

们不能有效地推断不同细胞类型的活性生物网络，同时进行细胞聚类，并且阐明这些复杂网络对特定细胞类型中外部刺激的反应的能力有限。

最近，图神经网络（GNN）通过在全局细胞图 [11, 13] 中传播邻居细胞特征和构建细胞关系，在学习单个细胞的低维表示方面表现出了优势。例如，我们的内部工具 scGNN，一个 GNN，基于大规模 scRNA-seq 数据 [16]，该模型已经证明了优越的细胞聚类和基因插补性能。此外，具有不同类型节点和边的异构图已被广泛用于对多关系知识图进行建模。它为整合 scMulti-omics 数据和学习潜在的细胞类型特异性生物网络提供了一个自然的表示框架。此外，用于建模和整合异构关系的注意力机制的最新发展使深度学习模型具有可解释性，并使细胞类型特异性生物网络的推理成为可能。

选择复现这篇论文的原因有两点。一是研究方向契合，我的研究方向正是与这篇论文密切相关的领域——细胞测序与异构图网络。论文涉及的主题与我当前的研究兴趣高度一致，这将使得复现工作不仅能够加深我对该领域的理解，也能够对我未来的研究提供宝贵的经验和启示。二是 transformer 应用，transformer 是这篇论文的一个重要框架，而我的研究中同样需要运用这种架构来提取细胞、基因之间的关系。通过复现这篇论文，我将深入学习并掌握论文中采用的提取数据特征的方法，为我的研究提供实用经验和参考。

2 相关工作

2.1 数据预处理和整合方法

2.1.1 数据标准化

数据标准化是预处理过程中的首要步骤，目的是调整不同细胞的表达水平，使之可比较。常用的方法包括对数转换（例如 log-normalization）[5]，它能够减少数据的偏态并稳定方差。在 Seurat 中，NormalizeData 函数通过计算每个细胞的表达量的均值和标准差，然后对每个基因的表达量进行缩放，以实现数据的标准化。

2.1.2 特征选择

特征选择旨在从成千上万的基因中筛选出最具代表性和变异性的特征，这些特征对于后续的分析至关重要。Seurat 中的 FindVariableFeatures 函数可以识别高变异性基因，这些基因在不同细胞间表达差异大，可能是区分不同细胞状态或类型的关键。

2.1.3 数据降维

数据降维技术如主成分分析（PCA）[1] 被用来减少数据的维度，提取最重要的变异来源。这一步骤有助于揭示数据中的潜在结构，并为后续的聚类和可视化打下基础。在 Seurat 中，RunPCA 函数执行 PCA，将高维数据投影到较低维的空间中，同时保留最多的变异信息。

2.1.4 数据整合

对于来自不同实验条件或技术平台的多模态数据，数据整合是将它们映射到一个共同的分析空间中，以便进行比较和联合分析。Seurat 的 IntegrateData 函数使用锚点（anchors）方

法，通过找到不同数据集间的共同特征（即锚点），将它们整合到一个统一的低维空间中，从而允许跨数据集的比较和分析。

2.2 单细胞测序技术

单细胞测序技术通过捕获单个细胞的基因表达信息，揭示细胞间的异质性。这项技术包括多种方法，如单细胞全转录组测序（scRNA-seq）[7] 和单细胞全基因组测序（scWGS）等。它允许科学家在分子层面上精细区分不同细胞类型，并研究它们的分子机制。单细胞测序技术面临的挑战包括技术难度高、数据量大且质量低、噪音高、变异性大。这些挑战需要专业的统计方法和算法来处理 and 解释。此外，单细胞测序的结果需要与生物学背景和问题相结合，需要多学科的交叉和合作，以及更多的验证实验来支持。

2.3 异构图模型

2.3.1 图注意力神经网络（GAT）

在 PINNACLE 模型中，利用图注意力神经网络在一个包含细胞类型、组织及其相互关系的异构图上进行深入学习，以此捕捉和理解不同细胞类型和组织间的复杂生物学关系和层次结构。这种方法实现了具有上下文感知能力的蛋白质表示，为蛋白质相互作用网络提供了一种丰富的、上下文化的 3D 视角。

2.3.2 异构图模型的构建与应用

异构图 [17] 模型通过整合不同类型的节点（如细胞和基因）和边（表示细胞间或基因间的关系）来构建一个多模态的生物网络。在论文中提到的 DeepMAPS 平台就是一个基于异构图 Transformer 框架，用于单细胞多组学数据的生物网络推理。这种模型能够同时识别细胞和基因的联合嵌入，并在完整的框架中推断特定于细胞类型的生物网络以及细胞类型本身。

2.4 嵌入学习

论文中提到的 scMDC 方法是一种端到端的深度模型，它能够显式地表征不同的数据源，并联合学习深度嵌入的潜在特征以进行聚类分析。这种方法通过深度学习框架整合多模态数据，并学习细胞的低维表示，为后续的聚类和网络推断提供了基础。

2.5 细胞聚类和注释

2.5.1 细胞聚类方法

论文中提到的 DeepMAPS 工具利用异构图和多头图变换器（HGT）来学习细胞与基因之间的关系，并进行细胞聚类。这种方法能够在局部和全局上下文中以鲁棒的方式识别细胞间的邻域信息，从而进行有效的细胞聚类。

2.5.2 多模态数据聚类

scMDC 模型是一个端到端的深度模型，它能够明确表征不同的数据源，并联合学习深度嵌入的潜在特征用于聚类分析。scMDC 在不同的单细胞多模态数据集上优于现有的单细胞单

模态和多模态聚类方法，且运行时间的线性可扩展性使得 scMDC 成为分析大规模多模态数据集的有前景的方法。

2.5.3 细胞注释方法

单细胞数据分析中，使用 SingleR 进行细胞注释是一种常见的方法。通过将聚类结果与已知的参考数据集进行比较，可以对未知的细胞群进行注释。这种方法可以帮助研究者识别和理解不同细胞类型的功能和状态。

2.5.4 异构图 Transformer 在细胞聚类中的应用

DeepMAPS 利用异质图 Transformer 对单细胞多组学数据进行建模，这种模型可以捕捉细胞间的复杂关系，并用于推断细胞类型特异性的生物网络。这种方法在细胞聚类和生物网络构建方面的表现优于现有工具。

2.5.5 聚类算法的选择

在单细胞测序中，常用的聚类算法包括 Leiden 和 Louvain 算法 [2]。由于 Louvain 算法不再维护，推荐使用 Leiden 算法进行聚类分析。

3 本文方法

3.1 本文方法概述

本文提出了 DeepMAPS 框架，这是一个深度学习框架，它实现了异构图表示学习和图变换器，专门用于从单细胞多组学数据 (scMulti-omics) 中推断生物网络。DeepMAPS 通过以下几个关键步骤进行网络构建和嵌入学习：

1. 数据预处理：首先对单细胞数据进行质量控制，移除低质量的细胞和低表达的基因，并根据不同的数据类型应用适当的归一化方法，生成一个综合的细胞-基因矩阵，以表示每个基因在每个细胞中的活性。

2. 异构图构建：基于预处理后的数据，构建一个包含细胞和基因为节点的异构图，节点之间的边表示基因在细胞中的活性存在情况。

3. 异构图变换器 (HGT) 模型：利用 HGT 模型联合学习细胞和基因的低维嵌入，并生成注意力分数，指示基因对细胞的重要性，从而区分基因贡献并增强生物学解释性。

4. 细胞聚类与功能基因模块预测：基于 HGT 学习到的嵌入和注意力分数进行细胞聚类及功能基因模块预测，进一步推断出不同细胞类型中的生物网络，例如基因调控网络 (GRN) 和基因关联网。

5. 网络服务器部署：为了提升 scMulti-omics 数据分析的可用性和可重复性，研究人员还部署了一个 DeepMAPS 网页服务器，提供多种功能和可视化工具。

DeepMAPS 框架的优势在于其能够构建一体化的异构图，同时捕捉细胞和基因之间的邻近和全局拓扑特征，用以构建细胞-细胞和基因-基因关系。此外，HGT 模型中的注意力机制可以评估基因对特定细胞的重要性，无需依赖基因共表达约束，能够推断其他工具通常无法

识别的基因调控关系。通过这种方式，DeepMAPS 不仅能够推断出特定细胞类型的生物网络，还能够提供对细胞异质性和复杂分子机制的深入理解。整体模型图如图 1 所示：

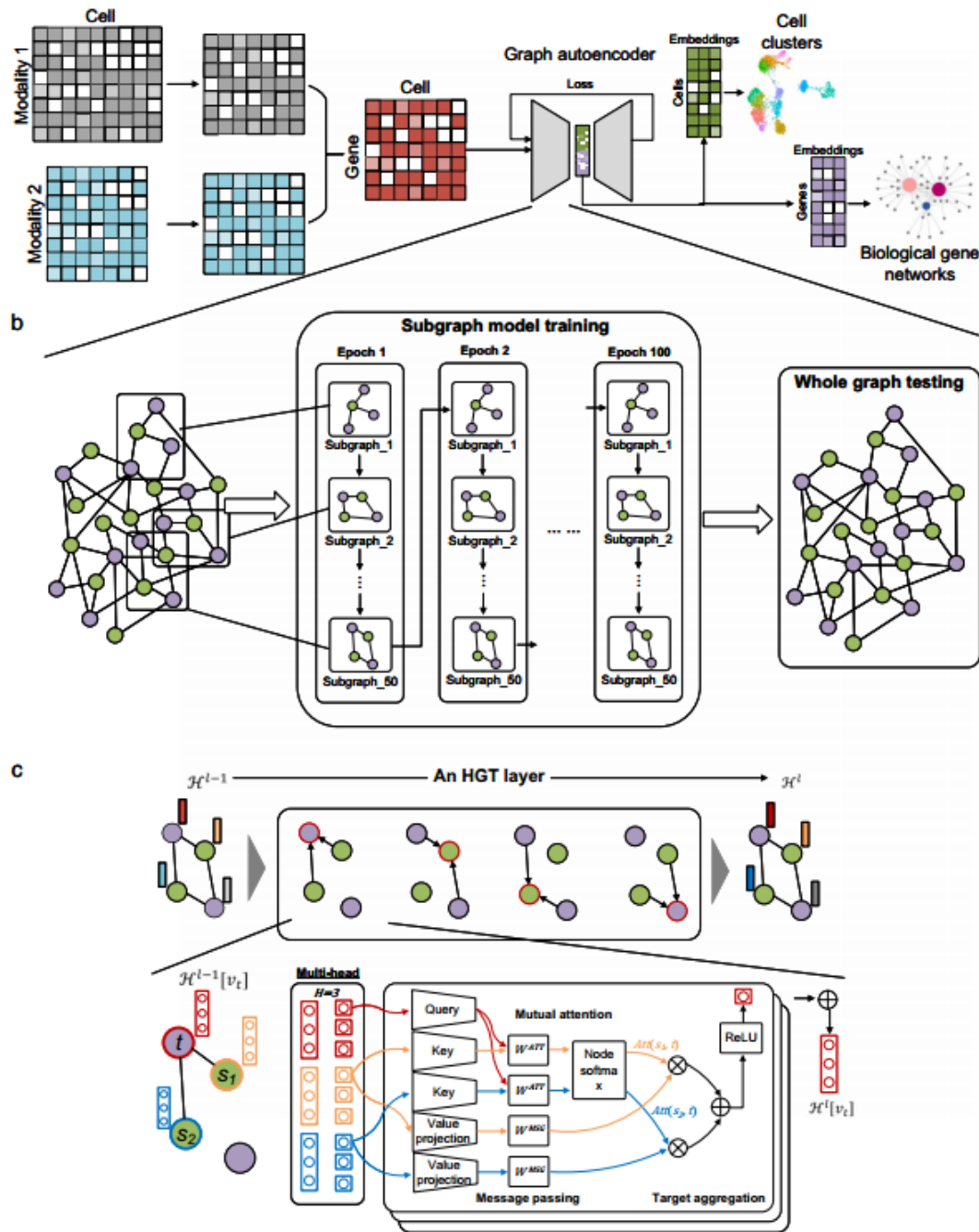


图 1. 模型示意图

3.2 异质图变换器（HGT）模型

3.2.1 节点类型和边类型的嵌入

在 HGT 模型中，数据首先通过嵌入层进行转换，将不同类型的节点和边映射到统一的低维向量空间。与传统变换器处理同质图不同，HGT 特别考虑了节点和边的异质性。每种类型的节点会有一个独立的嵌入层，用于将其映射到一个共享的嵌入空间。假设在单细胞生物网络中，可能包含不同类型的节点，如基因节点、细胞类型节点、疾病节点等，每种节点的嵌

入层会根据其类型独立学习权重。这意味着不同类型的节点有不同的嵌入表示，能够捕捉其特定的生物学特征。类似地，HGT 也为每种类型的边定义了独立的嵌入层。这些边通常代表不同类型的关系，如基因-基因相互作用、细胞-基因关系、细胞-细胞相互作用等。每种边类型的嵌入层会根据边的类型学习不同的映射，帮助模型理解不同类型边的生物学意义。

3.2.2 变换器的自注意力机制

HGT 模型的核心机制是基于变换器的自注意力 (Self-Attention) 机制。传统的变换器模型通常假设所有的节点在信息传播时是等价的，忽略了图中不同节点类型间的差异。为了解决这一问题，HGT 模型在每个变换器层中使用了异质自注意力机制 (Heterogeneous Self-Attention)。具体而言：类型感知的自注意力：HGT 通过为每种类型的节点和边设计独立的自注意力头，确保不同类型的节点在信息传播时能够得到适当的处理。例如，基因节点的自注意力计算方式与细胞节点的计算方式可能不同，因为它们的上下文信息和关系具有不同的生物学含义。节点间的信息传播：在自注意力机制中，每个节点根据其类型和邻接节点的类型，计算出与其邻接节点的权重。自注意力机制可以动态地调整节点之间的信息传播强度，使得某些类型的节点对图中的其他节点产生更强的影响。多头注意力机制：HGT 采用多头自注意力机制 (Multi-head Attention)，每个头可以学习不同的关系模式，从而提高模型的表达能力。在图数据中，不同的关系可能对推断任务有不同的重要性，多头注意力机制有助于捕捉到这些复杂的关系。

3.2.3 异质图卷积 (Heterogeneous Graph Convolution)

HGT 模型在自注意力机制之后，还引入了异质图卷积的概念。异质图卷积用于对节点的特征进行进一步的聚合和更新。与传统的图卷积网络 (GCN) [6] 不同，HGT 中的卷积操作考虑到了节点类型和边类型的异质性。在每一层的卷积中，HGT 将节点的特征与其邻居节点的特征进行融合，通过加权聚合得到更新后的节点表示。具体来说：

邻居信息聚合：每个节点的特征通过与其邻接节点的信息融合来更新。在融合过程中，考虑了邻接节点的类型和边的类型，这使得聚合操作能够针对不同类型的边采用不同的加权策略。

节点特征更新：经过聚合后的邻居节点信息会被传递回当前节点，以更新其特征表示。这一过程在多层变换器中进行，使得节点能够逐渐融合更多层级的上下文信息。

3.2.4 跨层信息融合与节点表示

HGT 模型通过多个变换器层的堆叠实现深层次的特征抽取和信息融合。在每一层中，节点的特征表示会基于邻居节点的特征进行更新，并通过自注意力机制加强重要特征的特征表示。最后，通过多层堆叠，节点能够获得从局部到全局的上下文信息，从而更好地捕捉到图中各节点之间的复杂关系。

节点表示融合：每一层的输出会作为下一层的输入，并通过变换器的自注意力机制逐步融合来自不同类型节点和边的特征。这使得每个节点能够综合考虑所有类型的信息，并最终获得一个全面的节点表示。

全局表示学习：为了进一步增强模型的表达能力，HGT 还设计了跨层信息融合的机制，

使得每个节点的表示不仅仅依赖于局部邻域，还能够捕捉到全局图结构的信息。这有助于推断细胞之间的全局关系以及复杂的基因交互模式。

4 复现细节

4.1 与已有开源代码对比

复现过程中使用了作者已经开源的代码，但是我对原始模型进行了创新和优化，主要体现在以下几个方面，尤其是在节点和边的嵌入方式进行了定制化改进。以下将对比现有的开源代码，突出我们所做的创新和提升。

与现有的开源实现（如标准的 Heterogeneous Graph Transformer, HGT）相比，本文复现过程中对节点和边的嵌入进行了改进，主要体现在对节点类型的独立嵌入设计。原论文中采用了基于类型的普通嵌入策略，但我们引入了类型感知嵌入，使得每种类型的节点和边都有独立的表示方式。我们在原有嵌入层基础上引入了类型感知的嵌入，即对每种类型的节点和边设计独立的嵌入表示。这样做可以显著提高模型对不同类型节点间关系的理解，尤其是当图中包含多个复杂的生物学节点（如基因、细胞、疾病等）时，这种类型感知的方式能够更精确地捕捉到它们之间的交互。且许多开源的异质图变换器（HGT）实现将不同类型的节点和边统一嵌入到同一个空间中，可能会丧失不同类型信息的细粒度表达。相比之下，我们的方法通过为每个节点和边类型单独设计嵌入层，能更好地应对生物学任务中多样化的节点和边关系，提升了单细胞生物网络推断任务的准确性。没有优化前的细胞与基因关系如图 2 所示，优化后的如图 3。

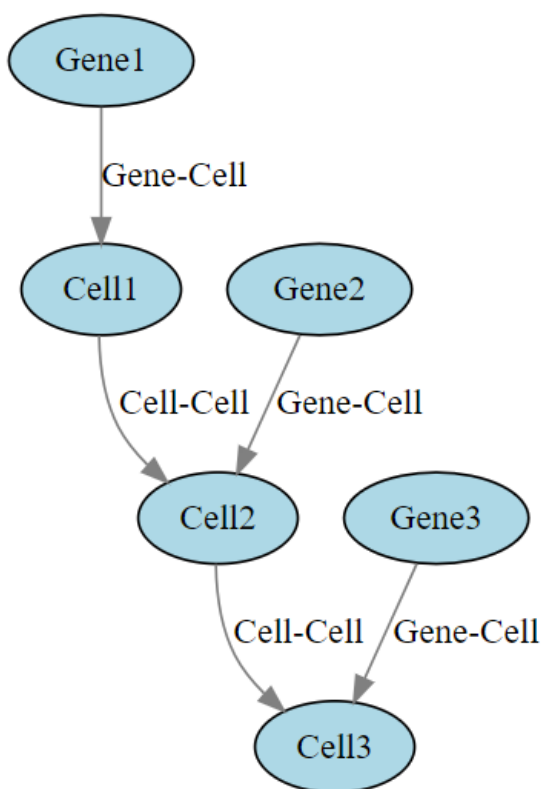


图 2. 没有优化前的细胞与基因关系图

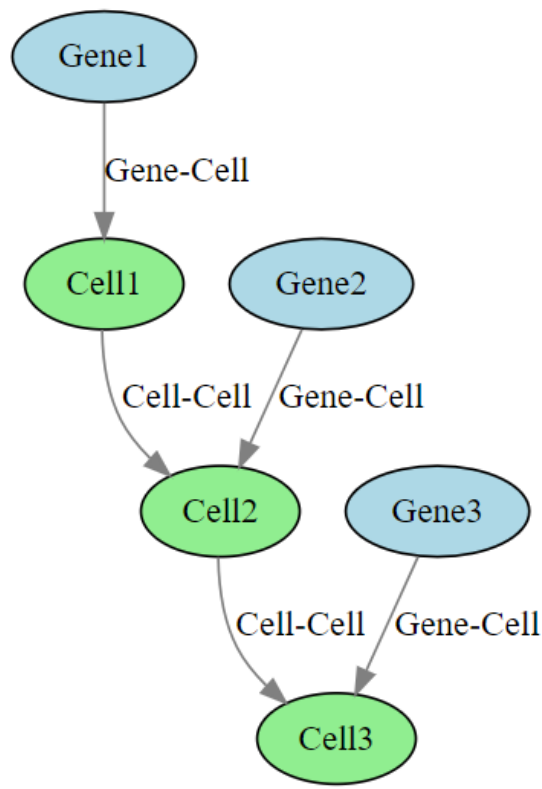


图 3. 优化后的细胞与基因关系图

上图主要的区别在于细胞基因的颜色表示，没有优化前虽然论文模型图中有讲基因和细胞分成不同颜色表示，通过颜色进行了区分，但是仍然通过统一的嵌入层进行处理的，没有进一步细化每种类型的嵌入层进行特定优化。优化后的模型采用了类型感知嵌入，即基因和细胞的处理方式变得更加精细，每种类型的节点和边类型都有独立的嵌入层和处理方式，不仅在显示上有所区别，而且在处理的计算层面也做了优化，使得模型能够更好地捕捉不同类型节点间的关系。

4.2 实验环境搭建

硬件环境	CPU	Intel(R) Xeon(R) Gold 6226R CPU @ 2.90GHz
	GPU	NVIDIA GeForce RTX 3090
	内存	377G
操作系统	OS	Ubuntu 22.04.4 LTS
开发环境	编程语言	Python 3.8.5
	深度学习框架	Pytorch
	开发工具	PyCharm 2021.2.3 (Community Edition)

4.3 创新点

我对原始模型进行了类型感知嵌入 (Type-aware Embedding) 的创新，使得模型在处理异质图数据时更加精细和有效。论文中的原始方法采用了统一的嵌入层来处理不同类型的节点和边（如基因节点、细胞节点和它们之间的关系）。然而，这种统一的嵌入方式无法充分捕捉不同类型节点和边的特征及其相互关系。因此，在复现过程中，我们引入了类型感知的嵌入方法，将不同类型的节点和边分别处理，并为它们设计独立的嵌入层。这一创新提升了模型的能力，使其能够更加精细地捕捉异质图中的复杂关系。我为基因节点和细胞节点分别设计了独立的嵌入层。通过这种方式，基因节点的表达模式可以与细胞状态和相互作用分开处理，使得两者之间的关系被更加准确地建模。对于图中的边（如基因-细胞边、细胞-细胞边等），我同样为不同类型的边设计了独立的嵌入层。这样，基因-细胞边和细胞-细胞边可以通过各自的嵌入层进行独立处理，避免它们被统一对待，从而更好地捕捉两者之间的生物学差异。

通过类型感知嵌入，我们能更加灵活地处理多种类型的节点和边，提升了模型在单细胞网络推断任务中的表现。

5 实验结果分析

原论文中使用来自 10X Genomics 的单细胞多组 ATAC+ 基因表达数据集。原始数据来自诊断为淋巴结弥漫性小淋巴细胞淋巴瘤 (DSLL) 的 14,566 个细胞。我同样使用了这些数据。数据展示如表1所示。

数据文件名称	文件类型	文件大小	说明
RNA + ATAC count matrix	.h5 文件	118 MB	RNA 和 ATAC 信号的计数矩阵，包含每个细胞的基因表达和 ATAC 信息。
ATAC fragments	.tsv.gz 文件	2.7 GB	ATAC-seq 数据的片段信息，提供有关染色质可接近性的详细信息。
ATAC fragments index	.tbi 文件	1 MB	ATAC 数据的索引文件，帮助高效查找特定的片段。
RNA velocity matrix	.csv.gz 文件	434 MB	RNA 流动性矩阵，用于预测基因表达的动态变化，通常用于推测细胞的发育过程。

表 1. 数据集文件详情

使用单细胞 RNA 测序 (scRNA-seq) 数据然后使用训练好的模型进行复现论文中的一些实验，论文细胞聚类如图 4所示，使用优化后的方法细胞聚类如图 5。可以看出使用类型感知的嵌入之后对于细胞聚类与原本的聚类并没有太多的改变，主要是左下角 Naive CD4+ T cells、Central memory-like CD4+Tcells 等四种细胞的位置发生了改动，以及 NK 细胞与 Monocyte 更家疏远，这个改动主要是说明了细胞基因表达的差异，细胞聚类实质是说明越靠近的细胞簇类基因表达差异性越小，越远说明差异性越大，但是模型优化后总体的各个细胞还是保持紧靠在一起说明这种细胞聚类是正确的，且自己的模型可能察觉到了 TRM CD4+ T 细胞应该与 Monocyte 这些细胞簇更加远离，这将更可能的揭示细胞的异质性，同时更加细致捕获细胞簇间的关系，为进一步解释生物信息提供帮助。

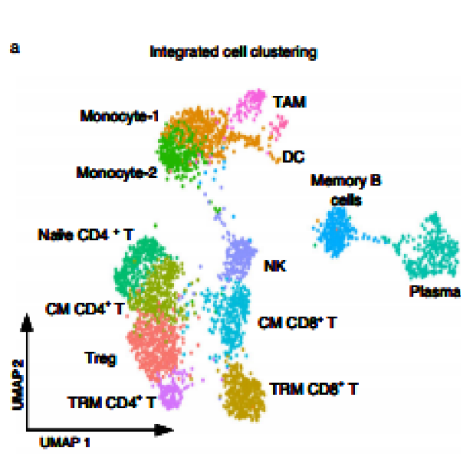


图 4. 细胞聚类图

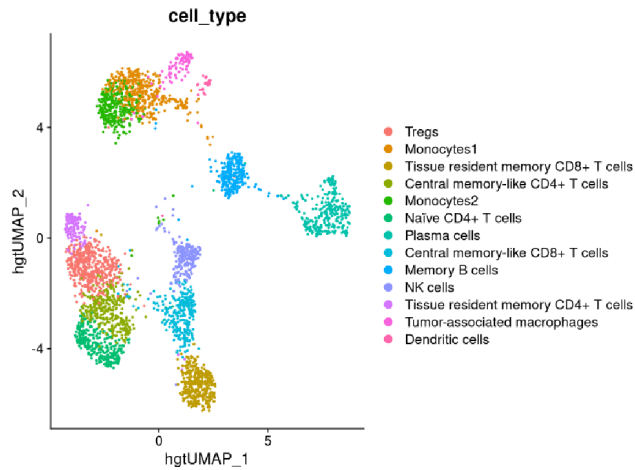


图 5. 细胞聚类图

继续使用该数据对基因的表达进行实验，论文中的热力图如图 6所示，优化后的细胞聚类和注释的基因热图如图 6所示，细胞聚类和注释的精选标记蛋白热图如图 7所示。可以看到细胞聚类和注释的基因热图和细胞聚类和注释的精选标记蛋白热图这两个图与原论文中的图并没有很大差别，基因热图有一些区域的颜色跟原图有了差别但是总体基因表达较高的区域并没有发生改变说明细胞基因高表达的并没有发生改变还是这些，有些低表达的基因颜色变淡可能是由于采用类型感知嵌入导致一些不是强表达仅仅是弱表达导致了一些基因表达在热力图上显示变淡。总体上对于高表达基因保持一样，对于低表达基因来说模型表现不是很好，

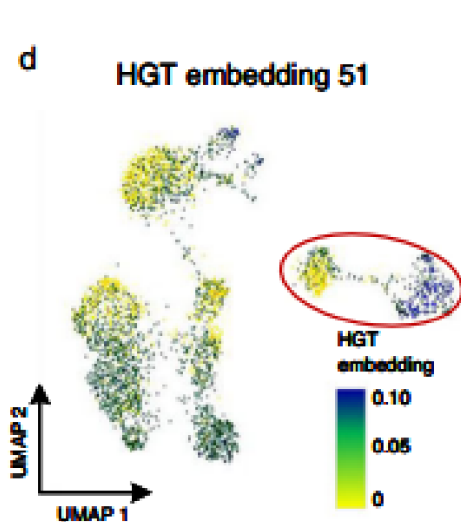


图 9. 论文原图

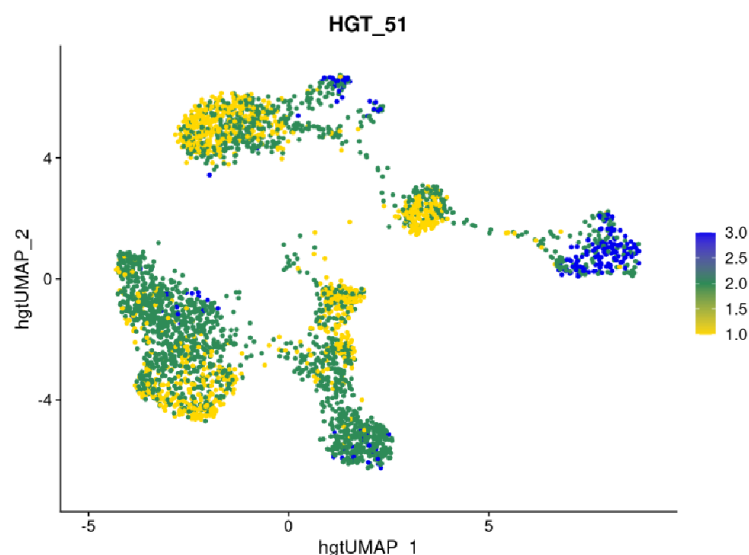


图 10. 优化后的嵌入图

由于论文中的图表几乎没有, 为了显示优化后的模型对于其他数据也有比较好的表现, 我使用 10X Genomics 免费提供的新鲜冷冻淋巴结 B 细胞淋巴瘤数据集 (14k 分类细胞核) 进行展示情况, 这些数据同样进行预处理删除低表达基因以及异常基因, 将矩阵交给训练好的模型。运行 HGT 模型后, DeepMAPS 会应用基于图的模型 Louvain, 根据从 HGT 模型返回的细胞特征约简矩阵对细胞进行聚类。分别标记基因 CD3E、CD4、CD8A 展示小提琴图如图 11 所示。

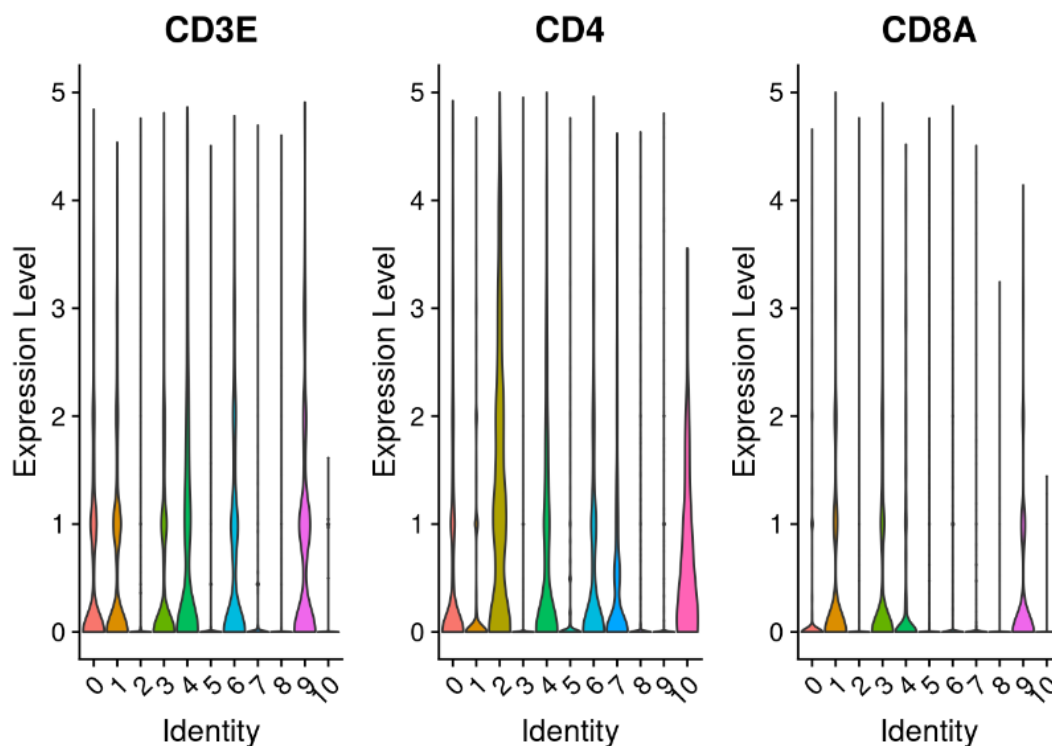


图 11. 标记基因的小提琴图

将细胞聚类之后, 并增加细胞类型注释然后显示 UMAP 如图 12 所示。可以看到对该数

据优化后的模型也具有较好的聚类功能，同时对于基因表达水平的测量也是可以展示出来。

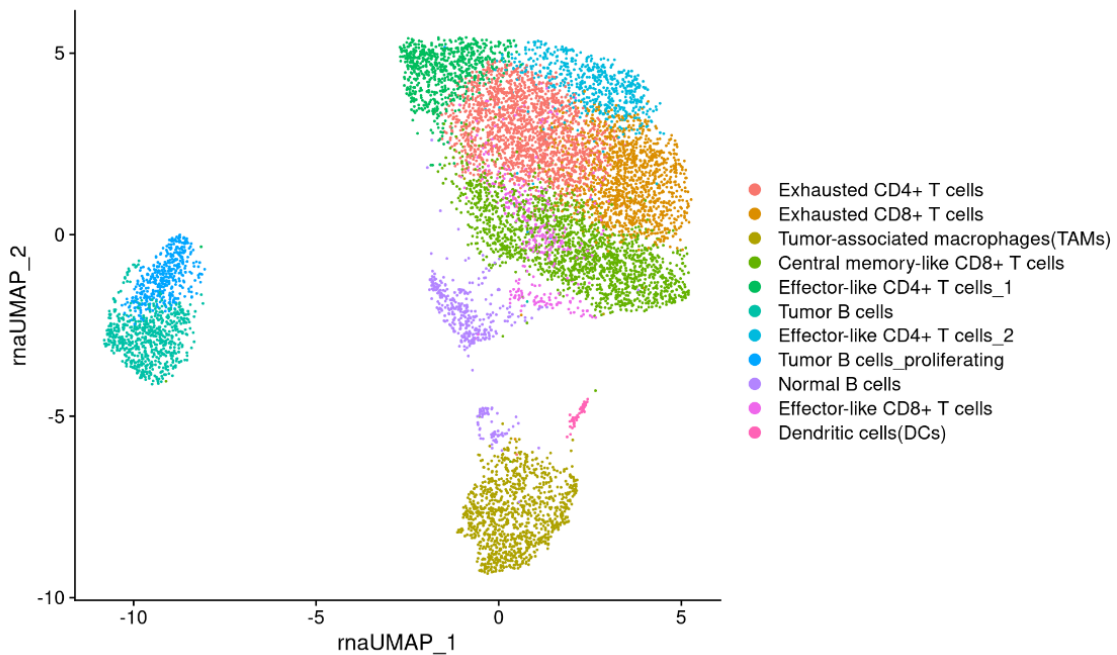


图 12. 细胞聚类图

6 总结与展望

在前面五个章节里，分别介绍了复现的论文的不同内容。第一章节介绍复现论文的研究背景、意义以及我选择复现这篇论文的原因；第二章节我对本论文涉及的相关工作进行了介绍，例如异构图和数据预处理的作用；在第三章节中，对复现论文的模型框架进行了大致的介绍，同时介绍模型中使用到的 HGT 模型发挥的作用；第四章节介绍了论文复现的细节，如复现环境、方法的创新点等；第五章节对我训练得到的模型进行结果分析，结果可视化等。

本文的核心工作包括主要是首先通过类型感知的嵌入优化了原始模型，在处理不同类型的节点和边时，采用了独立的嵌入层，从而使得模型能够更精细地捕捉异质图中节点间的复杂关系，通过为不同类型的节点和边设计独立的嵌入层，我们增强了模型对异质图结构的表达能力，使其能够更好地建模基因、细胞等不同生物学实体之间的复杂关系。

尽管本模型在性能上取得了不错的效果，但对于生物学家来说，模型的可解释性仍然是一个重要问题。未来的研究可以尝试通过图神经网络 (GNN) 中的可解释性方法（如节点影响力分析、重要性评分等）来增强模型的透明性，帮助生物学家理解模型的决策过程。同时，可以结合数据可视化技术，使模型的预测结果和生物学推理过程更加直观，便于生物学家进行进一步的实验验证。对于优化后的结果优化效果没有太明显的问题，推测是因为仅仅用独立的嵌入层可能对整体没有太多的影响，应该在自注意力那里继续修改，比如也用类型嵌入试试这样整体就全都用上了类型感知嵌入，这样将可能对整个模型结果有更明显的优化，预计在未来一段时间内对这部分实验进行完善后，如果如期所料即可将研究成果投稿至 BIBM。

参考文献

- [1] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.
- [2] Siti Haryanti Hairol Anuar, Zuraida Abal Abas, Norhazwani Mohd Yunos, Nurul Hafizah Mohd Zaki, Nurul Akmal Hashim, Mohd Fariddudin Mokhtar, Siti Azirah Asmai, Zaheera Zainal Abidin, and Ahmad Fadzli Nizam. Comparison between louvain and leiden algorithm for network structure: a review. In *Journal of Physics: Conference Series*, volume 2129, page 012028. IOP Publishing, 2021.
- [3] Ricard Argelaguet, Damien Arnol, Danila Bredikhin, Yonatan Deloro, Britta Velten, John C Marioni, and Oliver Stegle. Mofa+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome biology*, 21:1–17, 2020.
- [4] Ricard Argelaguet, Anna SE Cuomo, Oliver Stegle, and John C Marioni. Computational principles and challenges in single-cell data integration. *Nature biotechnology*, 39(10):1202–1215, 2021.
- [5] E Bornemann and JH Doveton. Log normalization by trend surface analysis. *The log analyst*, 22(04), 1981.
- [6] Wei-Lin Chiang, Xuanqing Liu, Si Si, Yang Li, Samy Bengio, and Cho-Jui Hsieh. Cluster-gcn: An efficient algorithm for training deep and large graph convolutional networks. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 257–266, 2019.
- [7] Saket Choudhary and Rahul Satija. Comparison and evaluation of statistical error models for scrna-seq. *Genome biology*, 23(1):27, 2022.
- [8] Adam Gayoso, Zoë Steier, Romain Lopez, Jeffrey Regier, Kristopher L Nazon, Aaron Streets, and Nir Yosef. Joint probabilistic modeling of single-cell multi-omic data with totalvi. *Nature methods*, 18(3):272–282, 2021.
- [9] Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, William M Mauck, Shiwei Zheng, Andrew Butler, Maddie J Lee, Aaron J Wilk, Charlotte Darby, Michael Zager, et al. Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587, 2021.
- [10] Ilya Korsunsky, Nghia Millard, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy Baglaenko, Michael Brenner, Po-ru Loh, and Soumya Raychaudhuri. Fast, sensitive and accurate integration of single-cell data with harmony. *Nature methods*, 16(12):1289–1296, 2019.
- [11] Yang Li, Anjun Ma, Ewy A Mathé, Lang Li, Bingqiang Liu, and Qin Ma. Elucidation of biological networks across complex diseases using single-cell omics. *Trends in Genetics*, 36(12):951–966, 2020.

- [12] Anjun Ma, Adam McDermaid, Jennifer Xu, Yuzhou Chang, and Qin Ma. Integrative methods and practical challenges for single-cell multi-omics. *Trends in biotechnology*, 38(9):1007–1022, 2020.
- [13] Qin Ma and Dong Xu. Deep learning shapes single-cell data analysis. *Nature Reviews Molecular Cell Biology*, 23(5):303–304, 2022.
- [14] Tim Stuart and Rahul Satija. Integrative single-cell analysis. *Nature reviews genetics*, 20(5):257–272, 2019.
- [15] Sarah Teichmann and Mirjana Efremova. Method of the year 2019: single-cell multimodal omics. *Nat. Methods*, 17(1):2020, 2020.
- [16] Juexin Wang, Anjun Ma, Yuzhou Chang, Jianting Gong, Yuexu Jiang, Ren Qi, Cankun Wang, Hongjun Fu, Qin Ma, and Dong Xu. scgcn is a novel graph neural network framework for single-cell rna-seq analyses. *Nature communications*, 12(1):1882, 2021.
- [17] Chuxu Zhang, Dongjin Song, Chao Huang, Ananthram Swami, and Nitesh V Chawla. Heterogeneous graph neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 793–803, 2019.