

学校代码: 10284
分类号: TP181
密级: 公开
UDC: 004.8
学号: MG21370043



南京大学

硕士学位论文

论文题目 基于高低频与最小显著
信息的图神经网络研究

作者姓名 熊昕

专业名称 计算机科学与技术

研究方向 图神经网络

导师姓名 申富饶 教授

2024年5月27日


答辩委员会主席 戴新宇 教授

评 阅 人 张 荆 高工

徐明华 教授

论文答辩日期 2024年5月16日

研究生签名: 熊 昕

导师签名: 

Research on Graph Neural Networks Based on Low-High Frequency and Minimal Noteworthy Information

by
Xiong Xin

Supervised by
Professor Shen Fu-Rao

A dissertation submitted to
the graduate school of Nanjing University
in partial fulfilment of the requirements for the degree of

MASTER

in

Computer Science and Technology



Nanjing University

May 27, 2024

南京大学研究生毕业论文中文摘要首页用纸

毕业论文题目：基于高低频与最小显著信息的图神经网络研究

计算机科学与技术 专业 2021 级硕士生姓名：熊昕

指导教师（姓名、职称）：申富饶 教授

摘 要

深度学习技术在处理结构化数据方面大放异彩，其应用领域涵盖了计算机视觉、自然语言处理以及时序信号处理等多个领域。然而，实际生活中的大部分数据都是非结构化的，例如蛋白质结构、社交网络、交通线路网络等。图神经网络的出现正是为了处理这类同时具有拓扑结构与特征信息的非结构化数据。图神经网络的研究主要集中在三个方向：图神经网络模型、图表示学习以及图表示学习的应用。因此，本文依据这三个方向展开了研究工作，提出了一种图神经网络模型：SGCN-MH，以及一种图表示学习方法：GMNI，并将这两项算法应用于医疗领域。

1. 频域图卷积网络存在着高频信息丢失、过平滑现象以及较大的计算量等问题，本文针对这些问题提出了基于频率自适应的简化频域图卷积网络模型，名为SGCN-MH。首先，本文使用了3种方法探究图中高低频信息的作用。SGCN-MH使用单跳及跨跳连接图卷积完成了自动学习高低频信息的目标。其次，本文探究了6种层间交互方式，旨在提升模型的表达能力并减轻过度平滑现象。最后，SGCN-MH简化了图卷积网络的传播模式，以减少模型的参数量和防止过拟合。在浅层和低参数量的情况下，SGCN-MH在4个数据集上取得了优于其他基线图卷积网络的表现，基线包括深层图卷积网络。

2. 许多无监督图对比学习方法存在着信息不足或信息冗余风险，没有达成对比学习中的最优视角。本文提出了一种基于最小显著信息的无监督图对比学习方法：GMNI，它通过对抗训练策略来生成共享最小显著信息的视图，去除了上述两种风险。此外，GMNI还引入了基于最小显著信息的随机性来增强视图的多样性并使模型对扰动更加稳定。对无监督和半监督条件下的14个数据集的节点分类和图分类的广泛实验表明，GMNI相较于自动和手动数据增强的其他图

对比学习方法具有显著优势，它实现了无监督条件下对最优视图的更好的近似。

3. 图表示学习因其可解释性良好，可以用于诸如医疗、金融等要求严谨的领域。本文将图表示学习应用于相似药物推荐任务。本文首先根据医疗病理数据集构建了药物-疾病关系图，改造 SGCN-MH 以及 GMNI 在该图上学习药物的专业性表征，并根据该表征来获得相似药物的推荐结果。实验结果表明，推荐结果具有可解释性与良好的泛化性能。此外，开发的相似药物推荐系统方便了用户查询和展示相似药物的推荐结果。

关键词：图卷积网络；图表示学习；图对比学习；数据增强

南京大学研究生毕业论文英文摘要首页用纸

THESIS: Research on Graph Neural Networks Based on Low-High Frequency and Minimal Noteworthy Information

SPECIALIZATION: Computer Science and Technology

POSTGRADUATE: Xiong Xin

MENTOR: Professor Shen Fu-Rao

ABSTRACT

The advent of deep learning has made remarkable strides in processing structured data, with applications spanning various fields, including computer vision, natural language processing, and time-series signal processing. However, most real-life data, such as protein structures, social networks, and transportation networks, is unstructured. The emergence of graph neural networks aims to tackle unstructured data that simultaneously possess topological structure and feature information. Research on graph neural networks primarily focuses on three directions: the architecture of graph neural network models, graph representation learning, and the application of graph representation learning. Accordingly, this paper unfolds its research endeavors based on these three directions. It proposes a novel graph neural network model, SGCN-MH, and a unique graph representation learning approach, GMNI, both of which are applied in healthcare.

1. Spectral graph convolutional networks face challenges like high-frequency information loss, oversmoothing phenomenon, and computational overhead. To tackle these challenges, this paper introduces SGCN-MH, a simplified frequency-adaptive graph convolutional network. Firstly, this paper investigates the roles of low and high-frequency information in graphs using three approaches. SGCN-MH achieves automatic learning of these frequencies through single-hop and multi-hop connection convolutional layers. Secondly, this paper explores six inter-layer interaction mechanisms to enhance model expressiveness and alleviate the oversmoothing phenomenon. Finally, SGCN-MH simplifies the propagation pattern of the graph convolutional network to reduce model parameters and prevent overfitting. Under shallow and low-parameter condi-

tions, SGCN-MH outperforms other baseline graph convolutional networks, including deep graph convolutional networks, across 4 datasets.

2. Many unsupervised graph contrastive learning methods face risks of insufficient information or information redundancy, failing to achieve the optimal views in contrastive learning. This paper proposes an unsupervised Graph contrastive learning method based on Minimal Noteworthy Information: GMNI. It employs adversarial training strategies to generate views sharing Minimal Noteworthy Information (MNI), thereby eliminating the risks above. Moreover, GMNI introduces randomness based on MNI to enhance view diversity and stabilize the model against perturbations. Extensive experiments on node classification and graph classification tasks across 14 datasets under unsupervised and semi-supervised settings demonstrate significant advantages of GMNI over other graph contrastive learning methods with automated and manual data augmentation. GMNI achieves a better approximation of the optimal view under an unsupervised setting.

3. Due to its excellent interpretability, graph representation learning can be applied to rigorous domains such as healthcare and finance. This paper applies graph representation learning to the task of recommending similar chemicals. This paper constructs a chemical-disease graph based on the Comparative Toxicogenomics Database and adapts SGCN-MH and GMNI to learn representations of chemicals. Leveraging these representations, recommendations for similar chemicals are obtained. Experimental results demonstrate that the recommendations are interpretable and exhibit good generalization performance. Moreover, a similar chemical recommendation system has been developed, facilitating user queries and showcasing the recommendation results.

KEYWORDS: Graph Convolutional Networks; Graph Representation Learning; Graph Contrastive Learning; Data Augmentation

目 录

中文摘要	I
ABSTRACT	III
目 录	V
插图目录	1
表格目录	3
第一章 绪论	5
1.1 研究背景	5
1.2 研究现状与挑战	6
1.3 研究内容与贡献	10
1.4 本文结构安排	12
第二章 相关工作	15
2.1 图卷积网络模型	15
2.1.1 卷积神经网络的图推广网络 Graph CNN	15
2.1.2 采样与聚合图卷积网络 GraphSAGE	17
2.1.3 图注意力网络 GAT	18
2.1.4 频率自适应图卷积网络 FAGCN	20
2.1.5 自适应图编码器 AGE	21
2.1.6 简化图卷积网络 SGC	23
2.2 图对比学习	24
2.2.1 互信息最小化原则 InfoMin 与互信息最大化原则 InfoMax	24
2.2.2 联合增强优化框架 JOAO	25

2.2.3	对抗图数据增强图对比学习 AD-GCL	26
2.2.4	自动图对比学习 AutoGCL	27
2.3	本章小结	28
第三章 基于频率自适应的简化频域图卷积网络		29
3.1	研究背景与动机	29
3.2	模型	31
3.2.1	符号定义	31
3.2.2	前置知识: 频域图卷积网络	33
3.2.3	图网络中的高低频信息配比	36
3.2.4	层间信息交互方式	39
3.2.5	跨跳连接的简化图卷积网络 SGCN-MH	42
3.3	实验与分析	44
3.3.1	数据集	44
3.3.2	对比基线及实验设置	45
3.3.3	对比实验	46
3.3.4	消融实验	47
3.3.5	高低频信息的选择实验	49
3.3.6	多层信息交互方式实验	52
3.3.7	合适信息传播模式实验	53
3.4	本章小结	53
第四章 基于最小显著信息的无监督图对比学习		55
4.1	研究背景与动机	55
4.2	模型	61
4.2.1	符号定义	61
4.2.2	自动视图生成器	62
4.2.3	视图对比模块	68
4.2.4	算法与时间复杂度	69
4.3	实验与分析	69
4.3.1	数据集	69

4.3.2	对比基线	71
4.3.3	实验设置	74
4.3.4	对比实验	75
4.3.5	消融实验	78
4.3.6	附加实验	79
4.4	本章小结	82
第五章	图表示学习在相似药物推荐中的应用	83
5.1	研究背景与动机	83
5.2	数据分析	85
5.3	需求分析	86
5.4	系统架构	88
5.4.1	关系图构建	88
5.4.2	获得药物嵌入的图表示学习模型	89
5.5	应用展示	91
5.6	本章小结	93
第六章	总结与展望	97
	参考文献	99
	致 谢	113
	简历与科研成果	115

插图目录

1-1	图神经网络框架及本文研究内容	9
1-2	本文整体组织结构	12
2-1	GraphSAGE 的算法步骤	17
2-2	GAT 的聚合过程	18
3-1	频域图卷积过程	33
3-2	修改参数 t , 模型在 AMusic 数据集上的表现	37
3-3	跨跳连接与单跳连接图卷积过程	38
3-4	SGCN-MH 模型结构	42
3-5	叠加多层 GCN 卷积层的测试准确率折线图 (平均准确率 Accuracy (%) \pm 标准差)	48
3-6	改变广义拉普拉斯卷积核参数 κ 的测试准确率折线图 (平均准确率 Accuracy (%) \pm 标准差)	49
3-7	2-Hop 跨跳连接图卷积层平衡系数变化折线图	50
4-1	图对比学习过程	56
4-2	用于 OCR 识别的图片	59
4-3	InfoMin 原则定义出的最优视图阐释	60
4-4	自动视图生成器生成重要性图的前向过程	62
4-5	自动视图生成器生成重要性图的优化过程	62
4-6	视图对比模块	68
4-7	重要性图的热力图可视化	80
4-8	GMNI 在 REDDIT-BINARY 数据集上的无监督图分类性能	81
4-9	Amazon-Photo 数据集上原始特征和不同基线学习到节点嵌入的 t-SNE 可视化图	82

5-1	美林牌布洛芬混悬液在美团的页面	84
5-2	药物与疾病关系图	85
5-3	药物、疾病节点度的直方图	86
5-4	相似药物推荐系统框架及流程图	87
5-5	通过图表示学习获得相似药物流程图	88
5-6	获取药物嵌入的图表示学习网络结构	89
5-7	药物库展示界面	92
5-8	药物/疾病搜索界面	92
5-9	药物/疾病关联查询界面	93
5-10	相似药物推荐页面，排序为 0 的药物为查询药物	94

表格目录

3-1	传统傅立叶变换与图傅立叶变换的对比	33
3-2	比较不同图卷积网络的卷积核	35
3-3	节点分类数据集	44
3-4	基线模型在有监督节点分类任务上的表现 (平均准确率 Accuracy (%))	47
3-5	SGCN 与 GCNII 的参数量对比	47
3-6	SGCN-MH 的消融实验 (平均准确率 Accuracy (%) \pm 标准差 (图卷积层层数))	48
3-7	跨跳连接图卷积核的作用 (平均准确率 Accuracy (%) \pm 标准差) .	49
3-8	不同层间交互方式 GCN 的测试准确率 (平均准确率 Accuracy (%) \pm 标准差 (图卷积层层数))	51
3-9	去除了 GCN 卷积层中的特征变换 (GCN-NW) 与非线性激活函数 (GCN-NA) 后的测试准确率 (平均准确率 Accuracy (%) \pm 标准差 (图卷积层层数))	52
4-1	节点分类数据集	72
4-2	图分类数据集	72
4-3	GMNI 的其他超参数	73
4-4	模型在无监督节点分类任务上的表现 (重复运行了 5 次之后的平均准确率 Accuracy (%) \pm 标准差)	74
4-5	模型在无监督图分类任务上的表现 (重复运行了 5 次之后的平均准确率 Accuracy (%) \pm 标准差)	75
4-6	模型在半监督图分类任务上的表现 (重复运行了 5 次之后的平均准确率 Accuracy (%) \pm 标准差)	76

4-7 GMNI 的消融实验（重复运行了 3 次之后的平均准确率 Accuracy （%） \pm 标准差）	77
5-1 CTD 中药物和疾病节点的度频数分布	85
5-2 相似药物 Top 10 推荐结果（中文）	95

第一章 绪论

1.1 研究背景

在信息时代，随着海量数据的迅速涌现和机器算力的不断提升，深度学习技术正迎来其黄金时代。深度学习模型在处理多种结构化数据方面展现出卓越的性能，涵盖语音、文本和图像等领域。其中，经典研究包括语音识别领域的 Deep Speech^[1]、LAS^[2]、WaveNet^[3]；自然语言处理领域的 Seq2Seq^[4]、Transformer^[5]、BERT^[6]；以及计算机视觉领域的 AlexNet^[7]、ResNet^[8]、ViT^[9]等。

然而，深度学习在处理实际世界中大多数数据时面临一个挑战，即这些数据无法简单地表示成结构化的序列或网格结构。例如，复杂的社交网络、多变的蛋白质结构和错综的交通路线都属于这类非结构化的数据。这些数据不仅规模庞大，如社交网络中的节点数量可能高达数十亿，而且结构复杂，如社交网络中的每个节点都可能与其他节点产生不同类型的交互，这为深度学习带来了巨大的挑战。结构化数据被称为欧几里得结构数据（Euclidean Structure Data），而非结构化数据被称为非欧几里得结构数据（Non-Euclidean Structure Data）。欧几里得结构数据被看作是在多维线性空间中的数据点，它们遵循欧几里得几何的基本公设。例如，一个大小为 $M \times N$ 的图像被视为在 $M \times N$ 维线性空间中的一个点。非欧几里得结构数据的特点是其结构不规整，难以简单地映射到固定维度的线性空间。在这类数据中，每个节点的邻居数量和连接情况可能都不同。这些非欧几里得数据中蕴含了丰富的拓扑结构和特征信息。为了有效地挖掘这些宝贵的信息，图神经网络（Graph Neural Network, GNN^[10-12]）应运而生，它被设计用来处理和分析这种复杂的非欧几里得数据。图是由节点和边组成的抽象数据结构，是一种更加广义和抽象的非欧几里得结构。

图神经网络的输入是一张图，经过多层的图神经网络层获取节点嵌入或再经过图池化层获取图嵌入。获取节点嵌入的过程被称为图节点表示学习，它的

核心思想是利用中心节点本身的特征以及其邻居节点的特征来更新中心节点的代表向量，常见的更新方式包括邻居节点特征的求和、求平均或取最大值等。模型经过前向传播获得节点嵌入或图嵌入后，将根据具体任务和数据情况设计的损失函数对模型进行反向传播学习。通常的图任务可以划分为节点级别、边级别和图级别。

节点级别的图任务，模型需要通过学习图节点之间的连接关系和节点自身的特征来预测节点的属性或标签，从而实现节点级别的分类、回归或其他预测任务。例如，研究者可以根据论文的引用关系图，论文节点特征以及论文的类别完成预测未知论文节点类别的分类任务。处理节点级别图任务的经典图网络包括：GAT^[11]，GraphSAGE^[12]和GIN^[13]等。边级别的图任务，模型需要根据节点连接关系、节点特征、边特征来预测边的属性或标签。其经典例子是链接预测任务，即预测图中节点间尚未出现的边是否有几率连接。这种任务可以帮助发现新的社交关系、分析推荐用户间的潜在联系，或者发现知识图谱中的潜在关联，这对于提升社交网络的挖掘能力、改进推荐系统的准确性以及构建更加完善的知识图谱都具有重要意义。处理边级别图任务的经典图网络包括：STGCN^[14]、LightGCN^[15]、NGCF^[16]等。图级别的图任务旨在对整个图结构进行深入分析和综合理解，而不仅仅是关注单个节点或边的特性。这类任务通常要求从全局角度预测或学习图的整体属性，如图的社区结构、连通性和密度等重要性质。图级别任务的一个经典例子是图分类问题，即给定一个图，预测其所属的类别或标签。处理图级别图任务的经典图网络包括：DNGR^[17]、DGCNN^[18]、TransE^[19]等。

考虑到非欧几里得数据在人们的日常生活中无处不在，而三种级别的图任务涉及到多种现实场景，图神经网络的应用范围变得异常广泛。从个性化推荐，到社交网络分析，再到生物信息学和金融欺诈检测等领域，图神经网络都发挥着重要的作用。它为解决现实世界中的复杂问题提供了有效的工具和方法，因此图神经网络的研究显得尤为必要。

1.2 研究现状与挑战

图神经网络的研究可以被分为以下几个主要方向：

1. 图神经网络模型 (Graph Neural Networks, GNN)：该研究着眼于设计更

加高效和灵活的网络结构，以适应不同类型的图数据和任务。常见的模型包括基于卷积神经网络 (Convolutional Neural Networks, CNN^[20]) 的模型：图卷积网络 (Graph Convolutional Network, GCN^[10])、基于注意力的模型：图注意力网络 (Graph Attention Network, GAT^[11])、基于循环神经网络 (Recurrent Neural Network, RNN^[21]) 的模型：门控图序列神经网络 (Gated Graph Sequence Neural Network, GGSNN^[22])、基于自动编码器的模型：图自编码器 (Graph Autoencoders^[23])、基于生成式的模型：图生成模型 (Graph Generative Models, 如 MolGAN^[24]) 等。

除了根据基本结构的不同来划分模型外，还可以根据输入图的数据形式来设计图网络。图可以呈现多种类型，包括具有单一类型边和节点的同质图 (Homogenous Graph)、具有多种类型边和节点的异质图 (Heterogeneous Graph)、以及单一节点类型和多类边关联的多维图 (Multi-dimensional Graph)、边可以关联任意数目节点的超图 (Hypergraph)，以及随时会加入新边和新节点的动态图 (Dynamic Graph) 等。之前举例的网络基本上用于处理同质图，而处理异质图的网络如 HAN^[25]、处理多维图的网络如 mGCN^[26]、处理超图的网络如 HGNN^[27]、处理动态图的网络如 WD-GCN^[28]等。这些网络的设计考虑到了不同类型图的特点，使得图神经网络能够更加灵活地应对各种图数据的形式和复杂性。

处理不同图数据的网络基本启迪于同质图网络，它们根据新加入的信息来改进网络以适应新的数据类型。这些网络的演进实际上关联于实际应用需求。例如，LightGCN^[15]用于完成推荐系统中的协同过滤算法，适用于二部图。二部图实际上是一种具有两类节点和一类边的异质图，它改进了 GCN 算法来完成用户-物品的喜好预估任务。

2. 图表示学习 (Graph Representation Learning, GRL)：该研究着眼于如何有效地将图中的节点和边表示为低维向量，以便进行后续的机器学习任务。图表示学习可以帮助研究者们将复杂的图结构转化为向量表示，从而实现对图数据的有效建模和分析。

基于矩阵分解的方法^[29-31]和基于随机游走的方法^[32-33]是用于图表示学习的经典方法。基于矩阵分解的方法通常通过对邻接矩阵进行矩阵分解来产生低维表示，以保留原始空间中的相似性。这种方法计算成本高，难以在大规模图上实现，并且无法利用特征信息。基于随机游走的方法通过在图上进行随机游走 (Random Walk) 建立多个节点序列，然后将诸如 Word2Vec^[34], skip-gram^[35]等语言

模型应用于这些序列以学习表示。DeepWalk^[32]搭配了随机游走与 skip-gram^[35]来学习节点表示。DeepWalk 的缺点是缺乏明确的优化目标,于是 LINE^[36]提出了两种目标函数:一阶邻域相似性 (First-order proximity) 和二阶邻域相似性 (Second-order proximity)。通过最大化这两种相似性,LINE 把网络的局部信息和全局信息分开学习,从而保留了图中的局部和全局信息。Node2Vec^[33]同样通过在图中执行随机游走,并将节点序列输入 Word2Vec 模型进行训练,从而学习节点的低维嵌入。Node2Vec 定义了一种灵活的随机游走策略,用于在图中生成节点序列。该策略通过控制两个参数,即返回参数 (return parameter) 和进出参数 (in-out parameter),来平衡广度优先搜索 (BFS) 和深度优先搜索 (DFS) 之间的权衡关系。然而,这些方法往往过分强调拓扑信息,而忽略特征信息。近年来,深度图表示学习^[37-38] 逐渐发展,这类方法往往具有更强的表达能力。SDNE^[39]将图结构表示为邻接矩阵,并利用自编码器将原始的高维图结构编码成低维稠密向量表示。在这个过程中,SDNE 采用了局部和全局约束来保持图的重要信息,最终通过最小化重构损失来训练网络。

图对比学习作为一种有前景的深度图表示学习方法逐渐崭露头角。这种方法的主要思想是通过学习图中节点或子图之间的相似性或差异性来进行表示学习。其核心目标是将相似的节点或子图映射到嵌入空间中的相近位置,而将不相似的节点或子图映射到嵌入空间中的远离位置。图对比学习的经典工作包括 DGI^[40]、GraphCL^[41]、MVGRL^[42]等。这些方法通过不同的对比损失函数或对比学习框架来实现图结构的表示学习。本文将在后面的章节中具体介绍图对比学习。

3. 图表示学习的应用: 该研究着眼于如何利用图表示学习解决实际的应用问题,如社交网络分析、推荐系统、生物信息学、金融欺诈检测等领域的任务。这些任务往往涉及对大规模、复杂图数据的处理和分析,一方面,图表示学习能够有效地将图中的节点映射到低维向量空间中,并保留图结构和图特征的重要信息;另一方面,图表示学习具有较强的可解释性,可以很好的应用于如金融欺诈检测这种严谨的场景中。

在社交网络分析领域,图表示学习用于发现社交网络中的社区结构、识别关键节点和评估社交影响力、预测用户行为等任务^[43]。在推荐系统领域,图表示学习用于学习用户和物品的表示,从而实现个性化推荐^[44-45]。通过将用户和物

品表示为图中的节点，并利用图结构信息来捕捉用户和物品之间的关系，可以提高推荐系统的效果和准确性。图网络可以很好的整合多种来源的信息，从而获取更完备的用户及物品表示。在生物信息学领域，图表示学习可以用于分析生物分子之间的相互作用关系、预测蛋白质的结构和功能等任务^[46-47]。通过将生物分子表示为图中的节点，并利用图结构信息来捕捉生物分子之间的相互作用关系，可以帮助研究人员更好地理解生物体系的复杂性。在金融欺诈检测领域，图表示学习用于分析金融交易网络中的欺诈行为、识别欺诈节点和异常交易模式等任务^[48-49]。通过将金融交易网络表示为图，利用图表示学习来挖掘交易网络中隐藏的欺诈模式和异常行为，可以帮助金融机构及时发现和防范欺诈风险。

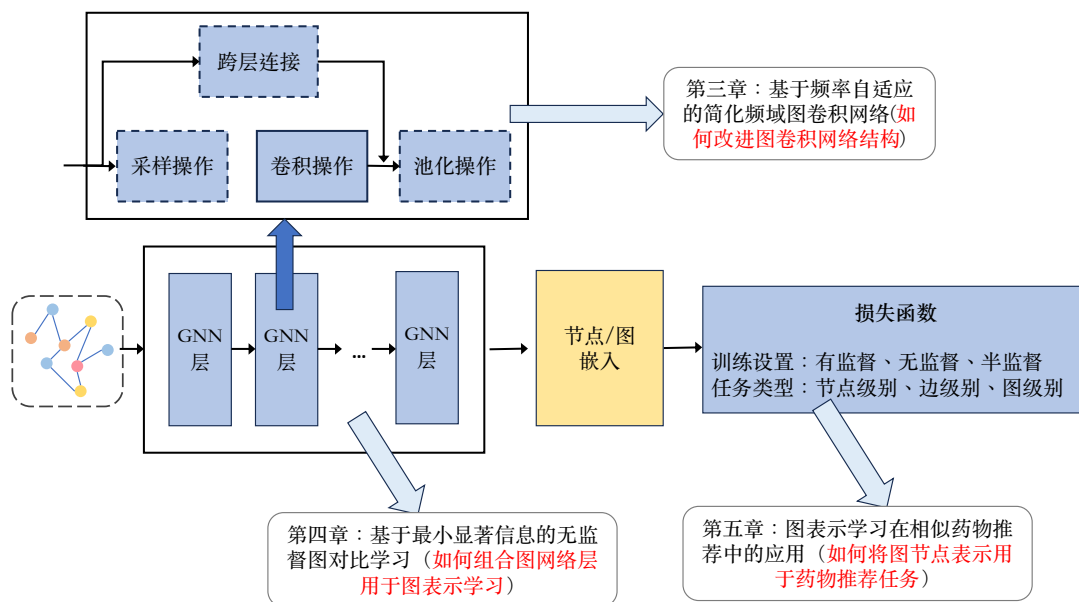


图 1-1 图神经网络框架及本文研究内容

图1-1展示了整个图神经网络的框架以及本文的研究内容。输入图经过多层的图网络层来获取节点表示或者再经过图池化层获得图表示，损失函数需要根据具体任务类型（节点级别，边级别，图级别）和数据形式（有监督，无监督，半监督）来设计。在图中的 GNN 层展开的结构中，虚线部分是可选的。采样操作（Sampling Operator）用于采样出部分节点和边，通过采样数据学习出的模型具有较好的泛化性；卷积操作（Conv Operator）是图卷积操作，用于聚合邻居信息并进行信息传播；跨层连接（Skip Connection）来源于 ResNet^[50]，用于实现图神经网络层间的信息交互；池化操作（Pooling Operator）用于根据节点嵌入来获得图嵌入。这张图展示了图神经网络的三个研究方向。首先，图神经网络模型的研究

涉及到图中 GNN 层的设计，即如何提出和组合这些不同的操作元。其次，图表示学习关乎如何获得节点嵌入 (Node Embedding) 与图嵌入 (Graph Embedding) 的方式，这涉及到如何组合不同的 GNN 层以及如何设计损失函数。最后，图表示学习的应用则是关注如何生成和充分利用学习到的节点与图嵌入。本文的研究依据这三个方向展开，第三章与第四章的算法将被融合用于第五章的应用。

除了以上提到的三个研究方向外，图神经网络还涉及到诸如图网络可解释性、图谱理论等的研究方向。这些方向在图数据分析和应用中也起着重要的作用。图网络可解释性关注如何解释图神经网络的输出结果，以便更好地理解模型的决策过程和输出结果，从而将结果应用于严谨的诸如医学和金融等领域^[13]。图谱理论研究图网络的理论基础，其研究不仅有助于深入理解图数据的本质，还能够为图神经网络的设计和应用提供理论指导^[51]。

下面将介绍图神经网络面临的研究挑战。首先，随着图神经网络的层数增加，节点不断聚合扩大范围的邻居信息，获得的节点嵌入可能会趋于一致，导致节点嵌入之间的区分度降低，这种现象被称为过平滑 (Over Smoothing)，过平滑会导致学习到的节点嵌入退化为无效。因此，研究者们面临着如何在提升模型表达能力和避免过平滑现象之间取得平衡的挑战。其次，图数据通常具有高度复杂的结构和大规模的节点/边数量，这使得处理图数据的计算和存储成本非常高昂。将模型泛化到大规模图上，除了考虑到计算成本外，还需要考虑到对大规模图的复杂结构进行充分学习。最后，如何将现实生活中的数据抽象为图是另一个挑战。如何在进行数据抽象时最大程度地保留重要信息，同时又不引入过多的噪音和不必要的信息，是一个需要仔细考虑和平衡的问题。在完成了图抽象之后，研究者们还需要考虑如何改造图模型以适配此图，保证学习充分且不遗漏重要信息。举例来说，如何为金融欺诈场景建立用户关系图，其中需要融入用户时序交互、用户特征等，并根据建立出的关系图如何使用图模型来定位黑灰产和防控风险。

1.3 研究内容与贡献

本文基于图神经网络展开研究，聚焦于图神经网络的三个关键研究方向：图神经网络模型、图表示学习以及图表示学习的应用。

首先, 本文研究了基于频域的图卷积网络, 并提出模型 **Simple Graph Convolutional Network with Multi-Hop connection (SGCN-MH)**。

1. 在节点分类任务的背景下, 本文通过三种方法来探究高频与低频信息对图卷积网络性能的影响: 叠加图卷积层的层数, 引入广义拉普拉斯卷积核, 以及增加跨跳连接图卷积网络层。利用这些方法, 本文验证了在不同的数据集上, 网络所需的高低频信息各不相同。基于这一发现, 本文选择采用自适应的方式, 让网络能够灵活地学习所需的高低频信息, 以获得最佳性能。其次, 本文总结了 6 种图网络的层间交互模式: GCN-ResPre, GCN-ResIni, GCN-Add, GCN-Cat, GCN-Dense 以及 GCN-Dense*, 增加层间交互可以提升网络的表达能力以及缓解过平滑现象。最终, 本文还探究了适合图网络的信息传播模式, 简化了图网络的计算, 加速了推理过程并减轻了网络的过拟合风险。本文提出的模型 **SGCN-MH** 在使用有限的网络层数下达到了与深层图网络相匹配的表现, 在 4 个公共数据集上获得了与基线模型相比卓越的表现。

其次, 本文研究了无监督条件下的图对比学习, 并提出模型 **Minimal Noteworthy Information for unsupervised Graph contrastive learning (GMNI)**。

2. 本文在无监督图对比学习中实现了一种良好的数据增强方案, 提出了一种更有效的方法来近似无监督条件下的 InfoMin 原则, 实现了图对比学习中的最优视角的近似。本文采用对抗训练策略生成了共享最小显著 (noteworthy) 信息的视图, 避免了使用任务相关信息来实现共享最小必要 (necessary) 信息的视图。将 InfoMax 原则应用于这些视图是无风险的, 因为它们减少了干扰信息并强调了显著信息, 以确保信息充足。本文提出了一个具有自动化数据增强的新型无监督图对比学习方法 **GMNI**。在 5 个节点分类数据集和 9 个图分类数据集上的广泛的无监督和半监督条件下的实验表明了近似最优视图在不同下游任务上的有效性以及 **GMNI** 相对于其他自动和手动数据增强的图对比学习方法的优越性。

最后, 本文将图表示学习应用于医疗专业领域, 旨在获取药物的专业性嵌入, 从而实现相似药物的推荐。本文展示了一个完整的图表示学习应用示例, 从图的构建到最终的应用阶段。

3. 本文利用 **Comparative Toxicogenomics Database (CTD)** 数据集中的药物-疾病关系建立了异质节点同质边图, 其中包括两类节点: 药物和疾病, 以及一类边: 药物-疾病关系。本文将模型 **GMNI** 与模型 **SGCN-MH** 融合用于获得药物的

专业性嵌入，并依据此来实现相似药物的推荐。实验结果表示出推荐结果具有可解释性与泛化性。本文还搭建了一个相似药物推荐系统，方便医疗工作者查询药物、疾病属性和进行相似药物挖掘。

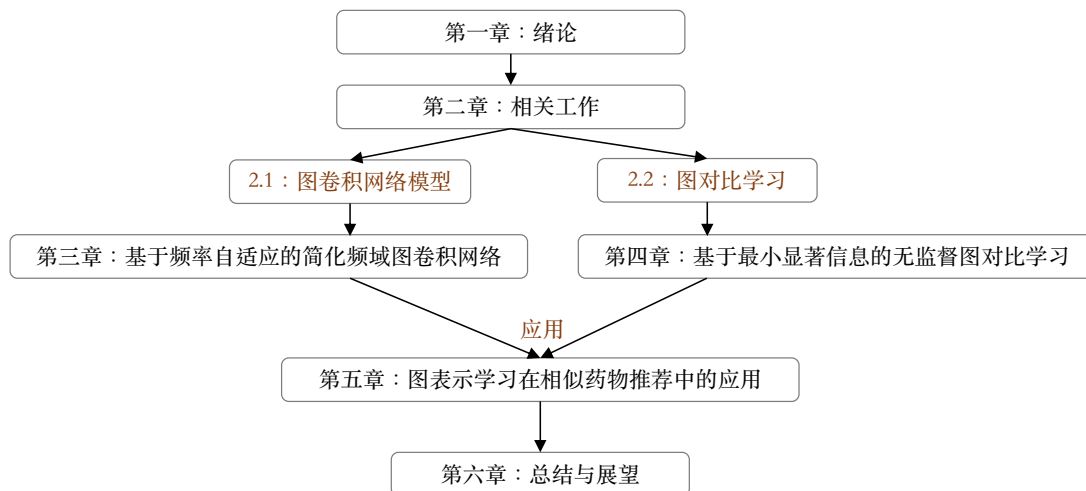


图 1-2 本文整体组织结构

1.4 本文结构安排

本文分为六个章节，图1-1及图1-2展示了本文主体的研究内容和整体组织结构。各章节的主要内容如下所示：

第一章为绪论，简要介绍了图神经网络的发展背景和实际意义。随后，介绍了图神经网络的三个主要研究方向：图神经网络模型、图表示学习以及图表示学习的应用，并探讨了相关的研究挑战。最后，指出了本文的研究内容与贡献，本文根据图神经网络的三个主要研究方向展开了研究工作。

第二章为相关工作，概述了图卷积网络模型和图对比学习的研究工作。图卷积网络模型部分：介绍了空域图卷积模型和频域图卷积模型中的高低频信息问题，并讨论了频域图卷积模型中的传播模式。图对比学习部分：探讨了无监督自动数据增强的图对比学习方法。这两个部分分别为后续章节“基于频率自适应的简化频域图卷积网络”和“基于最小显著信息的无监督图对比学习”提供了基础和铺垫。

第三章提出了一个基于频率自适应的简化频域图卷积网络模型：SGCN-MH。该模型涵盖了三个关键方面的讨论：如何有效利用高低频信息、如何更好地实现

层间信息交互以缓解过度平滑现象，以及如何简化模型的信息传递模式。SGCN-MH 通过高低频自适应、浅层网络设计和简化计算策略，实现了超越基线的优越准确率。

第四章提出了一种具有自动数据增强的无监督图对比学习方法：GMNI。GMNI 利用最小显著信息来实现对图对比学习中最优视角的近似，通过在这样的视角上应用 InfoMax 进行图对比，从而避免了信息不足和冗余信息的风险。GMNI 在 5 个节点分类数据集和 9 个图分类数据集上的无监督与半监督实验都超越了具有自动数据增强和手动数据增强的图对比学习方法。

第五章介绍了图表示学习应用于医疗领域的研究。具体而言，本章利用 CTD 数据集中药物与疾病之间的关联信息构建了一个图，并通过融合 GMNI 与 SGCN-MH 来生成了药物的专业性节点嵌入。本章将这些节点嵌入用于相似药物的推荐任务。实验结果显示相似药物的推荐结果具有可解释性与泛化性，同时本章还开发了系统来展示这些推荐结果，并增加了包括药物/疾病查询等的功能。

第六章为全文的总结和展望，总结了本文的内容和做出的贡献，并对未来的研究方向做出了展望。

第二章 相关工作

本章涉及了对图卷积网络与对图对比学习的讨论，分别服务于后续第三章与第四章的研究。主要介绍了图卷积网络工作模式，频域图卷积网络的研究重点，图对比学习的原理，以及几种经典的图对比学习模型。

2.1 图卷积网络模型

图卷积网络模型通常分为两大类：空域图卷积与频域图卷积。首先，本节对空域图卷积进行简要介绍，包括 Graph CNN^[52]，GraphSAGE^[12]，以及 GAT^[11]，以便更好地理解图网络的基本工作模式。其次，关于频域图卷积网络的工作原理将在第 3.2.2 章节中详细介绍，建议在阅读后续相关工作部分之前先阅读 3.2.2 章节。FAGCN^[53]和 AGE^[54]探讨了频域图卷积网络中高低频信息的作用，而 SGC^[55]则涉及了频域图卷积网络传播模式的讨论。关于频域图卷积网络中高低频信息的利用与图卷积网络传播模式的讨论将是本文的一个研究重点。在对这些相关工作进行公式描述时，本章会尽量使用和论文原文一致的符号。

2.1.1 卷积神经网络的图推广网络 Graph CNN

CNN^[20]通常用于处理规则的二维结构（如由像素点组成的图片），每个像素点有固定的 8 个邻居。CNN 利用一组可学习的卷积核在输入图像上进行滑动卷积操作，通常按照固定的卷积计算顺序（一般为从左上角到右下角），以提取图像的特征。然而，对于图数据而言，节点的邻居数量和顺序都是不确定的，这使得执行类似的卷积操作变得困难。

Graph CNN^[52]通过固定邻居的数量，并通过随机游走的访问频次来确定邻居的顺序，以模拟上述卷积过程。在 Graph CNN 中的卷积过程可分为两步：首先，构建节点固定大小的领域，并对邻居节点进行排序；其次，对领域中的节点

与卷积核参数进行内积操作。

Graph CNN 定义了随机游走概率转移矩阵 \mathbf{P} :

$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{S}, \quad (2-1)$$

其中, \mathbf{D} 为输入图的度矩阵, \mathbf{S} 为图的相似度矩阵, 原文中直接使用了邻接矩阵。 \mathbf{P} 中每个元素的值刻画了图中每一个节点被访问到的概率。继续定义多步转移期望矩阵 $\mathbf{Q}^{(k)}$:

$$\begin{aligned} \mathbf{Q}^{(0)} &= \mathbf{I}, \\ \mathbf{Q}^{(1)} &= \mathbf{I} + \mathbf{P}, \\ \mathbf{Q}^{(k)} &= \sum_{i=0}^k \mathbf{P}^i. \end{aligned} \quad (2-2)$$

其中上标 k 代表转移的步数, \mathbf{I} 为单位矩阵。

矩阵 $\mathbf{Q}^{(k)}$ 的第 i 行很自然地可以被用于获得节点 i 的邻居节点, 其值代表从节点 i 出发在 k 步到达不同节点的访问期望。对 $\mathbf{Q}^{(k)}$ 的第 i 行按从大到小进行排序, 取前 N 位元素: $\mathbf{Q}_{i\pi_i^{(k)}(1)} > \mathbf{Q}_{i\pi_i^{(k)}(2)} > \dots > \mathbf{Q}_{i\pi_i^{(k)}(N)}$, 其中 $\pi_i^{(k)}(c)$ 表示由节点 i 出发在 k 步内访问期望数第 c 大的节点。因此, 节点 i 的邻居集合可表示为: $\pi_i^{(k)}(1), \dots, \pi_i^{(k)}(N)$ 。

定义图的特征向量 $\mathbf{X} = (X_1, X_2, \dots, X_M)^T \in \mathbb{R}^{M \times F}$, 其中 M 代表节点个数, F 代表节点的特征维度。现在, 可以直接对定义好的邻居进行一次卷积操作:

$$\mathbf{Z}_i = \sum_{j=1}^N W_j X_{\pi_i^{(k)}(j)}. \quad (2-3)$$

其中, $W_j \in \mathbb{R}^{F \times F'}$ 为需要学习的参数矩阵, $\mathbf{Z}_i \in \mathbb{R}^{1 \times F'}$ 是获得的关于节点 i 的一次卷积结果。

经过多次卷积操作后, 得到最终的节点嵌入矩阵 \mathbf{Z} , 将其通过一层全连接层后可得到节点标签的预测 logits。最终的模型通过使用均方误差 (MSE) 损失函数进行训练。

需要提到的是, Graph CNN 受限于需要手动确认的邻居数量以及邻居顺序的需求, 但是为研究者们提供了一种在图上实现卷积的思路。

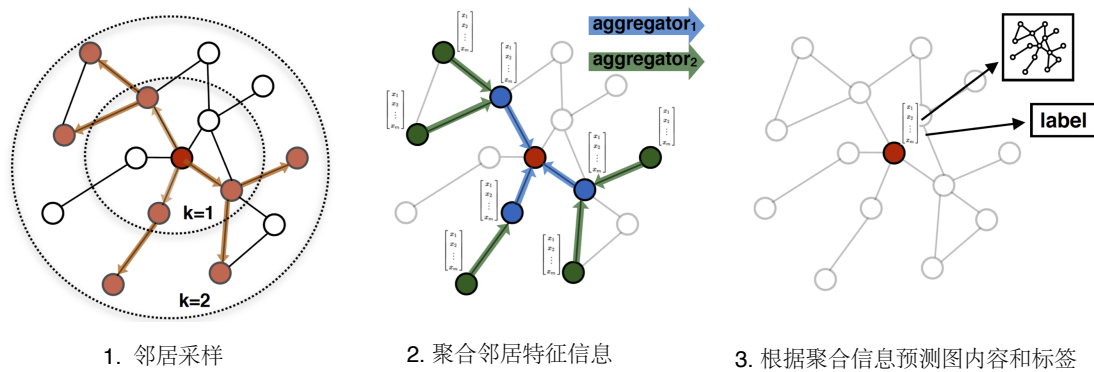


图 2-1 GraphSAGE 的算法步骤

2.1.2 采样与聚合图卷积网络 GraphSAGE

首先来介绍两个概念：transductive learning（直推式学习）和 inductive learning（归纳式学习）。Transductive learning 是指从当前学习的知识直接推广到给定的数据上。在训练阶段，模型可以利用测试节点的邻居信息和特征。Transductive learning 无法为从来没有见过的节点推理出嵌入结果。Inductive learning 是从已有数据中归纳出模式，以便将这些模式应用于新的数据。在训练时，测试数据是不可见的，模型必须通过训练数据来学习一般性的规律，以便在未知数据上进行推断和预测。这种方法允许模型泛化到未见过的数据，从而更加灵活地应对不同的情况。

GraphSAGE^[12]考虑到了引入新节点后重新训练模型所带来的大量计算消耗问题，于是它旨在学习一种嵌入模式，使其能够有效地对未曾见过的节点进行嵌入。这种嵌入模式的提出使得 GraphSAGE 成为了一种计算简便且具有强大可扩展性的归纳式学习的图卷积方法，它在处理大规模图数据时具有较高的效率和灵活性。

GraphSAGE 的卷积过程包括三个主要步骤：节点采样、节点信息聚合以及利用聚合信息进行标签预估。其详细步骤如图2-1所示，该图片来自原论文。在节点采样阶段，模型会为每个节点选择一定数量的邻居作为参考，而不是使用所有的邻居。在节点信息聚合阶段，模型会利用选定的邻居节点来聚合并更新目标节点的信息。聚合方式包括 Mean aggregator、LSTM aggregator 和 Max-pooling aggregator。这些聚合器确定了在节点信息聚合阶段如何有效地结合邻居节点的

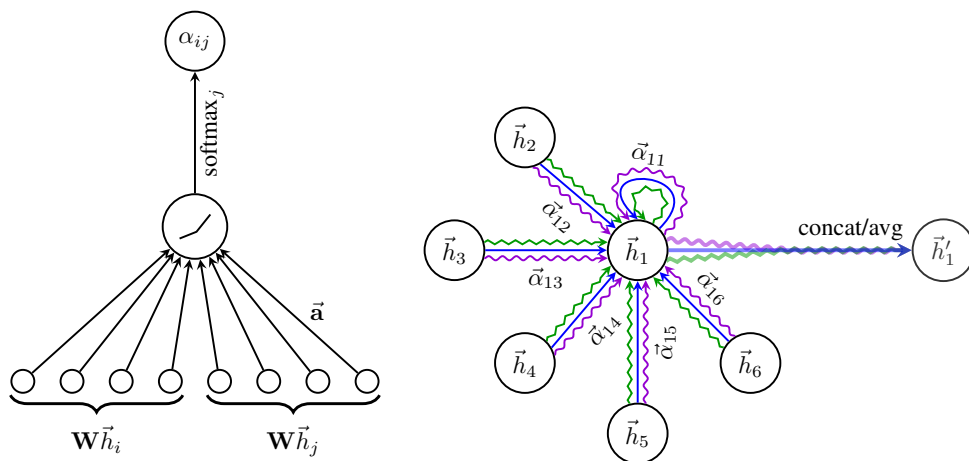


图 2-2 GAT 的聚合过程

信息以更新目标节点的表示。Mean aggregator 通过计算邻居节点的平均值来聚合信息，LSTM aggregator 利用长短期记忆网络来对邻居节点的信息进行序列建模，而 Max-pooling aggregator 则选取邻居节点中的最大值作为聚合结果。这些不同的聚合方式使得 GraphSAGE 模型能够灵活地适应不同的任务和数据特性，从而提高了模型的性能和泛化能力。最后，在利用聚合信息进行标签预估阶段，模型将聚合后的节点嵌入用于节点标签的预测。

GraphSAGE 的一个亮点在于其节点采样操作。与传统方法不同，节点采样只选择了节点的一部分邻居信息进行聚合和更新。这种采样策略不仅有效地降低了计算复杂度，使得 GraphSAGE 模型能够处理大规模图数据，而且由于节点聚合并没有使用所有邻居信息，从而增加了模型的鲁棒性，并使其成为一种归纳式学习的方法。

2.1.3 图注意力网络 GAT

注意力 (attention) 机制是一种在深度学习中常用的技术，用于改变模型对不同输入数据的关注程度。它的核心思想是在处理序列数据或者集合数据时，不是简单地对所有元素进行相同权重的处理，而是根据元素之间的相关性动态地分配权重，从而更加有效地利用输入信息。例如，在计算机视觉领域，注意力机制可以让模型集中注意力于图像中的特定区域，这种机制类似于人类在观察图像时会特别关注其中的某些部分。

GAT^[11]是一种将注意力机制引入图卷积网络的方法，旨在为节点的不同邻居分配不同的注意力权重，以更好地捕捉节点之间的关系和特征。

图注意力层是 GAT 中的单层结构。它的输入是节点特征 $\mathbf{h} = \{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_N\}$ ，其中 N 代表节点数， $\mathbf{h}_i \in \mathbb{R}^F$ 表示每个节点的特征向量， F 是特征的维度。该层的输出是新的节点特征 $\mathbf{h}' = \{\mathbf{h}'_1, \mathbf{h}'_2, \dots, \mathbf{h}'_N\}$ ，其中 $\mathbf{h}'_i \in \mathbb{R}^{F'}$ ， F' 是新特征的维度。

以计算节点 i 对于其邻居节点 j 的注意力系数 α_{ij} 的过程为示例：

$$\alpha_{ij} = \frac{\exp(\text{LeakyReLU}(\mathbf{a}^T[\mathbf{W}\mathbf{h}_i\|\mathbf{W}\mathbf{h}_j]))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(\mathbf{a}^T[\mathbf{W}\mathbf{h}_i\|\mathbf{W}\mathbf{h}_k]))}, \quad (2-4)$$

其中， $\mathbf{W} \in \mathbb{R}^{F \times F'}$ 为特征变换矩阵， $\mathbf{a} \in \mathbb{R}^{2F'}$ 为权重学习向量， $[\|\cdot\|]$ 代表向量拼接操作， N_i 代表节点 i 的邻居集合，归一化操作为了使得注意力系数在邻居节点之间具有可比性。注意力系数的计算过程如图2-2（左）所示，该图来源于 GAT 原文。

当获得了注意力系数后，通过加权聚合邻居的特征信息来获得聚合结果：

$$\mathbf{h}'_i = \sigma \left(\sum_{j \in N_i} \alpha_{ij} \mathbf{W}\mathbf{h}_j \right), \quad (2-5)$$

其中， $\sigma(\cdot)$ 为激活函数。

GAT 通过叠加多层图注意力层来工作，它还提出了多头注意力机制（multi-head attention）的聚合方式：

$$\mathbf{h}'_i = \parallel_{k=1}^K \sigma \left(\sum_{j \in N_i} \alpha_{ij}^k \mathbf{W}^k \mathbf{h}_j \right). \quad (2-6)$$

这种聚合方式增加了多个独立的注意力机制： K 个独立的 attention 完成信息聚合过程，它们聚合得到的结果将被拼接起来来获得最后的输出表示。其中 α_{ij}^k 代表被 k -th attention 机制计算出来的归一化后的注意力系数， \mathbf{W}^k 是相关的特征变换矩阵，最后的输出将会包含 KF' 维度的特征，该计算过程如图2-2（右）所示。

以上讨论了三种空域图卷积网络，而频域图卷积网络则是通过进行空域到频域再到空域的转换来完成卷积操作。由于这部分内容涉及较多公式推理，详

细介绍留到了 3.2.2 部分。

在这里，只需记住几个简要知识：定义图的邻接矩阵为 \mathbf{A} ，度矩阵 $\mathbf{D} : \mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$ 。最常用的频域图卷积网络 GCN^[10]的图卷积核为 $\mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}$ ，仅使用此图卷积核就可以完成从空域到频域再到空域的转变，拉普拉斯矩阵 $\mathbf{L} = \mathbf{D} - \mathbf{A}$ 。或者您可以选择阅读完毕 3.2.2 部分再继续此处的阅读。

2.1.4 频率自适应图卷积网络 FAGCN

大部分图神经网络倾向于处理图中的低频信息，FAGCN^[53]却在文中验证了仅使用低频信息存在一定局限性。使用低通滤波器的图神经网络有助于保留节点特征的共性，倾向于学习到的连通节点是相似的。然而，这类模型忽略了节点特征的差异性。实际上，对于不同的网络而言，连通节点并不一定具有相似的特征，例如蛋白质中的相互作用通常发生在不同类型的氨基酸之间。

相连节点具有相似特征的网络被称为同配网络 (assortative network)，而相连节点具有不相似特征的网络被称为异配网络 (disassortative network)。低频信息的平滑性适用于同配网络，而高频信息的差异性则适用于异配网络。基于此，FAGCN 中实现了一种频率自适应的图卷积网络。

首先，FAGCN 设计了一个低通滤波器 \mathcal{F}_L 与高通滤波器 \mathcal{F}_H 来从节点特征中分别分离低频与高频信息：

$$\begin{aligned}\mathcal{F}_L &= \epsilon \mathbf{I} + \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}, \\ \mathcal{F}_H &= \epsilon \mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}.\end{aligned}\tag{2-7}$$

其中， ϵ 为一个超参数， \mathbf{D} 为图的度矩阵， \mathbf{A} 为图的邻接矩阵。

使用 $\mathbf{h} = \{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_N\}$, $\mathbf{h}_i \in \mathbb{R}^F$ 作为层输入，FAGCN 利用了注意力机制来自适应地学习高低频的比例，其信息聚合过程如下：

$$\begin{aligned}\mathbf{h}'_i &= \alpha_{ij}^L(\mathcal{F}_L \cdot \mathbf{h})_i + \alpha_{ij}^H(\mathcal{F}_H \cdot \mathbf{h})_i \\ &= \epsilon \mathbf{h}_i + \sum_{j \in N_i} \frac{\alpha_{ij}^L - \alpha_{ij}^H}{\sqrt{d_i d_j}} h_j.\end{aligned}\tag{2-8}$$

其中， \mathbf{h}'_i 代表节点 i 进行一次信息聚合后的新特征， α_{ij}^L 和 α_{ij}^H 分别代表低频

和低频信息的比例， d_i 代表节点 i 的度。低频和高频信息的比例系数之和为 1： $\alpha_{ij}^L + \alpha_{ij}^H = 1$ 。

定义 $\alpha_{ij}^G = \alpha_{ij}^L - \alpha_{ij}^H$ ， α_{ij}^G 将通过注意力机制计算得到，同时考虑了节点本身特征和其邻居特征：

$$\alpha_{ij}^G = \tanh(\mathbf{W}[\mathbf{h}_i \parallel \mathbf{h}_j]), \quad (2-9)$$

其中 $\mathbf{W} \in \mathbb{R}^{2F \times 1}$ 为所需要学习的参数，可以获得最终的信息聚合公式：

$$\mathbf{h}'_i = \epsilon \mathbf{h}_i + \sum_{j \in N_i} \frac{\alpha_{ij}^G}{\sqrt{d_i d_j}} h_j. \quad (2-10)$$

2.1.5 自适应图编码器 AGE

本节将专注于探讨 AGE^[54]中涉及到图中高低频信息的部分。

将图特征表示为 $x \in \mathbb{R}^N$ ， N 为节点个数，为每个节点特征分配一个标量。Rayleigh quotient（瑞利商）是线性代数中的一个概念，用于描述矩阵的特征值与特征向量之间的关系。现将瑞利商引入来度量图特征之间的相似性：

$$R(\mathbf{L}, x) = \frac{x^T \mathbf{L} x}{x^T x} = \frac{\sum_{(i,j) \in \mathcal{E}} (x_i - x_j)^2}{\sum_{i \in \mathcal{V}} x_i^2}. \quad (2-11)$$

其中， \mathcal{E} 和 \mathcal{V} 分别代表图中的边集合与节点集合。因此，瑞利商较低的图，节点之间的差异更小，图信号更平滑。

图的拉普拉斯矩阵特征分解可以表示为 $\mathbf{L} = \mathbf{U} \Lambda \mathbf{U}^{-1}$ ，其中 $\mathbf{U} \in \mathbb{R}^{N \times N}$ 为特征向量矩阵， $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$ 为由特征值组成的对角矩阵。衡量特征向量 u_i 平滑程度的瑞利商可以表示如下：

$$R(\mathbf{L}, u_i) = \frac{u_i^T \mathbf{L} u_i}{u_i^T u_i} = \lambda_i. \quad (2-12)$$

上述公式说明了特征向量的平滑程度与其对应特征值的大小密切相关。当特征值较小时，特征向量的频率更低，更平滑。

以拉普拉斯矩阵的特征向量作为基对图特征进行分解：

$$x = \mathbf{U}\mathbf{P} = \sum_{i=1}^n p_i u_i, \quad (2-13)$$

其中， p_i 代表特征向量 u_i 的系数。进一步重写图信号的瑞利商：

$$R(\mathbf{L}, x) = \frac{x^T \mathbf{L} x}{x^T x} = \frac{\sum_{i=1}^n p_i^2 \lambda_i}{\sum_{i=1}^n p_i^2}. \quad (2-14)$$

AGE 提出了广义的拉普拉斯平滑滤波器，定义如下：

$$\mathbf{H} = \mathbf{I} - \kappa \mathbf{L}. \quad (2-15)$$

其中 κ 为实数，使用 \mathbf{H} 作为滤波器：

$$\begin{aligned} x' &= \mathbf{H}x \\ &= (\mathbf{I} - \kappa \mathbf{L})x \\ &= \mathbf{U}(\mathbf{I} - \kappa \mathbf{\Lambda})\mathbf{U}^{-1}\mathbf{U}\mathbf{P} \\ &= \sum_{i=1}^n (1 - \kappa \lambda_i) p_i u_i. \end{aligned} \quad (2-16)$$

可以将拉普拉斯矩阵替换为归一化拉普拉斯矩阵得到新的广义拉普拉斯滤波器：

$$\mathbf{H} = \mathbf{I} - \kappa \mathbf{D}^{-\frac{1}{2}} \mathbf{L} \mathbf{D}^{-\frac{1}{2}}. \quad (2-17)$$

设置 $\kappa = 1$ 就可以获得 GCN^[56]中使用的滤波器：

$$\begin{aligned} \mathbf{H} &= \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} \mathbf{L} \mathbf{D}^{-\frac{1}{2}} \\ &= \mathbf{I} - \mathbf{D}^{-\frac{1}{2}} (\mathbf{D} - \mathbf{A}) \mathbf{D}^{-\frac{1}{2}} \\ &= \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}. \end{aligned} \quad (2-18)$$

参数 κ 的选择需要考虑特征值的分布，将滤波后的特征 x' 代入公式 (2-14)：

$$R(\mathbf{L}, x') = \frac{x'^T \mathbf{L} x'}{x'^T x'} = \frac{\sum_{i=1}^n p'_i{}^2 \lambda_i}{\sum_{i=1}^n p'_i{}^2}, \quad (2-19)$$

$$p'_i = (1 - \kappa \lambda_i) p_i.$$

$p'_i{}^2$ 与 λ_i 共同决定图信号的平滑程度。定义拉普拉斯矩阵的最大特征值为 λ_{\max} ，AGE 给出了理论上的范围：

- 当 $\kappa > \frac{1}{\lambda_{\max}}$ 时，滤波器不是一个低通滤波器。
- 当 $\kappa < \frac{1}{\lambda_{\max}}$ 时，滤波器是不能去除所有的高频信息的低通滤波器。
- 当 $\kappa = \frac{1}{\lambda_{\max}}$ 时，滤波器是去除所有高频信息的标准低通滤波器。

2.1.6 简化图卷积网络 SGC

通过之前对图卷积网络的讨论，可以将节点特征的更新过程主要分为三步：特征聚合、特征变换和非线性激活。

以 GCN 举例，当前层的输入和输出分别记作 $\mathbf{h}^{(l)}$ 和 $\mathbf{h}^{(l+1)}$ ，当前层的输入同时也是上一层的输出：

$$\mathbf{h}^{(l+1)} = \sigma \left(\mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \mathbf{h}^{(l)} \mathbf{W}^{(l+1)} \right). \quad (2-20)$$

其中 $\mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$ 作为卷积核来完成节点的特征聚合， $\mathbf{W}^{(l+1)}$ 为用于特征变换的参数矩阵， $\sigma(\cdot)$ 为激活函数。

SGC^[55]认为 GCN 层间的非线性激活函数并不扮演主要角色。因此，文中删除了各个层间的非线性激活函数，仅保留了最终的 Softmax 函数，用于获得概率输出。

$$\begin{aligned} \mathbf{h}^{(1)} &= \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \mathbf{h}^{(0)} \mathbf{W}^{(1)}, \\ \mathbf{h}^{(l+1)} &= \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \mathbf{h}^{(l)} \mathbf{W}^{(l+1)}, \\ \hat{y} &= \text{Softmax} \left(\mathbf{h}^{(L)} \right). \end{aligned} \quad (2-21)$$

记 $\mathbf{H} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$ ，化简公式得到：

$$\begin{aligned} \hat{y} &= \text{Softmax} \left(\mathbf{H}^L \mathbf{h}^{(0)} \mathbf{W}^{(1)} \mathbf{W}^{(2)} \dots \mathbf{W}^{(L)} \right), \\ &= \text{Softmax} \left(\mathbf{H}^L \mathbf{h}^{(0)} \mathbf{W} \right). \end{aligned} \quad (2-22)$$

多个线性变换的组合等同于单个线性变换。SGC 的整体过程可以被看作是网络在各个层之间仅完成特征聚合，然后在输出层添加线性特征变换，并通过 Softmax 函数获得概率输出。

经过实验验证，SGC 在保证准确率或者获得更优准确率的前提下，能够实现更快的计算速度。LightGCN^[15]是图卷积网络在协同过滤算法上的应用。类似于 SGC，LightGCN 去除了图卷积层之间的特征变换函数和非线性激活函数。它在推荐任务中表现出了超越其他协同过滤算法的优异性能，并且具有快速的推理速度。

2.2 图对比学习

图对比学习包括两个重要的部分：视图生成以及视图对比。本节首先介绍指导最优视图生成的 InfoMin 原则与指导视图对比的 InfoMax 原则。其次本节介绍三个具有自动数据增强的无监督图对比学习方法 JOAO^[57]，AD-GCL^[58]，AutoGCL^[59]，并分析了它们的相关弊端。

2.2.1 互信息最小化原则 InfoMin 与互信息最大化原则 InfoMax

InfoMin 原则指出了对比学习中最优视角需要满足的条件：

$$\begin{aligned} (V_1^*, V_2^*) &= \operatorname{argmin}_{V_1, V_2} I(V_1; V_2), \\ \text{s. t. } &I(V_1; Y) = I(V_2; Y) = I(X; Y). \end{aligned} \quad (2-23)$$

其中 V_1 、 V_2 代表视图， V_1^* 、 V_2^* 代表最优视图， X 代表输入信息， Y 代表下游任务信息， $I(a; b)$ 代表 a 与 b 之间的互信息。简单来说，对比学习中的最优视图集合之间共享的信息只与下游任务相关，不应该共享任何冗余信息，且需要保证不遗漏下游任务信息。

InfoMax^[60]是对比学习中最常用的原则之一，其指的是最大化视图表示之间的互信息来学习编码器：

$$\max_f I(f(V_1); f(V_2)). \quad (2-24)$$

其中 f 表示编码器， $I(a; b)$ 表示 a 和 b 之间的互信息。

下面本节将举一些 InfoMax 的应用实例。DIM^[61]是对比学习领域的开创性工作之一，它通过利用 InfoMax 来实现对比学习的目标。DIM 为对比学习提供了理论基础，并指导了后续工作在正确选择对比损失函数方面的方向。DGI^[40]遵循了 DIM 的方法，首次将 InfoMax 应用于图数据。它通过节点洗牌 (Node Shuffling) 来破坏原始图以生成负样本，并利用经过 InfoMax 优化的鉴别器来区分正样本和负样本。GRACE^[62]通过随机去除边 (Removing Edges) 和节点特征掩码 (Masking Node Features) 生成两个视图，然后通过应用 InfoMax 优化编码器来学习节点表示。BGRL^[63]通过随机节点特征掩码 (Masking Node Feature) 和连边掩码 (Masking Edges) 生成视图，并通过最大化不同阶段视图表示的余弦相似度来进行学习，实质上也是应用了 InfoMax 进行优化。

总结一下，InfoMin 原则指出了对比学习中的最优视图，InfoMax 原则指出了对比学习中的最优编码器。

2.2.2 联合增强优化框架 JOAO

图对比学习的有效性很大程度上依赖于数据增强，大部分的图对比学习方法需要通过经验法或试错法来对每个数据集手动选择增强方法，这将会消耗很多的时间与算力资源。

JOAO^[57]提出使用一个双层优化框架来在图对比学习中自动地、可适应性地、动态地选择数据增强方案。该通用框架称为 JOint Augmentation Optimization (JOAO)，其实是最小化-最大化优化的应用。

JOAO 的优化框架可以形式化地写作如下形式：

$$\begin{aligned} & \min_{\theta} \mathcal{L}(G, A_1, A_2, \theta), \\ & \text{s. t. } \mathbb{P}_{(A_1, A_2)} \in \operatorname{argmax}_{\mathbb{P}_{(A'_1, A'_2)}} \left\{ \mathcal{L}(G, A'_1, A'_2, \theta) - \frac{\gamma}{2} \operatorname{dist}(\mathbb{P}_{(A'_1, A'_2)}, \mathbb{P}_{\text{prior}}) \right\}, \end{aligned} \quad (2-25)$$

其中， θ 代表图视角编码器的参数， A_1 和 A_2 代表图的数据增强方案，它们通过组合数据增强池中的数据增强方案得到。数据增强池中包含四种数据增强方案：随机丢弃节点 (NodeDrop)，生成子图 (Subgraph)，连边扰动 (EdgePert)，属性掩码 (AttrMask)，使用原图 (Identical)。 A_1 和 A_2 从分布 $\mathbb{P}_{(A_1, A_2)}$ 中得到， $\mathbb{P}_{\text{prior}}$ 代表其先验分布。 $\operatorname{dist}(\cdot)$ 代表距离函数，用于约束 $\mathbb{P}_{\text{prior}}$ 与 $\mathbb{P}_{(A_1, A_2)}$ 之间的距离不要过大。文章中设置 $\mathbb{P}_{\text{prior}}$ 为均匀分布，这样可以鼓励生成的数据增强方案更加多样化。

JOAO 的核心思想是产生最具有挑战性的视角有利于图对比学习进程，即视角集合需要与视角编码器的优化产生对抗。首先，这样的做法实际上减少了图视角之间共享的冗余信息，只满足了 InfoMin 中的部分条件： $\operatorname{argmin}_{V_1, V_2} I(V_1; V_2)$ ，并没有达到 InfoMin 原则中最优视角满足的条件。其次，JOAO^[57] 从固定的数据增强池中搜索最佳的数据增强策略组合。设计一个合适的数据增强池仍然需要专家领域知识，并且仅仅优化组合系数限制了数据增强的灵活性。

2.2.3 对抗图数据增强图对比学习 AD-GCL

AD-GCL^[58] 也成功地实现了自动数据增强的图对比学习。与 JOAO 类似，AD-GCL 采用对抗训练来生成数据增强方案。该方法遵循信息瓶颈 (Information Bottleneck, IB^[64-66]) 原则，以减少视图之间共享的冗余信息。

信息瓶颈原则旨在同时最小化 $I(X; Z)$ 和最大化 $I(Z; Y)$ ，其中 X 代表输入信息， Z 代表隐藏层输出， Y 代表下游任务信息。信息瓶颈原则指出编码器应该在最小化原始数据中的信息的同时，最大化与下游任务相关的信息。InfoMin 是信息瓶颈在对比学习中的成功应用，并用以产生最佳视图。图信息瓶颈原则 (Graph Information Bottleneck, GIB) 原则可以被表示为：

$$\max_f I(f(G); Y) - \beta I(f(G); G). \quad (2-26)$$

其中, $f(G)$ 代表在图 G 上的信息编码器。简单来说, GIB 原则期望最大化图编码信息与下游任务相关的信息, 最小化与原图之间共享的信息。

AD-GCL 的优化过程可形式化地表示如下:

$$\min_{T \in \mathcal{T}} \max_f I(f(G); f(t(G))), \text{ where } t(G) \sim T(G). \quad (2-27)$$

其中, \mathcal{T} 代表一组数据增强集合 $T_\phi(\cdot)$, $T_\phi(\cdot) \sim \mathcal{T}$ 是其中一个具有参数 ϕ 的数据增强方案。AD-GCL 把数据增强方案实例化为带有参数 ϕ 的随机边丢弃数据增强, 即利用参数 ϕ 生成伯努利分布来决定是否丢弃当前边。

首先, AD-GCL 仅在边上进行数据增强, 而忽略了关键的特征信息。其次, AD-GCL 通过这样的优化目标, 完成了一部分 GIB 原则, 和 JOAO 一样忽略了对与下游任务相关的关键信息保留。

2.2.4 自动图对比学习 AutoGCL

AutoGCL^[59] 同样是具有自动数据增强的图对比学习方法。它使用了一组可学习的图视角生成器来获得数据增强方案, 每个图视角生成器都学习一个输入图的概率分布。图视角生成器期望生成的视角保留原图中最具有表达能力的结构, 与此同时期望视角之间的差异尽可能的大。其优化过程可形式化地表示如下:

$$\min_{t_1, t_2} \text{sim}(t_1(G); t_2(G)) + \max_f I(f(t_1(G)); f(t_2(G))). \quad (2-28)$$

其中, $f(\cdot)$ 表示图编码器, $t_1(G), t_2(G)$ 表示经过数据增强获得的两个视图, $\text{sim}(\cdot)$ 代表相似度函数, 越大的值代表越相似。

AutoGCL 通过最小化两个视图之间的相似性并最大化两个视图表示之间的一致性来工作。首先, AutoGCL 通过最小化视图之间相似度来达成最小化视图之间共享信息的目标, 减少了视图共享的冗余信息, 仍然缺失了对重要信息的保留。其次, AutoGCL 仅在节点特征级别进行数据增强, 而忽略了图中的拓扑信息。

事实上, 以上包括 JOAO、AD-GCL 和 AutoGCL 在内的自动化数据增强的图对比学习方法均采用了一种最小化-最大化的训练策略。它们未能完全遵循

InfoMin 原则，其过度的信息减少可能导致潜在的必要信息缺乏的风险。

InfoGCL^[67]很好地遵循了 InfoMin 原则，以引导最佳视图的生成。然而，它在生成最佳视图、最佳视图编码器和最佳对比模式时依赖于下游任务特定的标签 Y 。此外，InfoGCL 中使用了手动设计的数据增强池，并通过穷举搜索获取最佳数据增强方法，这一过程耗时且计算成本极高。由于 InfoGCL 在训练过程中使用了标签信息来指导训练，它并不算入无监督图对比学习的范畴。

InfoMin^[68]的作者 Tian 等人也采用了上述最小化-最大化的训练优化策略，试图将 InfoMin 应用于无监督对比学习。他们给定了图像 X 以及经过数据增强变换后的图像 $g(X)$ ，并使用了两个编码器 f_1 和 f_2 ，模型的优化目标如下：

$$\min_g \max_{f_1, f_2} I(f_1(g(X)_1); f_2(g(X)_{2:3})). \quad (2-29)$$

其中， $\{g(X)_1, g(X)_{2:3}\}$ 代表 $g(X)$ 的分离通道图像，分别作为两个视图 V_1 和 V_2 。然而，Tian 等人指出，这种在无监督条件下寻找视图的方法严重破坏了 InfoMin 中的约束条件 $I(V_1; Y) = I(V_2; Y) = I(X; Y)$ ，导致 $I(V_1; V_2)$ 被过度减少。一种更好地在无监督条件下近似 InfoMin 原则的方法亟待提出。

2.3 本章小结

本章对经典的图卷积网络模型及图对比学习的原理展开介绍。首先，本章介绍了空域图卷积网络模型，其次本章探讨了频域图卷积网络模型与高低频信息以及网络传播模式相关的研究成果。随后，本章介绍了图对比学习中常用的 InfoMin 与 InfoMax 原则，最后本章介绍了几种具有自动数据增强的图对比学习方法。

第三章 基于频率自适应的简化频域图卷积网络

3.1 研究背景与动机

图卷积网络根据卷积操作的空间不同,可分为两个流派:空域图卷积与频域图卷积。空域图卷积借鉴了经典的卷积神经网络 CNN^[20]的思想,其操作直接在图的节点空间进行。该方法通常通过汇聚邻居节点的信息来更新每个节点的特征表示,从而实现图的局部信息提取。频域图卷积则利用了“频域上的乘积等同于空域上的卷积^[69]”的重要定理来实现卷积操作。该方法首先将原始图信息以及卷积核通过图傅立叶变换转换到频域,然后在频域中进行卷积核与图信息的乘积操作,最后再通过图傅立叶反变换回空域,从而完成卷积操作。简而言之,频域图卷积经历了空域到频域再到空域的转变过程。空域图卷积在处理局部信息时很高效,因为此类方法的核心是基于聚类-再聚类的局部卷积操作^[11-12]。除此之外,空域图卷积由于可以增加采样操作,所以在处理大规模图数据时具有优势。但是,如果整张图是具有良好全局性质的高维图,缺失聚类优势,那么利用空域图卷积反而会存在较大误差。总结来说,空域图卷积适用于大规模图数据处理,但可能受限于局部信息的表示能力。相反,频域图卷积在处理全局信息时表现出色,更能捕捉图的整体特征,并且其利用谱图理论来完成特征传播和信息聚合操作具有很深的理论背景。本章主要对频域图卷积网络进行相关的研究,并将研究聚焦在以下三个方面:

1. 图上的频率信息可分为低频信息与高频信息,低频信息保留了节点之间的相似性,而高频信息则保留了节点之间的差异性。大部分的频域图卷积网络^[55-56,70]认为连接的节点倾向于相似,于是往往利用低通滤波器来完成卷积过程。事实上,低频信息仅适用于同配网络,对于异配网络而言高频信息则是

必不可少的^[53]。相连节点具有相似特征的网络被称为同配网络，而相连节点具有不相似特征的网络被称为异配网络。FAGCN^[53]设计了一个低通滤波器与一个高通滤波器来分离低频与高频信息。本章将探讨如何在未知图是同配性或异配性、无需刻意设计低通与高通滤波器的情况下，实现高低频自适应。

2. 图卷积网络受限于过平滑现象，即随着网络层数的增加，节点表征倾向于一致，这将损害模型的表达能力。LR-GCCF^[71]参考 ResNet^[50]引入了残差连接，从而增加节点表示之间的差异性来缓解过平滑现象。过平滑现象致使图卷积网络无法变得更深，DeepGCNs^[72]引入了残差连接、密集连接和膨胀卷积来增加网络层数，并且在点云语义分割任务上取得了突出的表现。DropEdge^[73]认为稠密的网络连边造成了重复邻居信息，最终导致过平滑现象，于是它在训练的时候每次随机去掉图中的某些边。GCNII^[74]引入初始残差和等值映射来扩展传统的 GCN，从而缓解过平滑现象，并以此增加网络的深度。通过改进层间的交互方式来为节点表示增加差异性是一种缓解过平滑现象的常用做法，本章将全面地探究不同的层间交互方式对于过平滑现象的缓解能力。过平滑现象的缓解包括两个层面：一方面，当网络层数增加时，模型表现不会持续下降；另一方面，在同层次下，网络可以获得更高的表达能力，从而实现更好的模型表现。

3. 一般而言，空域图卷积网络在实际生产中的应用要优于频域图卷积网络。这是因为频域图卷积网络在信息聚合时涉及更多的矩阵计算，对计算资源的要求更高。因此，本章将探究如何简化频域图卷积网络，在不影响模型表达能力的前提下使用浅层和低参数量的网络完成任务，期望通过浅层网络达到其他深层网络相同的表达能力。

本章探究了三个主要问题：频域图卷积网络中的高低频信息利用问题、不同层间交互方式对过平滑现象的缓解问题，以及如何简化频域图卷积网络的信息传播模式问题。基于对以上问题的讨论，本章提出了一个基于频率自适应的简化频域图卷积网络：Simple Graph Convolutional Network with Multi-Hop connection (SGCN-MH)。SGCN-MH 通过使用单跳与跨跳连接图卷积网络层来完成高低频自适应，通过层间交互来缓解过平滑现象增加网络的表达能力，并通过去除层间的特征变换和非线性激活函数来完成网络的简化。

3.2 模型

本节将首先梳理频域图卷积网络的发展过程，分析其中存在的高频信息损失问题，并提出使用三种方式来改变图中的高低频信息比例。实验部分探究了不同高低频信息配比在不同数据集上对网络性能的影响。接着，本节探讨了六种图网络层间信息交互方式，探究哪种交互方式能够更好地缓解过平滑现象。最后，本节探讨了如何简化模型计算，并基于以上讨论提出了模型 SGCN-MH。

3.2.1 符号定义

定义无向图 \mathcal{G} ，它包括有 N 个节点的节点集合 $\mathcal{V} = \{v_1, \dots, v_N\}$ 和有 M 条边的边集合 $\mathcal{E} = \{\epsilon_1, \dots, \epsilon_M\}$ 。图 \mathcal{G} 的特征矩阵、邻接矩阵和度矩阵分别表示为 $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \in \mathbb{R}^{N \times F}$ 、 $\mathbf{A} \in \{0, 1\}^{N \times N}$ ： $\mathbf{A}_{ij} = \mathbb{I}((v_i, v_j) \in \mathcal{E})$ 和 \mathbf{D} ： $\mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$ 。其中 F 是特征维度， $\mathbf{x}_i \in \mathbb{R}^F$ 是节点 v_i 的特征。本节使用 $\mathcal{G} : (\mathbf{A}, \mathbf{X})$ 来表示这个图。图节点的 one-hot 标签为 $Y \in \mathbb{R}^{N \times C}$ ，其中第 i 个节点的标签为 Y_i ， C 为节点的类别数。

拉普拉斯矩阵是图的除了邻接矩阵之外另外一种重要的矩阵表示。它通常表示为 $\mathbf{L} = \mathbf{D} - \mathbf{A}$ 。给定图的特征矩阵 \mathbf{X} ，将拉普拉斯矩阵作用于该特征矩阵上，可以得到：

$$\begin{aligned} \mathbf{H} &= \mathbf{LX} = (\mathbf{D} - \mathbf{A})\mathbf{X}, \\ \mathbf{h}_i &= \sum_{v_j \in N_{v_i}} (\mathbf{x}_i - \mathbf{x}_j). \end{aligned} \tag{3-1}$$

其中， \mathbf{H} 代表了经过拉普拉斯矩阵变换的特征矩阵， \mathbf{h}_i 代表 \mathbf{H} 的第 i 行向量， N_{v_i} 代表节点 v_i 的邻居节点集合。可以看出， \mathbf{h}_i 的物理含义是节点 v_i 和它的邻居节点的特征的差异之和。

定义度归一化的拉普拉斯矩阵为：

$$\begin{aligned} \tilde{\mathbf{L}} &= \mathbf{D}^{-\frac{1}{2}}(\mathbf{D} - \mathbf{A})\mathbf{D}^{-\frac{1}{2}} \\ &= \mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{A}\mathbf{D}^{-\frac{1}{2}}. \end{aligned} \tag{3-2}$$

度归一化的目的在于消除节点度数对图操作的影响，以确保相对节点度数较高和较低的节点在进行操作时具有相似的权重。

定义图的平滑度 (Smoothness) 为:

$$\mathbf{X}^T \mathbf{L} \mathbf{X} = \frac{1}{2} \sum_{v_i \in \mathcal{V}} \sum_{v_j \in N_{v_i}} (\mathbf{x}_i - \mathbf{x}_j)^2. \quad (3-3)$$

平滑度的本质就是就是所有邻接节点的差异的平方和。直观上理解, 如果节点与邻接节点的差异小, 那么邻接节点相似, 图的平滑度将很小。

证明 1 度归一化的拉普拉斯矩阵 $\tilde{\mathbf{L}}$ 可相似对角化, 且对角阵上的元素为特征值。图 \mathcal{G} 为无向图, 邻接矩阵 \mathbf{A} 和度矩阵 \mathbf{D} 均为实对称矩阵, 所以拉普拉斯矩阵 $\mathbf{L} = \mathbf{D} - \mathbf{A}$ 也为实对称矩阵。

$$\begin{aligned} \tilde{\mathbf{L}}^T &= (\mathbf{D}^{-\frac{1}{2}} \mathbf{L} \mathbf{D}^{-\frac{1}{2}})^T \\ &= (\mathbf{D}^{-\frac{1}{2}})^T \mathbf{L}^T (\mathbf{D}^{-\frac{1}{2}})^T \\ &= \mathbf{D}^{-\frac{1}{2}} \mathbf{L} \mathbf{D}^{-\frac{1}{2}} \\ &= \tilde{\mathbf{L}}. \end{aligned} \quad (3-4)$$

证明了 $\tilde{\mathbf{L}}^T = \tilde{\mathbf{L}}$, 即 $\tilde{\mathbf{L}}$ 也为实对称矩阵。实对称矩阵一定有 N 个线性无关的特征向量, 且这些特征向量彼此正交, 即特征向量矩阵为正交矩阵。所以得到, 度归一化的拉普拉斯矩阵 $\tilde{\mathbf{L}}$ 可相似对角化, 且对角阵上的元素为特征值。

由以上的证明可知, 拉普拉斯矩阵 $\tilde{\mathbf{L}}$ 一定可以谱分解且分解后有特殊形式:

$$\begin{aligned} \tilde{\mathbf{L}} &= \mathbf{U} \mathbf{\Lambda} \mathbf{U}^{-1}, \\ &= \mathbf{U} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{bmatrix} \mathbf{U}^{-1}, \\ &= \mathbf{U} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{bmatrix} \mathbf{U}^T. \end{aligned} \quad (3-5)$$

其中 \mathbf{U} 和 Λ 分别为 $\tilde{\mathbf{L}}$ 的特征向量矩阵和特征值矩阵, $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_N\}$, 特征值从小到大依次排列, 可以证明 $\lambda_i \in [0, 2)$ 。 $\tilde{\mathbf{L}}$ 的特征向量构成了在图空间中的一组基, 图上的傅里叶变换将该特征向量看成是傅立叶变换基。表3-1对比了传统傅立叶变换与图傅立叶变换。

表 3-1 传统傅立叶变换与图傅立叶变换的对比

	传统傅立叶变换	图傅立叶变换
傅立叶变换基	$e^{-2\pi i x v}$	\mathbf{U}^T
傅立叶反变换基	$e^{2\pi i x v}$	\mathbf{U}
基维度	无穷	节点个数 N

3.2.2 前置知识：频域图卷积网络

$$G * \mathbf{X} = \mathbf{U} (\underbrace{\mathbf{U}^T G}_{\text{频域卷积核}} \cdot \underbrace{\mathbf{U}^T \mathbf{X}}_{\text{频域图信号}}).$$

空域卷积核
空域图信号
频域乘积

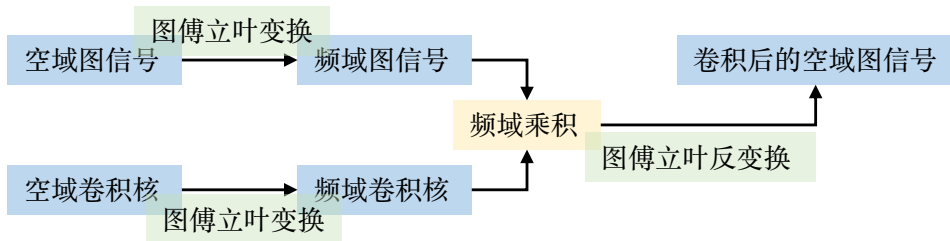


图 3-1 频域图卷积过程

定义卷积操作为 $*$, 乘积操作为 \cdot , 空域中的图信号为 \mathbf{X} , 频域中的图信号为 $\hat{\mathbf{X}}$, 空域图卷积核为 G , 频域图卷积核为 \hat{G} , 傅立叶变换函数为 $F(\cdot)$, 傅立叶反变换函数为 $F^{-1}(\cdot)$ 。定义傅立叶变换与傅立叶反变换形式如下:

$$\begin{aligned} \hat{\mathbf{X}} &= F(\mathbf{X}) = \mathbf{U}^T \mathbf{X}, \\ F^{-1}(\hat{\mathbf{X}}) &= \mathbf{U} \hat{\mathbf{X}}. \end{aligned} \tag{3-6}$$

根据“频域上的乘积等同于空域上的卷积”定理^[69], 有:

$$G * \mathbf{X} = \mathbf{U} (\mathbf{U}^T G \cdot \mathbf{U}^T \mathbf{X}). \tag{3-7}$$

图3-1简单地展示了这一卷积过程。

第一代频域图卷积网络^[75]，称它为 GCN-1。对于输入图信号 \mathbf{X} ，GCN-1 取以 $\theta \in \mathbb{R}^N$ 为参数的频域卷积核 \hat{G}_θ 进行卷积：

$$\begin{aligned} G * \mathbf{X} &= \mathbf{U} \left(\hat{G}_\theta \mathbf{U}^T \mathbf{X} \right) \\ &= \mathbf{U} \left(\begin{bmatrix} \theta_1 & & & \\ & \theta_2 & & \\ & & \ddots & \\ & & & \theta_N \end{bmatrix} \mathbf{U}^T \mathbf{X} \right). \end{aligned} \quad (3-8)$$

GCN-1 的计算复杂度极高。考虑到参数 $\theta \in \mathbb{R}^N$ ，如果图的节点数量达到了百亿级别，那么 GCN-1 的参数数量也将达到百亿级别。此外，GCN-1 的卷积操作缺乏局部连接性。

于是，第二代图卷积网络 ChebyNet^[76]提出使用切比雪夫多项式 (Chebyshev polynomials)^[77]来近似 GCN-1 中所需的众多参数，且通过这一近似操作达成了卷积的局部连接性。切比雪夫多项式在逼近理论中被广泛应用于多项式插值。一般的多项式插值方法是将 x 的零次方、一次方，一直到 x 的 K 次方作为基：

$$p(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \dots + \beta_K x^K. \quad (3-9)$$

切比雪夫多项式则是以递归方式定义的一系列正交多项式序列作为基来进行多项式插值：

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= x, \\ T_{n+1}(x) &= 2xT_n(x) - T_{n-1}(x), \\ p(x) &= \beta_0 T_0(x) + \beta_1 T_1(x) + \beta_2 T_2(x) + \dots + \beta_K T_K(x). \end{aligned} \quad (3-10)$$

ChebyNet 认为，频域卷积核的取值与特征值相关，可以利用 K 阶切比雪夫多项

式来近似 \hat{G}_θ :

$$\begin{aligned}
G * \mathbf{X} &= \mathbf{U} \left(\hat{G}_\theta \mathbf{U}^T \mathbf{X} \right) \\
&= \mathbf{U} \left(\sum_{k=0}^K \beta_k T_k(\Lambda) \mathbf{U}^T \mathbf{X} \right) \\
&= \sum_{k=0}^K \mathbf{U} \beta_k \left(\sum_{c=0}^k \alpha_{kc} \Lambda^k \right) \mathbf{U}^T \mathbf{X} \\
&= \sum_{k=0}^K \beta_k \left(\sum_{c=0}^k \alpha_{kc} (\mathbf{U} \Lambda \mathbf{U}^T)^k \right) \mathbf{X} \\
&= \sum_{k=0}^K \beta_k T_k(\mathbf{U} \Lambda \mathbf{U}^T) \mathbf{X} \\
&= \sum_{k=0}^K \beta_k T_k(\tilde{\mathbf{L}}) \mathbf{X}.
\end{aligned} \tag{3-11}$$

至此, ChebyNet 成功地将 GCN-1 的参数数量从 $O(N)$ 级别降低到 $O(K)$ 级别。GCN^[56]对 ChebyNet 进行了进一步的简化, 它直接令 $K = 2$, 得到:

$$G * \mathbf{X} = \beta_0 \mathbf{X} + \beta_1 \tilde{\mathbf{L}} \mathbf{X}. \tag{3-12}$$

继续做约束, $\beta = \beta_0 = -\beta_1$:

$$\begin{aligned}
G * \mathbf{X} &= \beta \mathbf{X} - \beta \tilde{\mathbf{L}} \mathbf{X} \\
&= \beta (\mathbf{I} - \tilde{\mathbf{L}}) \mathbf{X} \\
&= \beta \left(\mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \right) \mathbf{X}.
\end{aligned} \tag{3-13}$$

表3-2对比了前文讨论的三种图卷积网络的卷积核。GCN-1 的参数数量与图的规模

表 3-2 比较不同图卷积网络的卷积核

	卷积公式	相对参数数量
GCN-1	$\mathbf{U} \left(\hat{G}_\theta \mathbf{U}^T \mathbf{X} \right)$	N
ChebyNet	$\sum_{k=0}^K \beta_k T_k(\tilde{\mathbf{L}}) \mathbf{X}$	K
GCN	$\beta \left(\mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \right) \mathbf{X}$	1

成正比, 导致其可执行性不高。GCN 的参数数量比 ChebyNet 少, 但表达能力不及后者。然而, 由于 GCN 具有快速的收敛速度和较少的参数数量, 目前在实际应用

中被广泛采用。继续给出 GCN 的卷积结构：

$$\mathbf{Z} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}} \mathbf{X} \Theta. \quad (3-14)$$

其中， $\mathbf{Z} \in \mathbb{R}^{N \times C}$ 为卷积后的特征矩阵， $\Theta \in \mathbb{R}^{F \times C}$ 为卷积核的参数矩阵， C 为变换后的特征维度。

3.2.3 图网络中的高低频信息配比

ChebyNet 中 K 阶切比雪夫多项式的近似代表了聚合 $K - 1$ 跳邻居信息。当 $K = 1$ 时，聚合时仅使用了中心节点信息；当 $K = 2$ 时，聚合时使用了中心节点以及一跳邻居信息。值得注意的是，这种频域上的近似自身及多跳邻居信息的聚合方式与类似于 GraphSAGE^[12] 中的直接空域信息聚合方式有所不同。GraphSAGE 直接在空域中融合邻居节点的特征，而在频域图卷积网络中，信息的聚合是通过在频域中进行转换来实现的。

GCN 通过二阶切比雪夫多项式近似减少了参数数量：

$$\begin{aligned} \text{Conv}(\mathbf{X}) &= (\mathbf{I} - \tilde{\mathbf{L}}) \mathbf{X} \Theta \\ &= (\mathbf{U}(1 - \Lambda) \mathbf{U}^T) \mathbf{X} \Theta. \end{aligned} \quad (3-15)$$

其中， $1 - \Lambda$ 代表一个滤波器，其频率响应函数为 $1 - \lambda_i \in (-1, 1]$ 。该函数是一个线性收缩的函数，会对图信号产生低通滤波的作用，从而在计算过程中存在高频信息丢失的问题^[78]。这种低通滤波性对于平滑图数据和增强局部连接性是有益的，但是在处理具有复杂结构的图数据时会导致重要的高频信息丢失，从而影响模型效果。

通过叠加 GCN 的卷积核，可以获得低通效果更强的滤波器：

$$\text{Conv}_t(\mathbf{X}) = (\mathbf{I} - \tilde{\mathbf{L}})^t \mathbf{X} \Theta. \quad (3-16)$$

其中， t 代表叠加的卷积核幂次， t 值越大，低通效果越强。

本节做了一个简单的探究实验，选用 Amazon music (AMusic)^[79] 数据集。AMusic 是一个音乐点评数据集，包括不同的用户对不同的乐曲所做的评分。

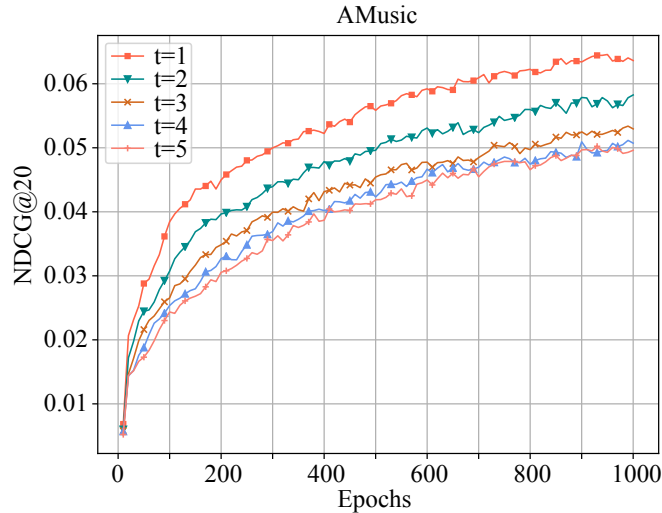


图 3-2 修改参数 t ，模型在 AMusic 数据集上的表现

AMusic 包括了 1776 个用户，12929 首乐曲以及 46087 条评分。使用 AMusic 数据集构建一个用户-乐曲的评分二部图：用户和乐曲作为图的节点，如果用户和乐曲之间存在评分，那么为该用户节点和乐曲节点之间增加一条边。如此构建的二部图属于异质图：具有两类节点和一类边，会形成较复杂的拓扑结构。

实验首先使用用户 ID 和乐曲 ID 来初始化用户节点和乐曲节点特征，之后对用户节点和乐曲节点分别使用 $\text{Conv}_t(\cdot)$ 进行卷积，以获得用户嵌入 \mathbf{E}_u 及乐曲嵌入 \mathbf{E}_m 。最终，可以获得预测的用户-乐曲评分为：

$$\hat{\mathbf{R}} = \mathbf{E}_u^T \mathbf{E}_m. \quad (3-17)$$

使用 BPR loss^[80]来训练该模型，存在连边的用户与乐曲对作为正样本对，否则为负样本对。以 NDCG@K^[81]作为评测指标，选用 $K = 20$ 是普遍设置，该指标的值越大代表模型的效果越好。图3-2展示了实验效果，可以看出当 t 越大时，模型的表现就越差。随着 t 的增加，模型的低通效果也增加，即越多的高频信号被去除。这种逐渐降低的模型表现验证了，在具有复杂拓扑的图数据集上，过强的低通滤波效果是有害的。

节点分类任务，被倾向于认为同类节点将具有相似属性，于是促使大部分图卷积网络使用低通滤波器来完成节点分类。接下来本节继续引入两种改变图中高低频信息占比的图卷积核：广义拉普拉斯卷积核以及跨跳连接图卷积核，并在后续实验部分验证了即使对于节点分类任务，高频信息也是不可忽略的。

广义拉普拉斯卷积核

通过引入频率响应系数 κ 到拉普拉斯卷积核中，将获得广义拉普拉斯卷积核：

$$\text{Conv}_{\kappa}(\mathbf{X}) = (\mathbf{I} - \kappa\tilde{\mathbf{L}})\mathbf{X}\Theta. \quad (3-18)$$

当 $\kappa = 1$ 时就是 GCN 的卷积核。

定义 $\tilde{\mathbf{L}}$ 中的最大特征值为 λ_{\max} 。当控制 $\kappa < \frac{1}{\lambda_{\max}}$ 时， $\text{Conv}_{\kappa}(\cdot)$ 仍然是一个低通滤波器，但是会保留部分的高频分量^[54]。于是可以通过设置 κ 来达到图网络中高频信息与低频信息的均衡。

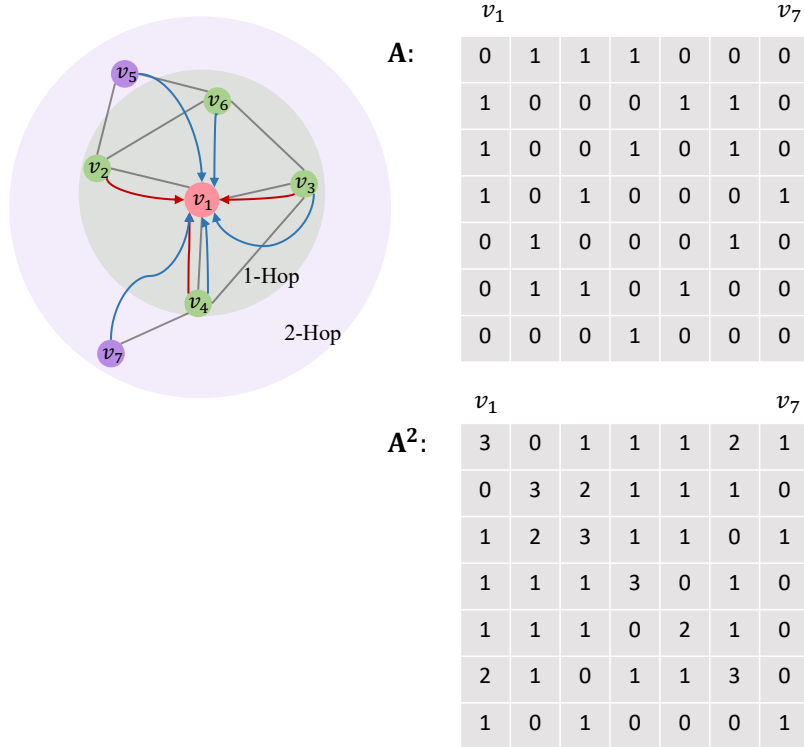


图 3-3 跨跳连接与单跳连接图卷积过程

跨跳连接图卷积核

在无权图中，邻接矩阵的幂次矩阵 \mathbf{A}^m 的元素 a_{ij}^m 表示从节点 v_i 到 v_j 有多少条长度为 m 的可行路径。这为实现跨跳连接提供了可能性。通过利用邻接矩阵的幂次矩阵，可以使得中心节点在卷积过程中一次性观察到更远处的节点。例如，利用 \mathbf{A}^2 进行卷积可以让中心节点一次性看到邻居节点的邻居节点。

计算经过原图变换定义出的 2 跳连接形成的新图 $\bar{\mathcal{G}}$ 的邻接矩阵及度矩阵： $\bar{\mathbf{A}} = \mathbf{A}^2, \bar{\mathbf{D}}_{ii} = \sum_j \bar{\mathbf{A}}_{ij}$ 。给出跨 2 跳连接的图卷积核形式：

$$\text{Conv2}(\mathbf{H}) = \bar{\mathbf{D}}^{-\frac{1}{2}} \bar{\mathbf{A}} \bar{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H} \Theta. \quad (3-19)$$

定义 $\mathbf{M} = \bar{\mathbf{D}}^{-\frac{1}{2}} \bar{\mathbf{A}} \bar{\mathbf{D}}^{-\frac{1}{2}}$ 及 $\hat{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$ 。

在卷积操作中引入这种跨跳连接，使得信息传播的范围更加广。当将邻接矩阵的幂次矩阵的拉普拉斯矩阵与邻接矩阵的拉普拉斯矩阵结合使用时，可以实现图卷积网络中整体性与局部性的均衡。这种组合操作有助于达到高频信息与低频信息的平衡，提高模型对图数据的全局和局部结构的感知。

图3-3展示了跨跳连接和单跳连接的图卷积网络在空域上对应的表现，图中仅展示了 v_1 节点作为中心节点进行信息聚合的过程。图中绿色节点为 v_1 的 1-hop 邻居，紫色节点为 v_1 的 2-hop 邻居，灰色连线代表原图的边。红色箭头连线代表单跳邻居信息聚合，蓝色箭头连线代表跨跳邻居信息聚合。其中 \mathbf{A} 和 \mathbf{A}^2 分别为该图的邻接矩阵以及邻接矩阵的平方。需要注意的是，实际的操作过程是在频域完成的，这样的图示仅为了方便理解。

3.2.4 层间信息交互方式

前文讨论了频域图卷积核的形式，这些卷积核用于完成图内的信息聚合。除此之外，叠加图卷积层的方式也是图神经网络的一个研究重点。随着图卷积层数的增加，模型的表达能力可以得到提升，但也可能引发图网络的过度平滑现象。因此，本章还探究了如何缓解图的过度平滑现象以及如何实现良好的层间信息交互。

图的过度平滑（over smoothing）现象是指在图神经网络中，随着网络层数的增加，节点的嵌入由于聚合了越来越多的邻居信息，逐渐变得模糊和相似，导致节点之间的差异性减弱。换句话说，即使是在原始图中具有明显差异的节点，在经过多层图卷积操作后，它们的特征表示会趋向于一致。这种现象会严重影响模型的性能，导致其在节点分类、图分类等任务中表现不佳。

定义当前任务为图节点分类，图卷积层的输入为 $\mathbf{H}^{(0)}$ ，经过 l 次图卷积得到

的输出为 $\mathbf{H}^{(l)}$:

$$\mathbf{H}^{(l)} = \text{ReLU} \left(\text{Conv} \left(\mathbf{H}^{(l-1)} \right) \right). \quad (3-20)$$

假定图网络一共有 L 层，一般的图卷积网络通常会使用最后一层的图卷积结果经过一层 MLP 和 Softmax 函数，即 $\text{Softmax}(f(\mathbf{H}^{(L)}))$ ，作为最终的预测结果。然而，通过不断叠加图网络层获得的最终输出很容易因为节点的区分度不够而导致过度平滑现象的发生。

可以通过改变层间交互方式来增加节点之间的差异性，从而减轻过度平滑现象。本章讨论了多种实现层间交互的方式如下：

第一种层间交互方式 GCN-ResPre 改进了每层图卷积层的结构， $\mathbf{H}^{(l)}$ 为第 l 层的图卷积层输出：

$$\mathbf{H}^{(l)} = \text{ReLU} \left(\text{Conv} \left(\mathbf{H}^{(l-1)} \right) + \mathbf{H}^{(l-1)} \right). \quad (3-21)$$

这种方式融合了上层的信息聚合结果和上层的直接输出结果。方法借鉴了 ResNet^[50] 及 ResGCN^[82] 的残差连接的思想。

第二种层间交互方式 GCN-ResIni 改进了每层图卷积层的结构：

$$\mathbf{H}^{(l)} = \text{ReLU} \left(\text{Conv} \left(\mathbf{H}^{(l-1)} \right) + \mathbf{H}^{(0)} \right). \quad (3-22)$$

这种方式融合了上层的信息聚合结果和首层的输入信息，即当前层的输出始终会保留输入的原始信息，从而增加节点之间的区分度。

第三种层间交互方式 GCN-Add 改进了图网络的最后输出：

$$\mathbf{H} = \text{Softmax} \left(f \left(\sum_{i=0}^L \mathbf{H}^{(i)} \right) \right). \quad (3-23)$$

其中， \mathbf{H} 代表图网络的最终输出结果， f 表示一层 MLP。

第四种层间交互方式 GCN-Cat 改进了图网络的最后输出：

$$\mathbf{H} = \text{Softmax} \left(f \left(\left[\mathbf{H}^{(0)}; \mathbf{H}^{(1)}; \dots; \mathbf{H}^{(L)} \right] \right) \right). \quad (3-24)$$

其中， $[\cdot]$ 代表矩阵拼接 (concatenation) 操作。GCN-Add 及 GCN-Cat 的输出结

果直接融合了之前所有图卷积层的输出以及初始输入信息。这样使得模型能够充分利用所有层的信息，从而提高了模型对不同粒度的图信息的感知，进而提升节点区分度及模型的性能。

第五种层间交互方式 GCN-Dense 同时修改了每层的图卷积层结构以及最终的网络输出形式，每层的图卷积层结构如下所示：

$$\begin{aligned}\mathbf{H}^{(l+1)} &= \text{ReLU} \left(\text{Conv} \left(\frac{1}{l+1} \sum_{i=0}^l \mathbf{H}^{(i)} \right) \right) \\ &= \text{ReLU} \left(\hat{\mathbf{A}} \left(\frac{1}{l+1} \sum_{i=0}^l \mathbf{H}^{(i)} \right) \mathbf{W}^{(l+1)} \right).\end{aligned}\quad (3-25)$$

最后的输出层：

$$\mathbf{H} = \text{Softmax} \left(f \left(\sum_{i=0}^L \mathbf{H}^{(i)} \right) \right).\quad (3-26)$$

一方面，每层的卷积层都融合了前面所有层的输出，另一方面，最后的预测结果也由所有层的输出组合而成。

第六种层间交互方式 GCN-Dense* 使用和 GCN-Dense 一致的卷积层，并修改了最后的输出层：

$$\mathbf{H} = \text{Softmax} \left(f \left(\mathbf{H}^{(0)} + \sum_{i=1}^L \hat{\mathbf{A}} \mathbf{H}^{(i)} \right) \right).\quad (3-27)$$

该输出层相当于对之前所有的卷积层输出结果再经过了一次信息聚合操作 $\hat{\mathbf{A}}(\cdot)$ 。

本章将在实验部分对这些层间交互方式进行比较和评估，以验证它们对模型性能的影响以及对过度平滑现象的缓解情况。基于上述对于卷积核和层间交互方式的讨论，本章提出了一种频域图卷积模型：Simple Graph Convolutional Network with Multi-Hop connection (SGCN-MH)。首先，SGCN-MH 是高低频自适应的，它会自动选择所需的高频和低频信息，并确定它们的配比。其次，SGCN-MH 是学习简便的，它仅保留了首层图卷积层中的特征变换矩阵，在无需叠加过多卷积层的情况下获得了与深度网络相匹配的模型性能。最后，它通过增加层间信息交互减轻了过平滑的风险。

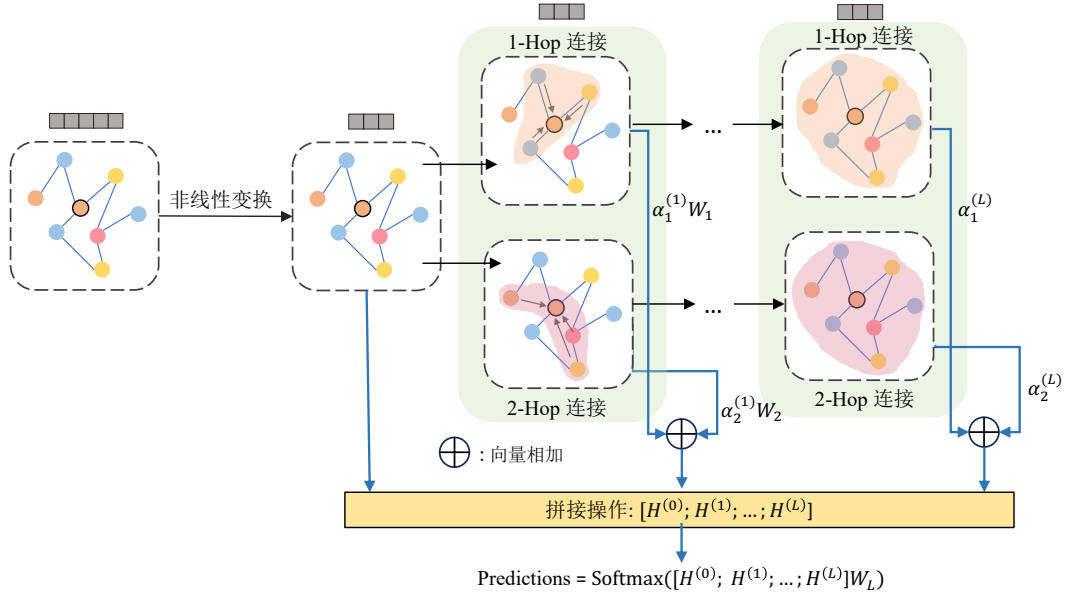


图 3-4 SGCN-MH 模型结构

3.2.5 跨跳连接的简化图卷积网络 SGCN-MH

本节将介绍 SGCN-MH 的整体结构。其结构如图3-4所示，图中不同颜色的节点代表不同的类别，信息聚合过程仅展示了具有黑色边框的橙色节点作为中心节点的聚合过程。

首先将原图的特征矩阵 \mathbf{X} 经过一层的特征变换矩阵与非线性激活函数来获得压缩后的特征矩阵：

$$\mathbf{H}^{(0)} = \text{ReLU}(\mathbf{X}\mathbf{W}). \quad (3-28)$$

其中， $\mathbf{W} \in \mathbb{R}^{F \times P}$ 代表需要学习的特征变换矩阵， P 为变换后的特征维度， $\mathbf{H}^{(0)} \in \mathbb{R}^{N \times P}$ 代表图卷积网络层的输入， (i) 代表所在图卷积层的层次。网络中还包含 Dropout 操作，为了简洁起见，后文不再重复说明。

将原始图卷积 $\text{Conv}(\cdot)$ 及跨跳连接 $\text{Conv2}(\cdot)$ 的图卷积组合使用来完成卷积操作，首层的卷积操作如下所示：

$$\begin{aligned} \mathbf{H}^{(1)} &= \text{ReLU} \left(\alpha_1^{(1)} \hat{\mathbf{A}} \mathbf{H}^{(0)} \mathbf{W}_1 + \alpha_2^{(1)} \mathbf{M} \mathbf{H}^{(0)} \mathbf{W}_2 \right), \\ \alpha_1^{(1)} &= \frac{\exp(\alpha^{(1),1})}{\sum_{i=1}^2 \exp(\alpha^{(1),i})}, \\ \alpha_2^{(1)} &= \frac{\exp(\alpha^{(1),2})}{\sum_{i=1}^2 \exp(\alpha^{(1),i})}. \end{aligned} \quad (3-29)$$

其中, $\mathbf{W}_1 \in \mathbb{R}^{P \times P}$, $\mathbf{W}_2 \in \mathbb{R}^{P \times P}$ 是首层需要学习的特征变换矩阵。特征变换矩阵输出维度与输入一致, 以便后续进行多层信息交互。 $\alpha^{(1)}$ 为模型首层端对端自动学习的平衡系数, 平衡系数维度为 2, $\alpha^{(1),t}$ 代表 $\alpha^{(1)}$ 的第 t 个维度, 其中 **Softmax** 操作用于将平衡系数转换到 $[0, 1]$ 范围内。

$\hat{\mathbf{A}}\mathbf{H}^{(0)}\mathbf{W}_1$ 代表单跳连接图卷积过程, 而 $\mathbf{M}\mathbf{H}^{(0)}\mathbf{W}_2$ 代表跨跳连接图卷积过程。通过 $\alpha^{(i)}$ 的学习, **SGCN-MH** 可以自动调整图中整体性与局部性的均衡。当 $\alpha^{(i),1}$ 更大, 模型更侧重于图上的局部性, 即关注低频信息; 当 $\alpha^{(i),2}$ 更大, 模型更侧重于图上的整体性, 即关注高频信息。

事实上, 也可以通过自适应地选择广义拉普拉斯卷积核的频率响应系数来完成图的高低频信息自适应。然而, 由于广义拉普拉斯卷积核中的参数 κ 修改会涉及到每次修改卷积核, 这将为网络学习和分析带来难度。因此, **SGCN-MH** 选择了采用跨跳连接的自适应方式。

之后的图卷积层去除了其中的非线性激活函数 $\text{ReLU}(\cdot)$ 以及特征变换矩阵:

$$\begin{aligned}\mathbf{H}^{(l)} &= \alpha_1^{(l)} \hat{\mathbf{A}}\mathbf{H}^{(l-1)} + \alpha_2^{(l)} \mathbf{M}\mathbf{H}^{(l-1)}, \\ \alpha_1^{(l)} &= \frac{\exp(\alpha^{(l),1})}{\sum_{i=1}^2 \exp(\alpha^{(l),i})}, \\ \alpha_2^{(l)} &= \frac{\exp(\alpha^{(l),2})}{\sum_{i=1}^2 \exp(\alpha^{(l),i})}.\end{aligned}\tag{3-30}$$

其中, $\mathbf{H}^{(l)}$ 表示第 l 层图卷积层的输出, $\alpha^{(l)}$ 表示第 l 层图卷积层的单跳及跨跳平衡系数。卷积操作 $\hat{\mathbf{A}}\mathbf{H}^{(l-1)}$ 和 $\mathbf{M}\mathbf{H}^{(l-1)}$ 仅包括信息聚合过程, 而不包含特征变换。这种变化去除了图网络中的冗余参数, 降低了过拟合的风险, 同时加速了模型的训练。本章将在后续的实验部分验证去除这种参数的有效性。

定义图卷积层的总层数为 L , 聚合图卷积层的多层输出结果如下:

$$\hat{\mathbf{Y}} = \text{Softmax} \left(\left[\mathbf{H}^{(0)}; \mathbf{H}^{(1)}; \dots; \mathbf{H}^{(L)} \right] \mathbf{W}_L \right).\tag{3-31}$$

其中, $\mathbf{W}^{(L)} \in \mathbb{R}^{((L+1)*P) \times C}$ 代表输出层的特征变换矩阵, 预测结果 $\hat{\mathbf{Y}} \in \mathbb{R}^{N \times C}$ 。拼接操作结合了每层图卷积层的输出, 并通过同一个特征变换矩阵进行学习, 从而实现了多层信息交互。

最终模型使用交叉熵损失函数进行训练：

$$L = -\frac{1}{N} \sum_i^N \mathbf{Y}_i \log \hat{\mathbf{Y}}_i. \quad (3-32)$$

算法 1 基于频率自适应的简化频域图卷积网络

输入： 节点特征矩阵 \mathbf{X} ；单跳连接图卷积核 $\hat{\mathbf{A}}$ ；跨跳连接图卷积核 \mathbf{M} ；需要学习的权重矩阵 $\mathbf{W}, \mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_L$ ；每层的平衡系数 $\alpha^{(l)}, l = 1, \dots, L$ ；激活函数 $\text{ReLU}(\cdot), \text{Softmax}(\cdot)$ ；节点标签 \mathbf{Y} 。

超参数： 学习率 β ，模型层数 L 。

输出： 预测的节点标签 $\hat{\mathbf{Y}}$ 。

```

1: for ep in #epochs do
2:    $\mathbf{H}^{(0)} = \text{ReLU}(\mathbf{X}\mathbf{W})$ ;
3:    $\mathbf{H}^{(1)} = \text{ReLU}(\alpha_1^{(1)}\hat{\mathbf{A}}\mathbf{H}^{(0)}\mathbf{W}_1 + \alpha_2^{(1)}\mathbf{M}\mathbf{H}^{(0)}\mathbf{W}_2)$ ;
4:   for l = 2 to L do
5:      $\mathbf{H}^{(l)} = \alpha_1^{(l)}\hat{\mathbf{A}}\mathbf{H}^{(l-1)} + \alpha_2^{(l)}\mathbf{M}\mathbf{H}^{(l-1)}$ ;
6:   end for
7:    $\hat{\mathbf{Y}} = \text{Softmax}([\mathbf{H}^{(0)}; \mathbf{H}^{(1)}; \dots; \mathbf{H}^{(L)}]\mathbf{W}_L)$ ;
8:   使用损失函数公式 (3-32) 更新模型;
9: end for
10: return 预测的节点标签  $\hat{\mathbf{Y}}$ .

```

SGCN-MH 的算法伪代码如算法 1 所示。

3.3 实验与分析

为了验证 SGCN-MH 的合理性与优越性，本节在四个数据集上进行了充分的节点分类任务的对比实验与消融实验。此外，本节还广泛地进行了图网络高低频信息利用情况、多层信息交互方式和合适信息传播模式的实验研究。

3.3.1 数据集

表 3-3 节点分类数据集

Dataset	#Nodes	#Edges	#Features	#Classes
Cora	2,708	5,429	1,433	7
Citeseer	3,327	4,732	3,703	6
Pubmed	19,717	44,338	500	3
Chameleon	2,277	36,101	2,325	4

本节使用了三个标准引文网络数据集：Cora, Citeseer, Pubmed^[83]以及一个 web networks 数据集：Chameleon^[84]。在标准引文网络数据集中，节点对应于文档，边对应于文档之间的引用关系；每个节点特征对应于文档的词袋表示，并隶属于一个学术主题。在 web networks 数据集中，节点和边分别代表网页和超链接；每个节点的特征是对应页面的词袋表示。数据集的具体统计数据见表3-3所示。

遵循 GCNII^[74]中的设置，实验对每个数据集进行了随机分割，将其分为 60%、20% 和 20% 用于训练集、验证集和测试集，并在 10 个随机分割上评估所有模型在测试集上的性能。

3.3.2 对比基线及实验设置

本节使用了如下的对比基线：

- GAT^[11]。在图卷积层中引入了注意力机制，对邻域中不同节点分配不同权重。GAT 在 inductive 和 transductive 设置下都表现良好。
- APPNP^[85]。利用 GCN 和 PageRank 之间的关系，提出了一种改进的基于个性化 PageRank 的消息传递模型。该模型将预测和传播过程解耦，解决了许多消息传递模型固有的邻居范围有限的问题，而且无需引入任何其他参数。
- JKNet^[86]。引入了影响范围的概念，即不同节点需要考虑不同的邻域信息范围。为了适应具有不同局部结构的子图，文中探索了跳跃知识 (Jumping Knowledge) 架构，该架构灵活地利用每个节点的不同邻域范围，以实现更好的结构感知表示。
- DropEdge^[73]。分析了图神经网络无法加深的两个原因：过拟合和过平滑，并提出了 DropEdge 机制，通过在模型训练过程中随机删除原始图中的边来缓解这两个问题。作者还从理论上证明了 DropEdge 可以降低过平滑的收敛速度，并减少由过平滑引起的信息损失。
- IncepGCN^[87]。从计算机视觉领域中的 Inception Network 改造得到。IncepGCN 使用了多分支结构，每个分支使用不同的卷积层数，输出时对所有分支的输出做拼接。

- GCNII^[74]。探讨了 GCN 中的过度平滑现象，并通过引入初始残差 (Initial Residual) 和等值映射 (Identity Mapping) 来扩展传统的 GCN，从而能够构建深度图卷积网络。理论和实证结果表明，这两种操作有效地缓解了过度平滑问题。GCNII* 是 GCNII 的一个变体，它在当前层的嵌入和初始特征上使用了不同的权重矩阵。
- MMA^[88]。改进了图神经网络中的邻居信息聚合操作：引入多个聚合器，并为每个聚合器使用了不同的加权掩码以及聚合操作 (max、min、mean)，从而增强模型的表达能力。
- CGT^[89]。提出了图学习中的度偏差 (Degree Biases) 问题：低度节点具有稀疏的邻居难以进行学习，而高度的节点则具有过多的邻居。于是文中提出在图社区结构上使用可学习的图数据增强和图 transformer 来解决在信息传递过程中的度偏差问题。

实验将学习率固定为 0.01，Dropout 率固定为 0.5，并将隐藏层维度固定为 64。在 Cora、Citeseer、Pubmed 和 Chameleon 数据集上的 weight decay 分别为 $1e-4$ 、 $5e-6$ 、 $5e-6$ 和 $5e-4$ 。层间的激活函数均使用 $\text{ReLU}(\cdot)$ 。在所有数据集上都使用均匀分布初始化和 Adam 优化器^[90]。SGCN-MH 在 Cora、Citeseer 和 Pubmed 数据集上的所有图卷积层都不使用特征变换及非线性激活函数，即仅包括了卷积核的卷积过程，SGCN-MH 在 Chameleon 上仅首层图卷积层有特征变换及激活函数。

3.3.3 对比实验

基线模型在有监督节点分类任务上的实验表现如表3-4所示，表格中括号内的数字表示模型中使用的图卷积层的层数，表格仅标出了较深层网络的层数和 SGCN-MH 所使用的层数。最高和第二高的性能分别以**粗体**和下划线标示。其他基线模型的实验结果均来自于已发表的论文：GCN、GAT 的实验结果来自^[91]，APPNP、JKNet、JKNet(+DropEdge)、IncepGCN(+DropEdge)、GCNII 和 GCNII* 的实验结果来自^[74]，MMA 和 CGT 的实验结果来自^[88]和^[89]。

SGCN-MH 在四个数据集下取得了和 GCNII 相匹配的实验结果，并在 Chameleon 上获得了超越 GCNII* 6.2% 的表现。图卷积网络通过增加卷积层的深度和参数量

表 3-4 基线模型在有监督节点分类任务上的表现 (平均准确率 Accuracy (%))

	Cora	Citeseer	Pubmed	Chameleon
GCN	85.77	73.68	88.13	28.18
GAT	86.37	74.32	87.62	42.93
APPNP	87.87	76.53	89.40	54.3
JKNet	85.25 (16)	75.85 (8)	88.94 (64)	60.07 (32)
JKNet (+DropEdge)	87.46 (16)	75.96 (8)	89.45 (64)	62.08 (32)
IncepGCN (+DropEdge)	86.86 (8)	76.83 (8)	89.18 (4)	61.71 (8)
GCNII	88.49 (64)	77.08 (64)	89.57 (64)	60.61 (8)
GCNII*	88.01 (64)	<u>77.13 (64)</u>	90.30 (64)	<u>62.48 (8)</u>
MMA	85.80	76.30	86.00	-
CGT	87.10	76.59	86.86	-
SGCN-MH	<u>88.41 (5)</u>	77.27 (2)	<u>89.68 (1)</u>	66.34 (3)

表 3-5 SGCN 与 GCNII 的参数量对比

	Cora	Citeseer	Pubmed	Chameleon
GCNII	354k	499k	294k	181k
GCNII*	616k	761k	556k	214k
SGCN-MH	94k	238k	32k	158k

来提升模型的表达能力, 进而提升模型效果。SGCN-MH 在不增加卷积层深度的情况下, 通过引入跨跳连接的图卷积层增加了节点的卷积域。同时, SGCN-MH 去除了图卷积层之间的多余参数和非线性激活函数, 并实现了与深层图卷积网络 (GCNII 和 GCNII*) 相匹配的实验结果, SGCN 与 GCNII 的参数量对比如表3-5所示。SGCN-MH 实现了少参数数量和快速推理的目标, 并取得了优秀的实验表现。

3.3.4 消融实验

为了验证 SGCN-MH 中各组件的合理性, 本节进行了消融实验, 实验结果见表3-6, 表中不同数据集下的最好模型表现和次好模型表现分别以**粗体**和下划线标示。

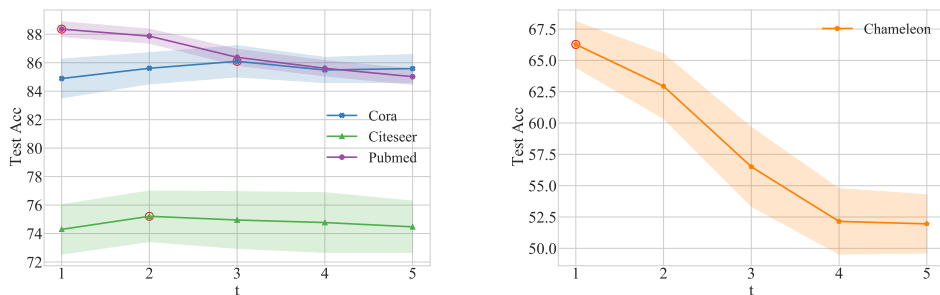
本节共使用了三个模型变体, 其中 SGCN 代表去除了卷积核的跨跳连接; SGCN-MH-nC 代表在 SGCN-MH 中去除了层间输出的拼接操作; GCN-MH 代表保留了图卷积核中的特征变换和非线性激活函数, 其余部分与 SGCN-MH 保持一致。首先, 对于 SGCN 来说, 当去除了卷积层间的跨跳连接后会降低模型表现, 在同层卷积下, 性能大致比 SGCN-MH 下降 1%-2% 左右。这是因为跨跳连

表 3-6 SGCN-MH 的消融实验 (平均准确率 Accuracy (%) \pm 标准差 (图卷积层层数))

Dataset	SGCN	SGCN-MH-nC	GCN-MH	SGCN-MH
Cora	85.41 ± 1.64 (1)	85.94 ± 1.09 (1)	87.71 ± 0.87 (1)	86.48 ± 1.81 (1)
	86.34 ± 1.53 (2)	86.12 ± 0.89 (2)	86.46 ± 1.22 (2)	87.93 ± 1.25 (2)
	85.75 ± 1.56 (3)	85.79 ± 0.75 (3)	86.86 ± 1.25 (3)	88.13 ± 1.22 (3)
	86.24 ± 1.28 (4)	85.71 ± 0.99 (4)	86.70 ± 1.00 (4)	88.27 ± 1.27 (4)
	86.30 ± 1.04 (5)	85.59 ± 1.15 (5)	86.74 ± 1.02 (5)	88.41 ± 1.18 (5)
Citeseer	75.49 ± 2.05 (1)	74.78 ± 1.42 (1)	75.95 ± 1.59 (1)	76.76 ± 1.60 (1)
	74.75 ± 2.00 (2)	74.05 ± 2.05 (2)	74.69 ± 2.18 (2)	77.27 ± 1.84 (2)
	74.03 ± 2.12 (3)	73.63 ± 1.40 (3)	74.12 ± 1.27 (3)	76.98 ± 1.60 (3)
	73.97 ± 1.72 (4)	73.34 ± 1.31 (4)	74.07 ± 1.91 (4)	76.56 ± 1.76 (4)
	73.94 ± 1.97 (5)	73.43 ± 1.84 (5)	74.49 ± 1.57 (5)	76.33 ± 1.54 (5)
Pubmed	89.45 ± 0.40 (1)	88.52 ± 0.47 (1)	89.66 ± 0.45 (1)	89.68 ± 0.44 (1)
	89.72 ± 0.47 (2)	87.89 ± 0.35 (2)	89.56 ± 0.59 (2)	89.65 ± 0.48 (2)
	89.76 ± 0.43 (3)	86.24 ± 0.49 (3)	89.74 ± 0.39 (3)	89.48 ± 0.36 (3)
	89.71 ± 0.34 (4)	85.14 ± 0.54 (4)	89.55 ± 0.57 (4)	89.57 ± 0.49 (4)
	89.72 ± 0.51 (5)	84.33 ± 0.41 (5)	89.69 ± 0.47 (5)	89.69 ± 0.45 (5)
Chameleon	65.18 ± 2.48 (1)	65.92 ± 2.48 (1)	66.10 ± 1.80 (1)	66.10 ± 1.80 (1)
	65.00 ± 2.54 (2)	59.01 ± 2.52 (2)	64.10 ± 2.75 (2)	65.15 ± 2.21 (2)
	64.67 ± 2.23 (3)	53.86 ± 3.84 (3)	66.43 ± 2.56 (3)	66.34 ± 2.18 (3)
	65.04 ± 2.32 (4)	50.18 ± 3.16 (4)	64.93 ± 3.12 (4)	65.90 ± 2.44 (4)
	64.56 ± 2.35 (5)	46.75 ± 4.22 (5)	65.29 ± 2.56 (5)	65.22 ± 3.11 (5)

接在同层卷积模型下提供了更多的邻居信息来获得节点表示,从而提升了模型的表达能力。其次,对于 SGCN-MH-nC 来说,当去除了层间的输出拼接操作后,模型的过度平滑现象变得明显,尤其是在 Citeseer、Pubmed 和 Chameleon 数据集上。当模型层数增加时,性能显著下降。最后,对于 GCN-MH 来说,在增加了特征变换矩阵和激活函数后,虽然增加了模型参数量,但并未提升模型效果,反而拖慢了模型的训练速度,并带来了过拟合的风险。

接下来本节将采用最简单的 GCN 模型作为骨干网络,并在此基础上研究高低频信息、多层信息交互方式和信息传播模式的选择。

图 3-5 叠加多层 GCN 卷积层的测试准确率折线图 (平均准确率 Accuracy (%) \pm 标准差)

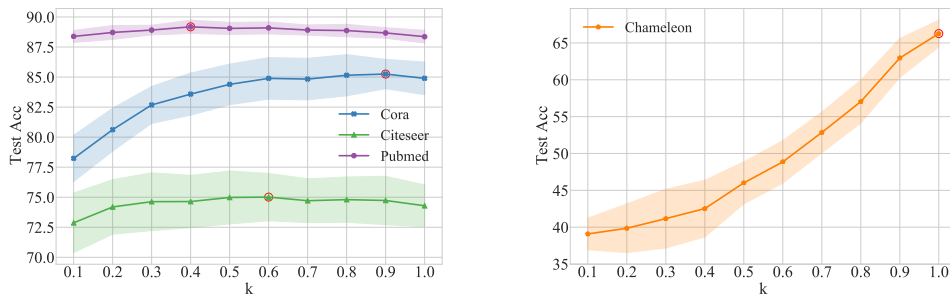


图 3-6 改变广义拉普拉斯卷积核参数 κ 的测试准确率折线图 (平均准确率 Accuracy (%) \pm 标准差)

表 3-7 跨跳连接图卷积核的作用 (平均准确率 Accuracy (%) \pm 标准差)

	Cora	Citeseer	Pubmed	Chameleon
GCN	84.89 ± 1.36	74.29 ± 1.74	88.36 ± 0.53	66.27 ± 1.82
GCN-2Hop	$85.94 \pm 1.09 (+1.24\%)$	$74.78 \pm 1.42 (+0.66\%)$	$88.52 \pm 0.47 (+0.18\%)$	$65.92 \pm 2.48 (-0.53\%)$
GCN-3Hop	$85.86 \pm 0.98 (+1.14\%)$	$74.61 \pm 1.24 (+0.43\%)$	$88.18 \pm 0.58 (-0.20\%)$	$66.78 \pm 2.40 (+0.77\%)$

3.3.5 高低频信息的选择实验

本节将探究通过不同的方式改变模型学习到的高低频信息对模型性能产生的影响。

首先, 本节通过叠加 GCN 的卷积核来获得低通效果更强的滤波器 Conv_t , 模型表现如图3-5所示, 图中横坐标 t 代表叠加的卷积核次数, 最好的模型表现处使用红色空心圆圈标识。可以获得如下结论:

(1) 对于数量级相对较大的 Pubmed 和 Chameleon 数据集来说, 增加叠加卷积核的次数会降低模型的性能。这验证了对于较复杂的图数据来说, 过强的低通滤波是有害的。

(2) 对于 Cora 和 Citeseer 数据集来说, 增加叠加卷积核的次数确实是有效的。然而, 当叠加的卷积核次数过多时, 模型性能也将迎来递减或齐平的趋势。

其次, 本节通过改变广义拉普拉斯卷积核的参数 κ 来探究引入高频信息对于模型性能产生的影响, 实验结果见图3-6, 图中横坐标 κ 代表参数 κ , 最好的模型表现处使用红色空心圆圈标识。当控制 $\kappa < \frac{1}{\lambda_{\max}}$ 时 ($\lambda_{\max} = 1$), 可以不使所有的高频成分去噪, 且当 κ 在此范围内越小时, 能够保留的高频成分就越多。可以获得如下结论:

(1) 对于 Cora, Citeseer 和 Pubmed 数据集来说, 测试准确率的折线图都呈现

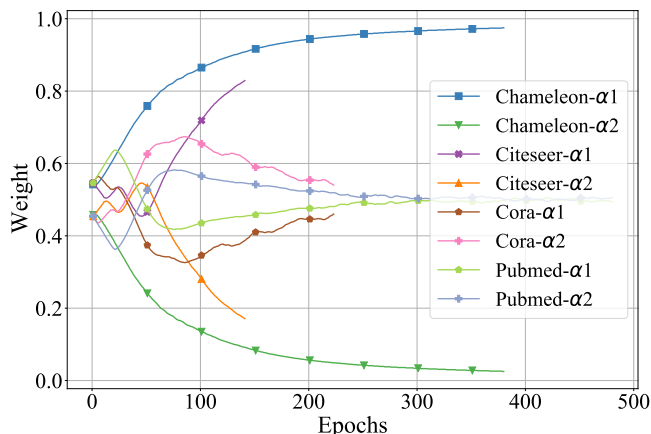


图 3-7 2-Hop 跨跳连接图卷积层平衡系数变化折线图

一种倒 U 型趋势，增加少量高频信息对于模型有效，但是当高频信息过多时又会导致模型性能下降。

(2) 对于 Chameleon 数据集来说，通过改变频率响应系数引入高频信息是对性能有害的，且高频信息越多，模型表现越差。结合叠加 GCN 的卷积核和广义拉普拉斯卷积核的实验，可以发现模型在 Chameleon 数据集上在 $t = 1, \kappa = 1$ 的情况下达到了最优。

再次，继续探究当增加跨跳卷积时会对模型产生什么影响，结果如表3-7所示，表中 GCN-2Hop 表示增加了跨两跳图卷积核，GCN-3Hop 表示增加了跨三跳图卷积核。可以获得如下结论：

(1) 对于 Cora 和 Citeseer 数据集来说，增加跨跳卷积会提升模型效果，GCN-2Hop 效果优于 GCN-3Hop 优于 GCN。

(2) 对于 Pubmed 数据集来说，GCN-2Hop 效果优于 GCN 优于 GCN-3Hop。

(3) 对于 Chameleon 数据集来说，GCN-3Hop 效果优于 GCN 优于 GCN-2Hop。

GCN-3Hop 会提供更远的视角，即增加模型注意到的全局性和高频信息。这种增加的高频信息对于不同的数据集影响不同。最后，本节可视化了 2-Hop 跨跳连接图卷积层中平衡系数 α_1 和 α_2 的训练变化，如图3-7所示，图中曲线从 epoch=1 处开始绘制，所有权重的初始值都为 0.5。图中中途停止的折线是由于设置的早停（early-stopping）机制，旨在防止模型过拟合。可以获得如下结论：

(1) 对于 Cora 和 Pubmed 数据集而言，1-Hop 跨跳连接在初期比 2-Hop 直接连接发挥更大的作用，之后呈现出趋近的趋势。Pubmed 数据集上的两条折线最后基本重合。

表 3-8 不同层间交互方式 GCN 的测试准确率 (平均准确率 Accuracy (%) \pm 标准差 (图卷积层层数))

Dataset	GCN	GCN-ResPre	GCN-ResIni	GCN-Add
Cora	84.89 \pm 1.36 (1)	85.25 \pm 1.41 (1)	85.73 \pm 1.31 (1)	85.41 \pm 1.18 (1)
	83.36 \pm 2.37 (2)	84.91 \pm 1.02 (2)	84.67 \pm 1.65 (2)	85.98 \pm 1.67 (2)
	82.92 \pm 1.29 (3)	83.10 \pm 1.95 (3)	83.60 \pm 2.54 (3)	85.21 \pm 1.50 (3)
	81.91 \pm 2.42 (4)	82.82 \pm 1.96 (4)	82.25 \pm 1.74 (4)	85.49 \pm 1.25 (4)
	82.13 \pm 2.23 (5)	82.72 \pm 1.83 (5)	82.96 \pm 2.41 (5)	85.15 \pm 1.38 (5)
Citeseer	74.29 \pm 1.74 (1)	75.34 \pm 1.31 (1)	75.38 \pm 1.49 (1)	75.56 \pm 1.46 (1)
	71.19 \pm 3.98 (2)	74.34 \pm 1.41 (2)	74.47 \pm 1.75 (2)	74.51 \pm 1.46 (2)
	65.21 \pm 8.85 (3)	71.86 \pm 2.55 (3)	73.17 \pm 1.66 (3)	73.27 \pm 2.35 (3)
	53.47 \pm 7.88 (4)	68.66 \pm 4.27 (4)	72.72 \pm 1.45 (4)	74.19 \pm 1.87 (4)
	50.36 \pm 8.12 (5)	63.99 \pm 5.14 (5)	72.34 \pm 1.74 (5)	73.84 \pm 1.86 (5)
Pubmed	88.36 \pm 0.53 (1)	88.67 \pm 0.47 (1)	89.25 \pm 0.46 (1)	89.21 \pm 0.42 (1)
	87.78 \pm 0.61 (2)	88.35 \pm 0.50 (2)	89.23 \pm 0.52 (2)	89.40 \pm 0.43 (2)
	86.31 \pm 0.53 (3)	87.57 \pm 0.49 (3)	89.05 \pm 0.52 (3)	89.29 \pm 0.48 (3)
	85.56 \pm 0.44 (4)	86.80 \pm 0.52 (4)	89.15 \pm 0.38 (4)	89.22 \pm 0.64 (4)
	84.52 \pm 0.77 (5)	86.54 \pm 0.47 (5)	89.07 \pm 0.56 (5)	89.39 \pm 0.69 (5)
Dataset	GCN-Cat	GCN-Dense	GCN-Dense*	
Cora	85.41 \pm 1.64 (1)	85.41 \pm 1.18 (1)	87.55 \pm 1.15 (1)	
	86.34 \pm 1.53 (2)	86.28 \pm 1.32 (2)	86.44 \pm 1.32 (2)	
	85.75 \pm 1.56 (3)	85.13 \pm 0.98 (3)	85.75 \pm 1.65 (3)	
	86.24 \pm 1.28 (4)	85.49 \pm 1.50 (4)	86.28 \pm 1.19 (4)	
	86.30 \pm 1.04 (5)	84.99 \pm 1.29 (5)	85.57 \pm 1.78 (5)	
Citeseer	75.49 \pm 2.05 (1)	75.56 \pm 1.46 (1)	75.91 \pm 1.91 (1)	
	74.75 \pm 2.00 (2)	74.85 \pm 1.58 (2)	75.02 \pm 1.33 (2)	
	74.03 \pm 2.12 (3)	74.69 \pm 1.86 (3)	73.70 \pm 1.87 (3)	
	73.97 \pm 1.72 (4)	73.97 \pm 1.83 (4)	73.70 \pm 1.58 (4)	
	73.94 \pm 1.97 (5)	73.66 \pm 2.07 (5)	73.99 \pm 1.98 (5)	
Pubmed	89.45 \pm 0.40 (1)	89.21 \pm 0.42 (1)	89.45 \pm 0.53 (1)	
	89.72 \pm 0.47 (2)	89.51 \pm 0.48 (2)	89.25 \pm 0.52 (2)	
	89.76 \pm 0.43 (3)	89.20 \pm 0.52 (3)	89.28 \pm 0.57 (3)	
	89.71 \pm 0.34 (4)	89.36 \pm 0.37 (4)	89.30 \pm 0.38 (4)	
	89.72 \pm 0.51 (5)	89.28 \pm 0.34 (5)	89.31 \pm 0.51 (5)	

(2) 对于 Citeseer 和 Chameleon 而言, 1-Hop 直接连接比 2-Hop 跨跳连接发挥更大的作用。且对 Chameleon 而言, 2-Hop 跨跳连接的系数一路走低, 最后贴近于 0, 它并不希望这样的跨跳连接。

在探讨 GCN 在不同数据集上的高低频信息影响后, 本节发现各数据集对于信息频率的需求存在差异。通过叠加 GCN 的卷积核, 可以获得具有更强低通效果的滤波器; 而通过调整广义拉普拉斯卷积核的参数 κ , 能够引入不同比例的高频信息; 通过引入跨跳连接卷积核和使用单跳连接卷积核, 可以灵活调节高低频信息的比例。

表 3-9 去除了 GCN 卷积层中的特征变换 (GCN-NW) 与非线性激活函数 (GCN-NA) 后的测试准确率 (平均准确率 Accuracy (%) \pm 标准差 (图卷积层层数))

Dataset	GCN	GCN-NW	GCN-NA
Cora	84.89 \pm 1.36 (1)	86.58 \pm 1.22 (1)	84.95 \pm 1.40 (1)
	83.36 \pm 2.37 (2)	87.42 \pm 0.74 (2)	84.49 \pm 1.21 (2)
	82.92 \pm 1.29 (3)	87.59 \pm 1.28 (3)	81.29 \pm 3.23 (3)
	81.91 \pm 2.42 (4)	87.22 \pm 1.00 (4)	77.83 \pm 8.31 (4)
	82.13 \pm 2.23 (5)	86.82 \pm 1.23 (5)	66.40 \pm 12.92 (5)
Citeseer	74.29 \pm 1.74 (1)	76.19 \pm 1.53 (1)	75.17 \pm 1.66 (1)
	71.19 \pm 3.98 (2)	76.26 \pm 1.68 (2)	72.88 \pm 1.59 (2)
	65.21 \pm 8.85 (3)	76.04 \pm 1.20 (3)	70.70 \pm 2.97 (3)
	53.47 \pm 7.88 (4)	75.56 \pm 1.30 (4)	69.60 \pm 4.25 (4)
	50.36 \pm 8.12 (5)	75.22 \pm 1.21 (5)	65.30 \pm 7.80 (5)
Pubmed	88.36 \pm 0.53 (1)	88.10 \pm 0.51 (1)	88.29 \pm 0.41 (1)
	87.78 \pm 0.61 (2)	87.42 \pm 0.50 (2)	87.87 \pm 0.35 (2)
	86.31 \pm 0.53 (3)	86.01 \pm 0.39 (3)	86.45 \pm 0.34 (3)
	85.56 \pm 0.44 (4)	84.96 \pm 0.36 (4)	85.60 \pm 0.46 (4)
	84.52 \pm 0.77 (5)	84.20 \pm 0.32 (5)	84.91 \pm 0.50 (5)

3.3.6 多层信息交互方式实验

在3.2.4小节中提出了六种实现层间交互的方式: GCN-ResPre, GCN-ResIni, GCN-Add, GCN-Cat, GCN-Dense, GCN-Dense*。本节将探讨它们对模型性能的影响,并研究它们是否有助于缓解图网络中的过度平滑现象,具体的实验结果如表3-8所示,相同图卷积层层数下最好的模型表现使用字体**加粗**标识。可以获得如下结论:

(1) 相较于未使用任何层间交互的原始 GCN 来说,增加任何一种层间交互方式都可以提升模型效果。原始 GCN 存在着较严重的过度平滑现象,当增加卷积层的层数时,模型性能持续下跌。

(2) GCN-ResPre 和 GCN-ResIni 均可以提升模型效果,两者都引入了残差连接的思想。在不同数据集的不同卷积层次下,GCN-ResIni 的表现基本上要优于 GCN-ResPre,且 GCN-ResIni 缓解过度平滑现象效果也要更优。这验证了直接保留原始输入信息到每层图卷积输入是一个较好的选择。GCN-ResPre 和 GCN-ResIni 还是存在着过度平滑现象,增加卷积层数并没有起到提升模型效果的作用。

(3) GCN-Add 和 GCN-Cat 通过直接改变图卷积网络的最后输出来实现层间交互,两者一定程度上缓解了过度平滑现象。GCN-Cat 的表现要优于 GCN-Add,

这是由于 GCN-Add 这样通过直接加和各层的输出为最后输出的方式可能会破坏每层学习到的特征嵌入。

(4) GCN-Dense 和 GCN-Dense* 从两个层面实现了多层信息交互，它们同时修改了图卷积核以及最后的输出层。GCN-Dense* 的模型效果要优于 GCN-Dense，额外为每层的输出结果增加信息聚合操作加入到最后的输出结果上可以提升模型效果。值得注意的是，GCN-Dense* 在每个数据集上只通过一层图卷积就得到了很好的效果，但是很可惜的是过平滑现象没有得到很好的缓解。

总结一下，GCN-Cat 和 GCN-Dense* 是相较于其他层间交互方式的更优选择，考虑到 GCN-Dense* 的额外信息聚合操作会带来更多的矩阵计算，最终 SGCN-MH 中选择采用 GCN-Cat 的方式实现层间交互。

3.3.7 合适信息传播模式实验

本节探究了 GCN 不同的信息传播模式对节点分类任务的影响，实验结果如表3-9所示，表中相同图卷积层层数下最好的模型表现使用字体**加粗**标识。其中 GCN-NW 去除了图卷积层内的特征变换矩阵，即 $\mathbf{H}^{(l)} = \text{ReLU}(\hat{\mathbf{A}}\mathbf{H}^{(l-1)})$ 。GCN-NA 去除了图卷积层内的非线性激活函数，即 $\mathbf{H}^{(l)} = \hat{\mathbf{A}}\mathbf{H}^{(l-1)}\Theta$ 。

实验结果表明，在节点分类任务下，去除图卷积层内的特征变换矩阵和非线性激活函数一般情况下能够提升模型的性能。此外，这一操作还有助于减轻模型的学习负担，降低过拟合的风险，并且加快模型的推理速度。

3.4 本章小结

本章深入探讨了频域图卷积网络的发展历程，并通过理论分析和实验验证了其中存在的高频信息丢失问题。本章使用了不同的方法，包括增加卷积核的叠加次数、使用广义拉普拉斯卷积核以及引入跨跳连接图卷积网络层，来探究在不同数据集上高低频信息的影响。此外，本章研究了六种不同的层间交互方式，并验证了它们对模型表达能力和缓解过度平滑现象的影响。SGCN-MH 通过共同使用跨跳连接图卷积核和单跳连接图卷积核来完成网络的高低频自适应，通过拼接层间输出来完成层间交互以及去除层间的特征变换矩阵来简化计算。SGCN-MH 在引文数据集和 web 网络数据集上仅使用浅层网络和少量的参数就超越了

深层的图卷积网络模型。

第四章 基于最小显著信息的无监督图 对比学习

4.1 研究背景与动机

给图打标签是一项非常具有挑战性和繁琐的任务，因为这通常需要掌握领域知识，并且图往往包含大量节点及复杂的关系。标记图不仅仅是简单地对每个节点进行分类，它需要考虑节点之间的连接方式以及它们在整个图中的位置和作用。这意味着标记过程需要进行深入的分析 and 判断，这将花费大量的时间和精力。因此，获得图的准确标签通常是困难的，并且一般需要领域专家的帮助。鉴于此，近年来无监督图表示学习^[92-94]受到越来越多的关注。无监督图表示学习的目标是在不使用标签信息的情况下学习出节点或整个图的低维表示。这些表示可以应用于各种下游任务，例如节点分类^[95]、图分类^[96]和图聚类^[97]等。

图对比学习 (Graph Contrastive Learning, GCL)^[57,98-99]由于其出色的数据利用效率和良好的泛化性能，已经在无监督图表示学习中显示出了巨大的潜力。在这里先引入一个概念：“视图”(View)。“视图”通常是对原始输入数据经过加工变换得到的新数据，该数据和原始数据的语义保持一致，但是和原始数据具有差别。例如，在图像数据中，视图可以是经过不同的图像变换，如旋转、缩放、平移等所生成的图像表示；在文本数据中，视图可以是通过不同的文本转换方法，如随机替换、随机插入、back translation 等所生成的新文本。而在图数据中，视图一般是通过加工图的特征矩阵和邻接矩阵所生成的新图。

对比学习，旨在通过比较视图之间的相似性和差异性来学习出有意义的表示。对比学习通常包括以下步骤：（1）视图生成：通过在原始图上使用数据增强 (Data Augmentation, DA) 来生成视图。（2）视图表示：使用编码器生成视图表示/Embedding。（3）视图对比：拉近正样本对视图表示，推远负样本

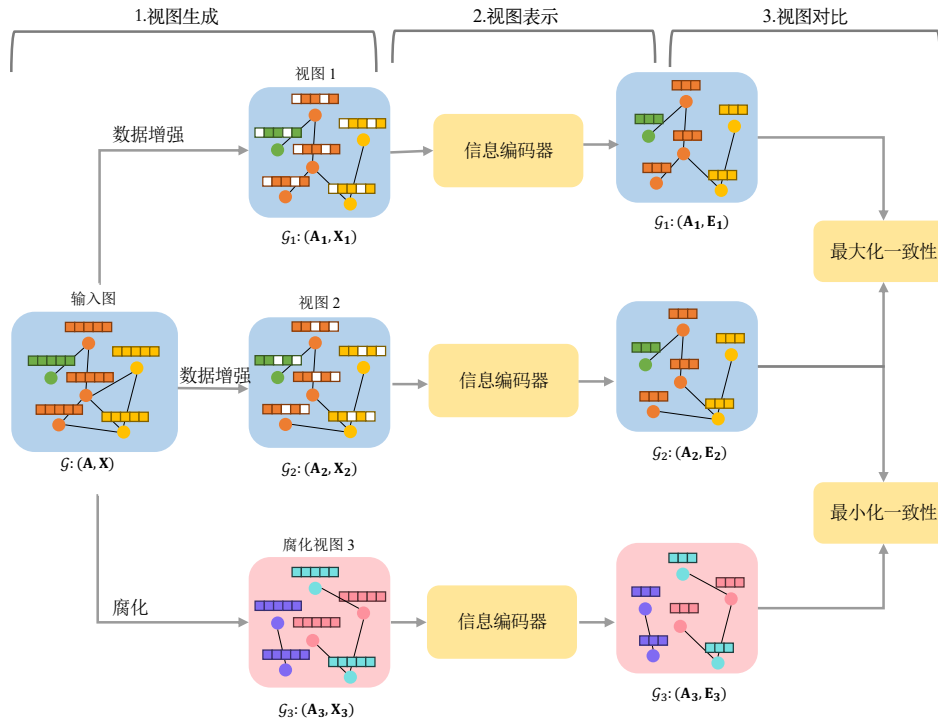


图 4-1 图对比学习过程

对视图表示。首先定义视图对的相似度或差异度，通常使用余弦相似度或其他度量方式。视图的相似度可以直接理解为视图一致性（View Agreement），并且通过互信息估计器（Mutual Information Estimator）来获得。互信息估计器包括：Noise-Contrastive Estimation (NCE)^[100-101]，Jensen-Shannon estimator (JSD)^[102]，Normalized Temperature-scaled cross-entropy (NT-Xent)^[103]等。图4-1展示了图对比学习的基本流程，图中蓝色背景的视图对为“正样本对”，粉色背景的视图对为“负样本对”。视图对比通过定义出的视图相似度来拉近来自相同分布的视图表示，同时将来自不同分布的视图表示推开，以优化编码器。来自相同分布的视图对被定义为“正样本对”，而来自不同分布的视图对被定义为“负样本对”。实际上，在图4-1中提到了一种称为腐化（corruption）的操作。这种操作旨在通过改变图的拓扑结构和特征信息来扰乱原始图的语义，从而生成负样本。然而，对于对比学习而言，使用腐化方法是不必要的^[63]。仅通过最大化正样本对之间的一致性就能达到足够好的效果。因此，本次研究的重点是图对比学习中的数据增强用于视图生成。

图对比学习对数据增强有着极大的依赖，如果选择不当的数据增强方式，将会导致模型性能严重下降。图数据增强需要综合考虑复杂的图拓扑结构以及特

征信息。目前已经涌现出众多图数据增强方法，可以根据它们的增强方式分为三大类别^[104]：结构导向型^[40,105-108]，特征导向型^[41,109-110]，以及标签导向型^[111-112]。

结构导向型图数据增强方法主要关注邻接矩阵，涉及改变节点、连边或直接通过邻接矩阵来提取图信息。这类方法的示例包括节点掩盖 (Node Masking)^[105]，连边扰动 (Edge Perturbation)^[40,106]，以及图扩散 (Graph Diffusion)^[107-108]等。其中节点掩盖和连边扰动是指随机掩码图中的节点及修改图中的边，修改边包括随机增加边或者减少边。邻接矩阵将边解释为纯粹的二元关系，即它们要么存在，要么不存在。然而，真实的关系要复杂得多。例如，在一个社交网络中，可能有一些联系紧密的好朋友，还有许多只见过一面的人，或者是好朋友的好朋友。这些不同类型的关系对于理解社交网络中的连接模式和社群结构至关重要。因此，可以尝试通过图扩散来重新建立这些复杂的关系。图扩散过程中首先将所有关注点集中在所考虑的节点上。随后，它持续地将一部分关注传递给节点的邻居，使得关注从起始节点开始逐渐分散开来。一段时间后，它停止关注点的传递，此时在节点上的关注点分布就确定了从起始节点到其他节点的新边。这种方法有助于捕捉节点之间更复杂的关系和连接模式。PageRank^[107]算法由 Google 创始人 Larry Page 提出，它是一种评估网页重要性的方法，主要针对网页进行排名，优化搜索引擎的搜索结果。PageRank 实际上利用了图扩散的思想，是有向图上的随机游走 (Random Walk)^[113]模型。该模型是一个一阶马尔可夫链，它描述了一个随机游走者如何沿着图的边随机移动，从一个节点扩散到另一个节点。在满足迭代次数或是两次迭代结果差异低于阈值时，随机游走停止，模型最终收敛到一个平稳分布并且获得了新的图关系分布和图中节点的打分。结构导向型图数据增强考虑图中的拓扑结构，重点学习网络连接模式。

特征导向型图数据增强专注于修改节点特征，例如特征破坏 (Feature Corruption)^[109-110]和特征掩盖 (Feature Masking)^[41]等。特征破坏和特征掩盖分别是指随机对图特征增加噪声或者是随机掩码一部分的特征。特征导向型图数据增强期望通过修改图特征来增加模型对节点特征的鲁棒性和泛化能力。标签导向型图数据增强则是通过修改标签来生成新样本，例如伪标签 (Pseudo-Labeling)^[111]和标签混合 (Label Mixing)^[112]等。伪标签图数据增强利用模型在未标记数据上的预测结果生成伪标签，将其与未标记数据合并形成扩充的数据集。标签混合图数据增强将首先选择两个或多个节点，之后将这些节点的标签进行混合来生成新

的节点加入数据集进行模型训练。标签导向型图数据增强的目标是增强模型对不同标签和样本之间关系的理解和学习能力，从而提升在图分析任务中的表现。由于本次讨论聚焦于无监督图对比学习，因此不会过多讨论标签导向型图数据增强方法。

由于图的拓扑关系和特征信息的复杂性，图数据增强与图像数据增强和文本数据增强有着明显的区别，并因此具有多种多样的定义方式。在选择合适的图数据增强方法时，需要综合考虑图的拓扑结构和特征信息，以及下游任务的需求和特点。实际上，图数据增强的好坏严格关联于下游任务与具体的图数据^[41]，这促使许多图对比学习方法在选择数据增强方案时依赖于试错法或者经验法。GraphCL^[41]中提出了四种图数据增强方案：Node Dropping、Edge Perturbation、Attribute Masking 以及 Subgraph。GraphCL 通过这些数据增强方案的俩俩配对共 25 种组合方式（包括直接使用原图作为视图的 Identical 方案，共 5*5 种组合）来生成对比视图，实验结果表示不同数据集上的最优数据增强方案互不相同。这表明了数据增强方案的良好与否取决于各种图数据集的特定特征。MVGRL^[42]探究了图对比学习中的数据增强方案、互信息估计器以及视图对比模式的组合在包括节点分类和图分类的不同下游任务上的表现。MVGRL 研究了四种结构导向型图数据增强方案，包括直接使用对称归一化的邻接矩阵、Personalized PageRank (PPR)^[114]、heat kernel^[108]以及通过 Floyd-Warshall^[115]算法计算得到的 pair-wise distance matrix。其中直接使用对称归一化的邻接矩阵的数据增强方案捕捉了图上的局部信息，而其余三种方案则是捕捉了图上的全局信息。MVGRL 验证了综合局部信息和全局信息的数据增强方案将取得更好的实验效果。MERIT^[116]中使用了四种数据增强方案：Graph Diffusion (GD)、Edge Modification (EM)、Subsampling (SS) 以及 Node Feature Masking (NFM)。MERIT 对这四种数据增强方案进行组合，使用 SS, EM, 以及 NFM 来生成第一个视图，以及叠加使用 SS + GD + NFM 来生成第二个视图。GCA^[106]通过使用人为定义的三种中心度：度中心度 (Degree Centrality)、特征向量中心度 (Eigenvector Centrality)、PageRank 中心度 (PageRank Centrality) 来生成数据增强方案，并通过试错法根据模型在下游任务中的表现来选择最佳的数据增强方案。中心度反应了图中节点的重要性，通过中心度来定义数据增强方案可以保留下图中重要的拓扑和特征信息，以及去除图中的冗余信息。一方面，以上方法需要精心设计数据增强池，以确保其

中包含有效的数据增强策略。另一方面，以上方法根据模型在不同数据集和下游任务中的表现，通过试错法选择最佳的数据增强方案是必要的。这一过程涉及大量的实验和计算成本，并且实际上在这个过程中，下游任务的标签信息被违规使用来参与无监督图对比学习的决策。因此，在图对比学习中实现自动化生成数据增强方案具有重要意义，可以有效地提高模型的性能并降低实验成本。

互信息最大化原则 (Mutual Information Maximization, InfoMax)^[60]是图对比学习中最常用的原则之一^[40,62-63]，其指的是最大化视图表示之间的互信息来学习编码器，即 $\max_f I(f(V_1); f(V_2))$ ，其中 V_1 、 V_2 表示视图， f 表示编码器， $I(a; b)$ 表示 a 和 b 之间的互信息。通过 InfoMax，对比模型可以辨别来自同一节点或图的成对样本，从而学习图的基本拓扑结构和特征信息。



图 4-2 用于 OCR 识别的图片

然而，如果不加以区分地应用 InfoMax 原则来最大化视图之间的互信息，会导致模型学习到与下游任务无关的冗余信息。这种冗余信息会导致模型学习出脆弱的嵌入表示，并可能严重降低模型在下游任务上的性能^[117]。举例来说，在训练光学字符识别模型 (Optical Character Recognition, OCR) 时，颜色信息是冗余且有害的，应该在训练之前就被移除。如图4-2所示，OCR 模型的任务是准确识别具体的数字和字母。在图中，存在多个数字“9”和“2”。研究者们期望 OCR 模型能够准确地将它们识别为“9”和“2”。然而，如果模型学习到了颜色信息，可能会错误地将不同颜色的“9”和“2”识别为不同的数字。这显然是有害的，因为数字本身应该是不受颜色影响的。因此，在实施 InfoMax 进行视图对比之前，应该剥离视图之间共享的冗余信息。现在新的问题是，如何定义视图中的冗余信息。

计算机视觉领域中的互信息最小化原则 (Mutual Information Minimization, InfoMin)^[68]指出，对比学习中的最优视图集合应该共享下游任务所需的最少信

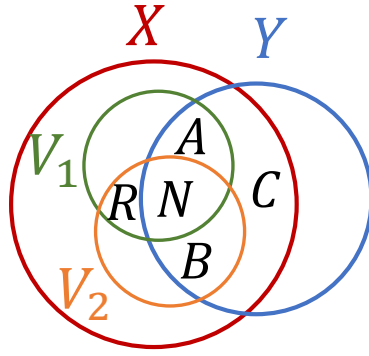


图 4-3 InfoMin 原则定义出的最优视图阐释

息，其余信息都为冗余信息。InfoMin 原则可以形式化地表示为：

$$\begin{aligned} (V_1^*, V_2^*) &= \operatorname{argmin}_{V_1, V_2} I(V_1; V_2), \\ \text{s. t. } I(V_1; Y) &= I(V_2; Y) = I(X; Y). \end{aligned} \quad (4-1)$$

其中 V_1^*, V_2^* 代表最优视图， X 代表输入信息， Y 代表下游任务信息。图4-3展示了在信息熵背景下，输入信息 X 、视图 V_1, V_2 以及下游任务信息 Y 之间的关系的文氏图 (venn's diagram)。A, B, C, N, R 为重合区域。视图 V_1, V_2 为输入信息 X 的函数，因此被 X 所包含。当 V_1, V_2 为最优视角时，A, B, C, R 将均为空。N 同时等同于 $I(X; Y)$, $I(V_1; Y)$ 以及 $I(V_2; Y)$ 。R 为冗余信息，在视图最优时为空。InfoMin 原则的另外一种理解是最优视图之间共享的信息只与下游任务相关，且包含了原始信息内的所有下游任务可用信息，即 $I(V_1^*; V_2^* | Y) = 0$, $I(X; Y) = I(V_1^*; V_2^*)$ 。本文将这样的共享信息称为最小必要信息。

实际上，采用手动数据增强的图对比学习方法本质上是在预先定义的数据增强池中寻找接近最佳视角的增强方法，以最好地符合 InfoMin 原则。然而，由于预定义的数据增强池中的数据增强方案数量有限，而且无法根据数据和下游任务的特性自适应地进行调整，因此所得到的视角无法达到最优且缺乏普适性。其次，许多自动化数据增强的图对比学习方法未能完全遵循 InfoMin 原则。AutoGCL^[59], AD-GCL^[58]和 JOAO^[57]等方法采用了一种最小化-最大化的优化框架，目标是寻求相似度最小的视图或者最具挑战性的视图。然而，通过最小化视图之间的相似度或者最小化视图表示之间的一致性得到的视角远非最佳。这些方法减少了冗余信息，但并未强调保留必要信息，可能导致信息的过度去除，进

而使模型面临着信息不足的风险。应用 InfoMin 获取最佳视图成为问题的关键：去除冗余信息，同时保留与下游任务相关的必要信息。

然而，在无监督环境中，研究者们无法获取下游任务的信息，也因此无法准确测量与任务相关的信息量，从而难以确定 InfoMin 所需的最小必要信息。举例来说，像 InfoGCL^[67]和 LP-A3^[118]等方法都依赖于下游任务标签 Y 来指导最优视图的获取。目前，在无监督条件下获得 InfoMin 中最佳视图的研究仍然相对有限。尽管如此，本文发现针对不同的下游任务，最小显著信息可作为最小必要信息的近似，从而避免了使用下游任务相关的信息。本章将这样的信息称为最小显著信息。

本章提出了一种名为 **Minimal Noteworthy Information for unsupervised Graph contrastive learning (GMNI)** 的全新方法。GMNI 实现了无监督图对比学习中 InfoMin 和 InfoMax 原则的结合。GMNI 自动化了数据增强，该数据增强同时应用于拓扑和特征，通过数据增强获得的视图在无需下游任务信息的情况下近似了 InfoMin 原则中的最优视角。将 InfoMax 应用于该视图集合就可以避免上述冗余信息和信息不足的风险。

4.2 模型

GMNI 包含两个关键模块，分别是自动视图生成器和视图对比模块。自动视图生成器生成一个用于生成两个视图的重要性图，生成的两个视图将被送入视图对比模块。

4.2.1 符号定义

图 \mathcal{G} 包括 N 个节点的节点集合 $\mathcal{V} = \{v_1, \dots, v_N\}$ 和有 M 条边的边集合 $\mathcal{E} = \{\varepsilon_1, \dots, \varepsilon_M\}$ 。图 \mathcal{G} 的特征矩阵、邻接矩阵和度矩阵分别表示为 $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_N\} \in \mathbb{R}^{N \times F}$ 、 $\mathbf{A} \in \{0, 1\}^{N \times N} : \mathbf{A}_{ij} = \mathbb{1}((v_i, v_j) \in \mathcal{E})$ 和 $\mathbf{D} : \mathbf{D}_{ii} = \sum_j \mathbf{A}_{ij}$ 。其中 F 是特征维度， $\mathbf{x}_i \in \mathbb{R}^F$ 是节点 v_i 的特征。使用 $\mathcal{G} : (\mathbf{A}, \mathbf{X})$ 来表示这个图。

GMNI 在未使用任务相关信息 Y 的情况下实现了对 InfoMin 原则的近似。为了验证这种近似策略在不同下游任务中的有效性，本章讨论了两种下游任务，即节点分类任务和图分类任务。对于节点分类任务，其输入是单个图 \mathcal{G} ，旨在学习

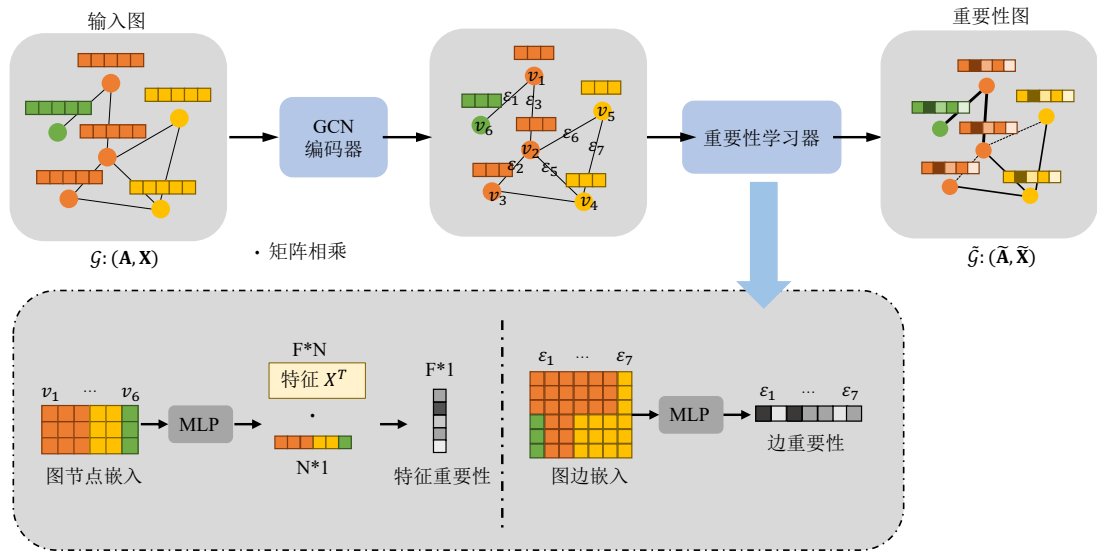


图 4-4 自动视图生成器生成重要性图的前向过程

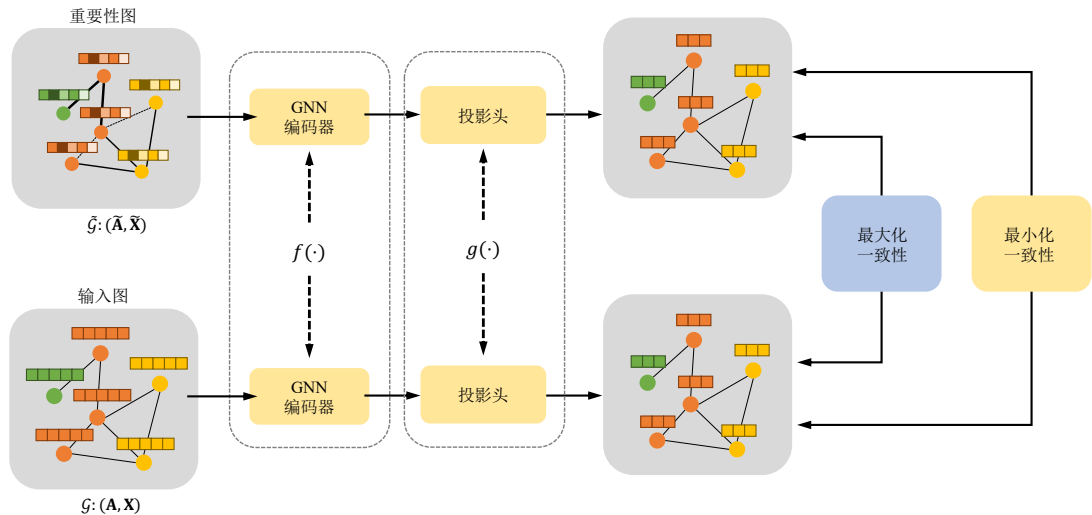


图 4-5 自动视图生成器生成重要性图的优化过程

一个编码器 $f_n: \mathcal{G} \rightarrow \mathbb{R}^{N \times d}$ ，以获得低维节点嵌入 $f_n(\mathcal{G})$, $d \ll F$ 。对于图分类任务，其输入是一组图 $\{\mathcal{G}_i\}_{i=1}^Q$ ，旨在学习一个编码器 $f_g: \mathcal{G}_i \rightarrow \mathbb{R}^d$ ，以获得低维图嵌入 $f_g(\mathcal{G}_i)$, $i = 1, \dots, Q$ 。

4.2.2 自动视图生成器

自动视图生成器的结构如图4-4和图4-5所示。在重要性图中，颜色较深的特征和较粗的连边更为重要。原始图被输入到 GCN 编码器和重要性学习器中，以生成一个重要性图。图4-5展示了在使用共享的 GNN 编码器和投影头获得了重要性图和原始图的 Embedding 之后，进行最大化-最小化优化。“最大化一致性”

部分优化图4-4中的蓝色 GCN 编码器和重要性学习器，而“最小化一致性”部分则优化图4-5中的黄色 GNN 编码器和投影头。

InfoMin 原则指出了对比学习中的最佳视图需要满足的条件，本文的自动视图生成器旨在无监督条件下近似这些最佳视图。在无监督条件下应用 InfoMin 原则是具有挑战性的，因为无法量化与任务相关的信息。因此，本章提出获取在各种下游任务中共享最小显著信息的视图集合。这种方法有效地与 InfoMin 原则相对应，而无需使用任务相关数据。

定义 1 (重要性图 / Importance Graph) 重要性图 $\tilde{\mathcal{G}} = T(\mathcal{G}) = (\tilde{\mathbf{A}}, \tilde{\mathbf{X}})$ 。通过给原始图的边和特征引入重要性值来获得重要性图。这里， $T(\cdot)$ 表示将原始图转换为重要性图的函数。边的重要性和特征的重要性分别用 $\mathbf{P}_e \in \mathbb{R}^{M \times 1}$ 和 $\mathbf{P}_f \in \mathbb{R}^{F \times 1}$ 表示。

定义 2 (最小显著信息 / Minimal Noteworthy Information) 图信息包括拓扑结构信息和特征信息。因此，最小显著信息由图中的关键边和特征组成。对于图 \mathcal{G} 及其对应的重要性图 $\tilde{\mathcal{G}}$ ，最小显著信息指的是从 \mathcal{G} 中选取出的边和特征的最小子集，它们在 $\tilde{\mathcal{G}}$ 上显示出较高的重要性。

重要性图基本公式

重要性图通过最大化-最小化优化得到：

$$\begin{aligned} \max_T \min_{f,g} I(g(f(\mathcal{G})); g(f(T(\mathcal{G})))) - |T(\mathcal{G})|, \\ \text{s.t. } I(g(f(\mathcal{G})); g(f(T(\mathcal{G})))) \geq \zeta, \end{aligned} \quad (4-2)$$

其中， $f(\cdot)$ 是用于获得节点嵌入的图编码器 (Graph Encoder)， $g(\cdot)$ 是用于提高模型表达能力的投影头 (Projection Head)。需要注意的是， $f(\cdot)$ 和 $g(\cdot)$ 不是视图对比模块的编码器。正则化项 $|T(\mathcal{G})|$ 是边重要性和特征重要性的归一化和。通过最小化 $|T(\mathcal{G})|$ ，只有显著的边和特征才会具有较高的重要性值。引入约束条件 $I(g(f(\mathcal{G})); g(f(T(\mathcal{G})))) \geq \zeta$ 是为了防止 $g(f(\cdot))$ 退化为边缘情况，即退化为 $I(g(f(\mathcal{G})); g(f(T(\mathcal{G})))) = 0$ 。一旦优化达到互信息下限 ζ ，模型将停止对 $f(\cdot)$ 和 $g(\cdot)$ 的优化。然而，寻找到 $g(f(\cdot))$ 的退化情况对于优化器来说是很困难的。实验表明， $g(f(\cdot))$ 的退化并没有发生。

结论 1 公式 (4-2) 等同于同时减少 $I(\mathcal{G}; T(\mathcal{G})|\mathcal{J})$ 并增加 $I(T(\mathcal{G}); \mathcal{J})$, 其中 \mathcal{J} 是显著信息。

$\max_T I(g(f(\mathcal{G})); g(f(T(\mathcal{G}))))$ 的最优解是 $T(\mathcal{G}) = \mathcal{G}$ 。然而, 由于 $-|T(\mathcal{G})|$ 充当正则化项, 它有效地约束了显著边和显著特征的数量。因此, 最小显著信息包括两个组成部分: $-\max_T |T(\mathcal{G})| = \min_T |T(\mathcal{G})|$ 的最小信息和 $\max_T I(g(f(\mathcal{G})); g(f(T(\mathcal{G}))))$ 的显著信息。为了最大化重要性图 $T(\mathcal{G})$ 与原始图 \mathcal{G} 之间的互信息, 对于 $T(\mathcal{G})$ 来说, 在其有限的信息量内选择性地保留最关键的信息变得至关重要。换句话说, 优化公式 (4-2) 可提高 $T(\mathcal{G})$ 中显著信息的比例, 并减少非显著信息的比例。 $\min_{f,g} I(g(f(\mathcal{G})); g(f(T(\mathcal{G}))))$ 的目的是获得一个具有挑战性的信息编码器 $g(f(\cdot))$ 。预期在具有挑战性的信息编码器 $g(f(\cdot))$ 下获得的最小显著信息将表现出更强的鲁棒性。

当获得了重要性图 $\tilde{\mathcal{G}} = T(\mathcal{G})$ 后, 就可以确定出最小显著信息。图上的最小显著信息是 \mathcal{G} 上的在 $\tilde{\mathcal{G}}$ 中展现出高重要性的最小子集。 $T(\mathcal{G})$ 上的这些关键边和关键特征可能与不同的任务高度相关。例如, 在图分割 (Graph Partition)^[119-120]、图分类 (Graph Classification)^[10,121] 和链路预测 (Link Prediction)^[122-123] 等任务中, 连接不同簇的关键边在图中起着至关重要的作用, 这些边是提供了图内社群关系和依赖性的基本连接。

重要性图构建

(1) 特征重要性计算。

为了获取特征的重要性, 首先需要获得节点的重要性。首先, 使用图卷积层^[56] 计算节点表示:

$$\mathbf{E} = \sigma(\mathbf{A}\mathbf{X}\mathbf{W}), \quad (4-3)$$

其中, σ 是激活函数, \mathbf{W} 是参数矩阵。考虑到节点的度对于重要性计算也是必要的, \mathbf{A} 是未进行度归一化的邻接矩阵。根据节点表示 \mathbf{E} 计算节点重要性

$\mathbf{P}_n \in \mathbb{R}^{N \times 1}$:

$$\begin{aligned} \mathbf{P}_n &= \text{Gumbel}(h_\phi(\mathbf{E})), \\ \text{Gumbel}(t) &= \text{Sigmoid}\left(\frac{\log \eta - \log(1 - \eta) + t}{\tau}\right), \\ \eta &\sim \text{Uniform}(0, 1). \end{aligned} \quad (4-4)$$

其中, $h_\phi(\cdot)$ 是一个简单的多层感知器 (MLP), $\text{Gumbel}(t)$ 是 Gumbel-Max 重参数化函数 (Gumbel-Max Reparametrization Function^[124])。 τ 是温度参数, τ 越接近于 0, \mathbf{P}_n 越接近于二值化。 Gumbel-Max 重参数化函数确保了上述过程可以反向传播, 并使 \mathbf{P}_n 具有概率意义。

现在, 特征重要性 $\mathbf{P}_f \in \mathbb{R}^{F \times 1}$ 可以通过节点重要性 $\mathbf{P}_n \in \mathbb{R}^{N \times 1}$ 计算得到:

$$\mathbf{P}_f = \mathbf{X}^T \mathbf{P}_n, \quad (4-5)$$

其中, $\mathbf{X}^T \in \mathbb{R}^{F \times N}$ 是特征矩阵 \mathbf{X} 的转置。 如果特征取值是介于 0 和 1 之间的实数, 则对 \mathbf{P}_f 而不是 \mathbf{P}_n 执行 Gumbel-Max 重参数化技巧, 以避免 \mathbf{P}_f 变得过小。

(2) 边重要性计算。

受 AD-GCL^[58] 的启发, 边 $\varepsilon_k = (v_i, v_j) \in \mathcal{E}$ 的重要性 $\mathbf{p}_{e,k}$ 可以用以下方式计算:

$$\mathbf{p}_{e,k} = \text{Gumbel}(h_\psi([\mathbf{E}[v_i]; \mathbf{E}[v_j]])), \quad (4-6)$$

其中 $[\cdot; \cdot]$ 是张量拼接操作, $h_\psi(\cdot)$ 是一个简单的 MLP, $\mathbf{E}[v_i]$ 是节点 v_i 的嵌入。 最终可以获得所有的边重要性 $\mathbf{P}_e = [\mathbf{p}_{e,1}, \dots, \mathbf{p}_{e,M}]^T \in \mathbb{R}^{M \times 1}$ 。

重要性图定义为 $\tilde{\mathcal{G}} = (\tilde{\mathbf{A}}, \tilde{\mathbf{X}})$ 。 将 $\mathbf{P}_f \in \mathbb{R}^{F \times 1}$ 广播到 $\mathbf{P}'_f \in \mathbb{R}^{N \times F}$ 。 $\tilde{\mathbf{X}}$ 由以下公式给出:

$$\tilde{\mathbf{X}} = \mathbf{X} \odot \mathbf{P}'_f. \quad (4-7)$$

替换邻接矩阵 \mathbf{A} 中的边值为 \mathbf{P}_e 中的边值来获得 $\tilde{\mathbf{A}}$ 。

接下来将使用共享的图编码器和投影头来获得重要性图和原始图的嵌入表

示 (Embedding)。节点 v_i 在 \mathcal{G} 中的嵌入表示为：

$$\mathbf{z}_{i,1} = g_\xi(f_\theta(\mathcal{G}(v_i))), \quad (4-8)$$

在 $\tilde{\mathcal{G}}$ 的嵌入表示为：

$$\mathbf{z}_{i,2} = g_\xi(f_\theta(\tilde{\mathcal{G}}(v_i))). \quad (4-9)$$

为了简化模型， f_θ 是一个 1 层或 2 层的图编码器， g_ξ 是一个 1 层或 2 层的 MLP 作为投影头。

InfoNCE^[125] 作为互信息的一个下界，被广泛应用于估计互信息。 $\mathbf{z}_{i,1}$ 和 $\mathbf{z}_{i,2}$ 之间的互信息可以被估计为：

$$\begin{aligned} \hat{I}(\mathbf{z}_{i,1}; \mathbf{z}_{i,2}) &= \frac{1}{2} \left(\hat{I}_0(\mathbf{z}_{i,1}; \mathbf{z}_{i,2}) + \hat{I}_0(\mathbf{z}_{i,2}; \mathbf{z}_{i,1}) \right), \\ \hat{I}_0(\mathbf{z}_{i,1}; \mathbf{z}_{i,2}) &= \log \frac{\exp(s(\mathbf{z}_{i,1}, \mathbf{z}_{i,2})/\epsilon)}{\sum_{j=1, j \neq i}^N \exp(s(\mathbf{z}_{i,1}, \mathbf{z}_{j,2})/\epsilon)}, \end{aligned} \quad (4-10)$$

其中 $s(\cdot, \cdot)$ 是余弦相似度， ϵ 是温度系数， $\hat{I}(\mathbf{z}_{i,1}; \mathbf{z}_{i,2})$ 的对称设计考虑了 \mathcal{G} 和 $\tilde{\mathcal{G}}$ 的对等地位。

\mathcal{G} 和 $\tilde{\mathcal{G}} = T(\mathcal{G})$ 的嵌入表示之间的互信息被估计为：

$$\begin{aligned} I(g_\xi(f_\theta(\mathcal{G})); g_\xi(f_\theta(T(\mathcal{G})))) &\rightarrow \hat{I}(g_\xi(f_\theta(\mathcal{G})); g_\xi(f_\theta(T(\mathcal{G})))) \\ &= \frac{1}{N} \sum_{i=1}^N \hat{I}(\mathbf{z}_{i,1}; \mathbf{z}_{i,2}). \end{aligned} \quad (4-11)$$

正则化项 $|T(\mathcal{G})|$ 被定义为：

$$|T(\mathcal{G})| = \lambda \left(\frac{\sum_{i=1}^F \mathbf{P}_f^{(i)}}{F} + \frac{\sum_{i=1}^M \mathbf{P}_e^{(i)}}{M} \right), \quad (4-12)$$

其中， λ 为正则化权重系数， $\mathbf{P}_f^{(i)}$ 是 \mathbf{P}_f 的第 i 维。

最终，重要性图的优化策略为：

$$\begin{aligned} \max_{\mathbf{W}, \phi, \psi} \min_{\theta, \xi} & \hat{I}(g_{\xi}(f_{\theta}(\mathcal{G})); g_{\xi}(f_{\theta}(T_{\mathbf{W}, \phi, \psi}(\mathcal{G})))) - \lambda \left(\frac{\sum_{i=1}^F \mathbf{P}_f^{(i)}}{F} + \frac{\sum_{i=1}^M \mathbf{P}_e^{(i)}}{M} \right), \\ \text{s.t.} & \hat{I}(g_{\xi}(f_{\theta}(\mathcal{G})); g_{\xi}(f_{\theta}(T(\mathcal{G})))) \geq \zeta. \end{aligned} \quad (4-13)$$

视图生成

上一节中计算得到了特征重要性 \mathbf{P}_f 和边重要性 \mathbf{P}_e 。对比学习，旨在通过比较视图之间的相似性和差异性来学习出有意义的表示。实际上，可以直接使用特征重要性和边重要性作为 Soft Features 和 Soft Edges，并以这些 Soft Features 和 Soft Edges 来生成视图集合。但这样做会导致视图集合具有一致的拓扑和特征，从而限制视图的多样性。

因此，在生成视图的时候本文引入了基于重要性的随机性，以增强视图的多样性并使得模型对扰动更加稳定。需要生成两个视图， $\mathcal{G}_1 : (\mathbf{A}_1, \mathbf{X}_1)$ 和 $\mathcal{G}_2 : (\mathbf{A}_2, \mathbf{X}_2)$ ，视图修改了原始图的拓扑和特征。以生成 \mathcal{G}_1 为例，生成 \mathcal{G}_2 的方法类似。

\mathcal{G}_1 的边集 \mathcal{E}_1 是 \mathcal{E} 的子集。使用一个随机变量 $p_{\varepsilon_k} \sim \text{Bernoulli}(1 - p_{d, \varepsilon_k})$ 来选择边 ε_k ，即如果 $p_{\varepsilon_k} = 1$ ，则 $\varepsilon_k \in \mathcal{E}_1$ ，否则 $\varepsilon_k \notin \mathcal{E}_1$ 。至此， \mathbf{A}_1 就可以通过 \mathcal{E}_1 推导出来。受 GCA^[106] 的启发， $\mathbf{P}_{d, \varepsilon} = [p_{d, \varepsilon_1}, \dots, p_{d, \varepsilon_M}]$ 被定义如下：

$$\mathbf{P}_{d, \varepsilon} = \min \left(\frac{\mathbf{P}_e^{\max} - \mathbf{P}_e}{\mathbf{P}_e^{\max} - \mathbf{P}_e^{\text{avg}}} \cdot p_{s1}, p_t \right), \quad (4-14)$$

其中， $\min(a, b)$ 表示选择 a 和 b 中较小的一个， p_{s1} 和 p_t 是介于 0 和 1 之间的超参数。 p_{s1} 用于调整边的整体丢弃率， p_t 用于截断丢弃率，以使得每条边都有被包含在视图中的机会。 \mathbf{P}_e^{\max} 和 $\mathbf{P}_e^{\text{avg}}$ 分别是 \mathbf{P}_e 中的最大值和平均值。通过以上的定义，具有更高重要性的边 \mathbf{P}_e 具有更小的丢弃率 $\mathbf{P}_{d, \varepsilon}$ ，所以它在 \mathcal{E}_1 中被包含的机会也更大。

\mathcal{G}_1 的特征矩阵可以表示为 $\mathbf{X}_1 = \mathbf{X} \odot \mathbf{M}$ ，其中 $\mathbf{M} \in \mathbb{R}^{N \times F}$ 是从 $\mathbf{m} \in \{0, 1\}^F$ 广播而来的。使用随机变量 $p_{f_k} \sim \text{Bernoulli}(1 - p_{d, f_k})$ 来选择原图中第 k 个维度的特征，即如果 $p_{f_k} = 1$ ，则 $\mathbf{m}_k = 1$ ，否则 $\mathbf{m}_k = 0$ 。这里 \mathbf{m}_k 表示 \mathbf{m} 的第 k 个

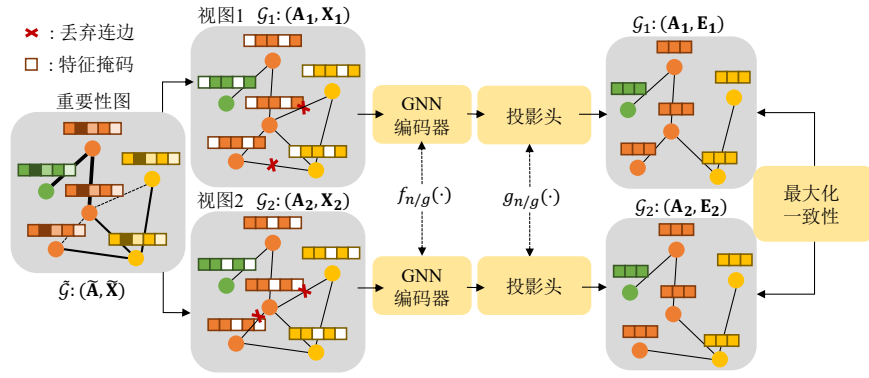


图 4-6 视图对比模块

维度。 $\mathbf{P}_{d,f} = [p_{d,f_1}, \dots, p_{d,f_F}]$ 被定义为:

$$\mathbf{P}_{d,f} = \min \left(\frac{\mathbf{P}_f^{\max} - \mathbf{P}_f}{\mathbf{P}_f^{\max} - \mathbf{P}_f^{\text{avg}}} \cdot p_{s2}, p_t \right), \quad (4-15)$$

其中, p_{s2} 是一个介于 0 和 1 之间的超参数, 用于控制特征的整体掩码率。

4.2.3 视图对比模块

上节生成的视图避免了冗余信息的干扰或重要信息的丢失风险。为了验证通过最小显著信息获得的近似最优视图的有效性, 本文采用了最简单的视图比较模块, 即基于 InfoMax 原则直接拉近视图表示来学习图编码器。图4-6展示了视图对比模块的基本架构, 该模块根据重要性图给出的节点特征重要性和边重要性, 通过特征掩码 (Mask Feature) 和丢弃连边 (Drop Edge) 来生成两个视图。之后通过共享的 GNN 编码器和投影头计算视图的 Embedding。如果下游任务是图分类, 网络需要进一步增加读出函数 (Readout Function) 来获取图表示。

节点分类任务旨在学习一个编码器 $f_n: \mathcal{G} \rightarrow \mathbb{R}^{N \times d}$ 。 \mathcal{G}_1 和 \mathcal{G}_2 是 \mathcal{G} 的视图, 视图对比的目标是:

$$\max_{f_n, g_n} I(g_n(f_n(\mathcal{G}_1)); g_n(f_n(\mathcal{G}_2))), \quad (4-16)$$

这里 $f_n(\cdot)$ 是一个 GCN^[56], $g_n(\cdot)$ 是一个简单的 1 或 2 层 MLP, 使用 InfoNCE (公式 (4-10)) 作为互信息的估计器。

图分类任务旨在学习一个编码器 $f_g: \mathcal{G}_i \rightarrow \mathbb{R}^d$ 。当给定一批 (batch) 图

$\{\mathcal{G}_i\}_{i=1}^K$ 时, $\mathcal{G}_{i,1}$ 和 $\mathcal{G}_{i,2}$ 是 \mathcal{G}_i 的视图, 视图对比的目标是:

$$\max_{f_g, g_g} \frac{1}{K} \sum_{i=1}^K I(g_g(f_g(\mathcal{G}_{i,1})); g_g(f_g(\mathcal{G}_{i,2}))), \quad (4-17)$$

其中, $f_g(\cdot)$ 是一个 GIN^[13], $g_g(\cdot)$ 是一个简单的 1 或 2 层 MLP, 后面跟随着一个读出函数 (Readout Function)。这里的读出函数是一个简单的求和。 $f_n(\cdot)$ 和 $f_g(\cdot)$ 可以替换为任何图编码器。最后, 节点嵌入 $f_n(\mathcal{G})$ 和图嵌入 $f_g(\mathcal{G}_i), i = 1, \dots, Q$ 被用于下游任务测试。

4.2.4 算法与时间复杂度

算法 2 是 GMNI 的自动视图生成器的算法伪代码。算法 3 和算法 4 分别是 GMNI 在节点分类任务和图分类任务上的视图对比模块的算法伪代码。

GMNI 的自动视图生成器的 FLOPs (Floating-Point Operations) 为 $O(N^2F + (M + N)F^2)$, 而视图对比模块的 FLOPs 为 $O(N^2F + NF^2)$ 。AD-GCL、JOAOv2 和 AutoGCL 的 FLOPs 分别为 $O(N^2F + (M + N)F^2)$ 、 $O(N^2F + NF^2)$ 和 $O(N^2F + NF^2)$ 。GMNI 的自动视图生成器具有与 AD-GCL 相当的渐近复杂度, 而其视图对比模块具有与 JOAO 和 AutoGCL 相当的渐近复杂度。首先, GMNI 的自动视图生成器的训练可以在很少的 epochs 内完成。其次, GMNI 的视图对比模块的训练仅包含了最大化的优化, 而 AD-GCL、JOAOv2 和 AutoGCL 的优化则涉及了同时最大化和最小化。因此 GMNI 不会带来过高的计算负载和过长的计算时间。

4.3 实验与分析

为了验证 GMNI 的有效性, 本节在半监督和无监督条件下进行了 5 个节点分类数据集和 9 个图分类数据集上的对比与消融实验。

4.3.1 数据集

本节使用了 5 个用于节点分类任务的数据集, 包括 Wiki-CS^[126]、Amazon-Computers^[127]、Amazon-Photo^[127]、Coauthor-CS^[127] 和 Coauthor-Physics^[127]。它们的详细统计数据如表 4-1 所示。Wiki-CS 具有密集的实数特征, 而其他数据集则

算法 2 自动视图生成器

输入: 图 $\mathcal{G} = (\mathbf{A}, \mathbf{X})$; 边集 \mathcal{E} , $|\mathcal{E}| = M$; 特征维度 F ; GCN 编码器 $f_{\mathbf{W}}(\cdot)$; 特征重要性 MLP $h_{\phi}(\cdot)$; 边重要性 MLP $h_{\psi}(\cdot)$; Gumbel-Max 函数 $\text{Gumbel}(\cdot)$; GNN 编码器 $f_{\theta}(\cdot)$; 投影头 $g_{\xi}(\cdot)$; 互信息估计器 $\hat{I}(\cdot; \cdot)$.

超参数: 正则化权重系数 λ ; view 1 和 view 2 的缩放参数 $p_{1,s1}, p_{2,s1}, p_{1,s2}, p_{2,s2}$; 截断参数 p_t ; 学习率 α .

输出: view 1 $\mathcal{G}_1 = (\mathbf{A}_1, \mathbf{X}_1)$; view 2 $\mathcal{G}_2 = (\mathbf{A}_2, \mathbf{X}_2)$.

- 1: **for** ep in #epochs **do**
- 2: $\mathbf{E} = f_{\mathbf{W}}(\mathbf{A}, \mathbf{X})$;
- 3: $\mathbf{P}_n = \text{Gumbel}(h_{\phi}(\mathbf{E}))$;
- 4: $\mathbf{P}_f = \mathbf{X}^T \mathbf{P}_n$;
- 5: **for** $\forall \varepsilon_k \in \mathcal{E}$ **do**
- 6: $\mathbf{p}_{e,k} = \text{Gumbel}(h_{\psi}([\mathbf{E}[v_i]; \mathbf{E}[v_j]])), \varepsilon_k = (v_i, v_j)$;
- 7: **end for**
- 8: $\mathbf{P}_e = [\mathbf{p}_{e,1}, \dots, \mathbf{p}_{e,M}]^T$;
- 9: 广播 $\mathbf{P}_f \in \mathbb{R}^{F \times 1}$ 到 $\mathbf{P}'_f \in \mathbb{R}^{N \times F}$;
- 10: $\tilde{\mathbf{X}} = \mathbf{X} \odot \mathbf{P}'_f$;
- 11: 通过 \mathbf{P}_e 导出 $\tilde{\mathbf{A}}$;
- 12: $\tilde{\mathcal{G}} = (\tilde{\mathbf{A}}, \tilde{\mathbf{X}})$;
- 13: $\mathcal{L} = -\hat{I}(g_{\xi}(f_{\theta}(\mathcal{G})); g_{\xi}(f_{\theta}(\tilde{\mathcal{G}})))$;
- 14: $\mathcal{R} = \frac{\sum_{i=1}^F \mathbf{P}'_f^{(i)}}{F} + \frac{\sum_{i=1}^M \mathbf{P}_e^{(i)}}{M}$;
- 14: /* 更新重要性学习器 */
- 15: $\mathbf{W} \leftarrow \mathbf{W} - \alpha \nabla_{\mathbf{W}}(\mathcal{L} + \lambda * \mathcal{R})$;
- 16: $\phi \leftarrow \phi - \alpha \nabla_{\phi}(\mathcal{L} + \lambda * \mathcal{R})$;
- 17: $\psi \leftarrow \psi - \alpha \nabla_{\psi}(\mathcal{L} + \lambda * \mathcal{R})$;
- 17: /* 更新 GNN 编码器和投影头 */
- 18: $\theta \leftarrow \theta + \alpha \nabla_{\theta}(\mathcal{L})$;
- 19: $\xi \leftarrow \xi + \alpha \nabla_{\xi}(\mathcal{L})$;
- 20: **end for**
- 21:
$$\mathbf{P}_{d,\varepsilon} = [p_{d,\varepsilon_1}, \dots, p_{d,\varepsilon_M}] = \min \left(\frac{\mathbf{P}_e^{\max} - \mathbf{P}_e}{\mathbf{P}_e^{\max} - \mathbf{P}_e^{\text{avg}}} \cdot p_{1,s1}, p_t \right)$$
;
- 22: 从 $\text{Bernoulli}(1 - p_{d,\varepsilon_k}), k = 1, \dots, M$ 采样 p_{ε_k} ;
- 23: 使用 $p_{\varepsilon_k}, k = 1, \dots, M$ 生成 \mathbf{A}_1 ;
- 24:
$$\mathbf{P}_{d,f} = [p_{d,f_1}, \dots, p_{d,f_F}] = \min \left(\frac{\mathbf{P}'_f^{\max} - \mathbf{P}'_f}{\mathbf{P}'_f^{\max} - \mathbf{P}'_f^{\text{avg}}} \cdot p_{1,s2}, p_t \right)$$
;
- 25: 从 $\text{Bernoulli}(1 - p_{d,f_k}), k = 1, \dots, F$ 采样 p_{f_k} ;
- 26: 使用 $p_{f_k}, k = 1, \dots, F$ 生成 \mathbf{X}_1 ;
- 27: 类似地生成 $\mathbf{A}_2, \mathbf{X}_2$;
- 28: **return** $\mathcal{G}_1 = (\mathbf{A}_1, \mathbf{X}_1)$; $\mathcal{G}_2 = (\mathbf{A}_2, \mathbf{X}_2)$;

算法 3 节点分类任务的视图对比模块

输入: 图 \mathcal{G} ; GNN 编码器 $f_\omega(\cdot)$; 投影头 $g_\rho(\cdot)$; 互信息估计器 $\hat{I}(\cdot; \cdot)$.

超参数: 学习率 β .

输出: Trained $f_\omega(\cdot)$.

- 1: $(\mathcal{G}_1, \mathcal{G}_2) = \text{算法 2}(\mathcal{G})$;
- 2: **for** ep in #epochs **do**
- 3: $\mathcal{L} = -\hat{I}(g_\rho(f_\omega(\mathcal{G}_1)); g_\rho(f_\omega(\mathcal{G}_2)))$;
- 4: $\omega \leftarrow \omega - \beta \nabla_\omega \mathcal{L}$; $\rho \leftarrow \rho - \beta \nabla_\rho \mathcal{L}$;
- 5: **end for**
- 6: **return** Trained $f_\omega(\cdot)$;

算法 4 图分类任务的视图对比模块

输入: 图集 $\{\mathcal{G}_i\}, i = 1, \dots, Q$; GNN 编码器 $f_\omega(\cdot)$; 投影头 $g_\rho(\cdot)$; 互信息估计器 $\hat{I}(\cdot; \cdot)$.

超参数: 学习率 β ; batch 大小 B ; batches 数目 BN .

输出: Trained $f_\omega(\cdot)$.

- 1: **for** $i = 1$ to BN **do**
- 2: **for** \mathcal{G}_j in sampled minibatch $\{\mathcal{G}_j\}_{j=1}^B$ **do**
- 3: $(\mathcal{G}_{i,(j,1)}, \mathcal{G}_{i,(j,2)}) = \text{算法 2}(\mathcal{G}_{i,j})$;
- 4: **end for**
- 5: **end for**
- 6: **for** ep in #epochs **do**
- 7: **for** $i = 1$ to BN **do**
- 8: $\mathcal{L} = -\frac{1}{B} \sum_{j=1}^B \hat{I}(g_\rho(f_\omega(\mathcal{G}_{i,(j,1)})); g_\rho(f_\omega(\mathcal{G}_{i,(j,2)})))$;
- 9: $\omega \leftarrow \omega - \beta \nabla_\omega \mathcal{L}$; $\rho \leftarrow \rho - \beta \nabla_\rho \mathcal{L}$;
- 10: **end for**
- 11: **end for**
- 12: **return** Trained $f_\omega(\cdot)$;

具有稀疏的 one-hot 特征。按照 GCA^[106]的方法，实验在 Wiki-CS 数据集上使用公开的训练、验证和测试集进行模型评估。对于其他四个数据集，实验将数据集随机分割为三个部分：80% 用于训练，10% 用于验证，10% 用于测试。

实验使用了来自 TUDataset^[128]的 9 个图分类数据集，它们的详细统计数据如表4-2所示。实验按照 JOAO 的方法^[57]，采用 10 折交叉验证使用整个数据集进行图表示的学习，并将图表示输入到下游分类器中进行验证。

4.3.2 对比基线

针对节点分类任务，GMNI 将与以下图对比学习方法进行对比：

- (1) 不具有数据增强的方法：DGI^[40]，GMI^[129]；

表 4-1 节点分类数据集

Dataset	#Nodes	#Edges	#Features	#Classes
Wiki-CS	11,701	216,123	300	10
Amazon-Computers	13,752	245,861	767	10
Amazon-Photo	7,650	119,081	745	8
Coauthor-CS	18,333	81,894	6,805	15
Coauthor-Physics	34,493	247,962	8,415	5

表 4-2 图分类数据集

Dataset	Avg. #Graphs	Avg. #Nodes	Avg. #Edges	#Classes
MUTUG	188	17.93	19.79	2
PROTEINS	1,113	39.06	72.82	2
DD	1,178	284.32	715.66	2
NCI1	4,110	29.87	32.30	2
COLLAB	5,000	74.49	2457.78	3
GITHUB	12,725	113.79	234.64	2
IMDB-BINARY	1,000	19.77	96.53	2
REDDIT-BINARY	2,000	429.63	497.75	2
REDDIT-MULTI-5K	4,999	508.52	594.87	5

- (2) 具有手动数据增强的方法: MVGRL^[42], GCA^[106];
- (3) 具有自动数据增强的方法: AD-GCL^[58], JOAOv2^[57], AutoGCL^[59]。

针对图分类任务, GMNI 将与以下图对比学习方法进行对比:

- (1) 不具有数据增强的方法: InfoGraph^[96];
- (2) 具有手动数据增强的方法: GraphCL^[41]; MVGRL^[42]; GCA^[106];
- (3) 具有自动数据增强的方法: AD-GCL^[58]; JOAOv2^[57]; AutoGCL^[59]。

以下将对每个方法做一个简单介绍:

- InfoGraph^[96]。通过最大化图级表示和不同尺度子结构 (nodes, edges, triangles) 表示之间的互信息来学习图表示。通过这样的方式, 图表示可以编码跨不同尺度子结构之间共享的数据。
- DGI^[40]。通过最大化局部表示 (patch representations) 和相应的高级图表示 (high-level summaries of graphs) 之间的互信息来学习图表示。其中局部表示获得了中心节点周围的兴趣子图的信息, 因此 DGI 很适用于节点级别的学习任务。
- GMI^[129]。提出了一个新概念: 图互信息 (Graph Mutual Information, GMI), 旨在测量输入图和高级隐藏表示之间的相关性。GMI 将传统的互信息计算思想从向量空间推广到图域。文中提出了一种无监督模型, 通过最大化图

表 4-3 GMNI 的其他超参数

Dataset	view 1 p_{s1}	view 2 p_{s1}	view 1 p_{s2}	view 2 p_{s2}	VG lr	VC lr
Wiki-CS	0.2	0.4	0.1	0.1	0.001	0.01
Amazon-Computers	0.6	0.3	0.2	0.1	0.001	0.001
Amazon-Photo	0.3	0.5	0.1	0.1	0.001	0.001
Coauthor-CS	0.1	0.1	0.3	0.2	0.001	0.0001
Coauthor-Physics	0.4	0.4	0.1	0.4	0.01	0.001
MUTAG	0.3	0.2	0.3	0.2	0.001	0.001
PROTEINS	0.3	0.2	0.3	0.2	0.001	0.001
DD	0.3	0.2	0.3	0.2	0.001	0.001
NCI1	0.1	0.2	0.4	0.2	0.001	0.001
COLLAB	0.3	0.2	0.3	0.2	0.001	0.001
IMDB-B	0.3	0.2	0.3	0.2	0.001	0.001
REDDIT-B	0.3	0.2	0.3	0.2	0.001	0.001
REDDIT-M-5K	0.4	0.2	0.3	0.2	0.001	0.001

信息编码器的输入和输出之间的 GMI 来学习图表示。

- GraphCL^[41]。引入了一种无监督图对比学习框架，设计了四种图数据增强方法进行视图生成，并探究了不同数据增强方法的组合对不同数据集上的影响。该框架通过最大化视图表示之间的互信息来学习图表示。
- MVGRL^[42]。通过最大化从一阶邻居和图扩散（Graph Diffusion）获得的视图的互信息的来学习图表示。文中还验证了增加视图数量到三个以上以及对不同维度（Scale）的视图是无效的。
- GCA^[106]。声明了图对比学习中的数据增强方案应该保留图中的重要特征和拓扑信息。GCA 通过定义不同的中心度（Degree Centrality, Eigenvector Centrality, PageRank Centrality）来获得数据增强方案，并利用最大化视图之间和视图内部的互信息来学习图表示。
- AD-GCL^[58]。通过对抗优化来学习数据增强方案，避免了获得包含冗余信息的视图。其中最小化优化用于获得去除冗余信息的视图集合，最大化优化用于获得良好的图编码器。
- JOAOv2^[57]。提出了一个统一的双层优化框架，旨在以自动、适应性和动态的方式，在具体的图数据上应用 GraphCL 时选择出适当的数据增强方案。此外，JOAOv2 还引入了一个新的机制，即增强感知投影头（augmentation-aware projection head），通过这个机制，可以根据不同的增强方法使得输出的特征经过不同的投影头进行处理。
- AutoGCL^[59]。集成了一组可训练的视图生成器，用于自动数据增强，每个

生成器根据输入获取图的概率分布。此外，AutoGCL 通过修改联合训练策略，实现了端对端地训练可学习的视图生成器、图编码器和分类器，从而在生成对比样本时同时获得拓扑异质性和保持语义相似性。

4.3.3 实验设置

对于无监督学习任务，实验首先利用无标签数据训练模型生成节点嵌入或图嵌入，然后使用这些嵌入训练下游分类器。节点分类任务的下游分类器是学习率为 0.01 的 l_2 正则化逻辑回归模型。图分类任务的下游分类器是带有参数 C 的 SVM， C 在参数范围 [0.001, 0.01, 0.1, 1, 10, 100, 1000] 内进行网格搜索。在所有数据集上都采用了 Xavier 初始化^[130]和 Adam 优化器^[90]。其余超参数如表4-3所示，表中 VG 代表视图生成模块，VC 代表视图对比模块，lr 代表学习率。

在 GMNI 中，将节点分类的输入设计为两个视图和原始图的节点表示之和，即 $f_n(\mathcal{G}_1) + f_n(\mathcal{G}_2) + 2 * f_n(\mathcal{G})$ ，或者仅使用两个视图的节点表示之和，即 $f_n(\mathcal{G}_1) + f_n(\mathcal{G}_2)$ 。这样的设计考虑到融合视图表示可以产生更加通用的节点表示。至于图分类任务，GMNI 使用了原始图表示 $f_g(\mathcal{G}_i), i = 1, \dots, Q$ 作为输入。

AD-GCL、JOAOv2 和 AutoGCL 的原始论文未提供它们在节点分类任务中的结果，而 GCA 的原始论文也未提供其在图分类任务上的结果。因此，本节通过删除或添加池化层来获取它们的节点表示和图表示，以进行实验评估。对于 GCA，本节测试了其三种变体 GCA-DE、GCA-EVC 和 GCA-PR，并选取了最佳结果作为比较基线。

表 4-4 模型在无监督节点分类任务上的表现（重复运行了 5 次之后的平均准确率 Accuracy (%) \pm 标准差)

Type	Model	Wiki-CS	Amaz-Comp	Amaz-Photo	Coauthor-CS	Coauthor-Phy
w/o GDA	DGI	75.35 \pm 0.14	83.95 \pm 0.47	91.61 \pm 0.22	92.15 \pm 0.63	94.51 \pm 0.52
	GMI	74.85 \pm 0.08	82.21 \pm 0.31	90.68 \pm 0.17	OOM	OOM
w/ Manual GDA	MVGRL	77.52 \pm 0.08	87.52 \pm 0.11	91.74 \pm 0.07	92.11 \pm 0.12	95.33 \pm 0.03
	GCA	<u>78.35 \pm 0.05</u>	<u>87.85 \pm 0.31</u>	<u>92.53 \pm 0.16</u>	<u>93.10 \pm 0.01</u>	<u>95.73 \pm 0.03</u>
w/ Automated GDA	AD-GCL	73.46 \pm 0.36	81.32 \pm 0.93	88.75 \pm 0.92	92.16 \pm 0.36	94.57 \pm 0.09
	JOAO-v2	75.36 \pm 0.47	85.96 \pm 0.98	91.15 \pm 0.55	91.33 \pm 0.27	OOM
	AutoGCL	73.66 \pm 0.59	86.44 \pm 1.24	91.98 \pm 0.58	92.26 \pm 0.32	OOM
	GMNI	79.19 \pm 0.13	89.29 \pm 0.05	93.52 \pm 0.33	93.61 \pm 0.15	95.91 \pm 0.10

表 4-5 模型在无监督图分类任务上的表现 (重复运行了 5 次之后的平均准确率 Accuracy (%) \pm 标准差)

Type	Model	MUTAG	PROTEINS	DD	NCII	COLLAB
w/o GDA	InfoGraph	89.01 \pm 1.13	74.44 \pm 0.31	72.85 \pm 1.78	76.20 \pm 1.06	70.65 \pm 1.13
w/ Manual GDA	GraphCL	86.80 \pm 1.34	74.39 \pm 0.45	78.62 \pm 0.40	77.87 \pm 0.41	71.36 \pm 1.15
	MVGRL	89.70 \pm 1.10	-	-	-	-
	GCA	<u>90.60 \pm 0.76</u>	75.53 \pm 0.22	<u>79.17 \pm 0.39</u>	75.87 \pm 0.96	76.67 \pm 0.43
w/ Automated GDA	AD-GCL	-	73.59 \pm 0.65	74.49 \pm 0.52	69.67 \pm 0.51	73.32 \pm 0.61
	JOAOv2	-	71.25 \pm 0.85	66.91 \pm 1.75	72.99 \pm 0.75	70.40 \pm 2.21
	AutoGCL	88.64 \pm 1.08	<u>75.80 \pm 0.36</u>	77.57 \pm 0.60	<u>82.00 \pm 0.29</u>	70.12 \pm 0.68
	GMNI	91.57 \pm 0.57	76.58 \pm 0.29	79.32 \pm 0.26	83.12 \pm 0.20	<u>76.19 \pm 0.71</u>

Type	Model	IMDB-B	RDT-B	RDT-M-5K
w/o GDA	InfoGraph	73.03 \pm 0.87	82.50 \pm 1.42	53.46 \pm 1.03
w/ Manual GDA	GraphCL	71.14 \pm 0.44	<u>89.53 \pm 0.84</u>	55.99 \pm 0.28
	MVGRL	74.20 \pm 0.70	84.50 \pm 0.60	-
	GCA	<u>74.97 \pm 0.40</u>	87.03 \pm 0.90	56.04 \pm 0.28
w/ Automated GDA	AD-GCL	71.57 \pm 1.01	85.52 \pm 0.79	53.00 \pm 0.82
	JOAOv2	71.60 \pm 0.86	78.35 \pm 1.38	45.57 \pm 2.86
	AutoGCL	73.30 \pm 0.40	88.58 \pm 1.49	<u>56.75 \pm 0.18</u>
	GMNI	75.54 \pm 0.29	91.29 \pm 0.41	57.02 \pm 0.14

4.3.4 对比实验

无监督节点分类任务

表4-4的实验结果显示了 GMNI 在无监督节点分类任务上表现出超越其他基线模型的优秀性能。表中除了 AD-GCL、JOAO-v2、AutoGCL 之外，其他基线模型的实验结果均来自于已发表的论文。在表格中，‘-’表示相关结果在论文中未提供，OOM 表示在 32GB GPU 上出现了内存溢出，表格将最高和次高的性能分别以**粗体**和下划线标示。

首先，GMNI 显著优于具有自动数据增强的图对比学习基线模型。这种提升的原因在于 GMNI 通过保留最小显著信息来生成视图，这些视图近似了 InfoMin 中的最优视图，而其他基线模型未达到这一要求。使用这样的近似最优视图可以防止对比模块学习到冗余信息，并使其集中于显著信息，从而实现更好的性能。其次，GMNI 在所有数据集上均优于具有手动数据增强的基线模型。具有手动数据增强的图对比学习方法严重依赖于预定义的数据增强池，导致对 InfoMin 原则的近似程度有限。如果预定义的数据增强池不够合理，则会严重影响模型的性能。考虑到时间成本、计算成本以及对 InfoMin 中的最优视图的近似能力等因素，GMNI 成为了替代手动数据增强图对比学习的方案。至此，GMNI 使得图

表 4-6 模型在半监督图分类任务上的表现 (重复运行了 5 次之后的平均准确率 Accuracy (%) \pm 标准差)

Dataset	GCA	GraphCL	JOAOv2	AD-GCL	AutoGCL
PROTEINS	73.85 \pm 5.56	74.21 \pm 4.50	73.31 \pm 0.48	73.96 \pm 0.47	75.65 \pm 2.40
DD	76.74 \pm 4.09	76.65 \pm 5.12	75.81 \pm 0.73	77.91 \pm 0.73	77.50 \pm 4.41
NCI1	68.73 \pm 2.36	73.16 \pm 2.90	74.86 \pm 0.39	75.18 \pm 0.31	73.75 \pm 2.25
COLLAB	74.32 \pm 2.30	75.50 \pm 2.15	75.53 \pm 0.18	75.82 \pm 0.26	77.16 \pm 1.48
GITHUB	59.24 \pm 3.21	63.51 \pm 1.02	<u>66.66 \pm 0.60</u>	-	62.46 \pm 1.51
IMDB-B	<u>73.70 \pm 4.88</u>	68.10 \pm 5.15	-	-	71.90 \pm 4.79
REDDIT-B	77.15 \pm 6.96	78.05 \pm 2.65	<u>88.79 \pm 0.65</u>	90.10 \pm 0.15	79.80 \pm 3.47
REDDIT-M-5K	32.95 \pm 10.89	48.09 \pm 1.74	52.71 \pm 0.28	<u>53.49 \pm 0.28</u>	49.91 \pm 2.70

Dataset	GMNI-Label	GMNI-Alter	GMNI-Unlabel-Fint	Ranks
PROTEINS	78.31 \pm 1.31	<u>76.46 \pm 0.67</u>	74.71 \pm 0.59	1
DD	80.69 \pm 0.47	<u>79.30 \pm 1.04</u>	76.38 \pm 0.91	1
NCI1	76.43 \pm 0.50	<u>76.13 \pm 0.59</u>	74.32 \pm 0.19	1
COLLAB	75.56 \pm 0.30	78.50 \pm 0.48	<u>77.21 \pm 0.47</u>	1
GITHUB	65.62 \pm 0.15	68.46 \pm 0.23	66.32 \pm 0.41	1
IMDB-B	74.16 \pm 0.57	71.96 \pm 0.84	69.66 \pm 1.00	1
REDDIT-B	82.68 \pm 0.57	87.96 \pm 0.35	85.91 \pm 1.94	3
REDDIT-M-5K	50.16 \pm 0.77	53.96 \pm 0.51	52.77 \pm 0.57	1

对比学习摆脱了繁琐的手动数据增强方案的选择过程。

无监督图分类任务

表4-5展示了模型在无监督图分类任务上的表现,除了 GCA 外,其他基线模型的实验结果均来自于已发表的论文。表中将最高和次高的性能分别以**粗体**和下划线标示。实验结果显示,GMNI 在 8 个数据集中的 7 个数据集上取得了无监督图分类任务下的最佳性能,并且在所有数据集上超过了具有自动数据增强的图对比学习方法。

无监督节点和图分类任务的结果验证了本文使用最小显著信息来近似最小必要信息的策略在不同的下游任务上是有效的。GMNI 近似了 InfoMin 的最优视图,而不需要使用任务相关信息。最后,表4-4和表4-5表明,具有数据增强的图对比学习方法通常优于没有数据增强的方法,凸显了数据增强在图对比学习中的重要性。

半监督图分类任务

按照 AutoGCL 的实验设置,本文在 TUDataset 上进行了图分类任务的半监督学习,采用了 10 折交叉验证。使用 ResGCN^[131]作为视图对比模块的 GNN 编

表 4-7 GMNI 的消融实验 (重复运行了 3 次之后的平均准确率 Accuracy (%) \pm 标准差)

Dataset	GMNI-Uni	GMNI-Feat	GMNI-Edge	GMNI
Wiki-CS	<u>79.30 \pm 0.00</u>	76.43 \pm 0.15	79.16 \pm 0.07	79.29 \pm 0.05
Amazon-Computers	88.12 \pm 0.25	86.74 \pm 0.28	88.87 \pm 0.17	<u>89.29 \pm 0.08</u>
Amazon-Photo	93.27 \pm 0.47	91.21 \pm 0.49	93.57 \pm 0.26	<u>93.68 \pm 0.35</u>
Coauthor-CS	93.32 \pm 0.05	93.41 \pm 0.10	93.33 \pm 0.02	<u>93.62 \pm 0.10</u>
MUTUG	90.79 \pm 0.28	91.13 \pm 0.77	90.24 \pm 0.42	<u>91.83 \pm 0.62</u>
PROTEINS	76.28 \pm 0.09	75.17 \pm 0.37	76.25 \pm 0.51	<u>76.61 \pm 0.37</u>
DD	79.37 \pm 0.59	76.62 \pm 0.48	<u>79.62 \pm 0.09</u>	79.43 \pm 0.34
COLLAB	72.30 \pm 1.06	74.00 \pm 0.28	<u>77.02 \pm 1.46</u>	76.08 \pm 0.86
IMDB-B	75.20 \pm 0.70	72.67 \pm 0.06	75.07 \pm 0.76	<u>75.70 \pm 0.26</u>
REDDIT-B	90.05 \pm 0.61	<u>92.00 \pm 0.48</u>	90.05 \pm 0.30	91.25 \pm 0.75
Dataset	GMNI-Simp	GMNI-ViewM	GMNI-Simult	GMNI
Wiki-CS	79.10 \pm 0.15	79.34 \pm 0.05	79.13 \pm 0.12	79.29 \pm 0.05
Amazon-Computers	89.12 \pm 0.16	89.04 \pm 0.08	89.00 \pm 0.32	89.29 \pm 0.08
Amazon-Photo	93.65 \pm 0.32	93.44 \pm 0.39	93.80 \pm 0.27	93.68 \pm 0.35
Coauthor-CS	93.44 \pm 0.08	93.48 \pm 0.02	93.39 \pm 0.12	93.62 \pm 0.10
MUTUG	91.53 \pm 1.02	90.45 \pm 0.54	91.35 \pm 0.61	91.83 \pm 0.62
PROTEINS	76.49 \pm 0.19	76.53 \pm 0.28	77.30 \pm 0.21	76.61 \pm 0.37
DD	79.20 \pm 0.22	79.00 \pm 0.38	78.83 \pm 0.05	79.43 \pm 0.34
COLLAB	70.39 \pm 0.95	75.91 \pm 0.10	74.12 \pm 0.63	76.08 \pm 0.86
IMDB-B	70.23 \pm 1.50	75.33 \pm 1.07	75.60 \pm 0.66	75.70 \pm 0.26
REDDIT-B	91.63 \pm 0.25	90.90 \pm 0.38	89.92 \pm 0.40	91.25 \pm 0.75

码器，隐藏层大小为 128，与 AutoGCL 的设置相同。

GMNI-Label 使用了 10% 的有标签数据进行训练和测试。GMNI-Alter 采用了多个交替的训练步骤，每个步骤包括 80% 的无标签数据的训练 (CL) 和 10% 的有标签数据的微调 (Fine-Tune, FT)。最后，在 10% 的有标签数据上进行测试。GMNI-Unlabel-Fint 首先在 80% 的无标签数据上进行训练，然后在 10% 的有标签数据上进行微调，最后在 10% 的有标签数据上进行测试。GMNI-Alter 在 CL 和 FT 之间交替进行，即 (CL1-FT1-CL2-FT2-....)，而 GMNI-Unlabel-Fint 则是在完成 CL 后进行 FT。

表4-6展示了模型在半监督图分类任务上的表现，最高和第二高的性能分别以**粗体**和下划线标示。实验结果显示，GMNI 在 8 个数据集中的 7 个上都优于基线模型，并获得了 1.25 的平均排名。有趣的是，无标签数据并非总是有效的，GMNI-Unlabel-Fint 的性能有时候甚至不如仅使用 10% 有标签数据进行训练的 GMNI-Label。这是因为直接使用标签来微调使用无标签数据训练出的模型，会干扰已学习的分布，从而导致性能下降。通过交替使用无标签数据和有标签数据训练出的 GMNI-Alter 要优于 GMNI-Unlabel-Fint。CL 和 FT 的交替训练有助于缓解无标签数据和有标签数据之间学习知识的差异，从而促进模型在这两个

数据来源之间的相互学习。

由于共享的视图生成器在所有图中的泛化能力有限,因此 GMNI 在 REDDIT-BINARY 数据集上的表现不如两种基线方法。如果根据每个批次都训练一个专用的视图生成器,则可以显著提升性能: GMNI-Label 的得分为 91.10 ± 0.36 , GMNI-Alter 的得分为 89.28 ± 1.83 , GMNI-Unlabel-Fint 的得分为 88.35 ± 0.23 。然而,由于计算限制,本研究并没有实施这种训练策略。

4.3.5 消融实验

为了验证 GMNI 中各组件的合理性,本节进行了消融实验。表4-7展示了在 10 个数据集(包括 4 个节点分类数据集和 6 个图分类数据集)上 GMNI 使用相同的超参数进行的消融实验结果。表中第一组(GMNI-Uni, GMNI-Feat, GMNI-Edge, GMNI)和第二组(GMNI-Simp, GMNI-ViewM, GMNI-Simult, GMNI)消融实验的最高性能分别以下划线和**粗体**标示。

自动视图生成器的合理性

本节针对自动视图生成器生成了 3 个变体来验证其有效性。GMNI-Uni 采用了均匀分布的边和特征重要性。GMNI-Feat 采用了均匀分布的边重要性,但保留了特征重要性的自动生成。GMNI-Edge 采用了均匀分布的特征重要性,但保留了边重要性的自动生成。GMNI-Uni, GMNI-Feat 和 GMNI-Edge 的其余部分都和 GMNI 保持一致。

GMNI 在除了 Wiki-CS 以外的所有数据集上的表现均优于 GMNI-Uni,而在 Wiki-CS 上其表现与 GMNI-Uni 相近。这一事实验证了 GMNI 的自动视图生成器的有效性。与此同时,GMNI 在 9 个数据集上表现均优于 GMNI-Feat,在 8 个数据集上超越 GMNI-Edge。结果表明,将边重要性和特征重要性相结合,而不是单独使用其中之一,能够更有效地提升模型性能。因此,本文强调在图数据增强过程中应该同时考虑拓扑结构和特征信息,以实现模型的最佳性能。

最大化-最小化优化的合理性

本节采用了 3 种变体来验证自动视图生成器中对抗优化的合理性。GMNI-Simp 将视图生成器中的可学习 GCN 编码器替换为无参数 GCN，即 $\mathbf{W} = \mathbf{I}$ 。GMNI-ViewM 通过最大化-最小化优化来获得视图生成器，即：

$$\max_T \max_{f,g} I(g(f(\mathcal{G})); g(f(T(\mathcal{G})))) - |T(\mathcal{G})|. \quad (4-18)$$

GMNI-Simult 同时训练视图生成器和视图对比模块，即视图生成器和视图对比模块共享 GNN 编码器。

首先，GMNI 在 9 个数据集上的性能优于 GMNI-Simp，这验证了视图生成器中可学习的 GCN 编码器的有效性。其次，GMNI 在 9 个数据集上的性能优于 GMNI-ViewM，显示出通过最小化训练得到的 GNN 编码器和投影头优于通过最大化训练得到的。通过最小化训练所获得的激进图编码器和投影头有助于生成更通用、更鲁棒的重要图，以及更精确地近似 InfoMin 中的最优视图。最后，GMNI 在 8 个数据集上击败了 GMNI-Simult，这是因为同时训练视图生成器和视图对比模块会阻碍更好的视图生成和更好的嵌入结果的产生。视图生成器强调生成保留显著信息并丢弃无用信息的视图，而视图对比模块强调学习到的图编码器的质量，但当它们一起训练时，会相互干扰训练过程。

4.3.6 附加实验

重要性图可视化

中心性 (Centrality) 反映了图中节点的重要性。在 GCA 中，边和特征的重要性由度、PageRank 和特征向量中心性来定义。为了验证 GMNI 中获得的重要性的合理性，本节在 MUTAG 数据集中随机选择了一张图来可视化由这些中心性定义的重要性以及由 GMNI 获得的重要性。MUTAG 数据集中的图是一组硝基芳香化合物，其中图的节点代表原子，节点之间的边代表相应原子之间的键。

图 4-7 展示了重要性图的热力图可视化结果，图中 DEGREE, PR, 和 EVC 分别代表度中心性、PageRank 中心性和特征向量中心性。为了方便展示，本节随机选择了来自特征的 100 个维度进行特征重要性的可视化。可视化结果表示

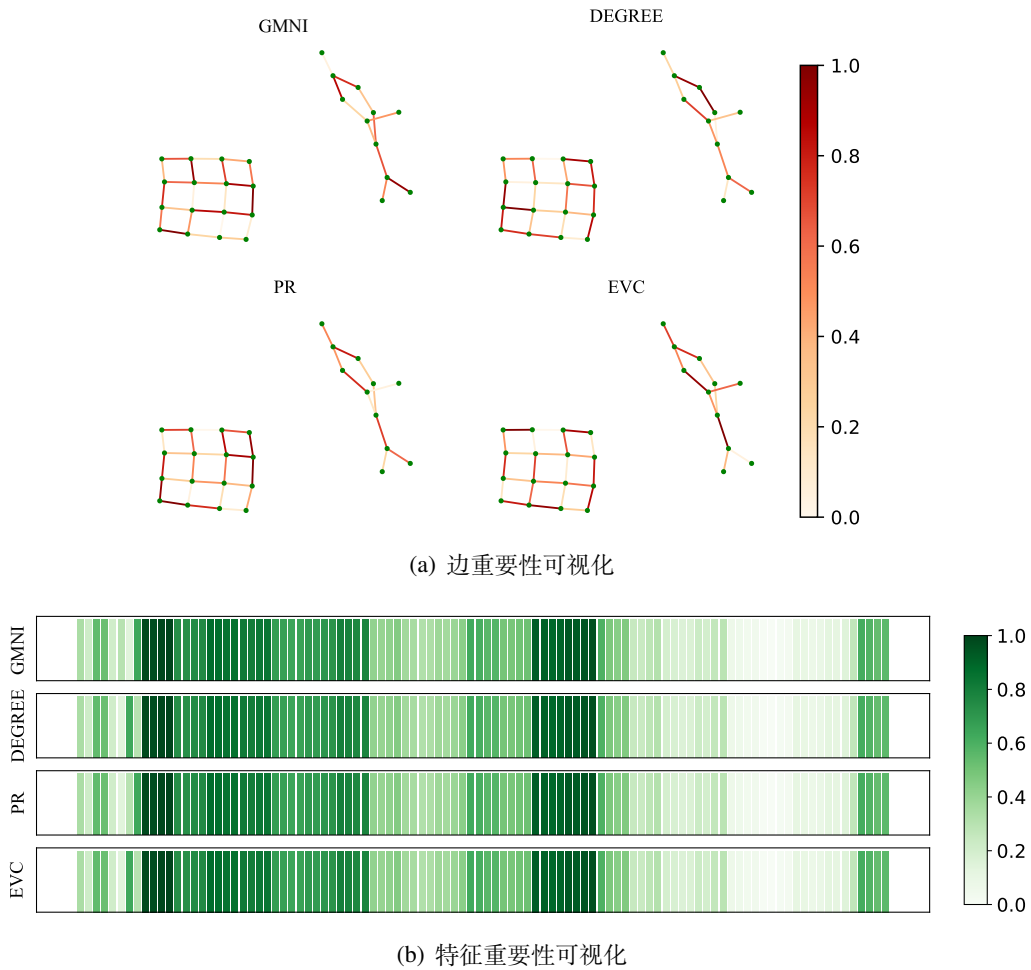


图 4-7 重要性图的热力图可视化

GMNI 获得的重要性特征与由中心性获得的重要性特征基本保持对齐，验证了 GMNI 准确地识别出了图中的关键边和特征。

参数敏感性分析

本节对 GMNI 的关键超参数 p_{s1} 和 p_{s2} 进行了敏感性分析。为了简化分析，本节将两个视图的 p_{s1} 和 p_{s2} 设置为相同，并用它们来调整边的整体丢弃率和特征的整体掩码率。实验从 $[0.0, 0.1, 0.3, 0.7, 0.9, 1.0]$ 中选择 p_{s1} 和 p_{s2} 的值，结果如图4-8所示。为了方便观察，图中展示了三维图的两个角度。

首先，可以观察到 GMNI 对 p_{s1} 的变化比对 p_{s2} 更为敏感：改变 p_{s1} 会导致更明显的性能变化。这表明在 GMNI 中，拓扑信息的作用比特征信息更为关键。其次，值得注意的是，过大的 p_{s1} 或 p_{s2} 会对性能产生不利影响。当 p_{s1} 变得过大时，图的拓扑信息几乎被完全破坏，导致大部分节点变成缺乏边的孤立节点。

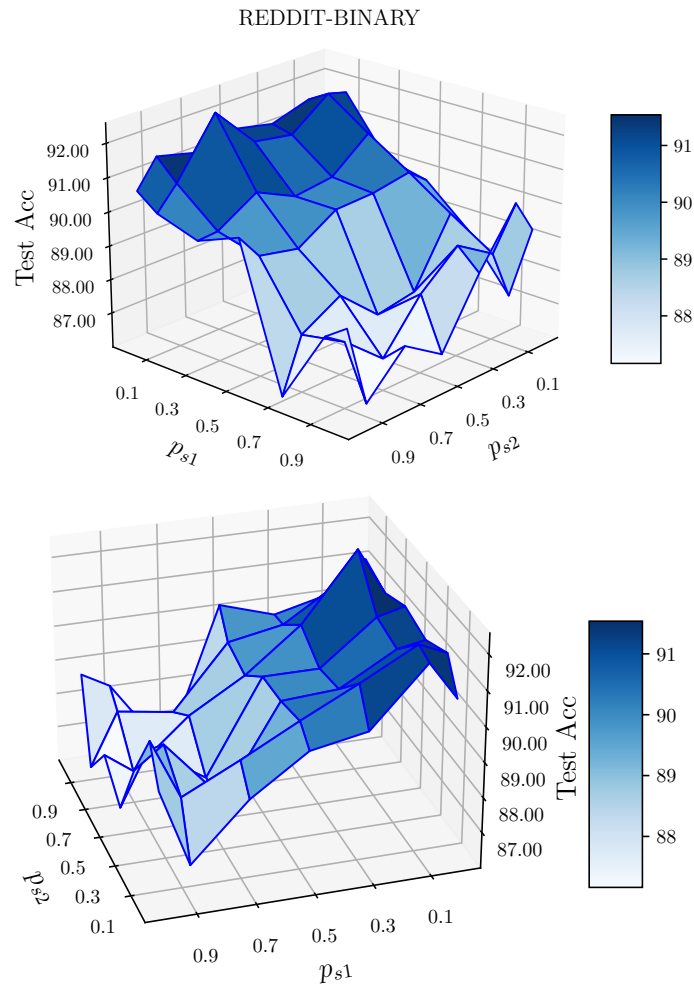


图 4-8 GMNI 在 REDDIT-BINARY 数据集上的无监督图分类性能

当 p_{s2} 变得过大时，图几乎退化为无特征的图。当两者同时变大时，性能往往会变得极差。第三，图4-8中接近 (0,0) 的点比 (0,0) 点的表现更好。这一观察表明，即使是在具有较小的丢弃率或掩码率的情况下，利用重要性生成视图也比仅仅使用原始图进行对比更有优势。

嵌入结果可视化

为了展示嵌入结果的质量,本节使用 PCA(2 components PCA)^[132]和 t-SNE^[133]来可视化原始特征、AD-GCL、JOAOv2、AutoGCL 和 GMNI 在 Amazon-Photo 数据集上的嵌入结果。图4-9展示了 t-SNE 可视化结果，图中不同的颜色代表不同的节点类别。可视化结果显示 GMNI 实现了更为明显的分类边界，突显了 GMNI 学习到的嵌入结果的优越性。

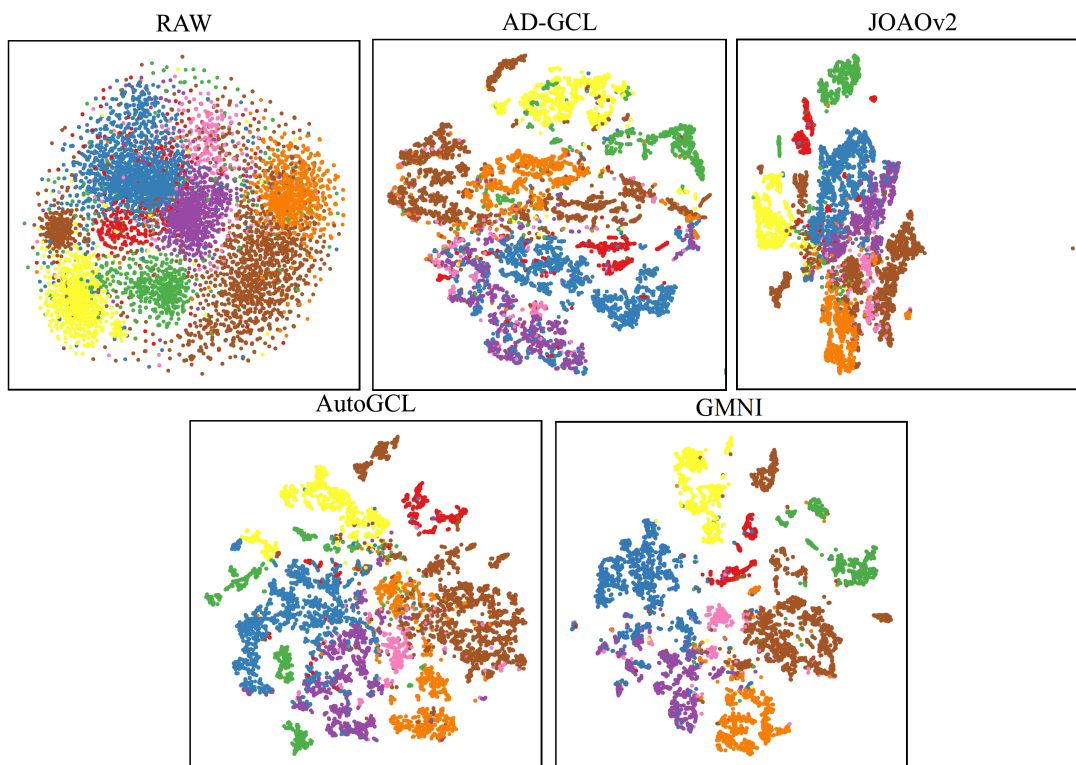


图 4-9 Amazon-Photo 数据集上原始特征和不同基线学习到节点嵌入的 t-SNE 可视化图

4.4 本章小结

本章首先分析了现存的手动数据增强的无监督图对比学习方法存在的问题，这些方法因为通过试错法或经验法对每个数据集来选择数据增强方案将带来大量的时间和计算成本。其次，本研究分析了现存的自动数据增强的无监督图对比学习方法存在的问题，即它们由于没有保留下游任务相关的信息，可能会造成信息不足的风险。为解决这些问题，本章提出了一种具有自动数据增强的无监督图对比学习方法：GMNI。GMNI 通过保留最小显著信息来降低模型编码冗余信息和信息不足的风险，从而更好地近似了 InfoMin 原则中的最优视角的条件。此外，GMNI 引入了基于重要性图的随机性来增强数据增强，提高了视图的多样性，并使模型对扰动更加稳定。在节点和图分类任务的 14 个数据集上的广泛实验验证了 GMNI 的卓越性能，GMNI 优于其他自动和手动数据增强的图对比学习方法。

第五章 图表示学习在相似药物推荐中的应用

5.1 研究背景与动机

医疗技术直接涉及到人类的生命健康和社会的稳定。随着医疗技术的不断创新和医学知识的积累，医疗领域持续发展，为人类提供更全面的健康保障。在这一进程中，计算机技术发挥着重要的作用。医学影像技术、智能辅助诊断等计算机技术应用为医疗领域注入了高效、精准和智能化的元素，对患者体验的提升、医疗质量的改善以及医学科研的推动起到了积极作用。

医疗领域的数据通常具有如下特征：

(1) 复杂的关系结构：医疗数据涉及到复杂的关系结构，如药物和疾病之间复杂的作用关系、药物内部复杂的化学结构等。

(2) 数据来源复杂且异质：医疗数据通常整合自多个不同的来源，如医学影像数据、基因组数据、临床数据等，而这些数据往往具有不同的数据形式。

(3) 数据稀疏性：医疗数据通常是稀疏的，即数据中存在大量缺失或不完整的信息。例如，某些基因表达数据可能缺失，这增加了数据分析和挖掘的难度。

此外，医疗领域的数据处理还需要技术具有高度的可解释性。考虑到医疗数据的上述特性，图神经网络是一个非常适合应用于该领域的技术。

从粗粒度的药物层面来看，药物、疾病和人类基因之间构成了复杂的关系网络。药物之间存在相互作用（相克、辅助）、药物与疾病之间有直接对应的诊疗关系，而人类基因组合与特定疾病绑定。从细粒度的药物层面来说，药物作为化学制品的集合，内部由原子和化学键构成，根据原子与键的构成科研工作者可以推断出药物的相关属性。事实上，无论是从细粒度还是粗粒度出发，它们都构成了图。举例来说，药物-疾病图、药物化学构成图等，即将药物本身视为节点，



(a) 商品界面

(b) 店内相似药物界面

图 5-1 美林牌布洛芬混悬液在美团页面

或将药物内部的原子视为节点。通过分析这样的图可以挖掘出药物隐藏的属性及用途，发现药物常用组合、排斥组合等。例如，人们应对感冒时，常常会组合使用维 C 银翘片以及重感冒灵片。专业地来说，药物组合图中这两种药物展现出强关联关系，因此一般建议患者同时服用这两种药物来缓解症状。

在医药领域，寻找相似药物具备广泛的应用前景。首先，通过寻找与已知药物相似的药物，可以发现新的药物用途，这有助于实现药物的再利用或重定位。其次，相似药物可用于预测新药的潜在副作用，为临床试验的设计和药物安全性评估提供重要支持。此外，根据相似药物来最大程度规避风险，个性化地定制药物治疗方案，有助于减少药物不良反应，实现患者的个性化治疗目标。以美团中的买药板块举例，如图5-1所示。图5-1(b)中展示了图5-1(a)中的美林牌布洛芬混悬液在店内的相似药物。这些相似药物实际上只是同种药物的不同规格或是不同品牌。然而，仅通过搜索同种药物的不同规格或不同品牌得到的相似药物，严重限制了结果的泛化，研究者们不会通过这种途径发掘到新的药物，也就无从达到上述寻找相似药物的目标。

在本章，将讨论如何将图表示学习应用于相似药物推荐。具体来说，如何使用图表示学习来获得药物的专业性嵌入，并将该嵌入结果用于寻找相似药物。通过将药物表示为图结构中的节点，研究者们可以融合药物、疾病、基因、功效等信息，从而更准确地捕捉药物之间的相似性和差异性。这种方法不仅可以考

虑药物的共同特征，还可以考虑到它们与不同疾病、基因、功效等的联系，因此提高了药物相似性的准确度和泛化性。

5.2 数据分析

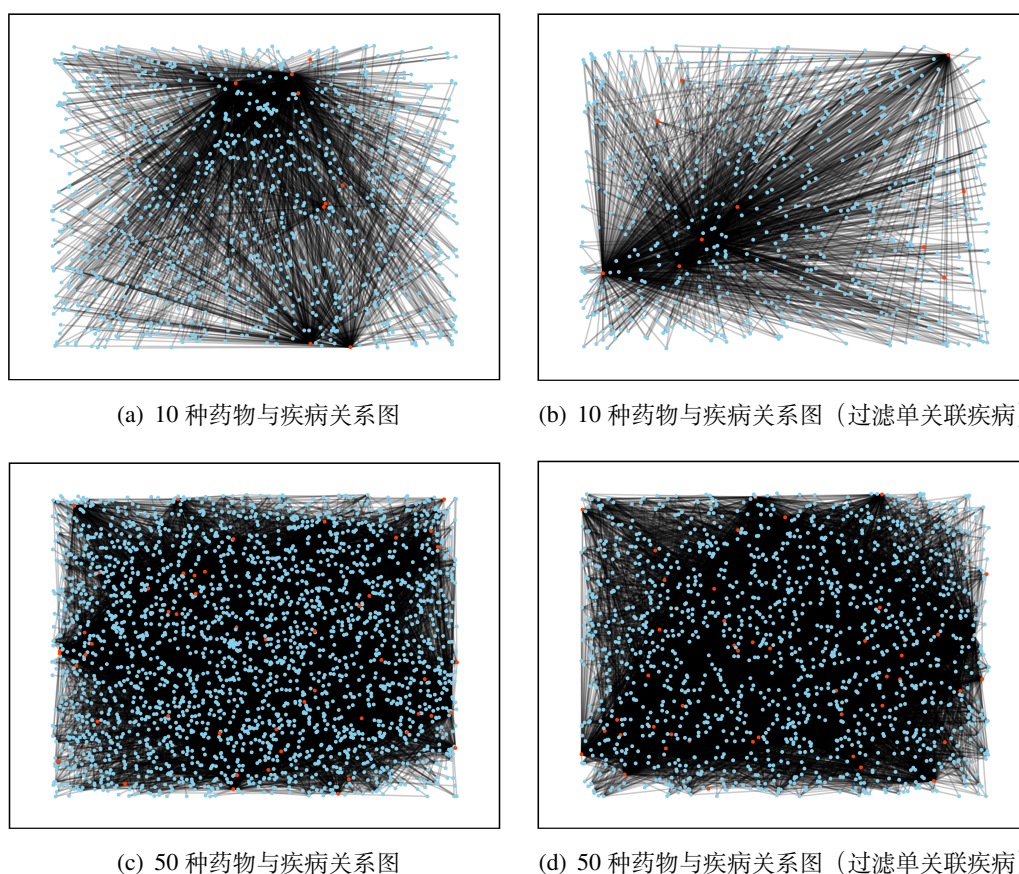


图 5-2 药物与疾病关系图

表 5-1 CTD 中药物和疾病节点的度频数分布

节点类型\度频数范围	(0,500]	(500,1000]	(1000,1500]	(1500,2000]	(2000,2500]	(2500,3000]
药物 Chemical	15668	1128	254	113	47	40
	(3000,3500]	(3500,4000]	(4000,4500]	(4500,5000]	(5000,5500]	(5500,6000]
	20	17	10	1	3	2
疾病 Disease	(0,1000]	(1000,2000]	(2000,3000]	(3000,4000]	(4000,5000]	(5000,6000]
	6509	274	159	132	96	39
	(6000,7000]	(7000,8000]	(8000,9000]	(9000,10000]	(10000,11000]	
	34	19	6	4	1	

Comparative Toxicogenomics Database (CTD, <https://ctdbase.org>) 是一个公开的医疗病理数据库，它提供了药物-疾病关系、基因-疾病关系、药物-基因关系的数据信息，它的构建目标是探究环境暴露对于人类健康的影响。

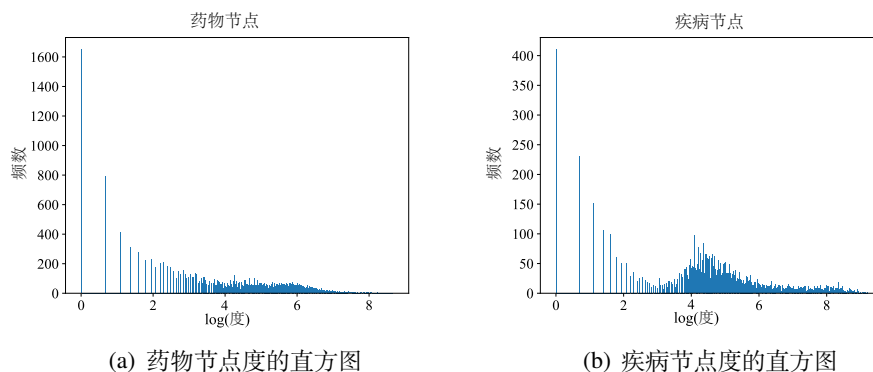


图 5-3 药物、疾病节点度的直方图

首先对 CTD 中药物与疾病的关系 (Chemical–Disease Associations) 进行分析：该数据集包含了 17304 种药物和 7274 种疾病之间的 8733910 条关联关系。为了可视化这些关联关系，图5-2随机选取了其中的 10 种和 50 种药物，并展示它们与相关疾病的关联网络。图中橙色节点代表药物，蓝色节点代表疾病，连边表示该药物可治疗该疾病。图5-2(a)和图5-2(b)展示了 10 种药物的关联网络，而图5-2(c)和图5-2(d)展示了 50 种药物的关联网络。图5-2(b)和图5-2(d)过滤了关联少于 2 种药物的疾病。从图中可以看出，即使是数量很少的药物节点，都能关联出极其复杂的网络图。这使得进行人为分析和标注药物关系图变得极为困难。

表5-1和图5-3展示了 CTD 中药物和疾病节点的度分布。考虑到度的数值差异极大，图5-3的直方图的横坐标为度的 \log 值，纵坐标为频数。首先，节点度分布呈现长尾特征，绝大部分节点为低度的。其次，高度节点为图中的热节点，将关联众多节点，成为图中的“枢纽”。最高度药物节点关联了 5811 种疾病，为：bisphenol A；最高度疾病节点关联了 10067 种药物，为：Breast Neoplasms (乳腺肿瘤)。

5.3 需求分析

本章期望构建一个相似药物推荐系统，该系统的使用用户为药物研究工作者，其设计目标主要包括两方面内容，第一是药物展示和查询功能，需要方便展示数据库内的所有药物，并可以对药物属性进行查询；第二是提供相似药物推荐功能，当用户选择了候选药物后，系统将会反馈推荐出的 Top 10 相似药物。

获取相似药物的推荐结果是系统的实现重点。相似药物通常被定义为具有

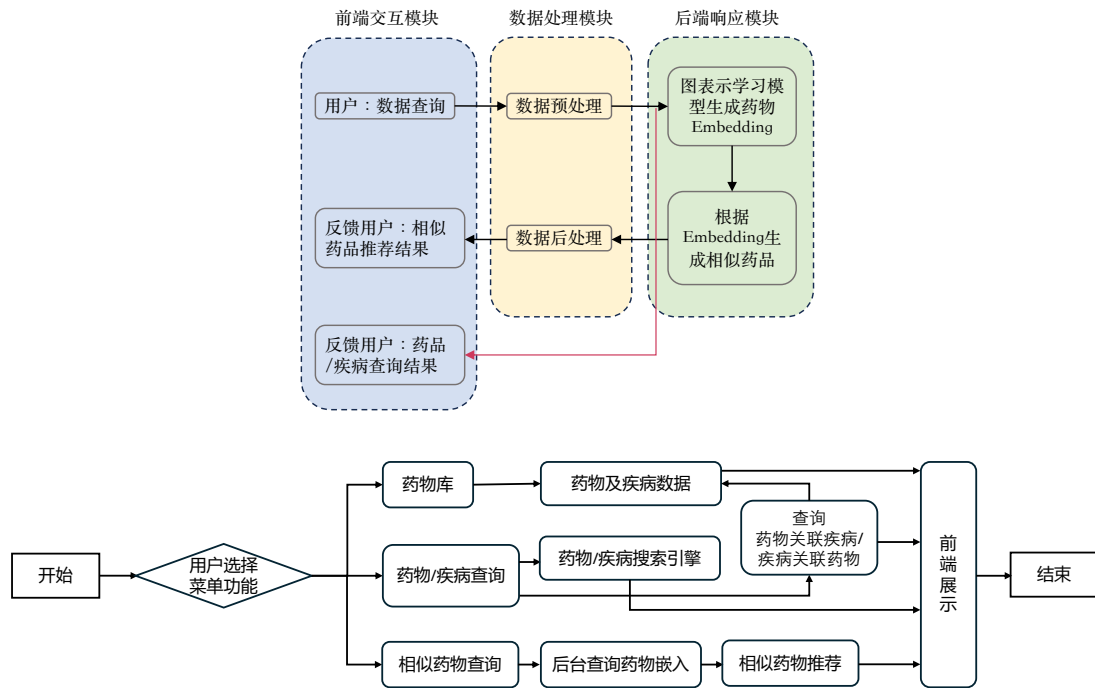


图 5-4 相似药物推荐系统框架及流程图

相似疗效的药物。事实上，简单地通过寻找具有共同可治疗疾病的药物就可以获得相似药物。例如，布洛芬和扑热息痛片都可以用于退烧，于是可以把它们定义为相似药物。然而，由于药物和疾病节点呈现长尾分布，大部分节点都是低度的，这使得泛化寻找相似药物变得困难。举例来说，对于只与唯一疾病 Ectoparasitic Infestations 关联的药物 Tenvermectin D，该疾病又没有其他关联药物，便无法利用这种方法找到其相似药物。因此，利用专业性的嵌入来帮助寻找相似药物是一个不错的选择。本章将基于 CTD 数据集，获得药物的专业性理解 (Embedding)，借助于这些 Embedding 来探究哪些药物为相似药物。

综合上述分析，相似药物推荐系统存在以下需求：

- 药物和疾病的信息查询：药物研究人员需要准确而完整的药物和疾病信息来支持科研工作。因此，系统需提供数据查询和数据分析功能，以确保用户可以方便地获取所需的信息。
- 可解释性的相似药物推荐结果：药物研究人员需要根据候选药物获得相似药物推荐结果。在医疗领域，严谨性和可解释性至关重要，系统提供的相似药物推荐结果必须有理可循。
- 简洁美观的系统页面：友好便捷的页面布局和操作方式能够降低用户的学

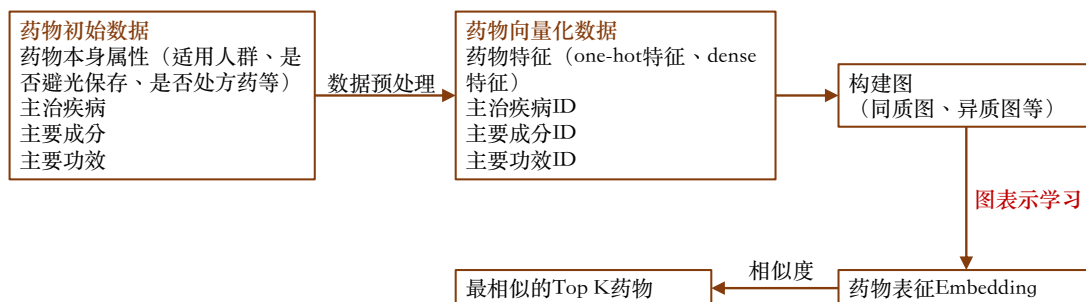


图 5-5 通过图表示学习获得相似药物流程图

习和使用难度。

5.4 系统架构

本章构建的系统架构如图5-4所示，使用 Flask 框架开发，主要包括三个关键部分：1. 前端交互模块：负责与用户进行交互，反馈用户的行为，响应用户的操作；2. 数据处理模块：该模块负责数据处理逻辑，并且存储了 CTD 数据库中药物-疾病的关联数据以及已经训练好的图表示学习模型；3. 后端响应模块：负责将处理好的数据传递给图表示学习模型来实现药物嵌入，并完成相似药物推荐的功能。

系统的核心功能是寻找与给定药物相似的其他药物，其流程如图5-5所示。主要成分及主要功效可用于加入构图，但由于数据集限制本系统只利用了药物-疾病关系来构图。本系统直接使用药物节点的 ID 嵌入来作为药物的向量化信息，组合使用了第三章中提出的基于频率自适应的简化频域图卷积网络 SGCN-MH 及第四章中提出的基于最小显著信息的无监督图对比学习网络 GMNI 用于获得药物向量嵌入。系统完成了从药物-疾病数据处理到相似药物推荐的全流程。

5.4.1 关系图构建

利用 CTD 数据集中药物和疾病之间的关联关系来构造异质节点同质关系图 $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ 。其中 \mathcal{V} 代表节点集合，包括两种类型的节点： M 个药物节点、 N 个疾病节点。 \mathcal{E} 代表边集合，边只有一种类型，假设包含 S 条边。当药物 v_i 可整治疾病 v_j 时， v_i 和 v_j 之间则存在一条边 e_{ij} ，定义该药物-疾病邻接矩阵为

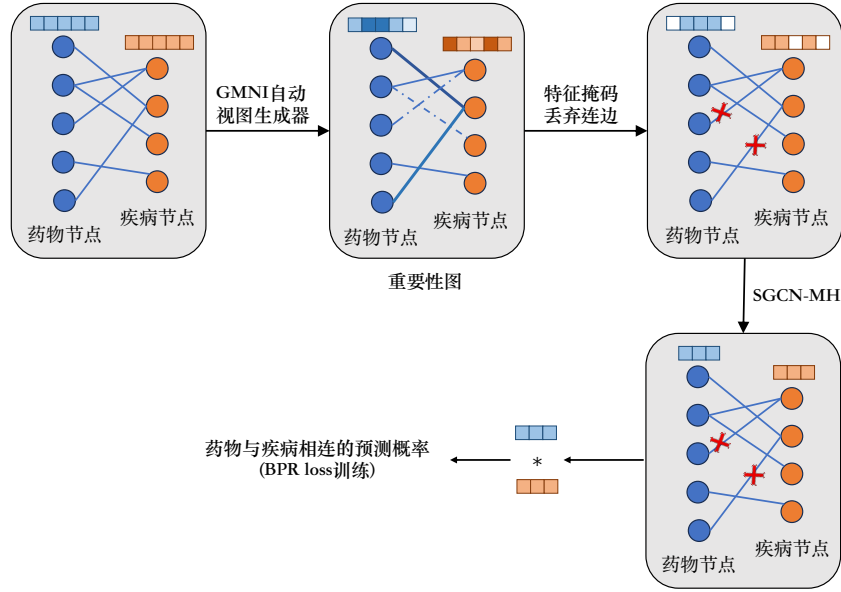


图 5-6 获取药物嵌入的图表示学习网络结构

$\mathbf{A}_{MD} = \{0, 1\}^{M \times N}$, 其度矩阵为 \mathbf{D}_{MD} 。构建关系图如图5-6的输入图所示。

5.4.2 获得药物嵌入的图表示学习模型

本章利用 GMNI 中的自动视图生成器及 SGCN-MH 来获得药物理解, 整体的网络结构如图5-6所示。定义药物节点 ID Embedding 为 $\mathbf{E}_M \in \mathbb{R}^{M \times F}$, 疾病节点 ID Embedding 为 $\mathbf{E}_D \in \mathbb{R}^{N \times F}$, 其中 F 为特征维度, 图网络的初始输入为 $\mathbf{E}_M^{(0)}$ 及 $\mathbf{E}_D^{(0)}$:

$$\begin{aligned} \mathbf{E}_M^{(0)} &= \mathbf{E}_M, \\ \mathbf{E}_D^{(0)} &= \mathbf{E}_D. \end{aligned} \quad (5-1)$$

下面通过 GMNI 中的自动视图生成器来获得重要性图, 并根据重要性图获得生成子图。药物节点的邻居只为疾病节点, 获得药物节点的嵌入过程如下:

$$\mathbf{E}_M^{(l)} = \sigma \left(\mathbf{A}_{MD} \mathbf{E}_M^{(l-1)} \mathbf{W}_M^{(l)} \right). \quad (5-2)$$

其中, $\sigma(\cdot)$ 为激活函数, $\mathbf{W}_M^{(l)}$ 为当前层的参数矩阵。根据药物节点嵌入获得药物节点重要性 $\mathbf{P}_M \in \mathbb{R}^{M \times 1}$:

$$\mathbf{P}_M = \text{Gumbel} \left(h_\phi \left(\mathbf{E}_M^{(L)} \right) \right). \quad (5-3)$$

定义使用的图卷积层为 L 层, $h_\phi(\cdot)$ 为一个简单的 MLP。继续获得药物节点特征重要性 $\mathbf{P}_{Mf} \in \mathbb{R}^{F \times 1}$:

$$\mathbf{P}_{Mf} = \mathbf{E}_M^{(0)T} \mathbf{P}_M. \quad (5-4)$$

对应的, 疾病节点的邻居只为药物节点, 因此其邻接矩阵为药物-疾病邻接矩阵的转置: $\mathbf{A}_{DM} = \mathbf{A}_{MD}^{-1}$ 。获得疾病节点的嵌入:

$$\mathbf{E}_D^{(l)} = \sigma(\mathbf{A}_{DM} \mathbf{E}_D^{(l-1)} \mathbf{W}_D^{(l)}). \quad (5-5)$$

与药物节点与特征重要性的计算过程类似, 可以获得疾病节点重要性: $\mathbf{P}_D \in \mathbb{R}^{N \times 1}$, 疾病节点特征重要性: $\mathbf{P}_{Df} \in \mathbb{R}^{F \times 1}$ 。边 $\varepsilon_k = (v_i, v_j) \in \mathcal{E}$ 的重要性 $\mathbf{p}_{e,k}$ 计算如下:

$$\mathbf{p}_{e,k} = \text{Gumbel} \left(h_\psi \left(\left[\mathbf{E}_M^{(L)}[v_i]; \mathbf{E}_D^{(L)}[v_j] \right] \right) \right), \quad (5-6)$$

其中, $h_\psi(\cdot)$ 是一个简单的 MLP。所有的边重要性表示为 $\mathbf{P}_e = [\mathbf{p}_{e,1}, \dots, \mathbf{p}_{e,S}]^T \in \mathbb{R}^{S \times 1}$ 。以上是重要性图的生成过程, 网络根据公式 (4-13) 来优化上述参数。

接下来, 根据特征重要性 \mathbf{P}_{Mf} , \mathbf{P}_{Df} 及边重要性 \mathbf{P}_e 通过伯努力分布经过特征掩码和丢弃连边来生成子图 $\mathcal{G}' = \{\mathcal{V}', \mathcal{E}'\}$, 新的药物和疾病节点特征为 $\mathbf{E}'_M, \mathbf{E}'_D$, 新的邻接矩阵为 $\mathbf{A}'_{DM}, \mathbf{A}'_{MD}$ 。子图 \mathcal{G}' 将会包括图中更加重要的边与特征信息, 有利于后续的嵌入学习过程, 一方面可以增加学习到节点的鲁棒性与精确性, 另一方面可以减少后续过程计算量。

在子图 \mathcal{G}' 上, 直接使用 SGCN-MH 来获取节点嵌入, 分别通过新的邻接矩阵 $\mathbf{A}'_{DM}, \mathbf{A}'_{MD}$ 计算它们的归一化邻接矩阵 $\hat{\mathbf{A}}'_{DM}, \hat{\mathbf{A}}'_{MD}$ 与归一化二跳连接矩阵 $\mathbf{M}'_{DM}, \mathbf{M}'_{MD}$ 。计算药物节点嵌入如下:

$$\begin{aligned} \mathbf{H}_M^{(0)} &= \text{ReLU}(\mathbf{E}'_M \mathbf{W}_M), \\ \mathbf{H}_M^{(1)} &= \text{ReLU} \left(\alpha_{M1}^{(1)} \hat{\mathbf{A}}'_{MD} \mathbf{H}_M^{(0)} \mathbf{W}_{M1} + \alpha_{M2}^{(1)} \mathbf{M}'_{MD} \mathbf{H}_M^{(0)} \mathbf{W}_{M2} \right), \\ \mathbf{H}_M^{(l)} &= \alpha_{M1}^{(l)} \hat{\mathbf{A}}'_{MD} \mathbf{H}_M^{(l-1)} + \alpha_{M2}^{(l)} \mathbf{M}'_{MD} \mathbf{H}_M^{(l-1)}. \end{aligned} \quad (5-7)$$

其中 $\mathbf{W}_M, \mathbf{W}_{M1}, \mathbf{W}_{M2}$ 为需要学习的参数矩阵, $\alpha_{M1}^{(l)}, \alpha_{M2}^{(l)}$ 为每层归一化平衡系

数。对应地，可以计算疾病节点嵌入如下：

$$\begin{aligned}\mathbf{H}_D^{(0)} &= \text{ReLU}(\mathbf{E}'_D \mathbf{W}_D), \\ \mathbf{H}_D^{(1)} &= \text{ReLU}\left(\alpha_{D1}^{(1)} \hat{\mathbf{A}}'_{DM} \mathbf{H}_D^{(0)} \mathbf{W}_{D1} + \alpha_{D2}^{(1)} \mathbf{M}'_{DM} \mathbf{H}_D^{(0)} \mathbf{W}_{D2}\right), \\ \mathbf{H}_D^{(l)} &= \alpha_{D1}^{(l)} \hat{\mathbf{A}}'_{DM} \mathbf{H}_D^{(l-1)} + \alpha_{D2}^{(l)} \mathbf{M}'_{DM} \mathbf{H}_D^{(l-1)}.\end{aligned}\quad (5-8)$$

假定网络的总层数为 K ，最终获得的药物节点嵌入为： $\mathbf{H}_M = [\mathbf{H}_M^{(0)}; \dots; \mathbf{H}_M^{(K)}]$ ，疾病节点嵌入为 $\mathbf{H}_D = [\mathbf{H}_D^{(0)}; \dots; \mathbf{H}_D^{(K)}]$ 。将该网络的学习目标定为链接预测，则药物节点与疾病节点相连的预测概率为：

$$\hat{\mathbf{R}} = \mathbf{H}_M (\mathbf{H}_D)^T. \quad (5-9)$$

其中， $\hat{\mathbf{R}} \in \mathbb{R}^{M \times N}$ 。使用 BPR loss 作为目标函数来训练网络：

$$\mathcal{L} = \sum_{m=1}^M \sum_{i \in \mathcal{N}_m} \sum_{j \notin \mathcal{N}_m} \ln \sigma(\hat{r}_{mi} - \hat{r}_{mj}). \quad (5-10)$$

其中， \hat{r}_{mi} 代表预测药物节点 v_m 与疾病节点 v_i 相连的概率， \mathcal{N}_m 代表药物节点 v_m 的邻居疾病节点的集合。在训练结束后，即可获得药物节点的嵌入表示 \mathbf{H}_M 。根据 \mathbf{H}_M 通过余弦相似度可以直接定义出候选药物最相近的前十个药物，其中排在第一位的药物是候选药物自身。

5.5 应用展示

图5-7为药物库界面，该页面展示了 CTD 中所有药物与疾病的名称和 ID 信息。图5-8及图5-9为药物-疾病的查询页面。一方面，药物-疾病的查询页面内嵌了 CTD 中的搜索引擎，方便用户对药物或疾病的属性进行查询；另一方面，用户可以选择下拉框中的药物或疾病，并从数据库中找到它们关联的疾病或药物。

图5-10展示了系统的相似药物预测结果。用户可以直接输入药物名称或在下拉框中选择药物来获得推荐结果。图5-10(前两张)为两个随机选择药物的 Top 10 相似药物的预测结果，该推荐结果十分合理，可以观察到它们大量重合的主治疾病。图5-10(后两张)展示了 Tenvermectin D 和 Versicolin 的 Top 10 相似药物

ChemicalName	ChemicalID	DiseaseName	DiseaseID
06-Paris-LA-66 protocol	C046983	Precursor Cell Lymphoblastic Leukemia-Lymphoma	MESH:D054198
10074-G5	C534883	Adenocarcinoma	MESH:D000230
10074-G5	C534883	Adenocarcinoma of Lung	MESH:D00077192
10074-G5	C534883	Alopecia	MESH:D000505
10074-G5	C534883	Androgen-Insensitivity Syndrome	MESH:D013734
10074-G5	C534883	Astrocytoma	MESH:D001254
10074-G5	C534883	Autistic Disorder	MESH:D001321
10074-G5	C534883	Breast Neoplasms	MESH:D001943
10074-G5	C534883	Breast Neoplasms, Male	MESH:D018567
10074-G5	C534883	Bulbospinal neuropathy, X-linked recessive	MESH:C537017
10074-G5	C534883	Burkitt Lymphoma	MESH:D002051
10074-G5	C534883	Carcinoma	MESH:D002277
10074-G5	C534883	Carcinoma, Ductal	MESH:D044584
10074-G5	C534883	Carcinoma, Hepatocellular	MESH:D006528
10074-G5	C534883	Carcinoma, Hepatocellular	MESH:D006528
10074-G5	C534883	Carcinoma, Merkel Cell	MESH:D015266
10074-G5	C534883	Carcinoma, Non-Small-Cell Lung	MESH:D002289

图 5-7 药物库展示界面

The screenshot shows the CTD website interface. At the top, it says 'Comparative Toxicogenomics Database' and 'Connect. Compare.' Below this, there's a search bar with the text '88,144,004 TOXICOGENOMIC RELATIONSHIPS!'. The search bar has a dropdown menu for 'All' and a search button. Below the search bar, there are sections for 'Updated Chemicals' and 'Updated Diseases'. The 'Updated Chemicals' section lists various chemicals like 1-Methyl-3-isobutylxanthine, arsenic trichloride, Arsenic Trioxide, bisphenol A, bisphenol S, butylbenzyl phthalate, cordycepin, Dexamethasone, di-n-octyl phthalate, Dibutyl Phthalate, Dietary Fats, diethyl phthalate, Diethylhexyl Phthalate, disodecyl phthalate, diisononyl phthalate, Escin, Indomethacin, mono-(2-ethylhexyl)phthalate, sodium arsenite, streptozocin, theaflavin-3,3'-digallate, and Tretinoin. The 'Updated Diseases' section lists Asthma, Cardiomyopathies, and Chemical and Drug Induced.

图 5-8 药物/疾病搜索界面

的预测结果。Tenvermectin D 和 Versicolin 仅有唯一的关联疾病，并且该关联疾病没有其他相关药物，因此它们的相似药物预测较为困难。Versicolin 用于治疗毛癣菌感染，推荐的最相似的药物 cholera vaccine CVD 103-HgR 用于治疗霍乱，为一种急性细菌性肠炎。Tenvermectin D 用于治疗体外寄生虫感染，推荐的最相似的药物是 1-(5-bromofur-2-yl)-2-bromo-2-nitroethene，用于治疗细菌感染和真菌病。需要提到的是，Tenvermectin D 和 Versicolin 的训练数据基本是缺失的。两者作为真菌与寄生虫感染的治疗药物，互相出现在对方的相似药品推荐列表中。为了更直白地展示相似药物的推荐结果，表5-2将推荐结果翻译为了中文。候选药物 gemeprost 与出血/腹痛/心脏问题相关，候选药物 thiopropazate 与帕金森病/运

© 2002–2012 MDI Biological Laboratory. All rights reserved.
© 2012–2024 NC State University. All rights reserved.
Data updated 2024年3月29日
Revision 1.2294M

选择查询药物:
orbifloxacin

选择查询疾病:
Fleck Retina, Familial Be

查询药物 (ID : Name)	关联疾病 (ID : Name)
13053 : orbifloxacin	18 : Cell Transformation, Neoplastic, 22 : Colonic Neoplasms, 50 : Neoplasms, 92 : Long QT Syndrome, 128 : Atrial Fibrillation, 272 : Muscular Atrophy, 479 : Torsades de Pointes, 647 : Arrhythmias, Cardiac, 4533 : Long Qt Syndrome 2, 4629 : Short QT Syndrome 1,

查询疾病 (ID : Name)	关联药物 (ID : Name)
5261 : Fleck Retina, Familial Benign	812 : 2,2',3',4',5-hexachlorobiphenyl, 1150 : 2,5',2',5'-tetrachlorobiphenyl, 1885 : 3-(4'-hydroxy-3'-adamantylbiphenyl-4-yl)acrylic acid, 2301 : 4,4'-bisphenol F, 3144 : 6F peptide, 3567 : Acetaminophen, 3752 : Aflatoxin B1, 4174 : Ammonium Chloride, 4472 : Arachidonic Acid, 4533 : Arsenic, 4648 : Atrazine, 4873 : benazepril, 4907 : Benzalkonium Compounds, 4941 : Benzo(a)pyrene, 4963 : benzo(a)pyrene, 5205 : Biological Factors, 5276 : bisphenol A, 5642 : Cadmium, 6066 : CDP 52608, 6268 : Choline, 6401 : Clozapine, 6474 : Clotrimazole, 6788 : Cuprozone, 7065 : dexamethrin, 7080 : Decitabine, 7362 : Dibutyl Phthalate, 7427 : Dietary Fats, 7452 : Diethylstilbestrol, 7529 : Dihydrotestosterone, 7652 : Dinoprostone, 7674 : Dioxins, 7836 : Doxorubicin, 7967 : Eicosanoids, 8084 : Environmental Pollutants, 8086 : enzalutamide, 8136 : epoxiconazole, 8221 : Estradiol, 8230 : estradiol 3-benzoate, 8272 : Ethanol, 8283 : Ethinyl Estradiol, 8800 : Folic Acid, 8812 : Formaldehyde, 9070 : Gentamicins, 9162 : Glucosamine, 9216 : glycidol, 9734 : Hydrogen Peroxide, 10005 : indoxam, 10065 : Iodine, 10196 : Isoliquiritigenin, 10593 : lead acetate, 10644 : Leukotriene C4, 10754 : Lipopolysaccharides, 10848 : LPM4870108, 11269 : Methamphetamine, 11284 : Methapyrene, 11298 : Methionine, 12184 : naphthalene, 12391 : N-Formylmethionine Leucyl-Phenylalanine, 13234 : Palm Oil, 13240 : Pam(3)CSK(4) peptide, 13313 : PCB 180, 13406 : pentanal, 13440 : Peptidoglycan, 13507 : Pesticides, 13832 : Piperonyl Butoxide, 13935 : Polychlorinated Biphenyls, 14160 : Progesterone, 14213 : propionaldehyde, 14243 : Propylthiouraci, 14829 : Rotenone, 14902 : S-(1,2-dichlorovinyl)l-cysteine, 15046 : Sarin, 15063 : SB 203347, 15294 : Silicon Dioxide, 15828 : T-2 Toxin, 16022 : tert-Butylhydroperoxide, 16031 : Testosterone, 16056 : Tetrachlorodibenzodioxin, 16525 : Trichloroethylene, 16599 : trimellitic anhydride, 17300 : Zymosan,

图 5-9 药物/疾病关联查询界面

动障碍相关，它们推荐出的药物同样具有相关的主治疾病。

5.6 本章小结

本章首先探讨了图表示学习在医疗领域的应用，接着对 CTD 数据集中的药物-疾病数据进行了分析，期望利用该数据集完成相似药物推荐的任务。该数据具有极其长尾的特点，并且药物与疾病之间的关系错综复杂。本章利用这些药物-疾病关系构建了一个二部图，并利用第三、四章提出的算法来获取药物节点的专业性表示，最终通过计算余弦相似度得到了相似药物的推荐结果。实验结果显示，推荐结果具有可解释性：推荐出的相似药物具有相同的治疗疾病或类似的疗效。依据以上主要功能，本章开发了一个相似药物推荐系统，方便医疗专业工作者查询药物/疾病属性，并获取相似药物推荐结果。

koniginin A

Rank	Chemical (ID : Name)	Disease (ID : Name)
0	10435 : koniginin A	1 : Adenocarcinoma, 7 : Breast Neoplasms, 11 : Carcinoma, 13 : Carcinoma, Hepatocellular, 16 : Carcinoma, Squamous Cell, 18 : Cell Transformation, Neoplastic, 21 : Cleft Palate, 22 : Colonic Neoplasms, 23 : Colorectal Neoplasms, 25 : Disease Models, Animal,
1	4245 : andrastin A	1 : Adenocarcinoma, 7 : Breast Neoplasms, 11 : Carcinoma, 13 : Carcinoma, Hepatocellular, 16 : Carcinoma, Squamous Cell, 18 : Cell Transformation, Neoplastic, 21 : Cleft Palate, 22 : Colonic Neoplasms, 23 : Colorectal Neoplasms, 25 : Disease Models, Animal,
2	15713 : sulbutamine	1 : Adenocarcinoma, 7 : Breast Neoplasms, 11 : Carcinoma, 13 : Carcinoma, Hepatocellular, 16 : Carcinoma, Squamous Cell, 18 : Cell Transformation, Neoplastic, 21 : Cleft Palate, 22 : Colonic Neoplasms, 23 : Colorectal Neoplasms, 25 : Disease Models, Animal,
3	2803 : 5-(3-(4-(4-benzylpiperazin-1-yl)-N-(4-phenylthiazol-2-yl)propanamide	1 : Adenocarcinoma, 7 : Breast Neoplasms, 11 : Carcinoma, 13 : Carcinoma, Hepatocellular, 16 : Carcinoma, Squamous Cell, 18 : Cell Transformation, Neoplastic, 21 : Cleft Palate, 22 : Colonic Neoplasms, 23 : Colorectal Neoplasms, 25 : Disease Models, Animal,
4	8079 : enterotoxin C, staphylococcal	0 : Precursor Cell Lymphoblastic Leukemia-Lymphoma, 1 : Adenocarcinoma, 6 : Autistic Disorder, 7 : Breast Neoplasms, 11 : Carcinoma, 13 : Carcinoma, Hepatocellular, 16 : Carcinoma, Squamous Cell, 18 : Cell Transformation, Neoplastic, 21 : Cleft Palate, 22 : Colonic Neoplasms, 23 : Colorectal Neoplasms, 25 : Disease Models, Animal,
5	4183 : ammonium perchlorate	1 : Adenocarcinoma, 7 : Breast Neoplasms, 11 : Carcinoma, 13 : Carcinoma, Hepatocellular, 16 : Carcinoma, Squamous Cell, 18 : Cell Transformation, Neoplastic, 21 : Cleft Palate, 22 : Colonic Neoplasms, 23 : Colorectal Neoplasms, 25 : Disease Models, Animal,
6	10705 : Lignin	1 : Adenocarcinoma, 5 : Astrocytoma, 7 : Breast Neoplasms, 11 : Carcinoma, 13 : Carcinoma, Hepatocellular, 16 : Carcinoma, Squamous Cell, 18 : Cell Transformation, Neoplastic, 21 : Cleft Palate, 22 : Colonic Neoplasms, 23 : Colorectal Neoplasms,
7	4094 : Ambroxol	1 : Adenocarcinoma, 7 : Breast Neoplasms, 11 : Carcinoma, 13 : Carcinoma, Hepatocellular, 16 : Carcinoma, Squamous Cell, 18 : Cell Transformation, Neoplastic, 21 : Cleft Palate, 22 : Colonic Neoplasms, 23 : Colorectal Neoplasms, 25 : Disease Models, Animal,
8	5836 : carboprostacyclin	1 : Adenocarcinoma, 6 : Autistic Disorder, 7 : Breast Neoplasms, 11 : Carcinoma, 13 : Carcinoma, Hepatocellular, 15 : Carcinoma, Non-Small-Cell Lung, 16 : Carcinoma, Squamous Cell, 18 : Cell Transformation, Neoplastic, 21 : Cleft Palate, 22 : Colonic Neoplasms,
9	10018 : Inorganic Chemicals	1 : Adenocarcinoma, 6 : Autistic Disorder, 7 : Breast Neoplasms, 11 : Carcinoma, 13 : Carcinoma, Hepatocellular, 16 : Carcinoma, Squamous Cell, 18 : Cell Transformation, Neoplastic, 21 : Cleft Palate, 22 : Colonic Neoplasms, 23 : Colorectal Neoplasms,
10	5647 : cadmium oxide	1 : Adenocarcinoma, 7 : Breast Neoplasms, 13 : Carcinoma, Hepatocellular, 22 : Colonic Neoplasms, 23 : Colorectal Neoplasms, 25 : Disease Models, Animal, 26 : Disease Progression, 29 : Fatty Liver, 33 : Hypertension, 38 : Insulin Resistance,

polychlorodibenzofuran

Rank	Chemical (ID : Name)	Disease (ID : Name)
0	13937 : polychlorodibenzofuran	2 : Adenocarcinoma of Lung, 7 : Breast Neoplasms, 25 : Disease Models, Animal, 33 : Hypertension, 34 : Hypertrophy, Left Ventricular, 36 : Infertility, Female, 37 : Infertility, Male, 38 : Insulin Resistance, 42 : Liver Neoplasms, 48 : Neoplasm Metastasis,
1	1514 : 2-iodobenzaniline	2 : Adenocarcinoma of Lung, 7 : Breast Neoplasms, 25 : Disease Models, Animal, 33 : Hypertension, 34 : Hypertrophy, Left Ventricular, 36 : Infertility, Female, 37 : Infertility, Male, 38 : Insulin Resistance, 42 : Liver Neoplasms, 48 : Neoplasm Metastasis,
2	14259 : prostaglandin F3alpha	2 : Adenocarcinoma of Lung, 7 : Breast Neoplasms, 25 : Disease Models, Animal, 33 : Hypertension, 34 : Hypertrophy, Left Ventricular, 36 : Infertility, Female, 37 : Infertility, Male, 38 : Insulin Resistance, 42 : Liver Neoplasms, 48 : Neoplasm Metastasis,
3	1922 : 3,5-dichlorobenzidine	2 : Adenocarcinoma of Lung, 7 : Breast Neoplasms, 25 : Disease Models, Animal, 33 : Hypertension, 34 : Hypertrophy, Left Ventricular, 36 : Infertility, Female, 37 : Infertility, Male, 38 : Insulin Resistance, 42 : Liver Neoplasms, 48 : Neoplasm Metastasis,
4	523 : 1,8-diaminonaphthalene	2 : Adenocarcinoma of Lung, 7 : Breast Neoplasms, 25 : Disease Models, Animal, 33 : Hypertension, 34 : Hypertrophy, Left Ventricular, 36 : Infertility, Female, 37 : Infertility, Male, 38 : Insulin Resistance, 42 : Liver Neoplasms, 48 : Neoplasm Metastasis,
5	1934 : 3',5'-dimethoxyflavone	2 : Adenocarcinoma of Lung, 7 : Breast Neoplasms, 25 : Disease Models, Animal, 33 : Hypertension, 34 : Hypertrophy, Left Ventricular, 36 : Infertility, Female, 37 : Infertility, Male, 38 : Insulin Resistance, 42 : Liver Neoplasms, 48 : Neoplasm Metastasis,
6	16050 : tetrabromophthalic anhydride	2 : Adenocarcinoma of Lung, 7 : Breast Neoplasms, 25 : Disease Models, Animal, 33 : Hypertension, 34 : Hypertrophy, Left Ventricular, 36 : Infertility, Female, 37 : Infertility, Male, 38 : Insulin Resistance, 42 : Liver Neoplasms, 48 : Neoplasm Metastasis,
7	1276 : 2-amino-1-phenylethanol	2 : Adenocarcinoma of Lung, 7 : Breast Neoplasms, 25 : Disease Models, Animal, 33 : Hypertension, 34 : Hypertrophy, Left Ventricular, 36 : Infertility, Female, 37 : Infertility, Male, 38 : Insulin Resistance, 42 : Liver Neoplasms, 48 : Neoplasm Metastasis,
8	690 : 1-nitrobenzo(a)pyrene	2 : Adenocarcinoma of Lung, 7 : Breast Neoplasms, 25 : Disease Models, Animal, 33 : Hypertension, 34 : Hypertrophy, Left Ventricular, 36 : Infertility, Female, 37 : Infertility, Male, 38 : Insulin Resistance, 42 : Liver Neoplasms, 48 : Neoplasm Metastasis,
9	153 : 1,2,7,8-tetrachlorodibenzo-p-dioxin	2 : Adenocarcinoma of Lung, 7 : Breast Neoplasms, 25 : Disease Models, Animal, 33 : Hypertension, 34 : Hypertrophy, Left Ventricular, 36 : Infertility, Female, 37 : Infertility, Male, 38 : Insulin Resistance, 42 : Liver Neoplasms, 48 : Neoplasm Metastasis,
10	5030 : benzyldiazine	2 : Adenocarcinoma of Lung, 7 : Breast Neoplasms, 25 : Disease Models, Animal, 33 : Hypertension, 34 : Hypertrophy, Left Ventricular, 36 : Infertility, Female, 37 : Infertility, Male, 38 : Insulin Resistance, 42 : Liver Neoplasms, 48 : Neoplasm Metastasis,

-- 选择查询药物 --

Rank	Chemical (ID : Name)	Disease (ID : Name)
0	15978 : tenvermectin D	7222 : Ectoparasitic Infestations,
1	430 : 1-(5-bromofur-2-yl)-2-bromo-2-nitroethene	4759 : Bacterial Infections and Mycoses,
2	6238 : cholera vaccine CVD 103-HgR	6590 : Cholera,
3	16946 : versicolin	7253 : Trichophyton infection,
4	11729 : morinidazole	5313 : Bacterial Infections, 6290 : Amebiasis, 7067 : Trichomonas Infections,
5	4531 : Arsenamide	5820 : Parasitic Diseases, 6290 : Amebiasis,
6	4378 : antimacrophage globulin	6261 : Rheumatic Fever,
7	9323 : GS-441524 triphosphate	6915 : Coronaviridae Infections,
8	15053 : satranidazole	6576 : Liver Abscess, Amebic,
9	12896 : octadecyl-(2-(N-methylpiperidino)ethyl)phosphate	7022 : Trypanosomiasis, African,
10	11542 : milbemycin oxime	6781 : Helminthiasis, Animal, 6823 : Dirofilariasis,

-- 选择查询药物 --

Rank	Chemical (ID : Name)	Disease (ID : Name)
0	16946 : versicolin	7253 : Trichophyton infection,
1	6238 : cholera vaccine CVD 103-HgR	6590 : Cholera,
2	15978 : tenvermectin D	7222 : Ectoparasitic Infestations,
3	430 : 1-(5-bromofur-2-yl)-2-bromo-2-nitroethene	4759 : Bacterial Infections and Mycoses,
4	4531 : Arsenamide	5820 : Parasitic Diseases, 6290 : Amebiasis,
5	4378 : antimacrophage globulin	6261 : Rheumatic Fever,
6	4558 : Arthotec	6305 : Autonomic Dysreflexia,
7	12896 : octadecyl-(2-(N-methylpiperidino)ethyl)phosphate	7022 : Trypanosomiasis, African,
8	11542 : milbemycin oxime	6781 : Helminthiasis, Animal, 6823 : Dirofilariasis,
9	11729 : morinidazole	5313 : Bacterial Infections, 6290 : Amebiasis, 7067 : Trichomonas Infections,
10	6154 : chlorine dioxide	6103 : Hepatitis A, 6378 : Swine Diseases, 6571 : Rotavirus Infections,

图 5-10 相似药物推荐页面，排序为 0 的药物为查询药物

表 5-2 相似药物 Top 10 推荐结果 (中文)

	Chemical	Diseases
候选药物	gemeprost	86 出血, 99 癫痫发作, 129 心血管疾病, 154 呕吐, 190 心肌梗塞, 305 中风, 461 心动过缓, 642 心绞痛, 934 休克, 心源性, 1206 妊娠并发症, 1221 腹泻, 1518 胎儿死亡, 2421 昏迷, 4796 子宫出血, 5191 腹痛
Top 1	isosorbide-5-mononitrate	33 高血压, 86 出血, 104 物质戒断综合征, 119 心肌缺血, 124 急性肾损伤, 190 心肌梗塞, 219 致癌, 251 高血压, 门静脉, 397 低血压, 461 心动过缓, 642 心绞痛, 726 肺栓塞, 939 心动过速, 室性, 1100 头痛, 1221 腹泻, 1361 无先兆偏头痛, 1555 无脑畸形, 1615 食管胃静脉曲张, 5191 腹痛, 5395 垂体中风
Top2	Enoximone	119 心肌缺血, 143 恶心, 177 心力衰竭, 181 缺血, 190 心肌梗塞, 231 冠状动脉疾病, 277 疼痛, 383 癫痫, 强直阵挛, 484 心肌病, 扩张, 567 心脏骤停, 642 心绞痛, 647 心律失常, 心脏, 956 潮红, 1221 腹泻, 1358 心输出量低
Top3	Ovosiston	422 麻痹, 664 脑血管疾病, 1100 头痛, 1200 肠系膜血管阻塞
Top4	ethylenebis(isothiocyanate)	1206 妊娠并发症, 1518 胎儿死亡
Top5	flosequinan	397 低血压, 1100 头痛
Top6	Methylegonovine	33 高血压, 93 神经系统疾病, 119 心肌缺血, 131 药物相关副作用和不良反应, 149 中毒, 177 心力衰竭, 178 心脏瓣膜疾病, 190 心肌梗塞, 223 心脏肥大, 231 冠状动脉疾病
Top7	batanopride	154 呕吐, 397 低血压, 1221 腹泻
Top8	6-(1-(4-fluoro-2-methylphenyl)-3-(trifluoromethyl)-1H-pyrazol-5-yl)-2H-1,4-benzoxazin-3(4H)-one	33 高血压, 80 子宫内异位症, 177 心力衰竭, 190 心肌梗塞, 484 扩张型心肌病, 519 心室功能不全, 左, 1584 假性醛固酮增多症, 3363 高血压, 早发型, 常染色体显性遗传, 严重恶化怀孕
Top9	Ioxaglic Acid	13 癌, 肝细胞, 29 脂肪肝, 73 脑损伤, 86 出血, 89 梗塞, 大脑中动脉, 92 长 QT 综合征, 96 肥胖, 124 急性肾损伤, 127 关节炎, 类风湿, 143 恶心
Top10	NSP 805	152 血栓形成, 177 心力衰竭, 190 心肌梗塞, 305 中风, 4864 短指合并高血压
候选药物	thiopropazate	147 帕金森病, 继发性, 508 帕金森病, 833 运动障碍, 药物诱发, 4875 舞蹈症, 5927 神经症
Top1	L-3-(3-hydroxy-4-pivaloyloxyphenyl)alanine	508 帕金森病, 833 运动障碍, 药物诱发
Top2	talampanel	99 癫痫发作, 508 帕金森病, 833 药物引起的运动障碍
Top3	valbenazine	88 运动过度, 93 神经系统疾病, 146 帕金森病, 147 继发性帕金森病, 继发性, 192 神经退行性变, 508 帕金森病, 1700 抽动秽语综合征, 6493 迟发性运动障碍
Top4	2,5-di-(2-phenethyl)pyrrolidine	66 苯丙胺相关疾病, 93 神经系统疾病, 146 帕金森病, 147 继发性帕金森病, 192 神经退行性变, 508 帕金森病
Top5	U 74500A	833 药物引起的运动障碍
Top6	Purinergic P1 Receptor Antagonists	833 药物引起的运动障碍, 1219 僵住症
Top7	BN82451	833 药物引起的运动障碍
Top8	seganserin	833 药物引起的运动障碍
Top9	levosulpiride	147 继发性帕金森病, 154 呕吐, 380 抑郁症, 437 震颤, 508 帕金森病, 833 药物引起的运动障碍, 6029 消化不良
Top10	deutetabenazine	88 运动过度, 93 神经系统疾病, 146 帕金森病, 147 继发性帕金森病, 192 神经退行性, 508 帕金森病, 1700 抽动秽语综合征

第六章 总结与展望

现实世界中存在大量的非结构化数据，例如社交网络、交通网络、生物分子网络等，这些数据具有复杂的拓扑结构和丰富的特征信息。传统的神经网络模型在处理这种非结构化数据时存在局限性，无法充分捕捉到数据之间的复杂关系和拓扑结构。因此，为了更有效地处理这类数据，图神经网络应运而生。图神经网络的主要研究方向包括：图神经网络模型、图表示学习以及图表示学习的应用。本文依据这三个主要方向出发，分别探索了频域图卷积网络模型、自动数据增强的无监督图对比学习方法以及图表示学习在相似药物推荐上的应用。

- 首先，频域图卷积网络中常用的低通滤波器存在一定的局限性。本文通过 3 种方法：叠加图卷积核幂次、使用广义拉普拉斯卷积核以及引入跨跳连接图卷积层，来实验验证不同的数据集需要不同的高低频信息配比。本文通过自适应学习跨跳连接及单跳连接图卷积核的系数来控制图网络引入的高低频信息比例。其次，对于图卷积网络而言，过平滑现象是另一个挑战。为了缓解这一现象，本文总结了 6 种层间信息交互方式，通过增加节点之间的区分度来缓解过平滑现象。最后，本文探究了如何简化图卷积核的信息传播方式，以减少模型的参数量并缓解过拟合现象。基于以上三点提出的跨跳连接的简化图卷积网络（SGCN-MH）在浅层和低参数量的情况下击败了深层的图卷积网络，具有较高的性能和效率。

现实场景中图的规模可达到亿级别，空域图卷积网络由于采样操作以及局部卷积方案可以应用于大图，而频域图卷积网络依赖全局邻接矩阵进行卷积操作。未来的研究可以探索在频域图卷积网络中引入采样操作和图切割方案，从而扩展其在大规模图数据中的应用。

- 对比学习中的数据增强方案至关重要，而其的选择与下游任务严格相关。InfoMin 原则中定义对比学习中的最优视角集合为共享最小信息且保留了下游任务所必须信息的视角。无监督设置下与下游任务相关的信息量无法

度量，因此造成了众多无监督图对比学习中经过数据增强得到的视角包含了过多冗余信息或是缺失了关键信息。于是，本文提出了一种基于最小显著信息的自动数据增强的无监督图对比学习方法（GMNI）。GMNI 使用对抗性的训练策略，它通过最小化优化去除了视角中过多的冗余信息，通过最大化优化保留了视角中的重要信息，这种基于最小显著信息的视角近似了 InfoMin 中的最优视角。在 5 个节点分类数据集和 9 个图分类数据集上的无监督与半监督实验验证了 GMNI 中的视角的优越性，它超越了其他的自动和手动数据增强的图对比学习方法。

在未来，本研究将进一步验证 GMNI 中这种近似最优视角在更多下游任务上的可行性，以及显著信息与下游任务必要信息的强相关性。此外，鉴于图对比框架的多样性，未来的研究还将探讨 GMNI 在不同图对比框架上的泛化性能。

- 除了理论研究外，本文将上述图卷积网络及图对比学习方法运用于医疗领域。本文根据医疗病理数据集从图构建开始完成对药物节点嵌入的学习，并将其用于相似药物推荐任务。系统提供了药物展示，药物/疾病查询及相似药物推荐功能，获得的相似药物推荐结果具有较好的可解释性与鲁棒性。与普通的直接推荐同药物不同品牌或同品牌不同规格药物相比，本相似药物推荐系统为药物重定位等提供了更多可能性。

除了药物与疾病的关联关系外，基因信息和药物分子结构等信息也至关重要。未来的研究计划是构建更加复杂的图，以整合上述信息从而获取更专业的药物嵌入。此外，除了相似药物推荐外，组合药物推荐也非常重要。组合药物指的是搭配使用的药物，未来的研究将尝试将图表示学习应用于组合药物推荐。

参考文献

- [1] HANNUN A, CASE C, CASPER J, et al. Deep speech: Scaling up end-to-end speech recognition[J]. ArXiv preprint arXiv:1412.5567, 2014.
- [2] CHAN W, JAITLY N, LE Q, et al. Listen, attend and spell: A neural network for large vocabulary conversational speech recognition[C]//2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2016: 4960-4964.
- [3] OORD A V D, DIELEMAN S, ZEN H, et al. WaveNet: A generative model for raw audio[J]. ArXiv preprint arXiv:1609.03499, 2016.
- [4] SUTSKEVER I, VINYALS O, LE Q V. Sequence to sequence learning with neural networks[C]//Advances in Neural Information Processing Systems: vol. 27. 2014.
- [5] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[Z]. 2023. ArXiv preprint arXiv: 1706.03762.
- [6] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding[Z]. 2019. ArXiv preprint arXiv: 1810.04805.
- [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]//Advances in Neural Information Processing Systems: vol. 25. 2012.
- [8] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.

- [9] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. ArXiv preprint arXiv:2010.11929, 2020.
- [10] KIPF T N, WELING M. Semi-supervised classification with graph convolutional networks[J]. ArXiv preprint arXiv:1609.02907, 2016.
- [11] VELIČKOVIĆ P, CUCURULL G, CASANOVA A, et al. Graph attention networks[Z]. 2018. ArXiv preprint arXiv: 1710.10903.
- [12] HAMILTON W, YING Z, LESKOVEC J. Inductive representation learning on large graphs[C] // Advances in Neural Information Processing Systems: vol. 30. 2017.
- [13] XU K, HU W, LESKOVEC J, et al. How powerful are graph neural networks? [J]. ArXiv preprint arXiv:1810.00826, 2018.
- [14] YU B, YIN H, ZHU Z. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting[J]. ArXiv preprint arXiv:1709.04875, 2017.
- [15] HE X, DENG K, WANG X, et al. LightGCN: Simplifying and powering graph convolution network for recommendation[Z]. 2020. ArXiv preprint arXiv: 2002.02126.
- [16] WANG X, HE X, WANG M, et al. Neural graph collaborative filtering[C] // Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2019: 165-174.
- [17] CAO S, LU W, XU Q. Deep neural networks for learning graph representations [C] // Proceedings of the AAAI Conference on Artificial Intelligence: vol. 30: 1. 2016.
- [18] ZHANG M, CUI Z, NEUMANN M, et al. An end-to-end deep learning architecture for graph classification[C] // Proceedings of the AAAI Conference on Artificial Intelligence: vol. 32: 1. 2018.

- [19] WANG Z, ZHANG J, FENG J, et al. Knowledge graph embedding by translating on hyperplanes[C] // Proceedings of the AAAI Conference on Artificial Intelligence: vol. 28: 1. 2014.
- [20] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2012, 60: 84-90.
- [21] ZAREMBA W, SUTSKEVER I, VINYALS O. Recurrent neural network regularization[Z]. 2015. ArXiv preprint arXiv: 1409.2329.
- [22] LI Y, TARLOW D, BROCKSCHMIDT M, et al. Gated graph sequence neural networks[Z]. 2017. ArXiv preprint arXiv: 1511.05493.
- [23] KIPF T N, WELING M. Variational graph auto-encoders[J]. ArXiv preprint arXiv:1611.07308, 2016.
- [24] DE CAO N, KIPF T. MolGAN: An implicit generative model for small molecular graphs[J]. ArXiv preprint arXiv:1805.11973, 2018.
- [25] WANG X, JI H, SHI C, et al. Heterogeneous graph attention network[C] // The World Wide Web Conference. 2019: 2022-2032.
- [26] MA Y, WANG S, AGGARWAL C C, et al. Multi-dimensional graph convolutional networks[C] // Proceedings of the 2019 Siam International Conference on Data Mining. 2019: 657-665.
- [27] FENG Y, YOU H, ZHANG Z, et al. Hypergraph neural networks[C] // Proceedings of the AAAI Conference on Artificial Intelligence: vol. 33: 01. 2019: 3558-3565.
- [28] MANESSI F, ROZZA A, MANZO M. Dynamic graph convolutional networks [J]. Pattern Recognition, 2020, 97: 107000.
- [29] AHMED A, SHERVASHIDZE N, NARAYANAMURTHY S, et al. Distributed large-scale natural graph factorization[C] // Proceedings of the 22nd International Conference on World Wide Web. 2013: 37-48.

- [30] CAO S, LU W, XU Q. GraRep: Learning graph representations with global structural information[C]//Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. 2015: 891-900.
- [31] OU M, CUI P, PEI J, et al. Asymmetric transitivity preserving graph embedding[C]//Proceedings of the 22th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 1105-1114.
- [32] PEROZZI B, AL-RFOU R, SKIENA S. DeepWalk: Online learning of social representations[C]//Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014: 701-710.
- [33] GROVER A, LESKOVEC J. Node2vec: Scalable feature learning for networks [C]//Proceedings of the 22th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 855-864.
- [34] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space[Z]. 2013. ArXiv preprint arXiv: 1301.3781.
- [35] GUTHRIE D, ALLISON B, LIU W, et al. A closer look at skip-gram modelling [C]//Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC' 06). 2006.
- [36] TANG J, QU M, WANG M, et al. Line: Large-scale information network embedding[C]//Proceedings of the 24th International Conference on World Wide Web. 2015: 1067-1077.
- [37] ZHOU Z, HU Y, ZHANG Y, et al. Multiview deep graph InfoMax to achieve unsupervised graph embedding[J]. IEEE Transactions on Cybernetics, 2022.
- [38] OYALLON E. Interferometric graph transform: A deep unsupervised graph representation[C]//International Conference on Machine Learning. 2020: 7434-7444.
- [39] WANG D, CUI P, ZHU W. Structural deep network embedding[C]//Proceedings of the 22th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016: 1225-1234.

- [40] VELICKOVIC P, FEDUS W, HAMILTON W L, et al. Deep graph infomax.[J]. International Conference on Learning Representations (poster), 2019, 2(3): 4.
- [41] YOU Y, CHEN T, SUI Y, et al. Graph contrastive learning with augmentations[C]// Advances in Neural Information Processing Systems: vol. 33. 2020: 5812-5823.
- [42] HASSANI K, KHASAHMADI A H. Contrastive multi-view representation learning on graphs[C]// International Conference on Machine Learning. 2020: 4116-4126.
- [43] LI W, DUAN Y, LU J, et al. Graph-based social relation reasoning[C]// European Conference on Computer Vision. 2020: 18-34.
- [44] WANG H, ZHAO M, XIE X, et al. Knowledge graph convolutional networks for recommender systems[C]// The World Wide Web Conference. 2019: 3307-3313.
- [45] 郭又铭. 基于图表示学习的下一个兴趣点推荐算法研究[D]. 电子科技大学, 2023.
- [46] ZHOU J, TROYANSKAYA O G. Predicting effects of noncoding variants with deep learning-based sequence model[J]. Nature Methods, 2015, 12(10): 931-934.
- [47] 张宇轩. 基于异构图表示学习的药物—靶蛋白关联关系预测研究[D]. 东北师范大学, 2023.
- [48] CHENG D, WANG X, ZHANG Y, et al. Graph neural network for fraud detection via spatial-temporal attention[J]. IEEE Transactions on Knowledge and Data Engineering, 2020, 34(8): 3800-3813.
- [49] 高桢. 基于图表示学习的虚拟货币犯罪账户识别研究[D]. 华东师范大学, 2024.
- [50] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [Z]. 2015. ArXiv preprint arXiv: 1512.03385.
- [51] GOULD R. Graph theory[M]. Courier Corporation, 2012.

- [52] HECHTLINGER Y, CHAKRAVARTI P, QIN J. A generalization of convolutional neural networks to graph-structured data[J]. ArXiv preprint arXiv:1704.08165, 2017.
- [53] BO D, WANG X, SHI C, et al. Beyond low-frequency information in graph convolutional networks[C]//Proceedings of the AAAI Conference on Artificial Intelligence: vol. 35: 5. 2021: 3950-3957.
- [54] CUI G, ZHOU J, YANG C, et al. Adaptive graph encoder for attributed graph embedding[C]//KDD ' 20: Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2020. DOI: 10.1145/3394486.3403140.
- [55] WU F, ZHANG T, de SOUZA JR. AU2 A H, et al. Simplifying graph convolutional networks[Z]. 2019. ArXiv preprint arXiv: 1902.07153.
- [56] WELLING M, KIPF T N. Semi-supervised classification with graph convolutional networks[C]//International Conference on Learning Representations (ICLR 2017). 2016.
- [57] YOU Y, CHEN T, SHEN Y, et al. Graph contrastive learning automated[C]//International Conference on Machine Learning. 2021: 12121-12132.
- [58] SURESH S, LI P, HAO C, et al. Adversarial graph augmentation to improve graph contrastive learning[C]//Advances in Neural Information Processing Systems: vol. 34. 2021: 15920-15933.
- [59] YIN Y, WANG Q, HUANG S, et al. AutoGCL: Automated graph contrastive learning via learnable view generators[C]//Proceedings of the AAAI Conference on Artificial Intelligence: vol. 36: 8. 2022: 8892-8900.
- [60] LINSKER R. Self-organization in a perceptual network[J]. Computer, 1988, 21(3): 105-117.
- [61] HJELM R D, FEDOROV A, LAVOIE-MARCHILDON S, et al. Learning deep representations by mutual information estimation and maximization[J]. ArXiv preprint arXiv:1808.06670, 2018.

- [62] ZHU Y, XU Y, YU F, et al. Deep graph contrastive representation learning[J]. ArXiv preprint arXiv:2006.04131, 2020.
- [63] THAKOOR S, TALLEC C, AZAR M G, et al. Bootstrapped representation learning on graphs[C]//ICLR 2021 Workshop on Geometrical and Topological Representation Learning. 2021.
- [64] TISHBY N, PEREIRA F C, BIALEK W. The information bottleneck method [J]. arXiv Preprint Physics/0004057, 2000.
- [65] GOLDFELD Z, POLYANSKIY Y. The information bottleneck problem and its applications in machine learning[J]. IEEE Journal on Selected Areas in Information Theory, 2020, 1(1): 19-38.
- [66] ALEMI A A, FISCHER I, DILLON J V, et al. Deep variational information bottleneck[J]. ArXiv preprint arXiv:1612.00410, 2016.
- [67] CHENG W, XU D, CHEN H. Information-aware graph contrastive learning[Z]. US Patent App. 17/728,071. 2022.
- [68] TIAN Y, SUN C, POOLE B, et al. What makes for good views for contrastive learning?[C]//Advances in Neural Information Processing Systems: vol. 33. 2020: 6827-6839.
- [69] OPPENHEIM A V. Discrete-Time signal processing[J]. 1999.
- [70] LI Q, WU X M, LIU H, et al. Label efficient semi-supervised learning via graph filtering[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 9582-9591.
- [71] CHEN L, WU L, HONG R, et al. Revisiting graph based collaborative filtering: A linear residual graph convolutional network approach[C]//Proceedings of the AAAI Conference on Artificial Intelligence: vol. 34: 01. 2020: 27-34.
- [72] LI G, MÜLLER M, THABET A, et al. DeepGCNs: Can GCNs go as deep as CNNs?[Z]. 2019. ArXiv preprint arXiv: 1904.03751.
- [73] RONG Y, HUANG W, XU T, et al. DropEdge: Towards deep graph convolutional networks on node classification[Z]. 2020. ArXiv preprint arXiv: 1907.1

- 0903.
- [74] CHEN M, WEI Z, HUANG Z, et al. Simple and deep graph convolutional networks[Z]. 2020. ArXiv preprint arXiv: 2007.02133.
- [75] BRUNA J, ZAREMBA W, SZLAM A, et al. Spectral networks and locally connected networks on graphs[Z]. 2014. ArXiv preprint arXiv: 1312.6203.
- [76] DEFFERRARD M, BRESSON X, VANDERGHEYNST P. Convolutional neural networks on graphs with fast localized spectral filtering[C]// Advances in Neural Information Processing Systems: vol. 29. 2016.
- [77] MASON J C, HANDSCOMB D C. Chebyshev polynomials[M]. CRC press, 2002.
- [78] NT H, MAEHARA T. Revisiting graph neural networks: All we have is Low-Pass filters[Z]. 2019. ArXiv preprint arXiv: 1905.09550.
- [79] DENG Z H, HUANG L, WANG C D, et al. DeepCF: A unified framework of representation learning and matching function learning in recommender system [Z]. 2019. ArXiv preprint arXiv: 1901.04704.
- [80] RENDLE S, FREUDENTHALER C, GANTNER Z, et al. BPR: Bayesian personalized ranking from implicit feedback[Z]. 2012. ArXiv preprint arXiv: 1205.2618.
- [81] GEHRING D, GRAEPEL T. Large margin rank boundaries for ordinal regression[C]// . 2000.
- [82] PEI Y, HUANG T, van IPENBURG W, et al. ResGCN: Attention-based deep residual modeling for anomaly detection on attributed networks[Z]. 2020. ArXiv preprint arXiv: 2009.14738.
- [83] SEN P, NAMATA G, BILGIC M, et al. Collective classification in network data [J]. AI Magazine, 2008, 29(3): 93.
- [84] ROZEMBERCZKI B, ALLEN C, SARKAR R. Multi-scale attributed node embedding[Z]. 2021. ArXiv preprint arXiv: 1909.13021.

- [85] GASTEIGER J, BOJCHEVSKI A, GÜNNEMANN S. Predict then propagate: Graph neural networks meet personalized PageRank[Z]. 2022. ArXiv preprint arXiv: 1810.05997.
- [86] XU K, LI C, TIAN Y, et al. Representation learning on graphs with jumping knowledge networks[Z]. 2018. ArXiv preprint arXiv: 1806.03536.
- [87] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[Z]. 2015. ArXiv preprint arXiv: 1512.00567.
- [88] SARIGÜN A, RIFAI OGLU A S. Multi-Mask aggregators for graph neural networks[C]//The First Learning on Graphs Conference. 2022.
- [89] HOANG V T, LEE O J. Mitigating degree biases in message passing mechanism by utilizing community structures[Z]. 2023. ArXiv preprint arXiv: 2312.16788.
- [90] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. ArXiv preprint arXiv:1412.6980, 2014.
- [91] PEI H, WEI B, CHANG K C C, et al. Geom-GCN: Geometric graph convolutional networks[Z]. 2020. ArXiv preprint arXiv: 2002.05287.
- [92] JU W, GU Y, LUO X, et al. Unsupervised graph-level representation learning with hierarchical contrasts[J]. Neural Networks, 2023, 158: 359-368.
- [93] LIU Y, ZHENG Y, ZHANG D, et al. Beyond smoothing: Unsupervised graph representation learning with edge heterophily discriminating[C]//Proceedings of the AAAI Conference on Artificial Intelligence: vol. 37: 4. 2023: 4516-4524.
- [94] MO Y, PENG L, XU J, et al. Simple unsupervised graph representation learning [C]//Proceedings of the AAAI Conference on Artificial Intelligence: vol. 36: 7. 2022: 7797-7805.
- [95] XIAO S, WANG S, DAI Y, et al. Graph neural networks in node classification: Survey and evaluation[J]. Machine Vision and Applications, 2022, 33: 1-19.
- [96] SUN F Y, HOFFMANN J, VERMA V, et al. InfoGraph: Unsupervised and semi-supervised graph-level representation learning via mutual information maximization[J]. ArXiv preprint arXiv:1908.01000, 2019.

- [97] PAN E, KANG Z. Multi-view contrastive graph clustering[C]//Advances in Neural Information Processing Systems: vol. 34. 2021: 2148-2159.
- [98] CHEN J, KOU G. Attribute and structure preserving graph contrastive learning [C]//Proceedings of the AAAI Conference on Artificial Intelligence: vol. 37: 6. 2023: 7024-7032.
- [99] GONG X, YANG C, SHI C. MA-GCL: Model augmentation tricks for graph contrastive learning[C]//Proceedings of the AAAI Conference on Artificial Intelligence: vol. 37: 4. 2023: 4284-4292.
- [100] GUTMANN M, HYVÄRINEN A. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models[C]//Proceedings of the 13th International Conference on Artificial Intelligence and Statistics. 2010: 297-304.
- [101] Van den OORD A, LI Y, VINYALS O. Representation learning with contrastive predictive coding[Z]. 2019. ArXiv preprint arXiv: 1807.03748.
- [102] NOWOZIN S, CSEKE B, TOMIOKA R. F-GAN: Training generative neural samplers using variational divergence minimization[C]//Advances in Neural Information Processing Systems: vol. 29. 2016.
- [103] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//International Conference on Machine Learning. 2020: 1597-1607.
- [104] DING K, XU Z, TONG H, et al. Data augmentation for deep graph learning: A survey[J]. ACM SIGKDD Explorations Newsletter, 2022, 24(2): 61-77.
- [105] MISHRA P, PIKTUS A, GOOSSEN G, et al. Node masking: Making graph neural networks generalize and scale better[J]. Computing Research Repository, 2020, abs/2001.07524. arXiv: 2001.07524.
- [106] ZHU Y, XU Y, YU F, et al. Graph contrastive learning with adaptive augmentation[C]//Proceedings of the Web Conference 2021. 2021: 2069-2080.

- [107] PAGE L, BRIN S, MOTWANI R, et al. The PageRank citation ranking: Bringing order to the web.[R]. Stanford InfoLab, 1999.
- [108] KONDOR R I, LAFFERTY J. Diffusion kernels on graphs and other discrete structures[C]//Proceedings of the 19th International Conference on Machine Learning: vol. 2002. 2002: 315-322.
- [109] FENG F, HE X, TANG J, et al. Graph adversarial training: Dynamically regularizing based on graph structure[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 33(6): 2493-2504.
- [110] YANG L, ZHANG L, YANG W. Graph adversarial self-supervised learning[C]//Advances in Neural Information Processing Systems: vol. 34. 2021: 14887-14899.
- [111] YOU Y, CHEN T, WANG Z, et al. When does self-supervision help graph convolutional networks?[C]//International Conference on Machine Learning. 2020: 10871-10880.
- [112] PARK J, SHIM H, YANG E. Graph transplant: Node saliency-guided graph mixup with local structure preservation[C]//Proceedings of the AAAI Conference on Artificial Intelligence: vol. 36: 7. 2022: 7966-7974.
- [113] LAWLER G F, LIMIC V. Random walk: A modern introduction: vol. 123[M]. Cambridge University Press, 2010.
- [114] PAGE L, BRIN S, MOTWANI R, et al. The PageRank citation ranking: Bringing order to the web[R]. Stanford Digital Library Technologies Project, 1998.
- [115] WEISSTEIN E W. Floyd-warshall algorithm[J]. [https://mathworld. Wolfram. Com/](https://mathworld.wolfram.com/), 2008.
- [116] JIN M, ZHENG Y, LI Y F, et al. Multi-scale contrastive siamese networks for self-supervised graph representation learning[J]. ArXiv preprint arXiv:2105.05682, 2021.
- [117] TSCHANNEN M, DJOLONGA J, RUBENSTEIN P K, et al. On mutual information maximization for representation learning[J]. ArXiv preprint

- arXiv:1907.13625, 2019.
- [118] YANG K, SUN Y, SU J, et al. Adversarial Auto-Augment with label preservation: A representation learning principle guided approach[C] // Advances in Neural Information Processing Systems: vol. 35. 2022: 22035-22048.
- [119] KARYPIS G, KUMAR V. A fast and high quality multilevel scheme for partitioning irregular graphs[J]. SIAM Journal on Scientific Computing, 1998, 20(1): 359-392.
- [120] NAZI A, HANG W, GOLDIE A, et al. Gap: Generalizable approximate graph partitioning framework[J]. ArXiv preprint arXiv:1903.00614, 2019.
- [121] HAN X, JIANG Z, LIU N, et al. G-mixup: Graph data augmentation for graph classification[C] // International Conference on Machine Learning. 2022: 8230-8248.
- [122] ZHANG M, CHEN Y. Link prediction based on graph neural networks[C] // Advances in Neural Information Processing Systems: vol. 31. 2018.
- [123] YANG S, HU B, ZHANG Z, et al. Inductive link prediction with interactive structure learning on attributed graph[C] // Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part II 21. 2021: 383-398.
- [124] MADDISON C J, MNIH A, TEH Y W. The concrete distribution: A continuous relaxation of discrete random variables[J]. ArXiv preprint arXiv:1611.00712, 2016.
- [125] OORD A V D, LI Y, VINYALS O. Representation learning with contrastive predictive coding[J]. ArXiv preprint arXiv:1807.03748, 2018.
- [126] MERNYEI P, CANGEA C. Wiki-cs: A wikipedia-based benchmark for graph neural networks[J]. ArXiv preprint arXiv:2007.02901, 2020.
- [127] SHCHUR O, MUMME M, BOJCHEVSKI A, et al. Pitfalls of graph neural network evaluation[J]. ArXiv preprint arXiv:1811.05868, 2018.

- [128] MORRIS C, KRIEGE N M, BAUSE F, et al. Tudataset: A collection of benchmark datasets for learning with graphs[J]. ArXiv preprint arXiv:2007.08663, 2020.
- [129] PENG Z, HUANG W, LUO M, et al. Graph representation learning via graphical mutual information maximization[C] // Proceedings of the Web Conference 2020. 2020: 259-270.
- [130] GLOROT X, BENGIO Y. Understanding the difficulty of training deep feedforward neural networks[C] // Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. 2010: 249-256.
- [131] CHEN T, BIAN S, SUN Y. Are powerful graph neural nets necessary? A dissection on graph classification[J]. ArXiv preprint arXiv:1905.04579, 2019.
- [132] ABDI H, WILLIAMS L J. Principal component analysis[J]. Wiley Interdisciplinary Reviews: Computational Statistics, 2010, 2(4): 433-459.
- [133] VAN DER MAATEN L, HINTON G. Visualizing data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(11).

致 谢

三年匆匆而过，转眼间就要到达研究生学业的终点。此刻，我心怀无限感激之情，想要向那些曾在我学术道路上悉心引导、无私支持的前辈们、朋友们、亲人们表达最深沉的谢意。是你们的鼓励与指引，让我得以勇往直前，克服重重困难，踏上追寻知识的旅程。

首先，我由衷感谢我的导师，申富饶教授。在我整个研究生生涯中，申老师如同一盏明灯，照亮我前行的道路。他不仅在学术上给予我无微不至的指导和支持，更是身体力行地引领着我探索知识的海洋。申老师不仅仅是一位导师，更是一位贴心的指引者，他每周精心安排的个人讨论让我不仅能够认清当前的学术问题，更激发了我对未来研究的热情。他的言传身教不仅仅在学术上使我有所成长，更在为人处事的方方面面塑造了我。申老师的悉心教导将永远激励着我，成为我不断前行的动力源泉。

我也要衷心感谢 RINC 实验室的所有同仁们，是你们的合作与支持让我的研究之旅更加丰富多彩、顺利而愉悦。不论是在组会的思绪碰撞，还是在共同上课的相伴，我们共同经历的这段青春岁月，已经被镌刻在我心中，成为我珍贵的宝藏，为我今后的人生之路增添了浓墨重彩的一笔。

我要特别感谢我的家人、室友以及朋友们，是他们给予了我无条件的支持和鼓励，让我在学业和生活中不再感到孤独。他们的陪伴和理解如同一缕暖阳，温暖着我的心灵，给予我无尽的勇气和力量。在他们的关怀和支持下，我始终坚定地向前迈进，不畏挑战，不懈努力，因为有他们的存在，我知道无论何时何地，我都不会独行。

简历与科研成果

基本信息

熊昕，女，苗族，1999年10月出生，湖南省怀化人。

教育背景

2021年9月—2024年6月 南京大学人工智能学院 硕士

2017年9月—2021年6月 山东大学机电与信息工程学院 本科

攻读硕士学位期间完成的学术成果

- Xin Xiong**, Xunkai Li, Youpeng Hu, Yixuan Wu, Jian Yin. Handling information loss of graph convolutional networks in collaborative filtering[J]. Information Systems, 2022, 109: 102051.
- Xin Xiong**, Furao Shen, Xiangyu Wang, Jian Zhao. GIMM: InfoMin-Max for automated graph contrastive learning[J]. ArXiv preprint arXiv:2305.17437, 2023.

攻读硕士学位期间参与的科研课题

- 科技部重大项目“基于神经可塑性的脉冲网络高效学习机制与类脑智能系统”(项目编号2021ZD0201300, 课题年限2021年9月—2024年6月)。
- 国家自然科学基金面上项目“面向增量式无监督学习的新型神经网络研究”(项目编号62276127, 课题年限2023年1月—2024年4月)。

