用于鲁棒构建基因调控网络的二次图注意力网络（Q-GAT）

Hui Zhang[a,*], Xuexin An[b,*], Qiang He[a], Yudong Yao[c], Yudong Zhang[e], Feng-Lei Fan[d,**], Yueyang Teng[a,**]

[a]东北大学医学与生物信息工程学院，中国沈阳 *110169*[b]哈尔滨工业大学数学学院，中国哈尔滨 *236500*[c]史蒂文斯理工学院电气与计算机工程系，美国新泽西州霍博肯 *07030*[d]香港中文大学数学系，香港。[e]莱斯特大学计算与数学科学学院，英国莱斯特

arXiv:2303.14193v2 [q-bio.MN] 4 Nov 2023

摘要

基因调控关系可抽象为基因调控网络（GRN），在刻画复杂的细胞过程和通路方面发挥着关键作用。近年来，图神经网络（GNN）作为一类深度学习模型，已成为从基因表达数据中推断基因调控关系的有效工具。然而，研究发现深度学习模型容易受到噪声影响，这极大地阻碍了深度学习在构建基因调控网络中的应用，因为基因表达测量过程中往往无法避免高噪声。我们能否设计一种更适合构建基因调控网络的鲁棒图神经网络？本文对此给出肯定的回答，提出了一种采用双重注意力机制的二次图注意力网络（Q-GAT）。我们通过引入不同程度的对抗扰动，研究了 Q-GAT 与 9 种最先进的基线方法在预测准确率方面的变化。在大肠杆菌和酿酒酵母数据集上进行的实验表明，Q-GAT 的鲁棒性优于最先进的模型。最后，我们通过信噪比（SNR）和可解释性分析剖析了 Q-GAT 鲁棒的原因。前者表明，二次神经元的非线性聚合能够放大有用信号并抑制不需要的噪声，从而提升鲁棒性；后者揭示，得益于双重注意力机制，Q-GAT 能在预测中利用更多特征，使其能够应对对抗扰动。我们已将代码分享在 https://github.com/Minorway/Q-GAT_，供读者评估_鲁棒_构建_的_GRN。

关键词： 基因表达、基因调控网络、图神经网络、稳定性分析

1. 引言

基因调控网络（GRN）是细胞调控系统中的有力工具，例如信号转导和转录调控；这些过程是许多细胞现象的分子基础[1, 2]。这些网络也是发育活动的图谱，组织调控性基因表达的空间分布，阐释生物过程，并为复杂疾病中的分子调控因子筛选候选基因 [3, 1]。构建基因调控网络的关键问题在于推断两个基因之间是否存在调控关系。

为推断基因调控关系，人们提出了大量机器学习方法，以便以构建基因调控网络（GRN） [2]。Salleh 等[4] 从全局交互图中随机提取子图后，采用多元线性回归方法推断基因调控关系，从而构建基因调控网络。Haury 等 [5] 提出了一种稳健的推断方法，将稳定性选择准则与最小角回归相结合。Margolin 等 [6] 采用信息论策略，剔除共表达方法识别出的多数间接关系。树集成基因网络推断方法（GENIE3）将针对p 个基因的调控网络预测问题转化为若干独立的回归问题，并使用随机森林求解[7]。Mordelet 等 [8] 提出的调控网络监督式推断方法（SIRENE）将推断问题分解为多个并行的二元分类问题，并通过

*共同第一作者。

**共同通讯作者：hitfanfenglei@gmail.com（Feng-Lei Fan）和 tengyy@bmie.neu.edu.cn （ Yueyang Teng ）

预印本已提交至《神经网络》 *2023*年*11*月*7*日

支持向量机（SVM）解决每个问题。

深度学习 [9, 10, 11] 目前是主流的机器学习方法。图神经网络（GNN）是一类专门处理图的深度学习模型。近年来，许多具有代表性的图神经网络（GNN）被用于解决基因调控网络的构建问题 [12, 13, 14, 15]。例如，Yuan 等人利用图卷积网络（GCN）从空间转录组学数据中推断基因相互作用 [16]。Wang 等人 [17] 利用提取的子图训练集成框架 GRGNN，以构建基因调控网络。Bigness 等人 [18]使用 GCN，根据组蛋白修饰和 Hi-C 数据预测基因表达（RNA 测序）。Karbalayghareh 等人开发的 GraphReg 模型通过图注意力网络（GAT）层，在三维相互作用图上传播从一维输入中学习到的局部表征，以预测基因组区域（BIN）中的基因表达（CAGE-SEQ）[19]。尽管这些图神经网络的表现令人满意，但与许多常用的深度学习方法一样[20, 21, 22, 23, 24, 25, 26, 27], ，即便面对微小扰动，这些图神经网络的预测结果也往往不稳定。由于生物学测量中不可避免地存在噪声，这种不稳定性极大地限制了图神经网络的广泛应用，因为图神经网络可能会对细胞现象以及疾病机制产生误导性解释。遗憾的是，尽管图神经网络的稳定性是构建基因调控网络时亟待解决的关键问题，但据我们所知，针对这一问题开展的研究寥寥无几，甚至可能完全没有。

近来，人们提出了一个名为 NeuroAI 的概念 [28] ，其核心思想是，大量神经科学知识能够极大地促进深度学习的发展，孕育新一代人工智能。受 NeuroAI 启发，并考虑到人脑中的各种智能行为都由神经元多样性赋能，Fan 等通过设计二次神经元，将神经元多样性引入深度学习 [29, 30]；这种神经元以简化二次函数取代了传统神经元中的内积。二次网络在现实问题中展现了当前最先进的性能，例如例如低剂量CT去噪 [31], 、异常检测 [32], 和轴承故障诊断 [33]。

本文提出一种采用二次神经元的全新图注意力网络（Q-GAT），并发现它能够增强构建基因调控网络时的鲁棒性。此外，得益于二次神经元更强的表征能力，Q-GAT 的预测性能也令人满意。接下来，我们通过向基因表达数据中引入对抗扰动，验证 Q-GAT 在构建基因调控网络时的预测能力和稳定性。我们采用联合指标，将 Q-GAT 与 9 种最先进的模型在大肠杆菌和酿酒酵母数据集上的表现进行比较。实验表明，在应对对抗扰动方面，Q-GAT 优于其他模型。最后，借助信噪比（SNR）分析和可解释性分析，我们剖析了 Q-GAT 鲁棒的原因。信噪比分析表明，二次神经元的非线性聚合能够提高信噪比，即放大有用信号并抑制不需要的噪声，从而促进鲁棒的特征提取。可解释性分析揭示，Q-GAT 在预测中能够利用比 GAT 更多的特征，这使 Q-GAT 不易受到对抗扰动的误导。我们的贡献主要有三点:

- 我们提出了用于基因调控网络重建任务的 Q-GAT，这是首个解决基因调控网络构建稳定性问题的工作。我们并未对网络结构进行修改，而是从神经元多样性的角度，在神经元层面进行了创新。
- 系统性实验表明，在不同程度的对抗噪声下，所提出的 Q-GAT 在构建基因调控网络时，稳定性优于其他方法。
- 我们并不满足于 Q-GAT 的优异性能，而是进一步借助信噪比和可解释性分析，探究 Q-GAT 鲁棒性更强的原因。

2. 相关工作

图神经网络。节点表示学习和图结构信息的传递是图神经网络基本结构的两个关键组成部分。图神经网络中的节点表示学习旨在为每个节点的特征提供向量表示。每个节点都可以从其原始特征向量开始。图结构中邻居节点之间的信息传递，是指图神经网络更新节点表示的过程。在每一层中，每个节点都会结合其邻居节点的特征来更新自身表示。经过多次迭代，节点可以逐步获得对其特征更全面且更符合上下文的表示。最终，节点表示可用于图级任务，包括图分类、节点分类和链接预测。为了捕获节点之间的相互连接和局部结构，提出了图卷积

网络（GCN），通过结合邻居的信息来更新节点的表示。为了捕获网络的整体结构和重要特征，图池化是指合并节点或子图以获得低维表示的过程。

图卷积网络（GCN）是较早的一类图神经网络。随后，GAT 通过自适应图注意力机制聚合节点的邻居信息，实现了高效的节点表示学习，并处理图级任务。深层图卷积网络（DeepGCN）采用残差连接/密集连接和膨胀卷积来训练非常深的图神经网络 [35]。图采样与聚合（GraphSAGE）是一种广义归纳框架，利用节点特征信息高效地为此前未知的数据构建嵌入表示 [36]。自回归移动平均滤波器（ARMA）采用图卷积层，不仅更加灵活、抗噪能力更强，而且比多项式更能保留全局图结构 [37]。

基于图卷积网络的图卷积层，深层图卷积神经网络（DGCNN）使用排序池化层对图的顶点进行排序。排序池化层在图卷积层与全连接层之间起到连接作用。通过记录输入的排序顺序并将梯度传回前一层，排序池化层不仅使前一层能够参与训练，还有助于从全局图的拓扑结构中学习[38]。可微池化（DiffPool）采用可微图池化模块来提供分层图表征，并可与多种图神经网络结构集成 [39]。自注意力图池化（SAGPool）采用一种自注意力图池化技术，同时考虑节点属性和图拓扑 [40]。Top-k池化（TopKPool）采用图U-Net提出的一种新型图池化方法，根据标量投影值选取节点，以构建较小的图 [41]。图多重集变换器（GMT）利用基于多头注意力的全局池化层，捕捉节点之间的交互及其结构依赖关系 [42, 43]。PiNet 将一种广义的、基于可微注意力的池化技术与图卷积操作相结合，用于图级分类 [44]。

与其他类型的深度学习模型一样，图神经网络往往容易受到微小扰动的影响。例如，除了针对特定图像的对抗性扰动之外， [45, 46], Moosavi-Dezfooli 等人 [47] 还指出，“通用扰动”可以欺骗网络分类器，使其对任意图像都产生错误判断。此外，Dai 等人 [48] 发现图神经网络方法在文本和金融数据集上的图级和节点级分类任务中都容易受到攻击。因此，将图神经网络应用于构建基因调控网络时，由于生物测量中出现的噪声，很难判断推断出的关系是否正确。鉴于基因调控网络至关重要，解决其构建过程中的稳定性问题十分必要。

值得注意的是，在无标签条件下重建基因调控网络方面已取得显著进展。首先，许多研究成功运用半监督学习技术，处理大量无标签数据和少量有标签数据。例如，为了更好地理解数据的内在结构，基于图的方法可以在有标签数据与无标签数据之间共享信息 [49, 50]。迁移学习和多任务学习方法能够从一个或多个相关任务中迁移知识，因此在有标签数据稀缺且分布不均时十分有用。Paul 等人的 [51]基于多任务学习的人类和小鼠基因调控网络重建研究表明，即使有标签样本与无标签样本的比例极低，所提出的方法几乎总是优于相应的单任务方法。图神经网络（GNN）展现了在具有挑战性且噪声较多的环境中分析生物数据的潜力 [52, 53, 54, 55]。这些方法能够自动检测隐藏特征和模式，从而识别基因调控网络中的细微相互作用。与此同时，集成方法的发展已在基因调控网络重建中取代了传统方法。结合多个模型或策略可以进一步提高模型的鲁棒性和准确性 [17, 56, 57]。此外，因果推断在这一领域的重要性日益凸显。因果推断可用于解释基因之间的因果相互作用 [58]。

高阶单元。对高阶单元的研究可以追溯到数据处理组合法（GMDH [59]），这本质上是一种多项式特征提取器：

$$Y(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n) = a_0 + \sum_{i}^{n} a_i \boldsymbol{x}_i + \sum_{i}^{n} \sum_{j}^{n} a_{ij} \boldsymbol{x}_i \boldsymbol{x}_j + \sum_{i}^{n} \sum_{j}^{n} \sum_{k}^{n} a_{ijk} \boldsymbol{x}_i \boldsymbol{x}_j \boldsymbol{x}_k + \cdots, \tag{1}$$

其中，$\boldsymbol{x}_i$是第i个输入变量，a_i、a^{ij}、$a^{ijk}\cdots$是系数。通常，GMDH 仅保留二阶项，以避免高维输入导致参数爆炸。近期，高阶单元得到了

重新审视 [60, 61]。 [60], 高阶单元被嵌入深度网络中，通过张量分解和因子共享降低单个单元的复杂度。这种网络在多个任务上取得了前沿性能。Liu 和 Wang [61]定义了所谓的 Gang 神经元，其递归形式为 $\mathbf{A}^l = \mathbf{W}^{l,l-1}\mathbf{A}^{l-1} \circ \boldsymbol{x}$ 和 $\mathbf{A}^0 = \boldsymbol{x}$。然而，为了提高参数效率，大多数多项式神经元研究都聚焦于二次神经元，如表 1 所示。在这些二次神经元设计中，Fan 等人的设计 [62] 最为全面，因为其他神经元设计（例如 [63, 64, 65, 66]）实际上都可以视为它的特例。

表 1：近期提出的二次神经元概览。$\sigma(\cdot)$表示非线性激活函数。$\odot$ 是 Hadamard 乘积。此处省略神经元中的偏置项。$\boldsymbol{x} \in \mathbb{R}^{n\times 1}$。$\mathbf{w}_i \in \mathbb{R}^{n\times 1}$ 和$\mathbf{W}_i \in \mathbb{R}^{n\times n}$ 分别是可学习权重向量和矩阵。

工作参考文献	神经元形式
Zoumpourlis 等人[67] Bu&Karpatne [68]	$\boldsymbol{y} = \sigma(\boldsymbol{x}^\top \mathbf{W}_1 \boldsymbol{x} + \mathbf{w}_1^\top \boldsymbol{x})$
Tsapanos 等人*etal.* [69]	$\boldsymbol{y} = \sigma((\mathbf{w}_1^\top \boldsymbol{x}) \odot (\mathbf{w}_2^\top \boldsymbol{x}))$
Fan 等人 [62]	$\boldsymbol{y} = \sigma((\mathbf{w}_1^\top \boldsymbol{x}) \odot (\mathbf{w}_2^\top \boldsymbol{x}) + \mathbf{w}_3^\top (\boldsymbol{x} \odot \boldsymbol{x}))$
江等 [63] Mantini&Shah [64]	$\boldsymbol{y} = \sigma(\boldsymbol{x}^\top \mathbf{W}_1 \boldsymbol{x})$
Goyal 等人 [65]	$\boldsymbol{y} = \sigma(\mathbf{w}_1^\top (\boldsymbol{x} \odot \boldsymbol{x}))$
Xu 等人 [66]	$\boldsymbol{y} = \sigma((\mathbf{w}_1^\top \boldsymbol{x}) \odot (\mathbf{w}_2^\top \boldsymbol{x}) + \mathbf{w}_3^\top \boldsymbol{x})$

3. 预备知识

基因调控的复杂关系和相互作用可以用基因调控网络来描述。在这些网络中，基因、蛋白质、转录因子等分子被建模为节点，而它们之间的相互作用、调控关系等则由边连接。基因调控网络有助于阐明基因之间的调控关系、信号通路，以及整个调控系统的结构和运作方式。例如，在转录因子-基因网络中，节点代表转录因子和基因，边则表示它们之间的调控连接。

称为转录因子的蛋白质会结合靶基因的启动子和增强子，以控制这些基因的转录活性。不同的转录因子可以通过结合不同的调控区域（这些区域通常含有特定的结合位点），以协同或竞争的方式影响靶基因的表达。因此，在转录因子-基因网络中，不同转录因子之间可能发生相互作用，但不同靶基因之间通常不存在直接相互作用。

将转录因子（TF）的集合记为T，靶基因的集合记为G，它们的基因表达数据记为 $R_{i,\ j}$，其中 $i \in \{T,\ G\}$，且$j \in [1, n]$，适用于每个T 和G，且有n 个数组。通常，基因调控网络（GRN）用.....表示，即$T \times （T + G）$矩阵。该矩阵有T 行，对应转录因子的数量。矩阵的列数为（$T + G$），矩阵中前 T 列表示转录因子之间的相互作用，而最后G 列表示转录因子与基因之间的相互作用。这种表示方式便于观察基因调控网络中基因与转录因子之间的关系，从而使我们能够运用线性代数分析，更好地理解网络的结构和动态特性。然而，由于这是一个线性静态模型，它可能无法充分体现生物系统的复杂性。

我们的任务是获得一个本质上为$\mathbb{R}^{T\times(T+G)}$ 矩阵的基因调控网络。E 表示转录因子与基因之间以及转录因子彼此之间调控关系对应的边。值得注意的是，一对基因之间不允许存在边，因为迄今尚未发现此类机制。如果 $(i, j) \in E$，则$A_{i,j} = 1$；否则，$A_{i,j} = 0$。

在本研究中，我们遵循常见做法，准备一个数据集来训练用于推断的分类器。具体流程如下：

i) 通过对皮尔逊相关系数设定阈值，从基因表达数据中获取候选链接；ii) 融合节点特征并提取包含这些节点的子图；iii) 基于标记链接生成数据集。

步骤 i) 和 ii) 可通过排除许多不可能的链接，显著提高图构建的效率。步骤 iii) 本质上是收集所有子图。下文将详细介绍步骤 i) 和 ii)：

候选连接。当两个基因存在调控关系时，它们的表达数据通常具有相关性，因此我们可以利用相关性建立候选连接。我们采用皮尔逊相关系数作为启发式方法，确定两个基因之间的相关性，并得到邻接矩阵$H_{T,T+G}$。给定一个阈值，二值邻接矩阵定义如下：

$$A_{i,j} = \begin{cases} 1 & \text{if } H(i,j) \geq \text{threshold} \\ 0 & \text{if } H(i,j) < \text{threshold}. \end{cases} \tag{2}$$

尽管这种初步方法大大减少了工作量，但必然会产生许多伪链接。因此，我们采用以下步骤进行细致判断。

节点特征的融合与子图提取。嵌入是一种连续的特征表示，通过映射函数将图中的节点表示为向量。学习到的节点嵌入反映

了图的拓扑结构。换言之，相似的图节点具有相似的低维特征模式。本研究采用 Node2vec [70] 方法学习嵌入特征。每个节点的原始基因表达数据揭示其生物学功能，并包含大量信息。因此，我们对每个基因在各种条件下的基因表达数据取平均值。该平均值作为节点的显式特征，随后与隐式嵌入特征结合，形成新的特征。

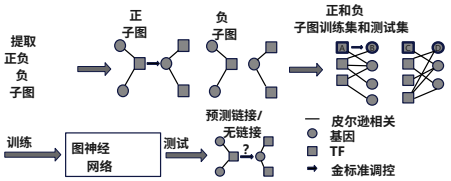


图 1：提取用于训练和测试的子图的流程。

接下来，如图 1 所示，我们针对每条连接边，从上述粗略构建的基因调控网络中提取子图，其中包括对应的两个节点及其h跳邻居。Wang等人指出，在基因调控网络中增加跳数并不会显著改善结果。因此，这里提取 1 跳子图。将存在连接的基因节点对的位置随机映射到金标准网络上，并划分为训练集和测试集。由于一些基因间的调控关系已由其他生物学工具预先确定，因此相应的子图可以标记为有连接或无连接。通常，为了获得平衡数据集，会使连接和空缺的数量大致相等。

最终的训练集和测试集由有连接和无连接的基因节点对组成。接下来，将皮尔逊相关网络中与测试集位置对应的连接屏蔽。然后，利用从金标准网络收集的训练集和测试集位置，在皮尔逊相关网络上提取子图。根据金标准网络中存在连接的基因对位置提取正子图，而根据不存在连接的位置提取负子图。

4. 二次图注意力网络（Q-GAT）

4.1. 二次神经元

假设输入向量为$\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top$，传统神经元可表示为

$$\sigma(l(\boldsymbol{x})) = \sigma\Big(\sum_{i=1}^{n} w_i x_i + b^r\Big) = \sigma(\boldsymbol{w}^\top \boldsymbol{x} + b), \quad (3)$$

其中，$\sigma(\cdot)$ 是非线性激活函数。受神经元多样性在生物神经系统中发挥的关键作用启发，人们提出了一种新型神经元，称为二次神经元 [71] ，以促进深度学习中的神经元多样性。二次神经元[71] 在进行非线性激活之前，对输入向量的两个内积和一个幂项进行整合。从数学上讲，二次神经元的计算方式为

$$\sigma(q(\cdot)) = \sigma\Big((\boldsymbol{w}_1^\top \boldsymbol{x} + b_1)(\boldsymbol{w}_2^\top \boldsymbol{x} + b_2) + \boldsymbol{w}_3^\top (\boldsymbol{x} \odot \boldsymbol{x}) + b_3\Big), \quad (4)$$

其中，$Q(\cdot)$ 是简化二次函数，$\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{w}_3 \in \mathbb{R}^n$ 是权重向量，$b_1, b_2, b_3 \in \mathbb{R}$是偏置。此外，二次层的运算可以表示为

$$\sigma(\boldsymbol{Q}(\boldsymbol{x})) = \sigma\Big((\mathbf{W}_1^\top \boldsymbol{x} + \boldsymbol{b}_1)(\mathbf{W}_2^\top \boldsymbol{x} + \boldsymbol{b}_2) + \mathbf{W}_3^\top (\boldsymbol{x} \odot \boldsymbol{x}) + \boldsymbol{b}_3\Big), \quad (5)$$

其中，$\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3 \in \mathbb{R}^{n_1 \times n^2}$ 是权重矩阵，$\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3 \in \mathbb{R}^{n_1}$ 是偏置向量。

二次网络在解决各种现实世界问题时展现了最先进的性能，例如低剂量 CT 去噪 [31], 异常检测 [32], 以及轴承故障诊断 [33]。与此同时，理论上也已证实，与传统神经元和网络相比，二次神经元和网络具有更强的表达能力 [29, 72]。这些积极的研究结果表明，二次网络将成为基因调控网络中的高性能模型。

4.2. qttention

注意力机制的核心是为输入生成一张重要性图，使从业者能够了解输入的哪些部分比其他部分更重要，从而促使网络有选择地利用输入信息。基于这一观察，早期研究 [33] 表明，可以通过基于规范注意力机制对所学习的二次函数进行因式分解，让二次神经元产生一种注意力机制，称为 *qttention*。

在数学上，我们将二次神经元的表达式（式（4））因式分解如下（为简洁起见，省略偏置项）：

$$\begin{aligned}
&\sigma((\boldsymbol{w}_1^\top \boldsymbol{x} + b_1)(\boldsymbol{w}_2^\top \boldsymbol{x} + b_2) + \boldsymbol{w}_3^\top (\boldsymbol{x} \odot \boldsymbol{x}) + b_3) \\
=&\sigma(\boldsymbol{w}_2^\top \boldsymbol{x}(\boldsymbol{w}_1^\top \boldsymbol{x} + b_1) + b_2 \boldsymbol{w}_1^\top \boldsymbol{x} + b_2 b_1 + \boldsymbol{w}_3^\top (\boldsymbol{x} \odot \boldsymbol{x})) \\
=&\sigma(\boldsymbol{w}_2^\top (\boldsymbol{w}_1^\top \boldsymbol{x} + b_1)\boldsymbol{x} + \boldsymbol{w}_1^\top b_2 \boldsymbol{x} + (\boldsymbol{x} \odot \boldsymbol{w}_3)^\top \boldsymbol{x}) \\
=&\sigma((\underbrace{\boldsymbol{x} \odot \boldsymbol{w}_3 + \boldsymbol{w}_2(\boldsymbol{w}_1^\top \boldsymbol{x})}_{qttention} + \underbrace{\boldsymbol{w}_2 b_1 + \boldsymbol{w}_1 b_2}_{bias})^\top \boldsymbol{x}),
\end{aligned} \quad (6)$$

我们排除了偏置项$\boldsymbol{w}_2 b_1 + \boldsymbol{w}_1 b_2$，因为它们对于不同的$\boldsymbol{x}$始终相同；因此，它们不能

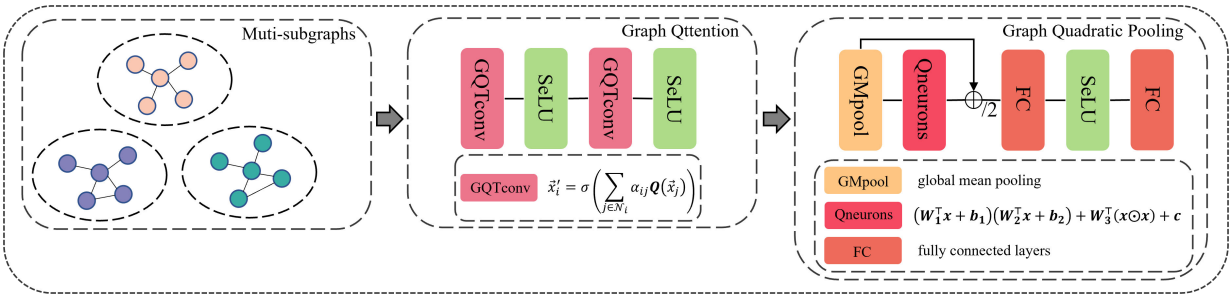


图 2：Q-GAT 的结构。

作为重要性分数发挥作用。接下来，我们令

$$Qttention(\boldsymbol{x}) = \boldsymbol{x}\odot\boldsymbol{w}_3 + \boldsymbol{w}_2(\boldsymbol{w}_1^\top\boldsymbol{x}). \tag{7}$$

4.3. 二次图注意力网络 *(Q-GAT)*

图神经网络方法可以利用图结构和节点特征，为节点或整个图学习一个表示向量 [73]。在本研究中，我们提出了二次图注意力模型 (Q-GAT)，利用图中的二次神经元处理特征 [74, 33]，并通过注意力机制聚合这些特征。Q-GAT 的关键创新并非对架构进行修改，而是图 *qttention* 层 (GQT)——一种借助二次神经元实现的双重注意力机制，以及基于图 qttention 的池化层。由于非线性的存在，二次运算不仅能在构建 GRN 时带来良好的预测性能，还能增强所构建 GRN 的鲁棒性。

4.3.1. 图 *Qttention*

基于嵌入在二次神经元中的 qttention 机制，我们提出了一种双重注意力图卷积结构——Graph Qttention (GQT)。如图 2 和公式 (8) 所示，输入数据经过二次层后，评分函数$e: \mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$将按如下方式作用于其上：

$$e_{ij} = \text{LeakyReLU}\left(\boldsymbol{a}^\top\cdot\left[\boldsymbol{Q}(\boldsymbol{x}_i)\|\boldsymbol{Q}(\boldsymbol{x}_j)\right]\right) \tag{8}$$

其中，$\boldsymbol{a}^\top\in\mathbb{R}^{2n}$是学习得到的，$\|$是拼接操作，注意力系数$e_{ij}$反映了$\boldsymbol{x}_j$对$\boldsymbol{x}_i$的重要性，而$\mathcal{N}_i$是节点$i$的邻居集合。在我们的模型中，只计算节点$j\in\mathcal{N}_i$对应的$e_{ij}$；因此，与直接计算所有节点之间的相对系数相比，可以节省更多计算资源。最后，应用 softmax 函数将e_{ij}归一化为 [0, 1]:

$$\alpha_{ij} = \text{softmax}_j\left(e_{ij}\right) = \frac{\exp\left(e_{ij}\right)}{\sum_{k\in\mathcal{N}_i}\exp\left(e_{ik}\right)} \tag{9}$$

为什么所提出的 *GQT* 是一种双重注意力机制？首先，每个输出节点的重要性可以通过其邻居以及得到的注意力系数α_{ij}来表示。另一方面，学习得到的二次映射$\boldsymbol{Q}$采用了 qttention机制，可以体现给定节点特征的重要性。

此外，还可以探索多头注意力。我们实现式 (10) 中 K 个相互独立的双注意力机制，然后将式 (11) 中得到的特征拼接起来：

$$\boldsymbol{x}_i' = \sigma\left(\sum_{j\in\mathcal{N}_i}\alpha_{ij}\boldsymbol{Q}(\boldsymbol{x}_j)\right) \tag{10}$$

$$\boldsymbol{x}_i' = \Big\|_{k=1}^{K}\sigma\left(\sum_{j\in\mathcal{N}_i}\alpha_{ij}^k\boldsymbol{Q}^K(\boldsymbol{x}_j)\right) \tag{11}$$

其中，α_{ij}^k和$\boldsymbol{Q}^K$分别是 K 项注意力系数和 qttention 映射。

4.3.2. 图二次池化

我们设计了图二次池化层，着重运用二次运算。其原因是为了类比基于注意力的池化层。通常采用全局平均池化来获得全局图表示。虽然直接取均值或最大值进行池化是可行的，但这会忽略节点差异。在第 4.2 小节中，我们讨论了二次神经元能够通过不同的特征权重为节点分配不同的权重。因此，当二次神经元的权重得到更新时，图二次池化就是一种基于注意力的池化。

首先，假设图 qttention 层的输出为$\boldsymbol{x}$，我们应用全局平均池化$\boldsymbol{GM}$，以获得一个未考虑节点差异的初步图表示：

$$\boldsymbol{x}_{\boldsymbol{GM}} = \boldsymbol{GM}(\boldsymbol{x}). \tag{12}$$

此外，图二次池化可表示为

公式 (13):

$$\boldsymbol{x}_{GQ} = \frac{1}{2}(\boldsymbol{x}_{GM} + \boldsymbol{Q}(\boldsymbol{x}_{GM})), \tag{13}$$

其中$\boldsymbol{Q}$ 表示二次函数。

由于二次神经元的注意力机制，式（13）能够为不同节点赋予不同权重，从而动态调节节点之间的关系。此外，保留$\boldsymbol{x}_{GM}$ 有助于提高网络的稳定性，并促进优化过程。二次神经元的注意力机制确实十分强大。不过，本文的重点并非讨论图二次池化的注意力机制，而是利用它来增强模型的鲁棒性。此外，考虑到图神经网络容易受到噪声影响，我们在池化层中通过具有特定初始化条件的二次神经元手动引入随机性，从而进一步提高鲁棒性。

5. 实验

5.1. 数据集

我们使用 DREAM5 挑战赛中的公开大肠杆菌和酿酒酵母 数据集，研究我们的方法与另外 9 种强大的图神经网络方法的预测准确率和鲁棒性 [1]。该数据集是一个基准数据集，包含一组已知的调控相互作用。详细统计数据见表 2。我们按 7:3 的比例将数据划分为训练集和测试集。构建初始基因调控网络的阈值为 0.8。嵌入特征是一个一维向量。

表 2：所用数据集的统计信息。

物种	TFs	靶基因	连接	样本
大肠杆菌	344	4177	2066	805
酿酒酵母	333	5617	3940	536

5.2. 对抗扰动

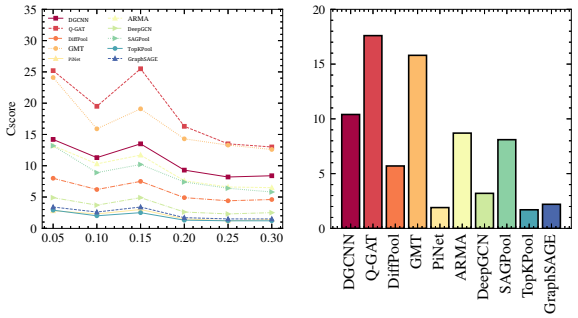
稳定性是指模型抵御对抗扰动的能力。尽管可以在训练阶段向数据集添加扰动，这称为数据投毒 [75, 76]，但通常是在模型已经使用训练集完成训练后，才向测试集引入扰动。

我们按如下方式向数据添加高斯扰动:

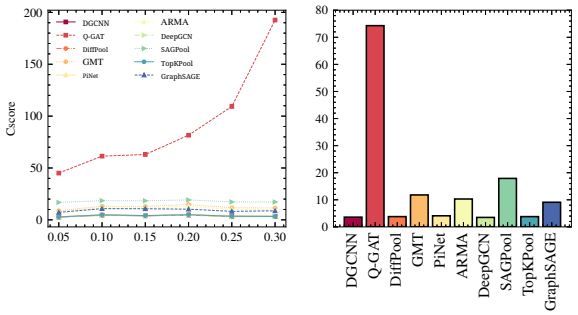
$$z^* = z + \varepsilon \tag{14}$$

其中 z 和 z^* 分别表示原始数据和受扰动数据，且 $\varepsilon \sim N\left(0, \sigma^2\right)$。众所周知，$P\{|\varepsilon| \leq 3\sigma\} = 0.997 \approx 1$; 因此，我们将$c = 3\sigma$ 记为绝对误差上限。

我们通过在$c = 0$中分别设置 0.05、0.1、0.15、0.2、0.25 和 0.3，加入不同程度的扰动，式（14）。我们重复上述步骤，得到受扰动的基因对，并使用我们的 Q-GAT 和其他图神经网络方法预测其标签。所有方法均在 30 次独立扰动下进行测试。随后，对均值分数及其标准差进行分析。



(a) 大肠杆菌



(b) 酿酒酵母

图3：(a) 左侧为大肠杆菌数据集在不同扰动水平下的 Cscore，右侧为所有方法的总 Cscore；(b) 左侧为酿酒酵母数据集在不同扰动水平下的 Cscore，右侧为所有方法的总 Cscore。

5.3. 评估指标

尽管本文重点关注稳定性，但为了模型的实用性，也应考虑其预测性能。为了评估不同模型在构建基因调控网络时的稳定性和预测性能，我们提出一种新的、更均衡的评估指标，如下所示:

$$\text{Cscore} = \frac{\text{Mean}}{\text{Bias}}, \tag{15}$$

其中

$$\text{Bias} = \sqrt{\frac{\sum_{i=1}^{N}(I_i - \bar{I})^2}{N-1}}, \tag{16}$$

且均值是同一扰动水平下N 次重复实验的平均准确率，I 是无扰动数据集的准确率，而$\bar{I}_i$ 表示第i个受扰动数据集的准确率。均值越大，且

表 3：图神经网络在不同扰动水平下对大肠杆菌数据集的多重扰动性能（均值± 偏差）。

方法	c =0.0	c =0.05	c =0.1	c =0.15	c =0.2	c =0.25	c =0.3
DGCNN	0.735	0.707±0.040	0.693±0.044	0.705±0.042	0.677±0.041	0.662±0.032	0.666±0.038
DiffPool	0.637	0.599±0.063	0.575±0.067	0.593±0.064	0.547±0.062	0.528±0.046	0.536±0.054
GMT	0.718	0.706±0.026	0.689±0.032	0.698±0.030	0.680±0.026	0.673±0.022	0.671±0.024
PiNet	**0.803**	0.662±0.184	0.622±0.191	0.674±0.179	0.513±0.171	0.460±0.134	0.496±0.165
ARMA	0.704	0.679±0.044	0.663±0.048	0.674±0.048	0.637±0.048	0.620±0.039	0.620±0.043
DeepGCN	0.748	0.674±0.114	0.633±0.124	0.678±0.116	0.570±0.116	0.535±0.089	0.554±0.108
SAGPool	0.650	0.637±0.046	0.613±0.056	0.622±0.053	0.590±0.051	0.570±0.038	0.565±0.043
TopKPool	0.782	0.660±0.184	0.578±0.201	0.630±0.195	0.465±0.172	0.439±0.151	0.443±0.151
GraphSAGE	0.773	0.664±0.161	0.620±0.172	0.667±0.162	0.515±0.147	0.479±0.119	0.477±0.118
Q-GAT（我们的方法）	0.726	**0.707±0.020**	**0.700±0.024**	**0.710±0.022**	**0.691±0.023**	**0.681±0.020**	**0.679±0.021**

表4：在不同水平下，针对大肠杆菌数据集，采用多重扰动的早停图神经网络和 Q-GAT 的性能（均值± 偏差）。

方法	c =0.0	c =0.05	c =0.1	c =0.15	c =0.2	c =0.25	c =0.3
PiNet	0.726	0.679±0.072	0.653±0.078	0.676±0.075	0.614±0.071	0.600±0.061	0.611±0.069
TopKPool	0.693	0.646±0.071	0.614±0.084	0.631±0.082	0.567±0.077	0.543±0.061	0.542±0.060
Q-GAT（我们的方法）	0.726	**0.707±0.020**	**0.700±0.024**	**0.710±0.022**	**0.691±0.023**	**0.681±0.020**	**0.679±0.021**

偏差越小，Cscore 就越大。请注意，也可以使用变异性（均值除以标准差）。我们不采用这种指标，因为当标准差较小时，其数值会变得非常大。然而，此时均值可能很低，这意味着模型仍然无法使用。

5.4. 稳定性结果

表5展示了不同模型所采用的超参数设置、训练迭代次数和损失函数。为了使每个模型都达到最佳性能，本文根据模型特定训练过程中的模型收敛情况，调整迭代次数和学习率。

表3和表6分别汇总了大肠杆菌 和酿酒酵母的对比结果。在大肠杆菌 数据集中没有扰动时，我们发现 PiNet 的准确率最高（ 0.803 ），DiffPool 的准确率最低（0.637）。DGCNN、GMT、ARMA 和 Q-GAT 的预测得分非常接近，分别为 0.735、0.718、0.704 和 0.726。所有模型的准确率在受到扰动后均有所下降。然而，随着扰动水平的升高，DiffPool、PiNet、Deep-GCN、SAGPool、TopKPool 和 GraphSAGE 的准确率大幅下降，而 Q-GAT 和 GMT 则能有效应对扰动。与 GMT 相比，Q-GAT 在每个扰动水平下都取得了最高得分；当c = 0.1 和 0.15 时，两者的差距尤为显著，这表明 Q-GAT 具有更强的稳定性。

观察发现，尽管 PiNet 和 TopKPool 在原始测试集上的准确率分别排名第一和第二，但即使受到极小的扰动，它们的表现也不佳。为考察这是否是过拟合所致，我们采用早停策略应用于 PiNet 和 TopKPool。表4展示了在 大肠杆菌 数据集上对 PiNet 和 TopKPool 进行早停后的测试结果。我们控制早停的程度，使 PiNet 和 TopKPool 在无扰动时的预测准确率与 Q-GAT 相近。可以看出，与完全收敛时的表现相比，应用早停后，PiNet 和 TopKPool 在所有扰动水平下的预测准确率均有所提高。然而，在每个扰动水平下，PiNet 和 TopKPool 的表现仍不如 Q-GAT。

对于酿酒酵母数据集，在无扰动情况下，GraphSAGE 和 SAG- Pool 分别取得最高和最低准确率（0.775 和 0.598），而 Q- GAT 的准确率为 0.765，接近 GraphSAGE。在不同扰动水平下，我们观察到 Q-GAT 的准确率始终最高，且其标准差最低，表明其具有很强的稳定性。此外，随着扰动水平升高，Q-GAT 的准确率仅适度下降。相比之下，其他所有方法的准确率均出现较大幅度的下降，其标准差通常也大于 Q-GAT。例如，当c = 0.3 时，Q-GAT 的准确率最高，为 0.764，而 SAGPool 的准确率最低，为 0.567。Q-GAT 的表现显著优于 SAGPool。

图3展示了所有模型在两个数据集上不同扰动水平下的C分数。对于大肠杆菌 数据集，随着扰动增加，总体而言，所有方法的C分数均呈下降趋势。但当c = 0.15时，各方法中不同模型的C分数略有上升。在所有扰动水平下，Q-GAT的C分数均为最高。对于酿酒酵母 数据集，随着扰动水平增加，Q-GAT的C分数始终显著高于其他方

表5：各模型所采用的超参数设置、训练迭代次数和损失函数。

方法	学习率	隐藏层	大肠杆菌上的迭代次数	酿酒酵母上的迭代次数	损失函数
DGCNN	10^{-4}	[32,32]	100	150	交叉熵损失
DiffPool	10^{-6}	[64,64]×3	50	50	交叉熵损失
GMT	10^{-4}	[32,32]	150	150	交叉熵损失
PiNet	10^{-4}	[64,64]	150	150	交叉熵损失
ARMA	10^{-4}	[128]	100	150	交叉熵损失
DeepGCN	10^{-4}	[64]×28	100	150	交叉熵损失
SAGPool	10^{-4}	[128,128,128]	150	150	交叉熵损失
TopKPool	10^{-4}	[128,128,128,64]	100	100	交叉熵损失
GraphSAGE	10^{-4}	[32,32,32]	100	150	交叉熵损失
Q-GAT（我们的方法）	10^{-5}	[256,256,128]	150	150	交叉熵损失

表6：不同水平多重扰动下图神经网络在酿酒酵母 数据集上的表现（均值±偏差).

方法	c =0.0	c =0.05	c =0.1	c =0.15	c =0.2	c =0.25	c =0.3
DGCNN	0.765	0.584±0.115	0.675±0.119	0.657±0.125	0.685±0.117	0.635±0.129	0.629±0.132
DiffPool	0.726	0.556±0.082	0.626±0.081	0.609±0.087	0.628±0.083	0.587±0.090	0.587±0.090
GMT	0.767	0.698±0.038	0.728±0.038	0.726±0.037	0.735±0.036	0.721±0.040	0.721±0.041
PiNet	0.756	0.587±0.100	0.668±0.104	0.653±0.108	0.676±0.101	0.633±0.108	0.635±0.108
ARMA	0.762	0.691±0.050	0.727±0.050	0.724±0.051	0.733±0.048	0.715±0.051	0.712±0.052
DeepGCN	0.761	0.587±0.123	0.666±0.126	0.658±0.129	0.676±0.125	0.606±0.130	0.627±0.136
SAGPool	0.598	0.566±0.009	0.569±0.009	0.569±0.009	0.570±0.008	0.567±0.008	0.567±0.008
TopKPool	0.765	0.592±0.113	0.680±0.110	0.659±0.117	0.685±0.109	0.633±0.118	0.621±0.117
GraphSAGE	**0.775**	0.693±0.046	0.730±0.049	0.729±0.048	0.727±0.050	0.708±0.052	0.714±0.053
Q-GAT(我们的方法)	0.765	**0.749±0.004**	**0.754±0.004**	**0.754±0.004**	**0.757±0.004**	**0.759±0.003**	**0.764±0.004**

法。具体而言，图3展示了大肠杆菌和酿酒酵母 数据集在不同扰动水平下的C分数。对于大肠杆菌 数据集，结果表明，Q-GAT的C分数为17.6，是最佳方法；而DGCNN、ARMA和SAGPool的C分数分别为10.4、8.7和8.1，彼此差异不显著。对于酿酒酵母，Q-GAT的C分数为74.3，远高于其他方法。

表7：在 酿酒酵母 数据集中，GMT 和 Q-GAT 在 c = 0.15 或 c = 0.3 与 c = 0 之间的预测标签相似度。

相似性	(c = 0.15, c = 0)	(c = 0.3,c = 0)
GMT	0.7216	0.7288
Q-GAT	0.9148	0.9013

随着扰动水平升高，仍能保持较高准确率固然出色，但原因可能是：未添加扰动时，原始预测标签被错误分类，而添加扰动后却被正确分类。为了实现稳定的构建，我们希望确保不同扰动水平下的分类结果尽可能接近无扰动时的原始预测。因此，这意味着该方法能够有效抑制噪声干扰。

为描述此类问题，这里引入一个相似度指数。首先预先定义函数 $f(x, y)$

$$f(x,y) = \begin{cases} 1, x = y \\ 0, x \neq y. \end{cases} \tag{17}$$

然后，相似度指数可定义为

$$\mathrm{Sim}(\boldsymbol{x}, \boldsymbol{y}) = \frac{\sum_{i=1}^{N} f(x_i, y_i)}{N}, \tag{18}$$

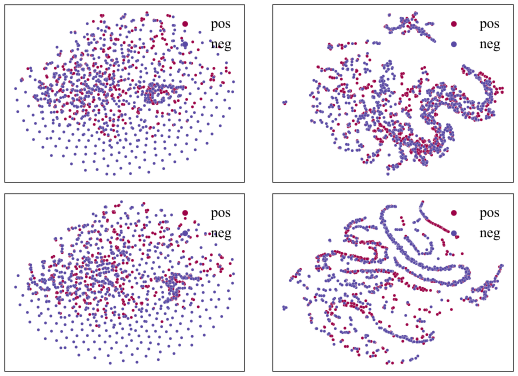
其中，$\boldsymbol{x} = \{x_1, .., x_N\}$ 和 $\boldsymbol{y} = \{y_1, .., y_N\}$ 表示模型（如 Q-GAT）在两个不同噪声水平下的预测标签，N 表示序列长度。

这里计算了酿酒酵母数据集中 GMT 和 Q-GAT 在酿酒酵母 数据集的c = 0.15、c = 0.3 与 c = 0 条件下预测标签的相似度，以比较 GMT 和 Q-GAT 能否有效抑制噪声干扰。之所以选择 GMT 和 Q-GAT，是因为它们在稳定性方面表现最佳。表7显示，噪声越强，Q-GAT 的预测结果与无噪声时的 Q-GAT 结果仍高度一致，即使在c = 0.3 时也是如此；而 GMT 的结果与无噪声时的 GMT 结果则存在显著差异。这表明，Q-GAT 不仅能更有效地抑制噪声的影响，而且比 GMT 更稳定。

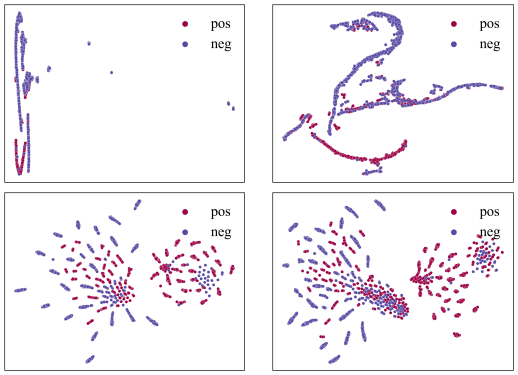
5.5. 可视化

为了更好地理解我们模型的表征能力，我们使用 t-SNE 可视化 Q-GAT 和 GMT 中间层生成的特征。t-SNE[77]是一种广泛使用的工具，可通过降维来可视化高维数据。图4展示了在扰动水平酿酒酵母 c = 0.3数据集上 Q-GAT 和 GMT 的可视化结果。在传递

卷积层之前，两个模型的特征相互混杂。特征经过卷积层处理后，便得到粗略分类。如图所示，在我们的 Q-GAT 中，同一类别的特征开始聚集，而 GMT 中的特征仍然高度混杂。在我们的 Q-GAT 中，同一类别的特征并不会聚集成一个簇，而是形成多个簇。这可能是一条分类捷径，因为将所有特征压缩到一个簇中更加困难。当特征经过池化层时，它们便得到了充分分类。在我们的 Q-GAT 中，两个类别的特征明显分离，而 GMT 生成的特征仍然混杂，这表明 Q-GAT 具有更强的表征能力。



(a) 卷积前　　(b) 池化前



(c) 池化后　　(d) 输出特征

图4：对于每个子图：第一行表示在$c = 0.3$时 GMT 在酿酒酵母数据集中的内部可视化；第二行表示在扰动水平为$c = 0.3$时，Q-GAT 在酿酒酵母数据集中的内部可视化。

5.6. 消融研究

尽管图*qttention* (GQT)和图二次池化 (GQP) 结合使用时具有出色的预测能力，但很难判断它们各自对整个模型的贡献有多大。为此，我们在酿酒酵母 数据集上进行了消融实验。将 GQT 中所有二次神经元替换为常规神经元后，我们得到了名为 Q-GAT（无 GQT）的模型。它可以视为 Q-GAT 的退化版本，即 GAT。另一方面，将 GQP 替换为全局平均池化后，即可得到 Q-GAT（无 GQP）。

表 8 显示，在不同扰动水平下，与 Q-GAT 相比，Q-GAT（无 GQT）和 Q-GAT（无 GQP）在酿酒酵母 数据集上的预测效果均有所下降。Q-GAT（无 GQT）的下降幅度最大。这说明 GQT 和 GQP 均有助于提升鲁棒性。此外可以看出，移除 GQT 后，我们的模型在酿酒酵母 数据集上的性能显著下降。移除 GQP 后，模型在酿酒酵母 数据集上的性能也有所下降，但降幅没有移除 GQT 时那么大。因此，GQT 对模型而言比 GQP 更重要。

6. 鲁棒性分析

6.1. 信噪比分析

现在，我们从信噪比（SNR）的角度阐明，为什么所提出的 Q-GAT 能够在噪声扰动下实现更高的预测性能和稳定性。信噪比是衡量信号质量的重要指标，定义为同一区域内信号功率与噪声功率之比。我们的核心观点是，二次网络之所以比其他模型，尤其是使用传统神经元的模型，能够做出更准确的预测并具备更强的噪声鲁棒性，是因为二次神经元中蕴含的非线性能够有效提升信号在各层传播过程中的信噪比。具体而言，在每一层中，二次运算都能放大信号并抑制噪声。此外，层组合的级联效应会进一步增强这一作用。信噪比提高后，有用信息会更多，不需要的噪声则会更少，从而有助于提取关键特征并识别不同特征类别。因此，网络的决策过程会更加准确、鲁棒。

神经元是神经网络（包括图神经网络）的基本组成部分。神经元的表征能力在很大程度上决定了网络的整体性能。与传统图神经网络中的线性神经元相比，二次神经元的突出特点在于简化二次函数中的非线性。Chapeau-Blondeau 等人[78] 证明，通常情况下，与线性滤波器相比，非线性器件能够提高信噪比。在同

表 8：不同扰动水平下 Q-GAT 在酿酒酵母 数据集上的消融结果（均值±偏差）。

方法	c =0.0	c =0.05	c =0.1	c =0.15	c =0.2	c =0.25	c =0.3
Q-GAT(**无 GQT**)	0.745	0.576±0.116	0.661±0.122	0.669±0.120	0.679±0.117	0.638±0.129	0.623±0.132
Q-GAT(**无 GQP**)	**0.769**	0.732±0.011	0.743±0.011	0.740±0.012	0.745±0.011	0.744±0.012	0.746±0.013
Q-GAT	0.765	**0.749±0.004**	**0.754±0.004**	**0.754±0.004**	**0.757±0.004**	**0.759±0.003**	**0.764±0.004**

理下，下面我们将从信号恢复的角度说明，二次神经元的处理能够获得比线性神经元更高的信噪比。

给定输入信号 $x(t) = s(t) + n(t)$，其中 $s(t)$ 表示纯信号，$n(t)$ 表示噪声。$x(t)$ 中纯信号 $s(t)$ 的功率可表示为

$$P_{x_s} = \frac{1}{T}\int_0^T |s(t)|^2 dt, \tag{19}$$

$x(t)$ 中噪声 $n(t)$ 的功率表示为

$$P_{x_n} = \frac{1}{T}\int_0^T |n(t)|^2 dt. \tag{20}$$

通常，信噪比定义为信号功率与噪声功率之比。为避免出现负值，输入信号的信噪比采用以下式（21）表示：

$$\mathcal{R}_{in} = 10\log\left(1 + \frac{P_{x_s}}{P_{x_n}}\right). \tag{21}$$

现在假设将信号 $x(t)$ 输入二次网络 Q，以恢复纯信号：

$$y(t) = Q[s(t) + n(t)]. \tag{22}$$

对于信号恢复任务，我们可将 $y(t)$ 改写为

$$\begin{aligned} y(t) &= Q[s(t) + n(t)] \\ &= s(t) + Q[s(t) + n(t)] - s(t) \\ &= s(t) + n'(t), \end{aligned} \tag{23}$$

其中，$n'(t)$ 被视为广义噪声。$y(t)$ 中纯信号 $s(t)$ 和噪声 $n'(t)$ 的功率分别可由式（24）和（25）表示。

$$P_{y_s} = \frac{1}{T}\int_0^T |s(t)|^2 dt \tag{24}$$

$$P_{y_n} = \frac{1}{T}\int_0^T |Q[s(t) + n(t)] - s(t)|^2 dt \tag{25}$$

因此，经过二次网络处理的信号的信噪比如下：

$$\mathcal{R}_Q = 10\log\left(1 + \frac{P_{y_s}}{P_{y_n}}\right) \tag{26}$$

如式（27）所示，V_Q 表示 $\mathcal{R}_Q$ 与 $\mathcal{R}_{in}$ 的比值，而 $V_Q > 1$ 表示输入信号信噪比的增加。

$$V_Q = \frac{\mathcal{R}_Q}{\mathcal{R}_{in}} = \frac{\log\left(1 + \frac{P_{y_s}}{P_{y_n}}\right)}{\log\left(1 + \frac{P_{x_s}}{P_{x_n}}\right)} \tag{27}$$

假设使用常规网络进行信号恢复，我们有

$$z(t) = L[s(t) + n(t)]. \tag{28}$$

其比值 V_L 表示为

$$V_L = \frac{\mathcal{R}_L}{\mathcal{R}_{in}} = \frac{\log\left(1 + \frac{P_{z_s}}{P_{z_n}}\right)}{\log\left(1 + \frac{P_{x_s}}{P_{x_n}}\right)}. \tag{29}$$

现在，我们分别训练全连接单隐藏层常规网络（NN）和二次网络（QNN），从信噪比的角度比较它们的信号恢复效果。我们设定 $s(t) = A\cos(\omega t + \varphi)$，其中 $A = 3$，以确保 $s(t)$ 不会被噪声淹没，$\omega = 2\pi$，且 $\varphi = 0$。噪声 $n(t)$ 为方差 σ^2 各不相同的高斯噪声，方差范围为 0.2 至 1。算法采用 Adam 训练 1000 个周期，初始学习率为 0.001。

表 9：特定噪声方差下，不同层数的 NN 和 QNN 的输出信噪比与输入信噪比之比 σ^2

σ^2	0.2	0.4	0.6	0.8	1
V_L（NN-1）	1.0210	0.6860	0.5523	0.7278	1.1925
V_L（NN-5）	1.0239	1.0427	1.1045	1.1342	1.1973
V_L(NN-25)	0.2107	0.2821	0.3538	0.3831	0.4135
V_Q(QNN-1)	**1.0240**	**1.0449**	**1.1055**	**1.1359**	**1.1984**

表9对比了不同噪声方差下、具有不同层数的神经网络和二次神经网络的 V_L 和 V_Q，其中 NN-5 和 NN-25 的每个中间层均有 5 个神经元。在某些噪声水平下，例如 $\sigma^2 = 0.4$、0.6，NN-1 的 V_L 远低于 1，这意味着传统神经网络几乎无法发挥作用。然而，在每个噪声水平下，V_Q 基本上始终大于 1。随着噪声水平升高，V_Q 总体呈上升趋势。这表明二次神经元能够抑制噪声，并显著提高信噪比。然而，无论层数有多少，神经网络提升信噪比的效果都不如 QNN-1，即使层数

达到 25 层也是如此。图5展示了 NN-3 和 QNN-3 重建的信号。QNN-3 能够很好地重建信号的周期性和振幅，而 NN-3 几乎无法完成重建。基于这项合成实验，我们认为二次神经元更有助于抑制噪声的影响。

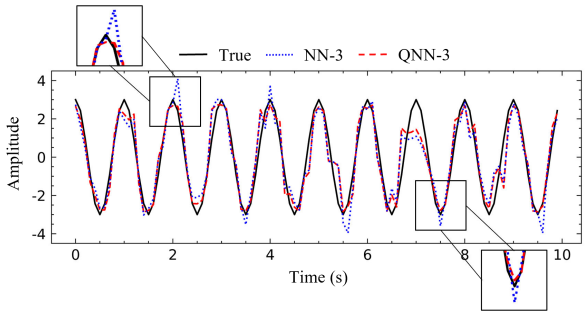


图5：3层 NN 与 QNN 重构信号的比较。

6.2. 可解释性分析

在此，我们从可解释性的角度剖析所提出的 Q-GAT 的鲁棒性。可解释性和鲁棒性是深度学习模型的两大主要弱点。前者指深度学习的预测难以解释，后者指深度学习的预测容易受到微小对抗样本的影响而改变。可解释性与鲁棒性密切相关。例如，显著性方法[79] 利用深度模型相对于输入的梯度来解释模型认为哪些内容很重要，是可解释性研究中的主流方法。有趣的是，这类梯度也被用于生成对抗样例。

我们认为，通过检查模型是否利用了正确的特征进行预测，可解释性研究能够揭示鲁棒性的来源。例如，一张含有北极熊的图片也可能包含雪或冰；模型可能会错误地利用雪或冰的信息来识别北极熊，而不是利用北极熊自身的特征。我们认为，模型如果能够尽可能多地利用有效特征，就能更好地抵御噪声扰动。沿着这一思路，众所周知，Q-GAT 是通过在 GAT 中用二次神经元替换传统神经元得到的。GAT 可以根据邻居的重要性为每个节点的邻居分配不同的权重，但它忽略了每个节点特征的重要性。然而，基因调控网络中的基因特征也包含丰富的重要信息，应当加以挖掘并适当地解读。与 GAT 相比，由于二次神经元中的*qttention* 机制，Q-GAT 还能够为节点特征分配权重，从而有助于利用更多特征信息。

图6展示了从大肠杆菌数据集中随机选取的一个子图：横轴为基因节点，纵轴为各节点的特征权重。即使在不同子图中，GAT 始终为每个节点特征分配相同的权重。然而，Q-GAT 通过分配不同的特征权重提取出一些关键节点，即某些特征有利于链接预测，而另一些则没有。这意味着，基因特征也被用于链接预测。因此，Q-GAT 是一种从特征到节点、再从节点到节点的多维图注意力模型。相比之下，GAT 只是一个从节点到节点的单层注意力模型。

接下来，我们考察 Q-GAT 挖掘出的关键节点是否确实至关重要。节点度通常用作衡量节点重要性的指标，它反映了边的连通性。节点度较高的节点通常被认为很重要。如图6底行所示，我们绘制了子图的节点度。但 Q-GAT 提取的关键节点与节点度并不十分吻合。

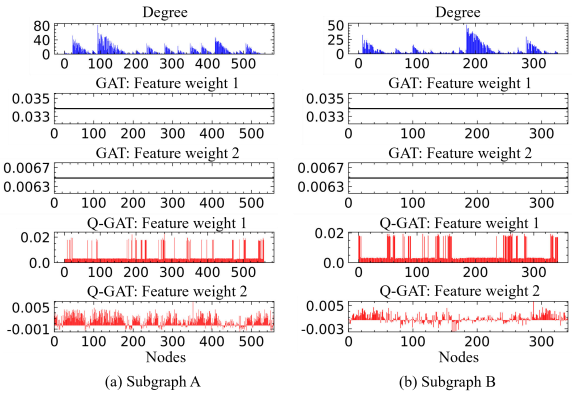


图6：子图A和子图B中 GAT 和 Q-GAT 的节点度与特征权重

因此，我们通过计算验证了 Q-GAT 提取的关键节点的重要性。在整个图的子图A中，我们将特征权重大于 0.0175 的节点设为 0，并在图中将数量相同、节点度最高的节点设为 0。随后，我们将结果与整个图的总体预测结果进行比较。表10显示，当抑制 Q-GAT 提取的关键节点的作用时，Q-GAT 的预测准确率会下降。然而，在抑制高阶节点的作用时，Q-GAT 的总体预测准确率变化不大。我们还在 GAT 模型中测试了 Q-GAT 提取的关键节点的重要性，也发现了相同的现象。这些结果表明，与节点度相比，Q-GAT 能更准确地提取重要节点。

总而言之，我们的可解释性分析表明，Q-GAT 能够在预测中同时利用特征重要性和

节点重要性。运用更多有用信息，使 Q-GAT 更有能力应对对抗性扰动。

表10：在扰动水平为 c = 0.1 时，抑制根据子图A中的qttention权重或节点度确定的关键基因所得的预测结果。

	Q-GAT	GAT
原始	0.7265	0.6997
qttention权重	0.7037	0.6810
节点度	0.7224	0.7037

7. 结论

本文提出了用于构建基因调控网络的 Q-GAT。为评估 Q-GAT 的稳定性，我们对基因表达数据添加了不同水平的扰动，然后测试了不同模型在大肠杆菌和酿酒酵母数据集上的推断性能。对比实验表明，在不同扰动水平下，Q-GAT 比现有先进方法更加稳定。此外，我们还从信号处理和可解释性的角度分析了 Q-GAT 鲁棒性更强的原因。未来，可以通过结合其他先进技术来升级 Q-GAT，进一步提升其鲁棒性。

致谢

本研究得到辽宁省自然科学基金（2022-MS-114）资助。

参考文献

[1] Daniel Marbach、James C Costello、Robert K¨uffner、Nicole M Vega、Robert J Prill、Diogo M Camacho、Kyle R Allison、Mano- lis Kellis、James J Collins 和 Gustavo Stolovitzky。群体智慧用于稳健的基因网络推断。*Nature Methods*，9(8):796–804，2012年。[2] Keiichi Mochida、Satoru Koda、Komaki Inoue 和 Ryuei Nishii。利用转录组数据集预测基因调控网络的统计与机器学习方法。*Frontiers in Plant Science*，9(1770):1–7，2018年。[3] Eric H. Davidson 和 Isabelle S. Peter。基因调控网络。载于*Genomic Control Process*，第41–77页。Elsevier，2015年。[4] Faridah Hani Mohamed Salleh、Suhaila Zainudin 和 Shereena M. Arif。用于解决级联误差问题的多元线性回归基因调控网络重建方法。*Advances in Bioinformatics*，2017年(4827171):1–14，2017年。[5] Anne-Claire Haury、Fantine Mordelet、Paola Vera-Licona 和 Jean-Philippe Vert。TIGRESS：利用稳定性选择进行可信的基因调控推断。*BMCSystemsBiology*，6(145):1–17，2012年。[6] Adam A Margolin、Ilya Nemenman、Katia Basso、Chris Wig- gins、Gustavo Stolovitzky、Riccardo Dalla Favera 和 Andrea Califano。ARACNE：一种用于在哺乳动物细胞环境中重建基因调控网络的算法。*BMC Bioinformatics*，7(S7):1–15，2006年。[7] Vˆan Anh Huynh-Thu、Alexandre Irrthum、Louis Wehenkel 和 Pierre Geurts。利用基于树的方法从表达数据推断调控网络。*PLOS ONE*，5(9):e12776，2010年。

[8] F. Mordelet 和 J.-P. Vert. SIRENE：调控网络监督式推断。*Bioinformatics*，24(16):76–82，2008年。[9] Daksh Chauhan、Emeka Anyanwu、Jacob Goes、Stephanie A. Besser、Simran Anand、Ravi Madduri、Neil Getty、Sebastian Kelle、Keigo Kawaji、Victor Mor-Avi 和 Amit R. Patel。比较机器学习与深度学习从心脏磁共振图像中识别视图的方法。*Clinical Imag- ing*，82:121–126，2022年。[10] Lu Zhang、Jianjun Tan、Dan Han 和 Hao Zhu。从机器学习到深度学习：理性药物发现中的机器智能进展。*Drug Discovery Today*，22(11):1680–1685，2017年。[11] Giang Nguyen、Stefan Dlugolinsky、Martin Bob´ak、Viet Tran、´Alvaro L´opez Garc´ıa、Ignacio Heredia、Peter Mal´ık 和 Ladislav Hluch´y。用于大规模数据挖掘的机器学习与深度学习框架和库：综述。*Arti- ficialIntelligenceReview*，52:77–124，2019年。[12] Davide Bacciu、Federico Errica、Alessio Micheli 和 Marco Podda。图深度学习入门。*Neural Networks*，129:203–221，2020年。[13] Ming Li、Zheng Ma、Yu Guang Wang 和 Xiaosheng Zhuang。用于图神经网络的快速哈尔变换。*Neural Net- works*，128:188–198，2020年。[14] Indro Spinelli、Simone Scardapane 和 Aurelio Uncini。自适应传播图卷积网络。*IEEE Transac-tions on Neural Networks and Learning Systems*，32(10):4755–4760，2021年。[15] Zonghan Wu、Shirui Pan、Fengwen Chen、Guodong Long、Chengqi Zhang 和 Philip S. Yu。图神经网络综合综述。*IEEE Transactions on Neural Networks andLearningSystems*，32(1):4–24，2021年。[16] Ye Yuan 和 Ziv Bar-Joseph。GCNG：用于从空间转录组学数据推断基因相互作用的图卷积网络。*Genome biology*，21(1):1–16，2020年。[17] Juexin Wang、Anjun Ma、Qin Ma、Dong Xu 和 Trupti Joshi。利用监督式和半监督式图神经网络对基因调控网络进行归纳推断。*Computational and structural biotechnology journal*，18:3335–3343，2020年。[18] Jeremy Bigness、Xavier Loinaz、Shalin Patel、Erica Larschan 和 Ritambhara Singh。整合长程调控相互作用，利用图卷积网络预测基因表达。*Journal of Computational Biology*，29 424，2022 5 409( ): –。[19] Alireza Karbalayghareh、Merve Sahin 和 Christina S Leslie。结合染色质相互作用信息的图注意力网络基因调控建模。*GenomeResearch*，32 944，2022 5 930( ): –。[20] Christian Szegedy、Wojciech Zaremba、Ilya Sutskever、Joan Bruna、Dumitru Erhan、Ian Goodfellow 和 Rob Fergus。神经网络的有趣特性。*arXiv preprint arXiv:1312.6199*，2014年。[21] Naveed Akhtar 和 Ajmal Mian。深度学习在计算机视觉中遭受对抗攻击的威胁：综述。*IEEE Access*，6(17677602):14410–14430，2018年。[22] Seyed-Mohsen Moosavi-Dezfooli、Alhussein Fawzi、Omar Fawzi 和 Pascal Frossard。通用对抗扰动。发表于*IEEE Conference on Computer Vision and Pattern Recogni- tion (CVPR)*，美国夏威夷，2017年。[23] Samuel G. Finlayson、Hyung Won Chung、Isaac S. Kohane 和 Andrew L. Beam。针对医学深度学习系统的对抗攻击。*arXiv preprint arXiv:1804.05296*，2019年。[24] Xingjun Ma、Yuhao Niu、Lin Gu、Yisen Wang、Yitian Zhao、James Bailey 和 Feng Lu。理解针对基于深度学习的医学图像分析系统的对抗攻击。*Pattern Recognition*，110(107332):1–11，2021年。

[25] Hanjun Dai、Hui Li、Tian Tian、Xin Huang、Lin Wang、Jun Zhu 和 Le Song。针对图结构数据的对抗攻击。*arXiv preprint arXiv:1806.02371*，2018年。[26] J. Ma、S. Ding 和 Q. Mei。迈向更实用的图神经网络对抗攻击。2020年。[27] M. Zhang、L. Hu、C. Shi 和 X. Wang。针对图神经网络的对抗标签翻转攻击与防御。发表于*2020 IEEE International Conference on Data Mining (ICDM)*，2020年。.[28] Anthony Zador、Blake Richards、Bence Olveczky、Sean Escola、Yoshua Bengio、Kwabena Boahen、Matthew Botvinick、Dmitri Chklovskii、Anne Churchland、Claudia Clopath 等。迈向下一代人工智能：推动神经人工智能革命。*arXiv preprint arXiv:2210.08340*，2022年。[29] Fenglei Fan、Jinjun Xiong 和 Ge Wang。二次深度网络的通用逼近。*Neural Networks*，124:383–392，2020年。[30] Feng-Lei Fan、Yingxin Li、Hanchuan Peng、Tieyong Zeng 和 Fei Wang。迈向神经人工智能：将神经元多样性引入人工神经网络。*arXiv preprint arXiv:230109245*，2023年。。[31] Fenglei Fan、Hongming Shan、Mannudeep K Kalra、Ramandeep Singh、Guhan Qian、Matthew Getzin、Yueyang Teng、Juergen Hahn 和 Ge Wang。用于低剂量CT去噪的二次自编码器（q-AE）。*IEEE transactions on medical imaging*，39(6):2035–2050，2019年。[32] Jing-Xiao Liao、Bo-Jian Hou、Hang-Cheng Dong、Hao Zhang、Jianwei Ma、Jinwei Sun、Shiping Zhang 和 Feng-Lei Fan。由二次神经元增强的异构自编码器。*arXiv preprint arXiv:2204.01707*，2022年。[33] Jing-Xiao Liao、Hang-Cheng Dong、Zhi-Qi Sun、Jinwei Sun、Shiping Zhang 和 Feng-Lei Fan。注意力嵌入二次网络（qttention）：实现有效且可解释的轴承故障诊断。*arXivpreprint arXiv:2206.00390*，2022年。[34] Feng-Lei Fan、Hang-Cheng Dong、Zhongming Wu、Lecheng Ruan、Tieyong Zeng、Yiming Cui 和 Jing-Xiao Liao。省下一枚神经元，便是赚得一枚神经元：论二次网络的参数效率。*arXivpreprintarXiv:2303.06316*，2023年。[35] Guohao Li、Matthias Muller、Ali Thabet 和 Bernard Ghanem。DeepGCNs：GCN能否像CNN一样加深？发表于*International Conferenceon Computer Vision (ICCV)*，韩国首尔，2019年。[36] William L. Hamilton、Rex Ying 和 Jure Leskovec。大型图上的归纳式表示学习。发表于*Proceedings of Thirty-first Conference on Neural Information Processing Sys- tems(NIPS)*，美国加利福尼亚州长滩，2017年。[37] Filippo Maria Bianchi、Daniele Grattarola、Lorenzo Livi 和 Cesare Alippi。采用卷积ARMA滤波器的图神经网络。*IEEE Transactions on Pattern Analysis and Machine In- telligence*，44(7):3496–3507，2022年。[38] Muhan Zhang、Zhicheng Cui、Marion Neumann 和 Yixin Chen。端到端深度学习图分类架构。发表于*Proceedings of Thirty-Second AAAI Conference onArtificialIntelligence (AAAI)*，美国路易斯安那州新奥尔良，2018年。[39] Zhitao Ying、Jiaxuan You、Christopher Morris、Xiang Ren、Will Hamilton 和 Jure Leskovec。基于可微池化的层次化图表示学习。发表于*Proceedings of Thirty-second Conference on Neural Information Processing Systems (NIPS)*，加拿大蒙特利尔，2018年。[40] Junhyun Lee、Inyeop Lee 和 Jaewoo Kang。自注意力图池化。*arXivpreprintarXiv:1904.08082*，2019年。[41] Hongyang Gao 和 Shuiwang Ji。图U-Net。*IEEETransactions on Pattern Analysis and Machine Intelligence*，(33999813):1–1，2021年。[42] Petar Velickovic、Guillem Cucurull、Arantxa Casanova、Adri- ana Romero、Pietro Lio’ 和 Yoshua Bengio。图注意力网络。载于《第六届国际学习表征会议论文集》（ICLR），加拿大温哥华，2018年。 [43] Jinheon Baek、Minki Kang 和 S ung Ju Hwang。利用图多重集池化准确学习图表征。arXiv预印本 arXiv:2102.11533，2021年。 [44] Peter Meltzer、Marcelo Daniel Gutierrez Mallea 和 Peter J. Bentley。PiNet：用于图分类的注意力池化。arXiv预印本 arXiv:2008.04575，2020年。[45] Christian Szegedy、Wojciech Zaremba、Ilya Sutskever、Joan Bruna、Dumitru Erhan、Ian Goodfellow 和 Rob Fergus。神经网络的有趣特性。arXiv预印本 arXiv:1312.6199，2013年。[46] Naveed Akhtar 和 Ajmal Mian。对抗攻击对计算机视觉深度学习的威胁：综述。IEEE Access，6:14410–14430，2018年。[47] Seyed-Mohsen Moosavi-Dezfooli、Alhussein Fawzi、Omar Fawzi 和 Pascal Frossard。通用对抗扰动。载于《IEEE计算机视觉与模式识别会议论文集》，第1765–1773页，2017年。[48] Hanjun Dai、Hui Li、Tian Tian、Xin Huang、Lin Wang、Jun Zhu 和 Le Song。针对图结构数据的对抗攻击。载于国际机器学习会议，第1115–1124页。PMLR，2018年。 [49] Xiaojin Zhu 和 Zoubin Ghahramani。利用标签传播从有标签和无标签数据中学习。ProQuest编号：致所有用户的信息，2002年。[50] Dengyong Zhou、Olivier Bousquet、Thomas Lal、Jason Weston 和 Bernhard Sch¨olkopf。利用局部与全局一致性进行学习。《神经信息处理系统进展》，16，2003年。[51] Paolo Mignone、Gianvito Pio、Savso Dvzeroski 和 Michelangelo Ceci。多任务学习：同时重建人类和小鼠基因调控网络。《Scientific Reports》，10(1):22295，2020年。[52] Junshui Ma、Robert P Sheridan、Andy Liaw、George E Dahl 和 Vladimir Svetnik。将深度神经网络用作定量构效关系方法。《Journal of Chemical Information and Modeling》，55(2):263–274，2015年。 [53] Jiarui Ding、Anne Condon 和 Sohrab P Shah。利用深度生成模型对单细胞转录组数据进行可解释的降维。《Nature Communications》，9 2002，2018 1( ): 。[54] Petar Velickovic、Guillem Cucurull、Arantxa Casanova、Adriana Romero、Pietro Lio 和 Yoshua Bengio。图注意力网络，国际学习表征会议。载于《国际学习表征会议论文集》，第1–2页，2018年。[55] Vladimir Gligorijevi´c、Meet Barot 和 Richard Bonneau。deepnf：用于蛋白质功能预测的深度网络融合。《Bioinformatics》，34(22):3873–3881，2018年。[56] Daniel Marbach、James C Costello、Robert K¨uffner、Nicole M Vega、Robert J Prill、Diogo M Camacho、Kyle R Allison、Manolis Kellis、James J Collins 等。群体智慧用于稳健的基因网络推断。《Nature Methods》，9(8):796–804，2012年。 [57] V̂ an Anh Huynh-Thu、Alexandre Irrthum、Louis Wehenkel 和 Pierre Geurts。使用基于树的方法从表达数据推断调控网络。《PLOS ONE》，5(9):e12776，2010年。 [58] Nadav Bar、Bahareh Nikparvar、Naresh Doni Jayavelu 和 Fabienne Krystin Roessler。对短期时间序列基因表达数据进行约束傅里叶估计，可降低噪声并改进聚类和基因调控网络预测。《BMC Bioinformatics》，23(1):1–21，2022年。[59] Alexey Grigorevich Ivakhnenko。复杂系统多项式理论。《IEEE系统、人类与控制论汇刊》，(4):364–378，1971年。

[60] Grigoris Chrysos、Stylianos Moschoglou、Giorgos Bouritsas、Jiankang Deng、Yannis Panagakis 和 Stefanos P Zafeiriou。深度多项式神经网络。*IEEE*《模式分析与机器智能汇刊》，2021年。[61] Gang Liu 和 Jing Wang。树突网络：用于分类、回归和系统辨识的白盒模块。*IEEE*《控制论汇刊》，2021年。[62] Fenglei Fan、Wenxiang Cong 和 Ge Wang。一种用于机器学习的新型神经元。《生物医学工程数值方法国际期刊》，34(2):e2920，2018年。[63] Yiyang Jiang、Fan Yang、Hengliang Zhu、Dian Zhou 和 Xuan Zeng。非线性卷积神经网络：通过二次卷积改进卷积神经网络。《神经计算与应用》，32 8516，2020 12 8507( ): – 。[64] Pranav Mantini 和 Shishr K Shah。卷积二次神经网络。载于*2020*年第*25*届国际模式识别会议（*ICPR*），第9819–9826页。IEEE，2021年。[65] Mohit Goyal、Rajan Goyal 和 Brejesh Lall。采用归一化激活的改进多项式神经网络。载于*2020*年国际神经网络联合会议（*IJCNN*），第1–8页。IEEE，2020年。[66] Zirui Xu、Fuxun Yu、Jinjun Xiong 和 Xiang Chen。QuadraLib：用于架构优化与设计探索的高性能二次神经网络库。《机器学习与系统论文集》，4:503–514，2022年。[67] Georgios Zoumpourlis、Alexandros Doumanoglou、Nicholas Vretos 和 Petros Daras。用于基于卷积神经网络学习的非线性卷积滤波器。载于*IEEE*国际计算机视觉会议论文集，第4761–4769页，2017年。[68] Jie Bu 和 Anuj Karpatne。二次残差网络：一类用于求解涉及偏微分方程的物理学正问题和逆问题的新型神经网络。载于第*2021*年*SIAM*国际数据挖掘会议（*SDM*）论文集，第675–683页。SIAM，2021年。[69] Nikolaos Tsapanos、Anastasios Tefas、Nikolaos Nikolaidis 和 Ioannis Pitas。采用抛物面决策边界的神经元可提升神经网络分类性能。*IEEE*《神经网络与学习系统汇刊》，30(1):284–294，2018年。[70] Aditya Grover 和 Jure Leskovec。Node2vec：面向网络的可扩展特征学习。载于第*22*届*ACM SIGKDD*国际知识发现与数据挖掘会议论文集，美国旧金山，2016年。[71] Fenglei Fan、Jinjun Xiong 和 Ge Wang。二次深度网络的通用逼近能力。《神经网络》，124:383–392，2020年。[72] Feng-Lei Fan、Mengzhou Li、Fei Wang、Rongjie Lai 和 Ge Wang。二次网络的表达能力与可训练性。*arXiv*预印本，2021年。[73] Jie Zhou、Ganqu Cui、Shengding Hu、Zhengyan Zhang、Cheng Yang、Zhiyuan Liu、Lifeng Wang、Changcheng Li 和 Maosong Sun。图神经网络：方法与应用综述。《*AI Open*》，1:57–81，2020年。[74] Fenglei Fan、Wenxiang Cong 和 Ge Wang。一种用于机器学习的新型神经元。《生物医学工程数值方法国际期刊》，34，2018年。[75] David J. Miller、Zhen Xiang 和 George Kesidis。针对深度神经网络分类的对抗学习：攻击防御综合综述。《*IEEE*论文集》，108(3):402–433，2020年。[76] Liang Chen、Jintang Li、Jiaying Peng、Tao Xie、Zengxu Cao、Kun Xu、Xiangnan He 和 Zibin Zheng。图上的对抗学习综述。*arXiv*预印本 *arXiv:2003.05730*，2020年。[77] Laurens van der Maaten 和 Geoffrey Hinton。可视化使用 t-SNE 对数据进行处理。机器学习研究期刊，9(86):2579–2605, 2008. [78] F. Chapeau-Blondeau 和 D. Rousseau. 非线性器件在噪声中的谐波信号上充当信噪比放大器。电路、系统与信号处理, (3), 2006. [79] Feng-Lei Fan、Jinjun Xiong、Mengzhou Li 和 Ge Wang. 关于人工神经网络可解释性的综述。*IEEE*辐射与等离子体医学科学汇刊，5(6):741–760, 2021年。