scGTN: Deep Siamese Graph Transformer Network for Single-cell RNA Sequencing Clustering

Jinke Wu[1], **Yifan Wang**[2], **Siyu Yi**[1†], **Caiyang Yu**[1], **Ziyue Qiao**[3], **Nan Yin**[4], **Jiancheng Lv**[1], and **Wei Ju**[1†]
[1]Sichuan University
[2]University of International Business and Economics
[3]Great Bay University
[4]The Education University of Hong Kong
w_rmsl@stu.scu.edu.cn, {siyuyi, lvjiancheng, juwei}@scu.edu.cn, yifanwang@uibe.edu.cn, {yucy324, ziyuejoe, yinnan8911}@gmail.com

Abstract

Single-cell RNA sequencing (scRNA-seq) serves a pivotal role in characterizing gene expression at the cellular level, enabling the identification of cell types and advancing the understanding of cellular heterogeneity. Despite the significant progress in scRNA-seq data clustering, we argue that current methods always ignore the sparsity and noise, as well as the complex intercellular structural information inherent in scRNA-seq data. Toward this end, in this paper, we propose a novel single-cell RNA-seq clustering framework via deep Siamese Graph Transformer Network (termed scGTN), which explicitly integrates gene expression profile and intercellular structural dependencies for cell clustering. In particular, we formulate scRNA-seq data as a graph and construct two augmented graph views that serve as dual views to capture complementary intercellular information. Then, a Siamese graph transformer network is employed to explicitly incorporate shortest-path information and node-wise distances for capturing richer structural relationships between cells. Finally, we employ an optimal transport strategy to guide the cell clustering in a self-supervised manner. Extensive experiments on multiple benchmark scRNA-seq datasets demonstrate that our scGTN consistently outperforms existing methods. Our code is available at https://github.com/W-RMSL/scGTN.

1 Introduction

Single-cell RNA sequencing (scRNA-seq) has emerged as a fundamental technology for profiling gene expression in bioinformatics, enabling high-resolution profiling of gene expression across individual cells. Based on the scRNA-seq data, cell clustering plays a central role by organizing cells into biologically meaningful groups based on their gene expression patterns, thereby revealing the intrinsic heterogeneity of cell populations and facilitating biological interpretation of their functions and interactions [Kiselev *et al.*, 2019; Peng *et al.*, 2020]. Accordingly, clustering scRNA-seq data is particularly valuable for a range of downstream tasks, including marker gene annotation [Dai *et al.*, 2022] and cell type identification [Hu *et al.*, 2020].

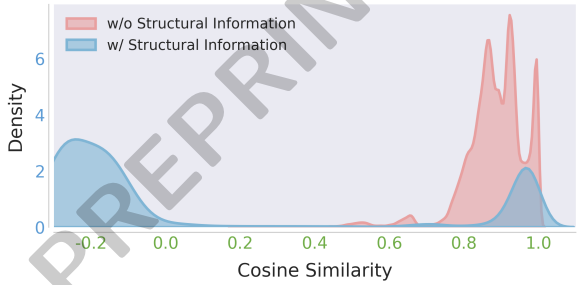


Figure 1: Similarity distributions of cell embeddings obtained by scGTN with and without structural information on the scHuman pancreas cells dataset.

For effectively scRNA-seq cell clustering, numerous methods have been proposed to achieve this goal. Early approaches primarily relied on classical clustering algorithms, such as k-means [Hicks *et al.*, 2021], hierarchical clustering [Johnson, 1967], and density-based methods [Petegrosso *et al.*, 2020], which are computationally efficient but are inherently limited in modeling the complex nonlinear characteristics of scRNA-seq data. With the rapid advances in deep learning, recent studies have shifted toward learning-based clustering frameworks that leverage graph neural networks (GNNs) [Ju *et al.*, 2025; Fan *et al.*, 2026] to extract informative representations from scRNA-seq data. By representing cells as nodes and their interactions as edges in a graph, GNN-based approaches jointly exploit gene expression features and structure characteristics to learn context-aware cell embeddings, enabling a more comprehensive cell clustering [Wang *et al.*, 2021; Gan *et al.*, 2022; Xu *et al.*, 2024; Xu *et al.*, 2025; Zhang *et al.*, 2026b].

Despite the effectiveness of these methods, we argue that two major challenges remain insufficiently addressed: ❶ ***Ig-***

†Corresponding authors.

noring the Sparsity and Noise inherent within the scRNA-seq Data. Arising from extensive dropout effects and measurement noise, scRNA-seq data present a high-dimensional expression matrix [Ou-Yang *et al.*, 2022; Wu and Ma, 2022], which substantially complicates the reliable construction of cell relationships and frequently leads to graphs that are highly sparse and noisy. ❷ ***Limited Exploitation of Complex Intercellular Structural Information.*** While existing GNN-based approaches primarily focus on local neighborhood aggregation, they often overlook richer structural cues embedded within the cell graph. Features such as shortest paths and node distances are rarely leveraged, resulting in an insufficient utilization of the graph's inherent information. As shown in Figure 1, cell similarities derived solely from gene expression features tend to be highly homogeneous and difficult to distinguish, whereas incorporating structural information such as positional relationships and shortest-path distances enables a more discriminative characterization of intercellular relationships for cell clustering.

Having realized the above challenges with existing methods, in this paper, we propose a novel deep Siamese **G**raph **T**ransformer **N**etwork for single-cell RNA sequencing (**scGTN**), which effectively exploits complementary graph information derived from gene expression data to learn discriminative and robust cell embeddings. Specifically, given the scRNA-seq data, we formulate it into a graph and construct two augmented graph views to serve as dual branches to capture and refine the intercellular information. Then, for each branch, we introduce a graph transformer network that incorporates shortest-path information and node-wise distances to capture the complementary graph structural dependencies, enabling a more expressive modeling of complex intercellular relationships. Based on the encoded information, we utilize an optimal transport-guided self-supervised clustering method to align cluster distributions for the clustering learning process.

To summarize, the contributions of this paper are:

❶ ***New Perspective:*** We study scRNA-seq data clustering by jointly considering gene expression profiles and cell graph structures, which effectively exploits complementary intercellular information beyond feature-level similarities.

❷ ***Novel Methodology:*** We propose a novel framework termed scGTN, which introduces a Siamese graph transformer network to incorporate richer graph structural information, including shortest paths and node-wise distances for the cell clustering task.

❸ ***Extensive Experiment:*** We conduct extensive experiments on multiple benchmark scRNA-seq datasets to evaluate the effectiveness of our scGTN. The results further show the superior clustering performance of our framework.

2 Methodology

2.1 Problem Definition

Let $\boldsymbol{X} = (x_{ij}) \in \mathbb{R}^{N \times D}$ denote the single-cell gene expression matrix, where each row $\boldsymbol{x}_i \in \mathbb{R}^D$ denotes the gene expression of the i-th cell, N and D denote the number of cells and gene respectively. Based on $\boldsymbol{X}$, we construct an undirected cell graph $G = (\mathcal{V}, \mathcal{E})$, where each node represents an individual cell and edges are established via a k-nearest neighbor (kNN) strategy to capture intercellular similarities quantified by the Pearson correlation coefficient [Benesty *et al.*, 2009]. The adjacency matrix of the constructed graph can be $\boldsymbol{A} = (a_{ij}) \in \mathbb{R}^{N \times N}$, where entry $a_{ij} = 1$ if $(v_i, v_j) \in \mathcal{E}$ for cell v_i and v_j, otherwise $a_{ij} = 0$. We associate the degree matrix of the graph as $\boldsymbol{D} = \mathrm{diag}(d_1, d_2, \ldots, d_N) \in \mathbb{R}^{N \times N}$, where $d_i = \sum_{j=1}^{N} a_{ij}$ is the degree of node v_i. The goal of scRNA-seq clustering is to partition the cells within the graph $\mathcal{G}$ into C disjoint clusters, with each node v_i belonging to a unique cluster according to its gene expression features and intercellular connectivity.

2.2 Framework Overview

In this work, we focus on scRNA-seq data clustering and propose a Siamese graph transformer network for the task. As shown in Figure 2, our scGTN consists of three modules: (1) *Dual Augmentation Module for scRNA-seq Data*, which jointly perturbs the gene expression features and the underlying cell graph structure to generate two complementary views, thereby facilitating robust representation learning; (2) *Siamese Graph Transformer Network Fusion Module*, which employs two weight-sharing Siamese encoders to process the augmented views while explicitly incorporating structural information of the cell graph; (3) *Optimal Transport Clustering Module*, which formulates clustering as a distribution alignment problem and employs optimal transport to guide the learning of cluster-consistent representations.

2.3 Dual Augmentation for scRNA-seq Data

To mitigate the effects of inherent sparsity and noise in scRNA-seq data, we introduce biologically reasonable augmentations by adding controlled noise to the gene expression profiles, generating two complementary views that improve the robustness of the learned representations.

Gene Expression Perturbation. Given the single-cell gene expression matrix $\boldsymbol{X}$, we employ Gaussian noise to perturb the gene expression profiles, which can simulate the natural variability of the gene expression profile as augmentation for the clustering task. The perturbation can be formulated as:

$$\tilde{\boldsymbol{X}} = \{\boldsymbol{x}_1 \odot \boldsymbol{m}, \ldots, \boldsymbol{x}_N \odot \boldsymbol{m}\}, \tag{1}$$

where $\boldsymbol{m}$ denotes the random mask drawn from a Gaussian distribution. In practice, we set the distribution as $\mathcal{N}(1, 0.1)$ following the previous work [Xu *et al.*, 2025]. The two augmented gene matrices can be generated as $\tilde{\boldsymbol{X}}^1$ and $\tilde{\boldsymbol{X}}^2$.

Intercellular Graph Structure Augmentation. In addition to perturbing the gene expression profiles, enhancing the intercellular structure is essential for improving the model's ability to capture the relationships between cells. Specifically, given the constructed cell graph, we remove spurious edges to refine the graph structure and preserve the most biologically relevant connections between cells as one view:

$$\tilde{\boldsymbol{A}}^1 = \boldsymbol{D}^{-\frac{1}{2}}(\boldsymbol{A} \odot \boldsymbol{M} + \boldsymbol{I})\boldsymbol{D}^{-\frac{1}{2}}, \tag{2}$$

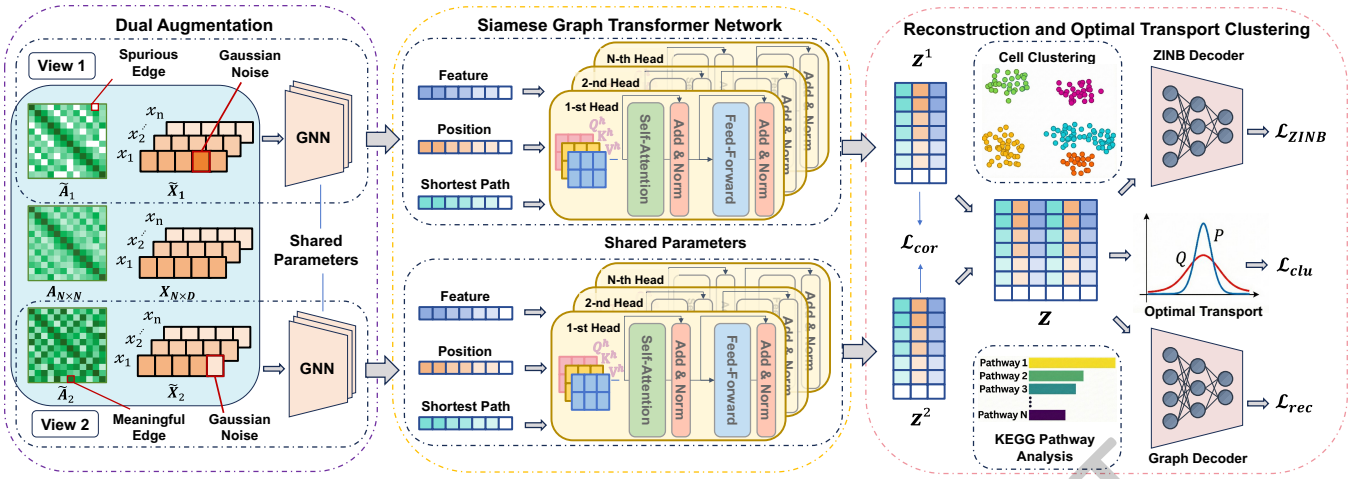


Figure 2: Illustration of the proposed scGTN, which consists of three components: (1) **Dual Augmentation for scRNA-seq Data**: Gene expression perturbation and intercellular graph structure augmentation to generate two complementary views. (2) **Siamese Graph Transformer Network Fusion**: We explicitly integrate gene expression and intercellular structure embeddings for the discriminative representation. (3) **Clustering with Optimal Transport**: We introduce a self-supervised optimal transport-guided manner to refine the cell clustering.

where $\boldsymbol{M}$ denotes the mask matrix based on pairwise cosine similarity. In practice, we identify and remove 10% of edges with the lowest similarity values using this mask. For another view, we enhance the intercellular structure by considering graph diffusion, which propagates information across the graph to strengthen meaningful cell connections:

$$\tilde{\boldsymbol{A}}^2 = \eta(\boldsymbol{I} - (1-\eta)\boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{A}\boldsymbol{D}^{-\frac{1}{2}})^{-1}, \tag{3}$$

where η denotes the diffusion coefficient to control the information propagation within the cell graph.

2.4 Siamese Graph Transformer Network Fusion

To further utilize the richer structural information from the intercellular graph, we introduce a Siamese graph transformer network [Wang *et al.*, 2025] for the two augmented views and adaptively fuse the encoded feature to learn discriminative cell representations.

Gene Expression Embedding. Since cells with similar gene expression profiles are assumed to be connected in the graph, we leverage the gene expression profiles of connected cells to enhance the learning of discriminative gene expression embeddings. In particular, we employ two Siamese graph encoders to produce the cell embedding as:

$$\boldsymbol{Z}^* = \mathcal{G}(\tilde{A}^*, \tilde{X}^*), * \in \{1, 2\}, \tag{4}$$

where $\mathcal{G}(\cdot,\cdot)$ denotes the Siamese graph encoder, which is shared across the two augmented views. The resulting encoded gene expression matrix $\boldsymbol{Z}^* \in \mathbb{R}^{N\times d}$ represents the learned feature embeddings for each cell, where d indicates the dimensionality of the feature space. Then, for each cell node with feature $\boldsymbol{z}_i^*$, we select the top-t neighbors with the most similar features and concatenate them into a sequence $\boldsymbol{S}_i^* \in \mathbb{R}^{(t+1)\times D}$. The sequence is then encoded as:

$$\boldsymbol{E}_i^* = \mathcal{F}_{ge}(\boldsymbol{S}_i^*) \in \mathbb{R}^{(t+1)\times d}, \tag{5}$$

where $\mathcal{F}_{ge}(\cdot)$ denotes the gene expression encoding function.

Explicit Intercellular Structure Embedding. To explicitly capture the intercellular structure within scRNA-seq data, we incorporate both the position embedding and shortest path embedding of the central node and its neighbors. Specifically, we define the relative positions of the central node as $\mathrm{Pos}(\boldsymbol{z}_i^*) = 0$ and assign position values to its neighbors based on their intimacy to the central node, with nodes that are more similar to the central node receiving position values closer to 0. The position embedding is formulated as:

$$\boldsymbol{P}_i^* = \mathcal{F}_{pos}(\mathrm{Pos}(\boldsymbol{z}_i^*)) \in \mathbb{R}^{(t+1)\times d}, \tag{6}$$

where $\mathcal{F}_{pos}(\cdot)$ denotes the position mapping function. In addition to the intimacy between two nodes, the shortest path characterizes the connection structure between them and explicitly the crucial intercellular structure. Specifically, the shortest path between two nodes within the constructed cell graph can be defined as $\mathrm{Sp}(\boldsymbol{z}_i^*, \boldsymbol{z}_j^*)$. Based on this, we formulate the shortest path embedding as:

$$\boldsymbol{H}_i^* = \mathcal{F}_{sp}(\mathrm{Sp}(\boldsymbol{z}_i^*, \boldsymbol{z}_j^*)) \in \mathbb{R}^{(t+1)\times d}, \tag{7}$$

where $\mathcal{F}_{sp}(\cdot)$ denote the shortest path mapping function.

Siamese Graph Transformer Refinement. We integrate the gene expression and intercellular structure embedding as:

$$\boldsymbol{Y}_i^* = \boldsymbol{E}_i^* + \boldsymbol{P}_i^* + \boldsymbol{H}_i^* \in \mathbb{R}^{(t+1)\times d}. \tag{8}$$

Then we leverage the attention mechanism to encode the integrated embedding, which can be formulated as:

$$\mathrm{Attention}(\boldsymbol{Y}_i) = \mathrm{Softmax}\left(\frac{\boldsymbol{W}_Q\boldsymbol{Y}_i(\boldsymbol{W}_K\boldsymbol{Y}_i)^\top}{\sqrt{d}}\right)\boldsymbol{W}_V\boldsymbol{Y}_i \tag{9}$$

where $\boldsymbol{W}_Q, \boldsymbol{W}_K, \boldsymbol{W}_V \in \mathbb{R}^{d\times d}$ denote the query, key and value matrices respectively. We employ the multi-attention mechanism with feed-forward networks (FFN) and residual connections as a transformer layer. Through stacking L-layers, the output feature can be $\boldsymbol{Y}_i^{L,*} \in \mathbb{R}^{(t+1)\times d}$. The final

cell node embedding can be refined as:

$$\hat{\boldsymbol{z}}_i^* = \frac{1}{t+1}\sum\nolimits_{t=0}^{t} \boldsymbol{Y}_i^{L,*}[t,:] \in \mathbb{R}^d. \tag{10}$$

We further introduce the correlation loss to prevent feature collapse and redundancy between two views, defined as:

$$\begin{aligned} \boldsymbol{R}_{ij} &= \frac{\hat{\boldsymbol{z}}_i^1(\hat{\boldsymbol{z}}_j^2)^\top}{\|\hat{\boldsymbol{z}}_i^1\|\|\hat{\boldsymbol{z}}_j^2)\|}, \\ \mathcal{L}_{\text{cor}} &= \sum\nolimits_i (1-\boldsymbol{R}_{ii})^2 + \sum\nolimits_i \sum\nolimits_{j\neq i} \boldsymbol{R}_{ij}^2. \end{aligned} \tag{11}$$

The final cell embedding $\hat{\boldsymbol{Z}}$ can be fused from two views as:

$$\hat{\boldsymbol{Z}} = \frac{1}{2}(\hat{\boldsymbol{Z}}^1 + \hat{\boldsymbol{Z}}^2). \tag{12}$$

2.5 Clustering with Optimal Transport

Given the refined cell embedding, we adopt a self-supervised strategy for the scRNA-seq data clustering. Specifically, we measure the similarity between cell embedding $\hat{\boldsymbol{z}}_i$ and the clustering center $\boldsymbol{\mu}_c$ as the Student's t-distribution [Maaten and Hinton, 2008], formulated as:

$$q_{ic} = \frac{(1+\|\hat{\boldsymbol{z}}_i - \boldsymbol{\mu}_c\|/\theta)^{-\frac{1+\theta}{2}}}{\sum_{c'}(1+\|\hat{\boldsymbol{z}}_i - \boldsymbol{\mu}_{c'}\|/\theta)^{-\frac{1+\theta}{2}}}, \tag{13}$$

where $\boldsymbol{Q} = [q_{ic}] \in \mathbb{R}_+^{N\times C}$ denotes the cluster assignment of the cell and θ is the degrees of freedom in the Student's t-distribution. Following the previous work [Bo *et al.*, 2020], we sharpen the cluster assignment as the target distribution $\boldsymbol{P} = [p_{ic}] \in \mathbb{R}_+^{N\times C}$, which can be calculated as:

$$p_{ic} = \frac{q_{ic}^2/f_c}{\sum q_{ic'}^2/f_{c'}}, \tag{14}$$

where $f_c = \sum_i q_{ic}$ denotes the soft cluster frequency. We further introduce the optimal transport strategy to align the cluster distribution with the mixing proportions, which ensures the robust clustering and prevents the degenerate solution that all cells allocate to a single label:

$$\begin{aligned} &\min_{\boldsymbol{P}} -\boldsymbol{P} * (\log \boldsymbol{Q}) \\ &\text{s.t. } \boldsymbol{P}\mathbf{1}_C = \mathbf{1}_N \ \text{ and } \ P^\top \mathbf{1}_N = N\boldsymbol{\pi}, \end{aligned} \tag{15}$$

where $\boldsymbol{P}$ and $-\log \boldsymbol{Q}$ denote the transport plan and cost matrix, respectively. Here $\boldsymbol{\pi}$ in the constraints represents the proportion of cells assigned to each cluster, which can be derived from the intermediate clustering results. In practice, we utilize the Sinkhorn distance method [Sinkhorn, 1967] for the optimization by introducing entropy constraint $H(\cdot)$ with a Lagrange multiplier λ as:

$$\begin{aligned} &\min_{\boldsymbol{P}} -\boldsymbol{P} * (\log \boldsymbol{Q}) - \frac{1}{\lambda} H(\boldsymbol{P}) \\ &\text{s.t. } \boldsymbol{P}\mathbf{1}_C = \mathbf{1}_N \ \text{ and } \ P^\top \mathbf{1}_N = N\boldsymbol{\pi}. \end{aligned} \tag{16}$$

Therefore, we can iteratively optimize the $\boldsymbol{P}$ as:

$$\hat{\boldsymbol{P}}^{(t)} = \text{diag}(\boldsymbol{u}^{(t)})\boldsymbol{Q}^\lambda \,\text{diag}(\boldsymbol{v}^{(t)}), \tag{17}$$

where we iteratively update $\boldsymbol{u}^{(t)}$ and $\boldsymbol{v}$ as $\mathbf{1}_N/(\boldsymbol{Q}^\lambda \boldsymbol{v}^{(t-1)})$ and $N_\pi/(\boldsymbol{Q}^\lambda \boldsymbol{u}^{(t)})$ at iteration t with N_π representing the the total mass of the target distribution and the initial value setting as $\boldsymbol{v}^{(0)} = \mathbf{1}_N$. During the training process, we fixed $\hat{\boldsymbol{P}}$ and align $\boldsymbol{Q}$ with the $\hat{\boldsymbol{P}}$ as the clustering loss:

$$\mathcal{L}_{\text{clu}} = \text{KL}(\hat{\boldsymbol{P}}\|\boldsymbol{Q}) = \sum\nolimits_{i=1}^{N}\sum\nolimits_{c=1}^{C} \hat{p}_{ic}\log\frac{\hat{p}_{ic}}{q_{ic}}, \tag{18}$$

2.6 Overall Optimization

To preserve the intercellular structure information for cell clustering, we reconstruct the cell graph with the learned cell embedding and the Mean Squared Error (MSE) loss can be:

$$\mathcal{L}_{\text{rec}} = \|\boldsymbol{A} - \boldsymbol{M}\|_F^2, \ \boldsymbol{M}_{ij} = \frac{1}{2}\left(\frac{\hat{\boldsymbol{z}}_i \cdot \hat{\boldsymbol{z}}_j}{\|\hat{\boldsymbol{z}}_i\|\|\hat{\boldsymbol{z}}_j\|} + 1\right). \tag{19}$$

Since scRNA-seq data exhibits sparsity and overdispersion, we further incorporate the Zero-Inflated Negative Binomial (ZINB) loss [Eraslan *et al.*, 2019; Xu *et al.*, 2025] to model the excess zeros and variability. Finally, the overall objective of our `scGTN` can be defined as:

$$\mathcal{L} = \mathcal{L}_{\text{clu}} + \alpha\mathcal{L}_{\text{cor}} + \beta\mathcal{L}_{\text{rec}} + \gamma\mathcal{L}_{\text{ZINB}}, \tag{20}$$

where α, β and γ denote the trade-off parameter to balance the contribution of each component.

3 Experiments

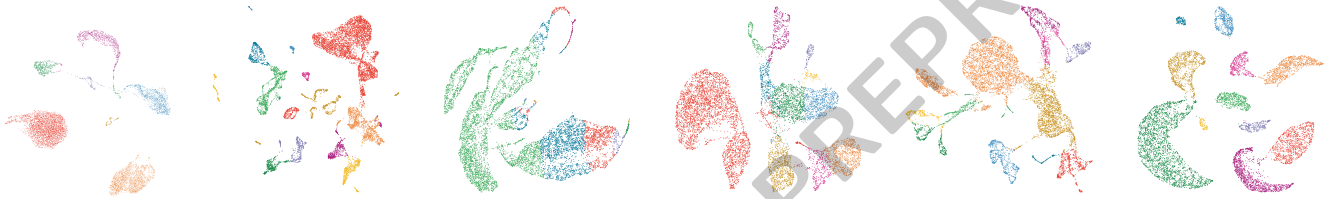
3.1 Experimental Setup

Benchmark Datasets. To evaluate the robustness of our model for clustering across species and diverse biological contexts, we conduct experiments on seven public single-cell benchmark datasets, covering both human and mouse samples and including pancreas, liver, and peripheral immune cell populations. In terms of data scale, the number of cells ranges from 1,064 to 8,617, the number of genes ranges from 2,000 to 20,125, and the expression matrices exhibit sparsity rates of approximately 73.02%–93.26%. To standardize the input dimensionality and focus on more informative features, we select either 1,500 or 2,000 highly variable genes (HVGs) for each dataset as the model input. Detailed statistics are summarized in the Table 5 of Appendix B.

Baselines. To highlight the advantages of `scGTN`, we compare it with ten representative clustering approaches. Specifically, pcaReduce [Žurauskienė and Yau, 2016] is included as a representative traditional clustering method. DEC [Xie *et al.*, 2016], contrastive-sc [Ciortan and Defrance, 2021], scNAME [Wan *et al.*, 2022], scDeepCluster [Tian *et al.*, 2019] and AttentionAE-sc [Li *et al.*, 2023] are deep learning–based clustering methods, which mainly learn representations via autoencoders or contrastive learning and then perform clustering. In addition, scGNN [Wang *et al.*, 2021], scDSC [Gan *et al.*, 2022], scCDCG [Xu *et al.*, 2024] and scSiamese [Xu *et al.*, 2025] are deep graph clustering methods that incorporate GNNs and structural information for representation learning and clustering. Detailed experimental settings and implementation details are provided in the Appendix B.

Datasets	Metric	pcaReduce	DEC	contrastive-sc	scNAME	scDeepCluster	scDSC	AttentionAE-sc	scGNN	scCDCG	scSiameseClu	Ours
Muraro Human Pancreas cells	ACC	50.56 ± 3.4	72.22 ± 5.5	81.85 ± 4.6	80.36 ± 7.9	74.70 ± 2.7	79.42 ± 1.4	93.84 ± 2.7	79.32 ± 4.0	92.65 ± 1.9	94.95 ± 0.1	**96.02 ± 0.5**
	NMI	54.83 ± 1.7	76.29 ± 2.7	77.73 ± 3.3	79.12 ± 2.7	79.32 ± 0.3	75.89 ± 0.6	87.80 ± 2.3	78.76 ± 5.6	86.81 ± 1.0	86.19 ± 0.3	**89.15 ± 1.1**
	ARI	36.61 ± 2.3	61.76 ± 5.6	75.22 ± 10.1	76.64 ± 10.4	64.59 ± 2.5	75.42 ± 1.2	90.74 ± 4.1	78.87 ± 3.2	91.37 ± 1.2	91.59 ± 0.3	**93.10 ± 0.9**
Human Pancreas cells 1	ACC	23.63 ± 0.2	40.64 ± 1.9	69.67 ± 7.8	84.20 ± 6.0	66.39 ± 3.8	73.10 ± 2.3	81.56 ± 1.8	55.10 ± 2.1	92.15 ± 0.8	95.12 ± 0.4	**97.21 ± 0.3**
	NMI	42.58 ± 0.4	62.35 ± 0.6	69.21 ± 3.3	65.74 ± 4.0	69.42 ± 12.5	60.50 ± 4.2	82.62 ± 3.0	64.03 ± 0.9	86.35 ± 0.9	89.50 ± 0.6	**93.23 ± 0.2**
	ARI	1.98 ± 0.5	26.07 ± 1.3	52.00 ± 8.7	63.31 ± 7.4	47.92 ± 3.0	69.81 ± 1.6	71.84 ± 8.5	38.32 ± 0.8	92.83 ± 0.6	93.40 ± 0.5	**96.61 ± 1.5**
Human Pancreas cells 2	ACC	37.56 ± 0.4	45.26 ± 1.8	60.01 ± 5.7	63.33 ± 2.0	62.88 ± 5.5	80.40 ± 4.0	80.40 ± 4.0	57.62 ± 2.3	84.66 ± 1.1	89.45 ± 0.9	**91.13 ± 0.9**
	NMI	50.81 ± 0.3	66.32 ± 2.1	62.25 ± 2.3	58.05 ± 1.6	73.97 ± 2.2	79.44 ± 1.7	82.16 ± 10.9	79.32 ± 2.3	83.44 ± 2.6	86.80 ± 1.2	**90.63 ± 1.2**
	ARI	19.23 ± 0.5	39.14 ± 1.8	47.10 ± 5.3	32.12 ± 1.5	64.56 ± 7.5	82.85 ± 6.5	73.20 ± 10.5	79.96 ± 1.7	85.30 ± 1.6	89.20 ± 1.1	**92.38 ± 1.4**
Human Pancreas cells 3	ACC	35.92 ± 0.4	43.56 ± 1.5	57.47 ± 5.0	83.20 ± 8.7	73.65 ± 3.1	80.12 ± 10.0	90.96 ± 2.4	67.94 ± 4.6	88.04 ± 0.3	89.45 ± 0.5	**95.31 ± 1.4**
	NMI	51.22 ± 0.3	64.51 ± 0.8	66.84 ± 2.3	80.56 ± 5.6	75.02 ± 1.2	75.87 ± 8.4	87.00 ± 1.8	62.60 ± 1.8	82.21 ± 1.1	86.20 ± 0.6	**91.37 ± 0.3**
	ARI	26.40 ± 4.5	42.73 ± 1.0	47.90 ± 3.4	81.08 ± 5.7	63.31 ± 2.3	82.85 ± 8.5	89.78 ± 2.5	58.41 ± 1.9	90.78 ± 0.6	92.34 ± 0.8	**95.25 ± 0.5**
Mouse Pancreas cells 2	ACC	28.97 ± 0.3	34.19 ± 2.3	59.55 ± 5.7	82.36 ± 8.0	69.91 ± 2.9	77.35 ± 2.3	81.92 ± 6.5	43.03 ± 3.0	93.97 ± 0.2	92.45 ± 0.4	**95.11 ± 1.2**
	NMI	48.93 ± 0.1	54.08 ± 0.9	62.34 ± 2.4	78.00 ± 3.9	59.58 ± 2.6	62.50 ± 0.4	81.88 ± 5.1	55.87 ± 1.9	88.02 ± 0.2	86.50 ± 0.5	**88.59 ± 0.6**
	ARI	12.80 ± 0.1	18.54 ± 1.1	49.75 ± 8.5	76.56 ± 11.4	60.14 ± 3.6	65.81 ± 1.4	82.14 ± 7.8	26.40 ± 2.5	92.61 ± 0.3	90.80 ± 0.6	**93.43 ± 0.8**
CITE-CMBC	ACC	28.19 ± 0.1	41.88 ± 5.4	52.70 ± 5.0	68.06 ± 13.3	70.80 ± 2.5	68.79 ± 0.9	61.18 ± 7.9	66.71 ± 4.0	71.45 ± 1.8	74.70 ± 0.5	**79.20 ± 0.5**
	NMI	28.36 ± 0.2	46.87 ± 9.2	64.83 ± 0.8	63.23 ± 13.2	72.53 ± 0.7	64.21 ± 3.3	62.02 ± 9.6	61.70 ± 3.1	74.77 ± 1.7	68.71 ± 0.7	**76.50 ± 0.4**
	ARI	5.10 ± 0.1	23.55 ± 10.3	47.88 ± 3.7	58.99 ± 21.3	56.23 ± 4.1	52.51 ± 2.2	44.20 ± 11.6	61.20 ± 3.1	61.46 ± 1.4	67.56 ± 1.1	**69.80 ± 0.5**
Human Liver cells	ACC	34.92 ± 0.2	46.32 ± 5.5	79.31 ± 2.1	74.55 ± 5.3	63.44 ± 0.7	70.90 ± 2.2	79.46 ± 6.7	68.65 ± 3.2	75.34 ± 1.7	88.33 ± 1.7	**94.50 ± 0.9**
	NMI	32.07 ± 0.6	52.24 ± 5.0	85.11 ± 0.1	77.42 ± 12.0	74.60 ± 2.5	71.63 ± 2.8	82.86 ± 4.5	62.33 ± 2.1	79.34 ± 2.6	88.82 ± 1.2	**90.15 ± 1.1**
	ARI	6.34 ± 14.4	31.04 ± 4.7	89.96 ± 0.1	79.91 ± 2.9	58.58 ± 10.1	75.34 ± 3.6	79.16 ± 11.8	65.41 ± 1.3	81.26 ± 2.7	91.90 ± 0.5	**94.10 ± 0.5**

Table 1: Clustering performance across seven benchmark datasets. The results are reported as mean ± standard deviation, where the best results are highlighted in **bold** and the second-best are underlined.



(a) scDeepCluster (b) scDSC (c) scGNN (d) scCDCG (e) scSiameseClu (f) scGTN

Figure 3: UMAP visualizations of scGTN and five baseline methods on the Human Liver cells dataset. Each point represents a cell, and colors indicate the predicted cell type.

Evaluation Metrics. To evaluate the effectiveness of the proposed method, we assess clustering performance using three widely adopted metrics: clustering accuracy (ACC), normalized mutual information (NMI) [Strehl and Ghosh, 2002], and adjusted Rand index (ARI) [Vinh *et al.*, 2009].

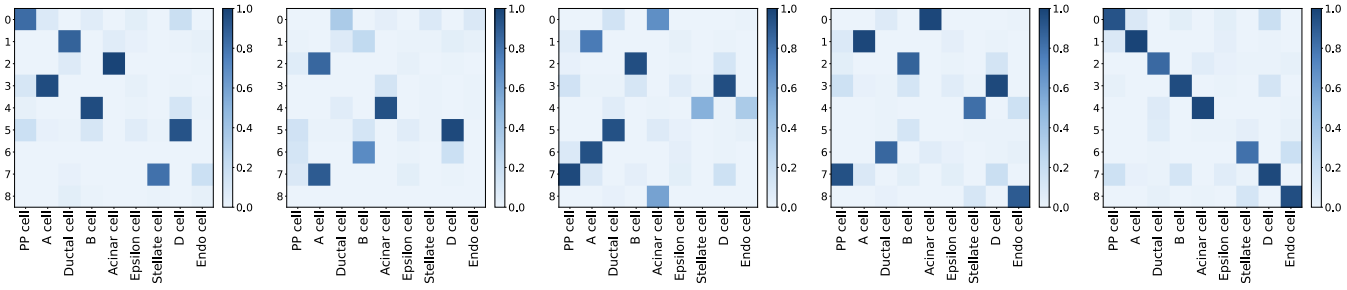
3.2 Overall Performance

Quantitative Analysis. Table 1 compares the clustering performance of scGTN with ten competitive baselines on seven benchmark datasets. From the results, we make the following observations. (1) scGTN achieves the best performance across all three metrics on all datasets, demonstrating its strong capability in handling high-dimensional and highly sparse scRNA-seq data. On average, compared with the second-best method, scGTN yields significant improvements of **3.00%**, **2.42%**, and **2.30%** in terms of ACC, NMI, and ARI, respectively, validating the effectiveness and generalizability of our approach. (2) Compared with deep clustering methods that rely solely on gene expression features, our scGTN consistently delivers substantial gains, highlighting the benefit of jointly leveraging expression features and cell–cell structural information. (3) Although GNN-based methods like scGNN incorporate cell graphs, they predominantly rely on local neighborhood aggregation, which renders them vulnerable to over-smoothing and limits their ability to capture global dependencies. scGTN overcomes this by encoding complex structural information via the Siamese Graph Transformer, yielding more discriminative representations. Analysis of runtime is provided in Appendix F.

Visualization. To further investigate the model's ability to capture the underlying structure of cells in the embedding space, we apply UMAP [McInnes *et al.*, 2020] to project the learned representations of the Human Liver dataset into two dimensions, as shown in Figure 3. We observe that our scGTN produces highly compact clusters with clear inter-cluster boundaries, faithfully reflecting the complex underlying tissue structure and effectively preserving the intrinsic heterogeneity among different cell types. In contrast, methods such as scDSC often yield projections with noticeable subtype mixing or diffuse distributions, making it difficult to separate rare cell populations from major groups.

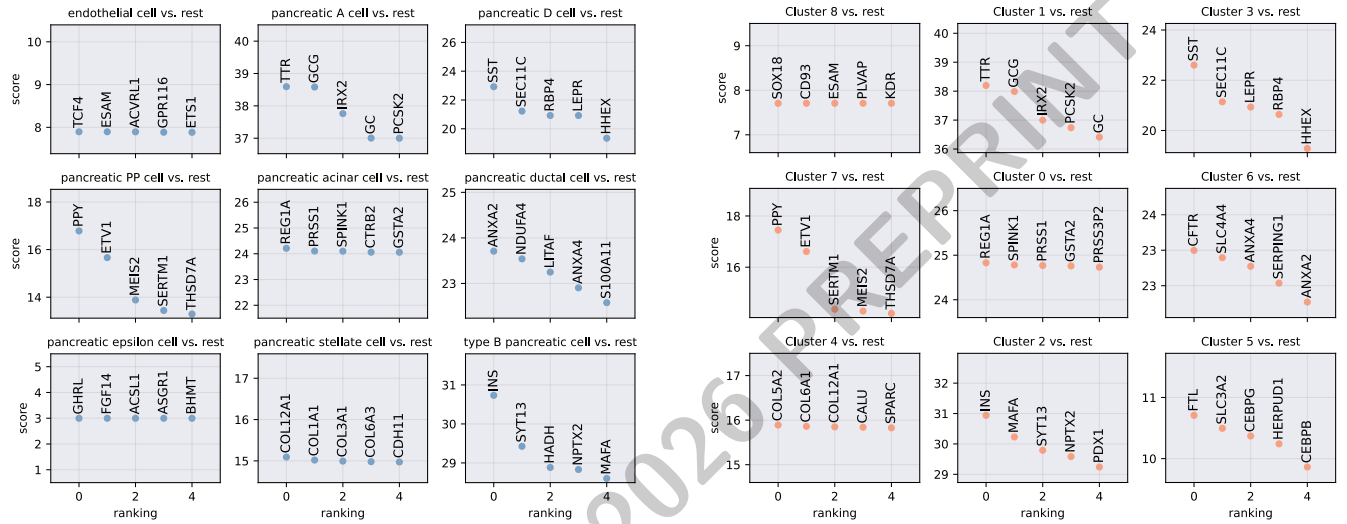
3.3 Biological Analysis

Differential Gene Expression Analysis. Differentially expressed genes (DEGs) serve as a critical metric to validate the biological identity of learned clusters. Using Seurat's "FindAllMarkers" [Butler *et al.*, 2018], we extract the top-100 DEGs for each cluster on the Muraro Human Pancreas dataset and quantitatively compute the overlap ratio with the corresponding ground-truth cell types. As shown in Figure 4, the



(a) scDSC (b) scGNN (c) scCDCG (d) scSiameseClu (e) scGTN

Figure 4: Cell type annotation. Heatmap of overlap between the top 100 DEGs in clusters detected by five methods and gold standard cell types (Cell types are abbreviated as PP, A, Ductal, B, Acinar, Epsilon, Stellate, D, and Endo).



(a) Gold Standard (b) scGTN

Figure 5: Top-5 DEGs Ranking. Comparison of the top 5 DEGs identified by the gold standard (a) and scGTN (b) for different cell types.

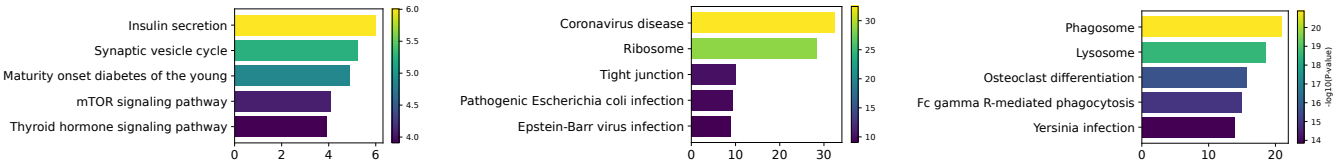
high-overlap regions of scGTN are predominantly concentrated along the diagonal, consistently surpassing 90% overlap in most clusters. This indicates that our clusters can be stably mapped to specific cell types with minimal confusion. In stark contrast, baseline methods frequently exhibit lower overlap ratios due to ambiguity and inter-cluster confusion. To further investigate the biological fidelity, we compare the rankings of the top 5 DEGs between scGTN and the Gold Standard as illustrated in Figure 5. Beyond mere overlap, we observe remarkable consistency in the dominant markers. For instance, in the cluster corresponding to Pancreatic PP cells (Cluster 7), our scGTN identifies *PPY* and *ETV1* as the top-two markers, perfectly matching the Gold Standard. This precise alignment in both gene identity and ranking confidence demonstrates that our scGTN does not merely capture global similarities but accurately prioritizes the key marker genes essential for distinct cell identification.

KEGG Pathway Enrichment Analysis. To validate the biological interpretability of the clustering results, we perform KEGG pathway enrichment analysis on representative clusters. As shown in Figure 6, our scGTN successfully resolves distinct functional subpopulations in pancreatic tissue. Cluster 3 is significantly enriched in "Insulin secretion" and critical diabetes-related pathways including MODY, while also exhibiting neuroendocrine signatures such as the "Synaptic vesicle cycle", confirming its biological identity as pancreatic β cells. Cluster 4 shows strong enrichment in "Ribosome" and protein synthesis-related pathways, highlighting the high protein synthesis activity characteristic of acinar cells. Finally, Cluster 5 is enriched in pathways such as "Phagosome" and "Lysosome", indicating the presence of tissue-resident immune cells such as macrophages. These results demonstrate that our scGTN can accurately identify KEGG pathways and distinct functional phenotypes.

Furthermore, we provide trajectory inference analysis in Appendix D to validate preserved developmental manifolds.

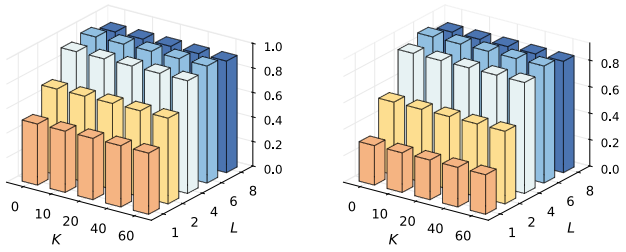
3.4 Ablation Study

Effectiveness of Major Components. We evaluate the contribution of each loss term on the Muraro dataset. Table 2 indicates that the exclusion of any component leads to a sharp



(a) Cluster 3 (b) Cluster 4 (c) Cluster 5

Figure 6: Top 5 KEGG pathways for each cluster (bar graph shows $-\log_{10}(q)$).



(a) ACC (b) NMI

Figure 7: The impact of two key hyperparameters (K and L) on clustering performance in terms of ACC and NMI.

$\mathcal{L}_{\text{clu}}$	$\mathcal{L}_{\text{cor}}$	$\mathcal{L}_{\text{ZINB}}$	$\mathcal{L}_{\text{rec}}$	ACC	NMI	ARI
✓	✓	✓	×	84.57	78.93	76.08
✓	✓	×	✓	84.18	78.36	75.62
✓	×	✓	✓	83.47	77.58	74.36
×	✓	✓	✓	81.24	74.79	70.68
✓	✓	✓	✓	**96.02**	**89.15**	**93.10**

Table 2: Ablation experiment results of major loss terms

performance degradation of over 10% in ACC, highlighting the non-redundancy of each module. Specifically, the absence of the optimal transport clustering loss $\mathcal{L}_{\text{clu}}$ results in the most severe drop, underscoring its critical role in guiding partition assignments. Similarly, removing the correlation term $\mathcal{L}_{\text{cor}}$ or the reconstruction objectives $\mathcal{L}_{\text{ZINB}}$ and $\mathcal{L}_{\text{rec}}$ significantly impairs the capability of the model to align views and preserve topology. Ultimately, the full scGTN achieves superior performance, validating the synergy of jointly optimizing for structural, feature, and distribution constraints. Dual-component ablation analysis is provided in Appendix G.

Impact of Embedding Components. We further dissect the embedding module to assess the contributions of Gene Expression (GE), Position (Pos), and Shortest-Path (SP) embeddings in Table 3. Results indicate that while GE provides the foundational semantics, integrating it with structural priors yields massive gains. Notably, adding Pos alone boosts ACC by over 12% compared to the GE baseline, and the full combination achieves peak performance, confirming that structural information is essential for distinguishing subtle cell states.

Model Variants	ACC	NMI	ARI
scGTN	**96.02**	**89.15**	**93.10**
GE + Pos	92.12	86.33	88.55
GE + SP	80.09	71.08	67.76
Pos + SP	27.52	12.13	5.06
Gene Expression (GE)	79.24	70.69	67.61
Position (Pos)	16.68	0.73	0.23
Shortest Path (SP)	26.53	12.14	6.80

Table 3: Ablation experiment results of embedding components.

3.5 Sensitivity Analysis

To evaluate the robustness of scGTN against hyperparameter variations, we conduct a sensitivity analysis on the Muraro Human Pancreas dataset. We examine neighborhood size K (5–60) and Transformer layers L (1–8). As illustrated in Figure 7, the model demonstrates high stability with respect to K, showing only marginal fluctuations across all metrics, indicating that our scGTN is robust to variations in the underlying graph construction. Conversely, performance depends more significantly on network depth L. Shallow networks where $L = 1$ yield suboptimal scores due to insufficient feature propagation and limited semantic capture. Increasing depth improves results until peaking at $L = 6$, where the model effectively balances local and global information. Further extending depth to 8 leads to a slight performance drop, likely due to over-smoothing. Further sensitivity analyses regarding other hyperparameters are detailed in Appendix E.

4 Conclusion

In this paper, we propose a novel deep Siamese Graph Transformer Network (scGTN), which effectively integrates gene expression profiles and intercellular structural information for scRNA-seq data clustering. Specifically, we formulate the scRNA-seq data as a graph and construct two augmented graphs as dual views to capture complementary intercellular relationships. Then, we introduce a Siamese graph transformer network which integrates richer graph structural information, i.e., shortest-path and node-wise distance, to explicitly capture intercellular relationships. To further refine the cell clustering, we incorporate an optimal transport mechanism in a self-supervised manner, enhancing representation learning while preserving biological relevance. Extensive experiments on scRNA-seq benchmark datasets verify the effectiveness and superiority of our proposed scGTN.

Contribution Statement

Jinke Wu and Yifan Wang contributed equally to this work.

Acknowledgments

This work is supported in part by the National Natural Science Foundation of China under Grant 62306014 and 12501344, Postdoctoral Fellowship Program (Grade A) of CPSF under Grant BX20250376 and BX20240239, China Postdoctoral Science Foundation under Grant 2024M762201, Sichuan Science and Technology Program under Grant 2025ZNSFSC1506 and 2025ZNSFSC0808, Sichuan University Interdisciplinary Innovation Fund.

References

[Benesty *et al.*, 2009] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. Pearson correlation coefficient. In *Noise reduction in speech processing*, pages 1–4. Springer, 2009.

[Bo *et al.*, 2020] Deyu Bo, Xiao Wang, Chuan Shi, Meiqi Zhu, Emiao Lu, and Peng Cui. Structural deep clustering network. In *Proceedings of the Web Conference*, pages 1400–1410, 2020.

[Butler *et al.*, 2018] Andrew Butler, Paul Hoffman, Peter Smibert, Efthymia Papalexi, and Rahul Satija. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology*, 36(5):411–420, 2018.

[Ciortan and Defrance, 2021] Madalina Ciortan and Matthieu Defrance. Contrastive self-supervised clustering of scrna-seq data. *BMC bioinformatics*, 22(1):280, 2021.

[Dai *et al.*, 2022] Min Dai, Xiaobing Pei, and Xiu-Jie Wang. Accurate and fast cell marker gene identification with cosg. *Briefings in bioinformatics*, 23(2):bbab579, 2022.

[Eraslan *et al.*, 2019] Gökcen Eraslan, Lukas M Simon, Maria Mircea, Nikola S Mueller, and Fabian J Theis. Single-cell rna-seq denoising using a deep count autoencoder. *Nature communications*, 10(1):390, 2019.

[Fan *et al.*, 2026] Boyang Fan, Hengchuang Yin, Siyu Yi, Yifan Wang, Zhicheng Li, Leijiyu Zhou, Jiancheng Lv, and Wei Ju. Cmgl: Confidence-guided multi-omics graph learning for cancer subtype classification. *arXiv preprint arXiv:2604.24201*, 2026.

[Gan *et al.*, 2022] Yanglan Gan, Xingyu Huang, Guobing Zou, Shuigeng Zhou, and Jihong Guan. Deep structural clustering for single-cell rna-seq data jointly through autoencoder and graph neural network. *Briefings in Bioinformatics*, 23(2), 2022.

[Hicks *et al.*, 2021] Stephanie C Hicks, Ruoxi Liu, Yuwei Ni, Elizabeth Purdom, and Davide Risso. mbkmeans: Fast clustering for single cell data using mini-batch k-means. *PLoS computational biology*, 17(1):e1008625, 2021.

[Hu *et al.*, 2020] Jian Hu, Xiangjie Li, Gang Hu, Yafei Lyu, Katalin Susztak, and Mingyao Li. Iterative transfer learning with neural network for clustering and cell type classification in single-cell rna-seq analysis. *Nature machine intelligence*, 2(10):607–618, 2020.

[Johnson, 1967] Stephen C Johnson. Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254, 1967.

[Ju *et al.*, 2023] Wei Ju, Yiyang Gu, Binqi Chen, Gongbo Sun, Yifang Qin, Xingyuming Liu, Xiao Luo, and Ming Zhang. Glcc: A general framework for graph-level clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 4391–4399, 2023.

[Ju *et al.*, 2025] Wei Ju, Siyu Yi, Yifan Wang, Zhiping Xiao, Zhengyang Mao, Hourun Li, Yiyang Gu, Yifang Qin, Nan Yin, Senzhang Wang, et al. A survey of graph neural networks in real world: Imbalance, noise, privacy and ood challenges. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025.

[Ju *et al.*, 2026] Wei Ju, Siyu Yi, Kangjie Zheng, Yifan Wang, Ziyue Qiao, Li Shen, Yongdao Zhou, Xiaochun Cao, and Jiancheng Lv. Compactness and consistency: A conjoint framework for deep graph clustering. In *The Fourteenth International Conference on Learning Representations*, 2026.

[Kingma, 2014] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[Kiselev *et al.*, 2019] Vladimir Yu Kiselev, Tallulah S Andrews, and Martin Hemberg. Challenges in unsupervised clustering of single-cell rna-seq data. *Nature Reviews Genetics*, 20(5):273–282, 2019.

[Li *et al.*, 2018] Qimai Li, Zhichao Han, and Xiao-Ming Wu. Deeper insights into graph convolutional networks for semi-supervised learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[Li *et al.*, 2023] Shenghao Li, Hui Guo, Simai Zhang, Yizhou Li, and Menglong Li. Attention-based deep clustering method for scRNA-seq cell type identification. *PLOS Computational Biology*, 19(11):e1011641, 2023.

[Lin *et al.*, 2017] Peijie Lin, Michael Troup, and Joshua WK Ho. Cidr: Ultrafast and accurate clustering through imputation for single-cell rna-seq data. *Genome biology*, 18(1):59, 2017.

[Luo *et al.*, 2021] Zixiang Luo, Chenyu Xu, Zhen Zhang, and Wenfei Jin. A topology-preserving dimensionality reduction method for single-cell rna-seq data using graph autoencoder. *Scientific reports*, 11(1):20028, 2021.

[Maaten and Hinton, 2008] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[McInnes *et al.*, 2020] Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2020.

[Ou-Yang *et al.*, 2022] Le Ou-Yang, Fan Lu, Zi-Chao Zhang, and Min Wu. Matrix factorization for biomedical link prediction and scrna-seq data imputation: an empirical survey. *Briefings in Bioinformatics*, 23(1):bbab479, 2022.

[Peng *et al.*, 2020] Lihong Peng, Xiongfei Tian, Geng Tian, Junlin Xu, Xin Huang, Yanbin Weng, Jialiang Yang, and Liqian Zhou. Single-cell rna-seq clustering: datasets, models, and algorithms. *RNA biology*, 17(6):765–783, 2020.

[Petegrosso *et al.*, 2020] Raphael Petegrosso, Zhuliu Li, and Rui Kuang. Machine learning and statistical methods for clustering single-cell rna-sequencing data. *Briefings in bioinformatics*, 21(4):1209–1223, 2020.

[Ren *et al.*, 2025] Tao Ren, Haodong Zhang, Yifan Wang, Wei Ju, Chengwu Liu, Fanchun Meng, Siyu Yi, and Xiao Luo. Mhgc: Multi-scale hard sample mining for contrastive deep graph clustering. *Information Processing & Management*, 62(4):104084, 2025.

[Sinkhorn, 1967] Richard Sinkhorn. Diagonal equivalence to matrices with prescribed row and column sums. *The American Mathematical Monthly*, 74(4):402–405, 1967.

[Stegle *et al.*, 2015] Oliver Stegle, Sarah A Teichmann, and John C Marioni. Computational and analytical challenges in single-cell transcriptomics. *Nature Reviews Genetics*, 16(3):133–145, 2015.

[Street *et al.*, 2018] Kelly Street, Davide Risso, Russell B Fletcher, Diya Das, John Ngai, Nir Yosef, Elizabeth Purdom, and Sandrine Dudoit. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics*, 19(1):477, 2018.

[Strehl and Ghosh, 2002] Alexander Strehl and Joydeep Ghosh. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of Machine Learning Research*, 3:583–617, 2002.

[Tian *et al.*, 2019] Tian Tian, Ji Wan, Qi Song, and Zhi Wei. Clustering single-cell RNA-seq data with a model-based deep learning approach. *Nature Machine Intelligence*, 1(4):191–198, 2019.

[Tian *et al.*, 2021] Tian Tian, Jie Zhang, Xiang Lin, Zhi Wei, and Hakon Hakonarson. Model-based deep embedding for constrained clustering analysis of single cell RNA-seq data. *Nature Communications*, 12(1):1873, 2021.

[Vinh *et al.*, 2009] Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information theoretic measures for clusterings comparison: Is a correction for chance necessary? In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 1073–1080, 2009.

[Wan *et al.*, 2022] Hui Wan, Liang Chen, and Minghua Deng. scNAME: Neighborhood contrastive clustering with ancillary mask estimation for scRNA-seq data. *Bioinformatics*, 38(6):1575–1583, 2022.

[Wang *et al.*, 2021] Juexin Wang, Anjun Ma, Yuzhou Chang, Jianting Gong, Yuexu Jiang, Ren Qi, Cankun Wang, Hongjun Fu, Qin Ma, and Dong Xu. scgnn is a novel graph neural network framework for single-cell rna-seq analyses. *Nature communications*, 12(1):1882, 2021.

[Wang *et al.*, 2025] Qianqian Wang, Haiming Xu, Zihao Zhang, Wei Feng, and Quanxue Gao. Deep multi-modal graph clustering via graph transformer network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 7835–7843, 2025.

[Wolf *et al.*, 2019] F Alexander Wolf, Fiona K Hamey, Mireya Plass, Jordi Solana, Joakim S Dahlin, Berthold Göttgens, Nikolaus Rajewsky, Lukas Simon, and Fabian J Theis. Paga: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome biology*, 20(1):59, 2019.

[Wu and Ma, 2022] Wenming Wu and Xiaoke Ma. Network-based structural learning nonnegative matrix factorization algorithm for clustering of scrna-seq data. *IEEE/ACM transactions on computational biology and bioinformatics*, 20(1):566–575, 2022.

[Xie *et al.*, 2016] Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *Proceedings of the 33rd International Conference on Machine Learning*, pages 478–487, 2016.

[Xu *et al.*, 2024] Ping Xu, Zhiyuan Ning, Meng Xiao, Guihai Feng, Xin Li, Yuanchun Zhou, and Pengfei Wang. sccdcg: efficient deep structural clustering for single-cell rna-seq via deep cut-informed graph embedding. In *International Conference on Database Systems for Advanced Applications*, pages 172–187. Springer, 2024.

[Xu *et al.*, 2025] Ping Xu, Zhiyuan Ning, Pengjiang Li, Wenhao Liu, Pengyang Wang, Jiaxu Cui, Yuanchun Zhou, and Pengfei Wang. scsiameseclu: A siamese clustering framework for interpreting single-cell rna sequencing data. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 7867–7875, 2025.

[Yu *et al.*, 2022] Zhuohan Yu, Yifu Lu, Yunhe Wang, Fan Tang, Ka-Chun Wong, and Xiangtao Li. Zinb-based graph embedding autoencoder for single-cell rna-seq interpretations. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 4671–4679, 2022.

[Zhang *et al.*, 2026a] Haodong Zhang, Xinyue Wang, Tao Ren, Yifan Wang, Siyu Yi, Fanchun Meng, Zeyu Ma, Qingqing Long, and Wei Ju. Fairgc: Fostering individual and group fairness for deep graph clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 28194–28202, 2026.

[Zhang *et al.*, 2026b] Wei Zhang, Siyu Yi, Lezhi Chen, Yifan Wang, Ziyue Qiao, Yongdao Zhou, and Wei Ju. Evidence-aware integration and domain identification of spatial transcriptomics data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 16352–16360, 2026.

[Žurauskienė and Yau, 2016] Justina Žurauskienė and Christopher Yau. pcareduce: hierarchical clustering of single cell transcriptional profiles. *BMC bioinformatics*, 17(1):140, 2016.