=\{x_{i1}+r_{i1}, x_{i2}+r_{i2}\cdots x_{in}+r_{in}\}$，这样能够引入随机性，使得模型在学习过程中，不会依赖原始特征分布，从而提高模型的泛化能力，避免过拟合。

k 值的选择采用网格搜索策略，通过在验证集上测试 $k\in\{5, 10, 15, 20\}$ 时的检测效果（以 AUC 为指标）确定最优值。如图 2 所示，当 $k=10$ 时四个数据集均达到局部最优。针对网络规模变化，设计动态调整机制：设节点密度 $\rho=N/E$（N 为节点数，E 为边数），当 $\rho>$阈值 θ 时，按公式 $k'=k\times(\rho/\rho_0)$ 缩放 k 值（ρ_0 为初始密度）。实验设定 $\theta=1.5$，确保稀疏区域获得更多连接。

2.2　特征学习

图神经网络是一种专门用于处理图结构数据的深度学习模型。它能够从图中提取特征，并在多种任务中表现出色。本文采用图神经网络学习属性网络的特征。

图神经网络的输入为：节点的特征和图结构。在本文中将增强后的网络结构 A' 和节点的特征矩阵 $\boldsymbol{X}$ 作为图神经网络的输入，二者共享参数 W_l，共享权重使模型能够学习一组统一的参数，这些参数能够有效地编码网络的属性特征和结构特征。

本文使用了两层的图神经网络，因为两层的网络不仅可以捕获每个节点的直接邻居（一阶），还可以捕获邻居的邻居（二阶），能够捕获更丰富和更复杂的网络特征。这个过程可以分为以下步骤：

首先，属性网络的特征矩阵 $\boldsymbol{X}$，经过两层图神经网络的处理后，得到最终的嵌入表示 $H'_{\text{attribute}}$，令 $H^0_{\text{attribute}}=\boldsymbol{X}$：

$$H^1_{\text{attribute}}=\sigma(\widetilde{A}\,H^0_{\text{attribute}}W^0) \tag{2}$$

$$H^2_{\text{attribute}}=\sigma(\widetilde{A}\,H^1_{\text{attribute}}W^1) \tag{3}$$

其中：$\widetilde{A}$ 是正规化的邻接矩阵，通常（$\widetilde{A}=D^{-1/2}A'D^{1/2}$），$W^0$ 和 W^1 分别是第一层和第二层的权重矩阵。经过第三层后的最终输出 $H^2_{\text{attribute}}$ 是节点属性的编码表示，记为 $H_{\text{attribute}}$。

接着，编码纯粹的结构信息，本文使用具有共享权重的相同 GNN 层，但这一次输入会有所不同，主要学习结构特征。初始化 $H^0_{\text{structure}}=A$：

$$H^1_{\text{structure}}=\sigma(\widetilde{A}H^0_{\text{structure}}W^0) \tag{4}$$

$$H^2_{\text{structure}}=\sigma(\widetilde{A}H^0_{\text{structure}}W^1) \tag{5}$$

其中：$\widetilde{A}$，W^0 和 W^1 表示的含义和公式（2）、（3）相同，经过第二层后的最终输出 $H^2_{\text{structure}}$ 是节点结构的编码表示，记为 $H_{\text{structure}}$。

为降低大规模网络的计算复杂度，采用邻居采样策略（Neighbor Sampling）：

每层随机采样 $[\log_2 N]$ 个邻居节点（N 为总节点数）

结合梯度压缩技术，将梯度张量稀疏率提升至 70% 如表 1 所示，在 Pubmed 数据集上提速 3.2 倍且 AUC 仅下降 0.4%。

表 1　计算效率对比

方法	训练时间(s/epoch)	AUC
原始 GNN	142.6	0.952 5
优化后	44.3	0.948 9

2.3　重构误差和典型特征向量计算

2.3.1　重构误差计算

为了从网络信息和属性信息的低维特征重构网络结构和节点属性，本文采用了双通道解码器。重构的过程包含以下步骤：

首先，属性解码器采用全连接的神经网络进行特征转换，重构节点属性。具体公式见公式（6）：

$$\hat{X}^l=Tanh[MLP(X^{l-1}W^{l-1},b)] \tag{6}$$

其中：$X^1=H_{\text{structure}}$，$\hat{X}^l$ 表示重构后的属性特征，$l\in\{1, 2, 3\cdots\}$ 表示解码器的层数。

接着，通过内积计算每个节点对的低维嵌入相似性，来重构网络拓扑结构的边。节点对的相似性通过属性网络的网络属性信息低维特征来进行计算。具体见公式（7）：

$$\hat{A} = \text{Sigmoid}(ZZ^T) \tag{7}$$

其中：Z 表示网络的低维属性特征，$Z = H_{\text{attribute}}$。使用（·）T 表示矩阵的转置，并使用 Sigmoid 激活函数做归一化处理，也即 Sigmoid（·）。

然后使用重构后的属性特征 $\hat{X}^l - X$ 作为属性重构误差记为 R_S，同理计算结构重构误差记为 R_A。

2.3.2 典型特征向量计算

步骤 1：KL 散度对齐。

本文使用 KL 散度对齐了节点属性 $H_{\text{attribute}}$ 和图结构 $H_{\text{structure}}$ 的潜在表示分布。主要目的是：最小化节点属性和图结构的潜在空间表示，优化异常检测的对齐分布：

$$L_{\text{KL}} = D_{\text{KL}}[P(H_{\text{structure}}) \mid P(H_{\text{attribute}})] \tag{8}$$

节点属性和图结构的潜在空间表示的概率分布分别是 P（$H_{\text{attribute}}$）和 P（$H_{\text{structure}}$），二者使用同样的共享参数进行特征压缩，因此属性和结构的低维表示在同样的特征空间，在相同的特征空间确保二者之间的可比性。该过程的下一阶段 SCCA，需要用到 KL 散度对齐的结果。

步骤 2：稀疏典型向量计算。

异常节点检测的主要依据是节点属性 $H_{\text{attribute}}$ 和图结构 $H_{\text{structure}}$ 的相关性，SCCA 可以得出二者的典型向量，再根据典型向量计算出相关性。给定两组变量 $H_{\text{attribute}}$ 和 $H_{\text{structure}}$，二者在同一特征空间，维度为 $n \times q$，其中 n 是节点数，q 是属性和结构潜在的空间维度。SCCA 的目标是确定两组典型变量，$W_{\text{attribute}}$，$W_{\text{structure}} \in R^q$，其中 q 表示节点属性和图结构的潜在空间维度，最大化节点属性 $H_{\text{attribute}}$ 和图结构 $H_{\text{structure}}$ 的相关性，同时添加稀疏性约束。优化问题可以描述为：

$$Max_{W_{\text{attribute}}, W_{\text{structure}}} Corr$$
$$(H_{\text{attribute}} W_{\text{attribute}}, H_{\text{structure}} W_{\text{structure}}) \tag{9}$$

其中：需要满足以下约束条件：

$$| W_{\text{attribute}} |_2 \leqslant 1, | W_{\text{structure}} |_2 \leqslant 1 \tag{10}$$

$$| W_{\text{attribute}} |_1 \leqslant c_{\text{attribute}}, | W_{\text{structure}} |_1 \leqslant c_{\text{structure}} \tag{11}$$

其中：L_1 范数约束 $| W_{\text{attribute}} |_1 \leqslant c_{\text{attribute}}$，$| W_{\text{structure}} |_1$。在典型向量中约束稀疏性，确保每个向量只利用其各自特征集中有限数量的重要特征。参数 $c_{\text{attribute}}$ 和 $c_{\text{structure}}$ 约束稀疏性。由于稀疏性约束，SCCA 中的优化问题可能具有挑战性，本文采用最小二乘法（ALS）进行参数优化，它在保持其他变量固定的情况下迭代优化一个变量。SCCA 的输出包括属性稀疏典型变量 $\boldsymbol{U}$ 和属性稀疏典型变量 $\boldsymbol{V}$，见公式（12）、（13）：

$$\boldsymbol{U} = H_{\text{attribute}} W_{\text{attribute}} \tag{12}$$

$$\boldsymbol{V} = H_{\text{structure}} W_{\text{structure}} \tag{13}$$

接着使用稀疏典型相关变量计算节点属性和结构的相关性，将节点属性和结构的偏差来作为检测异常节点的依据。具体计算见公式（14）：

$$e_i = 1 - \frac{U_i V_i}{| U_i | | V_i |} \tag{14}$$

其中：U_i 和 V_i 是节点 v_i 的稀疏典型变量得分，若节点结构和属性的相关性较小也即二者的偏差过大，则节点存在异常的可能性就越大。e_i 表示节点 v_i 结构和属性的偏差，e_i 越大则偏差越大。

2.4 异常节点检测

节点的异常值得分主要通过重构误差和稀疏典型向量相关性的偏差来进行计算，具体见公式（15）：

$$S_i = \alpha e_i + \beta \hat{R}_A^i + (1 - \beta)\ \hat{R}_S^i \tag{15}$$

其中：$\hat{R}_A^i$ 表示节点 v_i 的结构重构误差，$\hat{R}_S^i$ 表示节点的 v_i 的属性重构误差，e_i 为节点 v_i 属性和结构的偏差。α 和 β 用来调节稀疏典型向量相关性的偏差和重构误差的比例，α，$\beta \in$（0，1）。若 α 偏大则模型的泛化性更强，β 偏大则模型更加注重对当前数据集的检测。最终根据二者的确定每个节点的异常值得分 S_i，根据 S_i 的大小检测出属性网络中的异常节点。

2.5 损失函数

损失函数是模型训练优化的关键，是优化模型的重要部分。本文模型采用整体训练，见公式（16）：

$$L = \gamma Recon[(X, \hat{X}), (A, \hat{X})] + \eta D_{\text{KL}}(P \mid Q) + \mu(| W_{\text{attribute}} |_1) - (| W_{\text{structure}} |_1) - \lambda \text{corr}(H_{\text{attribute}} W_{\text{attribute}}, H_{\text{structure}}) W_{\text{structure}} \tag{16}$$

其中：γ，η，μ，λ 是权重参数，用来平衡损失函数中各个部分的贡献。

$Recon$［（X，$\hat{X}$），（A，$\hat{A}$）］是训练图神经网络的重构损失。D_{KL}（$P \mid Q$）是用于对齐属性和结构 KL 散度。$| W_{\text{attribute}} |_1 -$（$| W_{\text{structure}} |_1$，是 SCCA 的 L_1 正则化，约束稀疏性。

3 实验

3.1 数据集介绍

本文在四个广泛使用的属性网络异常检测基准数据集上进行了实验，分别是 BlogCatalog、Cora、Citeseer、Pubmed。

BlogCatalog：一个博客分享网站，节点表示网站用户，链接表示用户之间的关注关系。节点的属性由用户的个性化内容组成，例如用户发布博客、分享图片的标签、自定义的标签等。

Cora：该数据集由 2 708 篇论文组成，用 Node 表

示。有 5 429 个引用链接，用 Edges 表示。每篇论文的属性 Attributes 指示该论文是否具有某个关键词，属性的维度为 1 433。

Citeseer：该数据集由 3 327 篇科学论文组成，用 Node 表示。有 4 732 个引用链接，用 Edges 表示。每篇科学论文的属性 Attributes 的维度为 3 703。

Pubmed：该数据集由 19 717 份出版物组成，节点之间形成 44 338 个引用链接。每个出版物的属性 Attributes 的维度为 500。

实验数据集构建：

以上 4 个数据集的节点属性都使用词袋模型进行了向量化处理，每个节点的属性都由向量表示，向量维度的大小由词袋模型的字典大小确定。

由于上述 4 个数据集没有标注异常节点，并且目前也没有统一的异常节点标注标准，本文采用目前使用广泛的结构扰动和属性扰动异常注入方法[1,16-17]等。

通过扰动网络拓扑结构的方式注入结构异常节点。这种方法背后的原理为：在很多现实场景中，子图中很少有节点的连接是全连接结构，因此全连接结构被视为异常。首先，在网络中随机挑选 m 子图，每个子图的节点数量为 n；接着，将 m 个子图中的 n 个节点完全连接，形成全连接结构。这些子图中全连接的节点被视为结构异常节点[18]。

通过扰动节点属性的方式生成属性异常节点。用网络中和目标节点欧氏距离最大的节点对目标节点的属性进行扰动。首先，从网络中随机选择 $m\times n$ 个节点作为候选属性异常节点；接着，对候选节点中的每个目标节点，从除候选节点的数据中随机选择 k 个节点，计算目标节点和 k 个节点之间欧式距离，找出和目标节点欧式距离最大的节点 j；然后将目标节点的属性替换为节点 j 的属性；最后将 $m\times n$ 个节点生成为属性异常节点。

根据以上注入方法并且按照网络的规模，每个数据集注入网络规模 5%左右的异常节点。最终获得了扰动网络，并将异常的总数列在表 1 的最后一行中。在本实验中，所有的类别标签都被删除了，只有在推理阶段才能看到异常标签。

3.2　实验指标

本文采用曲线下面积（AUC，area under the curve）作为评估指标[1,16-17]等，来评价模型的性能。

AUC 是一种常见的用于评估二分类模型性能的指标，它代表了 ROC 曲线下的面积。其物理含义为：给定随机的正例和负例，模型将正例预测为正例的概率比将负例预测为正例的概率高的概率。因此，AUC 值越大，说明模型具有更好的性能，也即模型在区分正例和负例方面更为准确。

本文使用运行环境为 Windows 10，python 3.8，pytorch 1.1，硬件环境为 NVIDIA GeForce 显卡，显存 16 GB。

为了保证实验结果的公平性，本模型和对比实验采用相同的环境和设置：优化器采用 Adam，学习率取 0.01，节点最终嵌入维度设为 64，图神经网络的层数设置为 2 层，dropout 取值为 0.3。除以上公共参数外，对比实验的其他参数保持原文中的设置。

3.3　GNN 编码器选择

本小节分析不同 GNN 编码器对模型的影响，主要探究了常见图神经网络编码器 GCN，GAT，GIN 对模型性能的影响。在四个数据集上进行了实验验证，实验结果如图 2 所示。

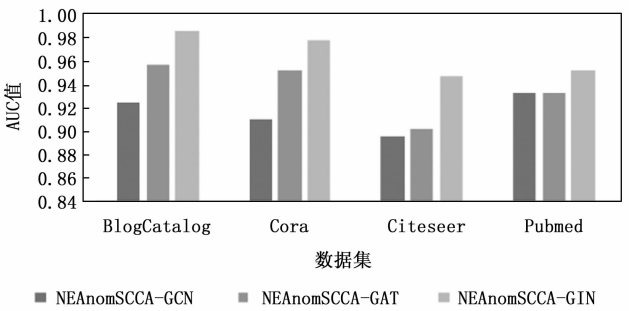

图 2　GNN 编码器结果影响图

根据实验结果，GIN 作为整个模型的编码器效果是最好的，因此选用 GIN 作为模型最终的编码器。GIN 在异常检测任务中，更加准确的区分出了正常节点和异常节点。GAT 在异常节点检测任务中虽不如 GIN，但相对于 GCN 还是有一定提升，这是由于 GAT 在聚合邻居信息时，考虑了不同邻居的权重贡献。特别是针对异常节点检测任务中的数据不平衡问题，GAT 聚合邻居信息的过程中，减缓了异常节点破坏网络的同质性。GCN 则在本任务中表现一般，其主要原因是数据集中存在连接不一致的边，也即正常节点和异常节点连接不一致，GCN 在聚合过程中引入了噪声数据，导致训练的图表示受到了噪声数据的干扰。

3.4　模型性能对比实验

本节主要探究了提出的 NEAnomSCCA 模型和其他六种基线模型的实验结果对比，下面介绍相关对比模型。

LOF[6]：基于属性的异常检测方法。通过观察同一社区的属性值，来判别节点是否为异常节点，只考虑了节点的属性信息，忽略了属性网络的结构异常。

DOMINANT[1]：基于图神经网络的异常检测架构。利用图卷积和自编码器来联合重构邻接矩阵和属性矩阵。通过计算重构误差来评估每个节点的异常程度，

该方法存在过平滑问题，GCN 聚合邻居信息时，会平滑异常信息。

AnomalyDAE[12]：将 GAT 应用于属性网络的嵌入，以学习不同邻居节点之间的重要性，从而提升异常检测的效果。

JAANE[13]：聚合过程中同时考虑目标节点及其邻居信息，融合二者学习节点的特征向量。通过重构误差和正则模块学习正常节点的超球面，在超球面之外的节点被视为异常节点。

ARISE[14]：该方法主要通过识别网络中的特定模式以检测异常，使用区域提议模块，检测网络中的密集模式作为异常节点存在的可疑区域，接着计算节点对之间的相似度，根据平均相似度检测异常节点。此外，使用图对比学习检测节点的属性异常。

PROPOSED[15]：基于稀疏典型分析和随机特征填充的异常节点检测模型，首先为每个节点添加随机特征，增强随机性，防止过拟合，接着提取结构和属性特征，通过 KL 散度将二者对齐到同样的潜在空间；然后通过 SCCA 计算属性和结构的稀疏典型向量，根据典型向量计算异常值得分，检测异常节点。

表 2 实验结果对比表

Methods	BlogCatalog	Cora	Citeseer	Pubmed
LOF	0.525 5	0.358 3	0.452 7	0.369 5
DOMINANT	0.748 6	0.881 8	0.742 2	0.832 4
Anomaly	0.753 7	0.897 8	0.768 9	0.841 7
JAANE	0.959 8	0.940 5	0.915 5	0.924 5
ARISE	0.955 7	0.914 4	0.925 4	0.896 4
PROPOSED	0.974 1	0.964 3	0.935 8	0.919 2
NEAnomSCCA	**0.985 4**	**0.977 8**	**0.946 9**	**0.952 5**

从表 1 中可以得出结论：NEAnomSCCA 模型在 4 个数据集上都取得了做好的结果，具体而言，NEAnomSCCA 在 BlogCatalog，Cora，Flickr，Pubmed 上分别取得了一定的提升，比对比模型最好的 AUC 值分别增加了：3.46%，1.04%，1.69%。这表明 NEAnomSCCA 通过 SCCA 稀疏性分析增强了异常得分的鲁棒性，弥补了重构误差不能检测边缘异常节点的不足，同时网络结构增强增强了模型信息传播能力。传统的异常节点检测方法在效果明显不如基于图神经网络的方法，这表明传统的机制无法同时捕获网络的属性和结构信息，处理网络的能力受到了限制。

3.5 消融分析

为了深入了解每个模块的作用，本小节对模型进行消融实验。主要探究以下模块是如何影响模型性能：1）网络结构增强的影响；2）SCCA 稀疏性分析的影响。

3.5.1 网络结构增强的影响

在该研究中，分析网络结构增强对模型性能的影响，本节将去掉节网络结构增强模块的模型和本文提出的模型进行对比实验。将去掉网络结构增强的模型记为 AnomSCCA。原模型和去掉网络结构增强模块模型在 4 个数据集上的实验结果如图 3 所示。

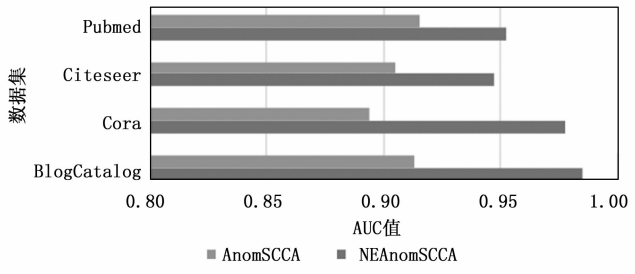

图 3 网络增强实验结果对比图

从数据中可以明显看出，NEAnomSCCA 在 4 个数据集上的 AUC 表现均为最佳。进一步分析可以得出：NEAnomSCCA 进行信息传播之前，先通过网络结构增强将网络中离散节点的结构补全，这样目标网络在经过图神经网络模型信息传播和学习节点特征时，能够增强离散节点的特征学习，促进网络的传播，更好的学习网络特征，增强检测效果。但是，NEAnomSCCA 去除了网络结构增强模块，在信息传播学习节点特征时，离散节点阻碍了网络的信息传播，导致学习到的节点特征不准确，影响异常节点检测的精度。

3.5.2 多模块联合分析

如表 3 所示，随机特征添加使 BlogCatalog 的 AUC 提升 1.7%（抗过拟合关键），双通道解码器在 Citeseer 上贡献 2.1%增益（解决边缘异常检测问题）。

表 3 模块消融影响

模块	BlogCatalog	Cora	作用
随机特征	+1.7%	+0.9%	提升泛化性
双通道解码器	+1.2%	+2.1%	捕获多维度异常

3.5.3 SCCA 稀疏性分析的影响

为了验证本文提出模型的 SCCA 稀疏性分析对属性网络异常节点检测是有帮助的，本节设计消融实验进行分析验证。将去掉 SCCA 稀疏性分析的模型和本文模型进行对比，去掉 SCCA 稀疏性分析的新模型表示为 NEAnom，两个模型在 4 个数据集上的实验结果如表 4 所示。黑色字体为 NEAnom 模型的结果，也即去掉 SCCA 模块。

表 4 SCCA 消融结果表

	BlogCatalog	Cora	Citeseer	Pubmed
NEAnomSCCA	0.985 4	0.977 8	0.946 9	0.952 5
NEAnom	0.912 8	0.894 0	0.904 7	0.915 3

从实验结果中可以看出加入 SCCA 稀疏性分析模块后，模型在 4 个数据集上的实验结果均有一定提升，AUC 值分别提升了 3.46%、8.92%、4.47%，实验充分证明了 SCCA 稀疏性分析在整个模型中的必要性；增加 SCCA 稀疏性分析模块能够减少模型过拟合的可能性，并且在检测异常节点时能够根据节点属性和结构的相关性计算异常值得分；若只根据重构误差检测检测异常节点，无法识别边缘异常节点，SCCA 稀疏性分析很好的弥补了这一缺陷。

3.6　鲁棒性验证实验

为评估模型在对抗攻击和数据噪声下的稳定性，设计对抗攻击测试如下。

采用 FGSM（Fast Gradient Sign Method）生成对抗样本：

攻击强度取值 0.1～0.4，得到如表 5 所示。

表 5　对抗攻击下的 AUC 变化

攻击强度 ε	BlogCatalog	Cora	性能保持率/%
0.1	0.976 3	0.968 2	98.9
0.3	0.961 8	0.948 7	95.2
0.4	0.942 1	0.927 6	89.7

4　结束语

鲁棒性局限与改进方向：

1）高强度噪声（λ＞40%）下需结合图结构修正（如 GNN-SVD）。

2）对抗训练使计算开销增加 23%（见表 6）。

表 6　计算开销对比

模型变体	训练时间(s/epoch)	AUC
基础模型	142.6	0.952 5
＋对抗训练	175.4（＋23%）	0.962 1

属性网络是现实世界常见的数据类型，如社交网络、金融网络、交通网络等，检测属性网络中的异常节点具有非常重要的意义。现有的属性网络异常节点检测方法主要使用图神经网络提取特征，计算重构误差，根据重构误差检测异常节点。然而，这类方法依赖原始属性网络的完备性，缺乏对异常检测任务的针对性设计。难以捕捉与正常节点相似度大的异常节点，泛化能力也较差。针对以上问题本文提出一种基于网络增强联合稀疏典型相关分析的异常节点检测模型，该模型首先采用 k 邻近算法生成 KNN 图，确保每个节点都有 k 个邻居节点，引入随机性缓解过拟合问题；接着使用权重共享的图神经网络编码器学习属性网络的结构和属性特征；然后根据学习到的特征计算重构误差，并利用 KL 散度将二者对齐到共享的潜在空间；随后使用双通道解码器计算网络的属性和结构重构误差，使用稀疏典型相关分析最大化正常节点结构和属性的相关性，集成 L_1 范数损失函数，约束 CCA 优化减少过拟合的可能性，使模型的泛化能力更强；最后，根据重构误差和节点结构与属性的相关性计算每个节点的异常值得分，以检测异常节点。

尽管 NEAnomSCCA 模型在实验中表现出优越的性能，但仍存在一定的局限性并面临特定场景的挑战。首先，在处理极大规模网络时，模型依赖于两层图神经网络进行特征学习，其计算复杂度随着网络规模增大而上升，可能影响实际应用效率；其次，当前模型主要针对静态属性网络设计，尚未充分考虑动态时变网络的特性，难以有效捕捉时间维度上的异常模式漂移；最后，虽然通过添加随机特征增强了模型对噪声的鲁棒性，但在面对高强度、系统性噪声干扰或精心设计的对抗性攻击时，模型的稳定性和检测精度可能受到影响。因此，未来研究将聚焦于提升模型的计算效率、扩展其处理动态网络的能力，并增强其在复杂噪声环境下的鲁棒性。

参考文献：

[1] DING K，LI J，BHANUSHALI R，etal. Deep anomaly detectionon attributed networks [C] //Proceedings of the2019 SIAM International Conference on Data Mining，2019：594－602.

[2] SEELAMMAL C，DEVI K V. Multi-criteria decision support for feature selection in network anomaly detection system [J]. International Journal of Data Analysis Techniques and Strategies，2018，10 (3)：334－350.

[3] THANG T M，KIM J. The anomaly detection by using dbscan clustering with multiple parameters [C] //2011 International Conference on Information Science and Applications. IEEE，2011：1－5.

[4] WANG H，TANG M，PARK Y，et al. Locality statistics for anomaly detection in time series of graphs [J]. IEEE Transactions on Signal Processing，2013，62 (3)：703－717.

[5] GOLDSTEIN M，DENGEL A. Histogram-based outlier score (hbos)：A fast unsupervised anomaly detection algorithm [J]. KI-2012：poster and demo track，2012 (1)：59－63.

[6] LEE J S，KANG B Y，KANG S H. The use of local outlier factor for improving performanceof independent component analysis basedstatistical process control [J]. Journal of the Korean Operations Researchand Management Science Society，2011，36 (1)：39－55.

[7] BHATTACHARYYA D K, KALITA J K. Network anomaly detection: A machine learning perspective [M]. Boca Raton: Crc Press, 2013.

[8] LIU Y, LI Z, PAN S, et al. Anomaly detection on attributed networks via contrastive self-supervised learning [J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33 (6): 2378 - 2392.

[9] ZHU D, MA Y, LIU Y. Anomaly detection with deep-graph autoencoders on attributed networks [C] //Proceedingsof the 2020 IEEE Symposium on Computers and Com-munications, 2020: 1 - 6.

[10] LUO X, WU J, BEHESHTI A, et al. ComGA: Community-aware attributed graph anomaly detection [C] // Proceedingsof the Fifteenth ACM International Conference onWeb Search and Data Mining, 2022: 657 - 665.

[11] PEI Y, HUANG T, et al. ResGCN: Attention- based deep residual modeling for anomalydetection on attributed networks [J]. Machine Learning, 2021, 111 (2): 519 - 541.

[12] FAN H, ZHANG F, LI Z. Anomalydae: Dual autoencoder for anomaly detection on attributed networks [C] // ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020: 5685 - 5689.

[13] FANH Y, et al. Deep joint adversarial learning for anomaly detection on attribute networks [J]. Information Sciences, 2024 (654): 119840.

[14] DUAN J, XIAO B, WANG S, et al. Arise: Graph anomaly detection on attributed networks via substructure awareness [J]. IEEE Transactions on Neural Networks and Learning Systems, 2024, 35 (12): 18172: 18185..

[15] KHAN W, ISHRAT M, KHAN A N, et al. Detecting anomalies in attributed networks through sparse canonical correlation analysis combined with random masking and padding [J]. IEEE Access, 2024 (12): 65555 - 65569.

[16] ZHAO M, LI X, TANG J. Structural and attribute perturbation for anomaly injection in attributed networks [J]. ACM Transactions on Knowledge Discovery from Data, 2021, 15 (4): 1 - 25.

[17] DUAN J, WANG S, ZHANG P, et al. Graph anomaly detection via multi-scale contrastive learning networks with augmented view [C] //Proceedings of the AAAI Conference on Artificial Intelligence. 2023, 37 (6): 7459 - 7467.

[18] SONG X, WU M, JERMAINE C, et al. Conditional anomaly detection [J]. IEEE Transactions on knowledge and Data Engineering, 2007, 19 (5): 631 - 645.

(上接第 233 页)

[13] SUBRAMANIAN B, MUTHUSAMY S, THANGARAJ K, et al. A novel approach using transfer learning architectural models based deep learning techniques for identification and classification of malignant skin cancer [J]. Wireless Personal Communications, 2024, 134 (4): 2183 - 2201.

[14] MARI C, MARI E. Deep learning based regime-switching models of energy commodity prices [J]. Energy systems, 2023, 14 (4): 913 - 934.

[15] OMURCA S L, EKINCI E, SEVIM S, et al. A document image classification system fusing deep and machine learning models [J]. Applied Intelligence, 2022, 53 (12): 15295 - 15310.

[16] YUNZHUO LIU, BO JIANG, TIAN GUO, et al. FuncPipe: A pipelined serverless framework for fast and cost-efficient training of deep learning models [J]. Performance Evaluation Review, 2023, 51 (1): 35 - 36.

[17] PRASHANTHI S. K, SAI ANUROOP KESANAPALLI, YOGESH SIMMHAN. Characterizing the performance of accelerated jetson edge devices for training deep learning models [J]. Performance Evaluation Review, 2023, 51 (1): 37 - 38.

[18] EDUARDO P F, DAMIAN C, FERNANDO M. A comparison of deep learning models applied to water gas shift catalysts for hydrogen purification [J]. International journal of hydrogen energy, 2023, 48 (64): 24742 - 24755.

[19] CHENG Y, CAO Z, ZHANG C D. Multi objective dynamic task scheduling optimization algorithm based on deep reinforcement learning [J]. Journal of Supercomputing, 2024, 80 (5): 6917 - 6945.

[20] CONG X, ZHANG W. Design of geometric flower pattern for clothing based on deep learning and interactive genetic algorithm [J]. Journal of Intelligent Systems, 2024, 33 (1): 57 - 76.

[21] LI Z, GE Y, YUE X, et al. MCAD: Multi-classification anomaly detection with relational knowledge distillation [J]. Neural Computing and Applications, 2024, 36 (23): 14543 - 14557.

[22] RUIXIANG TANG, YU-NENG CHUANG, XIA HU. The science of detecting LLM-Generated text, communications of the ACM [Z]. 2024, 67 (4): 50 - 59.