GBFlow: Grouping Belief-Guided Dual Normalizing Flows for Accurate Spatial Domain Delineation

Fengyi Zhou[1] , **Daoyuan Wang**[1] , **Wenlan Chen**[1] , **Cheng Liang**[2*] and **Fei Guo**[1*]
[1]School of Computer Science and Engineering, Central South University, Changsha, 410083, China
[2]School of Computer Science and Artificial Intelligence, Shandong Normal University, Jinan, 250358, China
{254711073, 244701040, 244701041}@csu.edu.cn, alcs417@sdnu.edu.cn, guofei@csu.edu.cn

Abstract

Existing spatial domain identification methods primarily use graph neural networks to model spatial and transcriptional relationships. However, their performance is highly sensitive to noisy affinity graphs. Moreover, graph autoencoders tend to over-constrain latent representations, which limits their ability to capture global variability in spatial multi-omics data. To overcome these issues, we propose a dual-flow latent refinement framework that simultaneously integrates structure-aware and structure-free transformations. Specifically, a graph normalizing flow is employed to enforce relational consistency, while a parallel vanilla normalizing flow preserves global distributional flexibility. Features learned by the two flows are then adaptively fused to obtain a robust and unified latent representations. In addition, we introduce a grouping belief-based affinity refinement strategy to suppress unreliable connections and strengthen confident neighborhood relationships, which provides a more stable structural prior for representation learning. Extensive experiments on multiple spatial multi-omics datasets show that the proposed method consistently outperforms state-of-the-art approaches and achieves more accurate and robust spatial domain identification. The supplementary material is publicly available at https://github.com/LiangSDNULab/GBFlow.

1 Introduction

Spatially resolved multi-omics technologies enable the joint measurement of molecular profiles and spatial coordinates, providing powerful means to investigate tissue organization, cellular heterogeneity, and microenvironmental interactions [Moses and Pachter, 2022; Duan *et al.*, 2023]. A key analytical task in this context is spatial domain identification, which aims to delineate coherent tissue regions by jointly leveraging molecular similarity and spatial proximity [Cai and Li, 2025; Pham *et al.*, 2023].

Most existing approaches rely on graph neural networks or graph autoencoders to model spatial and transcriptional relationships [Hu *et al.*, 2021; Li *et al.*, 2022]. These methods construct affinity graphs based on spatial proximity, molecular similarity, or both, and perform neighborhood aggregation to capture local context. Representative examples include HyperGCN [Ma *et al.*, 2024], which models higher-order spatial interactions via hypergraphs, and SpaMGCN [Zhang *et al.*, 2025b], which employs multi-scale adaptive graph convolutions for spatial multi-omics integration. Overall, these studies demonstrate that graph-based structural priors can substantially enhance spatial domain delineation.

However, the effectiveness of graph-based models critically depends on the quality of the predefined affinity graph [Kang *et al.*, 2025; Lu *et al.*, 2025]. In practice, spatial graphs are often constructed using heuristic rules such as k-nearest neighbors or fixed-radius thresholds, which may fail to reflect true biological neighborhoods [Gong *et al.*, 2025]. As a result, spurious cross-region connections frequently arise, especially near tissue boundaries or in heterogeneous regions with complex cellular compositions. Such noisy or misleading edges can propagate erroneous signals across distinct tissue domains, leading to representation distortion, over-smoothing, and degraded clustering performance[Jiang *et al.*, 2025b; Zhang *et al.*, 2025a]. Moreover, once the graph structure is fixed, most GNN-based methods lack mechanisms to adaptively correct unreliable connections during training, making them highly sensitive to graph construction choices. This issue is also widely observed in general clustering tasks, where separate graph construction and clustering learning limits performance and robustness[Jiang *et al.*, 2025a].

More fundamentally, graph autoencoders tend to over-regularize latent representations, biasing them toward the imposed topology and limiting their capacity to capture global distributional variability. Conversely, structure-agnostic models lack explicit mechanisms to enforce spatial coherence. Achieving a principled balance between structure-awareness and distributional flexibility remains an open challenge.

To address these challenges, we propose GBFlow, a grouping belief-guided dual normalizing flow framework for spatial domain identification. GBFlow refines latent representations through two complementary flows: a structure-aware flow that enforces relational consistency using a re-

*Corresponding author.

fined affinity graph, and a parallel structure-free flow that preserves global distributional flexibility. The outputs are adaptively fused to produce robust view-specific embeddings that capture both spatial coherence and molecular similarity. Additionally, a grouping belief-based affinity refinement (GBAR) strategy suppresses unreliable connections and strengthens confident neighborhoods. To handle multi-view spatial omics data, GBFlow integrates preliminary feature extraction, cross-graph feature integration, inter-omics attention, and dynamic graph learning, including structure-aware contrastive learning to enhance global consistency while preserving local diversity. These modules collectively enable progressive refinement of both view-specific and global affinity graphs, balancing structural constraints with distributional flexibility.

Our main contributions are summarized as follows:

- We propose a dual-flow latent refinement framework that jointly integrates structure-aware and structure-free normalizing flows, enabling a flexible yet consistent representation of spatial multi-omics data.
- We introduce a grouping belief-based affinity refinement strategy that effectively suppresses noisy connections and enhances reliable neighborhood relations.
- Extensive experiments on multiple spatial multi-omics datasets demonstrate that GBFlow consistently outperforms state-of-the-art methods in both accuracy and robustness.

2 Related Work

2.1 Single-omics Spatial Domain Identification

Early studies on spatial domain identification primarily relied on classical clustering methods applied to gene expression profiles, sometimes combined with spatial smoothing [Sun *et al.*, 2025; Xu *et al.*, 2025]. While simple and interpretable, these approaches often fail to explicitly model spatial relationships, resulting in fragmented or biologically inconsistent domain assignments.

More recent methods increasingly adopt graph neural networks or graph autoencoders to capture both spatial and transcriptional dependencies. Representative approaches include STAGATE [Dong and Zhang, 2022] and GraphST [Long *et al.*, 2023], as well as more advanced models such as SpaMGCN [Zhang *et al.*, 2025b] and HyperGCN [Ma *et al.*, 2024]. Although these methods achieve improved performance, their effectiveness is highly sensitive to the quality of the predefined affinity graph. Some works attempt to mitigate this issue through graph refinement or structure learning, such as UnGSL [Han *et al.*, 2025] and ARO-GNN [Lu and Lin, 2025], but they still rely on a single graph-based transformation and may over-regularize latent representations, limiting their ability to capture global variability.

2.2 Multi-omics Spatial Domain Identification

With the increasing availability of spatial multi-omics data, recent methods aim to integrate heterogeneous modalities to improve spatial domain delineation. A first line of work incorporates histological image information into spatial transcriptomics analysis. SiGra [Tang *et al.*, 2023] integrates gene expression, immunohistochemistry images, and spatial information via a hybrid graph transformer, while STMDA [Wang *et al.*, 2024] employs modality-specific GCNs followed by cross-modal distribution alignment to enhance cellular niche identification.

Beyond image-based integration, other approaches focus on jointly modeling multiple molecular modalities without relying on imaging data. SpatialGlue [Long *et al.*, 2024] constructs omics-specific graphs and integrates them through attention-based cross-modal fusion. PRAGA [Huang *et al.*, 2025] captures dynamic semantic relations across modalities via latent feature graphs, combining spatial priors with a prototype-based contrastive learning framework. While effective, existing spatial multi-omics methods often treat different graphs independently or rely on heuristic fusion strategies, and few explicitly address graph noise or the trade-off between structural consistency and distributional flexibility.

In contrast, our proposed GBFlow explicitly decouples structure-aware and structure-free transformations via dual normalizing flows, and introduces a grouping belief-based affinity refinement strategy to suppress unreliable connections, enabling more robust spatial domain identification.

3 Methods

Given a spatial multi-omics dataset consisting of V views, we denote the multi-view feature matrices as $\mathbf{X} = \{\mathbf{X}^{(v)} \in \mathbb{R}^{n\times d_v}\}_{v=1}^{V}$, where n is the number of spatial spots (or cells), and d_v is the feature dimension of the v-th view. The central objective of our framework is to learn a robust view-specific latent representations $\mathbf{Z}^{(v)} \in \mathbb{R}^{n\times l}$ for each omics view, along with a unified integrated representation $\mathbf{Z} \in \mathbb{R}^{n\times l}$. These representations are designed to effectively capture both spatial coherence and molecular similarity, thereby enabling accurate identification of spatial domains through clustering. At the core of the proposed framework (named GBFlow) lies the construction and progressive refinement of multiple graph structures. For each view v, we construct two initial graphs: the spatial proximity graph $\mathbf{A}_1^{(v)}$ and the feature similarity graph $\mathbf{A}_2^{(v)}$. These are subsequently refined and integrated into a fused graph $\mathbf{A}_3^{(v)}$, which is further evolved into the final refined affinity graph $\mathbf{G}^{(v)}$. Key learnable components of the model include: the dual-branch normalizing flow outputs $\hat{\mathbf{Z}}^{(v)}_{\text{graph}}$ and $\hat{\mathbf{Z}}^{(v)}_{\text{plain}}$, their adaptively fused representation $\tilde{\mathbf{Z}}^{(v)}$, and the dynamic self-expression graph $\mathbf{H}^{(v)}$ employed for subspace modeling.

The overall architecture of GBFlow integrates four major modules: (1) Preliminary Feature Learning and Integration, (2) Dual-flow Latent Refinement, (3) Grouping Belief-based Affinity Refinement, and (4) Dynamic Graph Learning. All components are jointly optimized under a unified loss function $\mathcal{L}$.

3.1 Preliminary Feature Learning and Integration

Feature Learning

To effectively capture the heterogeneous information in spatial transcriptomics data, we construct multiple complementary graph structures that jointly model spatial proximity and

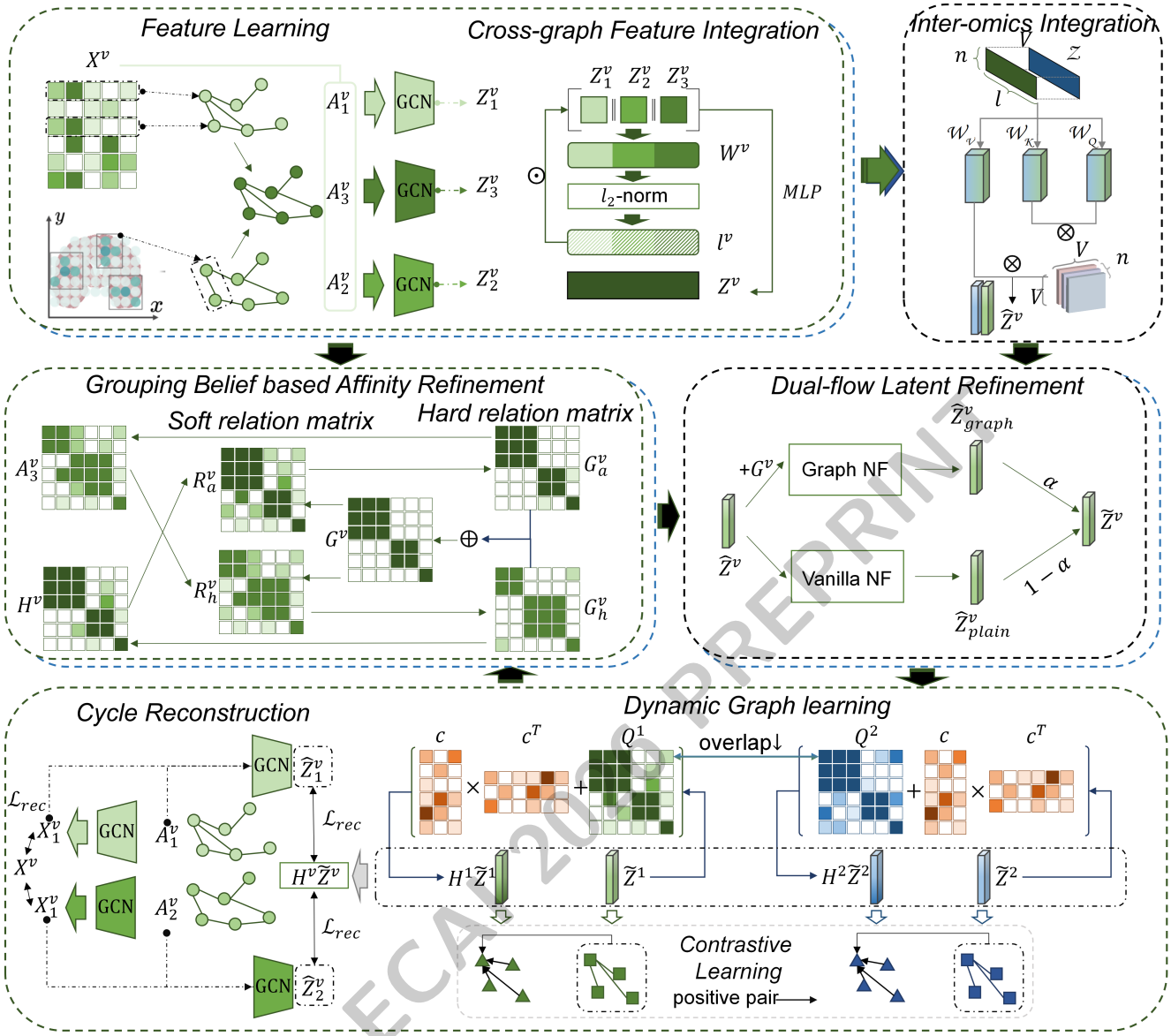

Figure 1: The framework of GBFlow. This model consists of seven core modules, namely Feature Learning, Cross-graph Feature Integration, Inter-omics Integration, Grouping Belief based Affinity Refinement, Dual-flow Latent Refinement, Dynamic Graph Learning, and Cycle Reconstruction. The Dynamic Graph Learning contains a contrastive learning module. The contrastive learning module is used to enhance global consistency and local diversity.

transcriptional similarity. For each view v, we first construct a spatial neighborhood graph $\mathbf{A}_1^{(v)}$ based on the Euclidean distances between spatial coordinates. Recognizing that spatial proximity does not always imply functional similarity, we additionally build a feature-based similarity graph $\mathbf{A}_2^{(v)}$ in the gene expression space. To mitigate spurious connections across distinct regions, we refine the spatial graph by incorporating transcriptional similarity:

$$\mathbf{A}_1^{(v)} = \mathbf{A}_1^{(v)} \odot \mathbf{A}_2^{(v)}, \tag{1}$$

where $\odot$ denotes element-wise multiplication.

We then obtain a fused graph that balances spatial continuity and expression consistency as follows:

$$\mathbf{A}_3^{(v)} = (1 - \eta^{(v)})\mathbf{A}_1^{(v)} + \eta^{(v)}\mathbf{A}_2^{(v)}, \tag{2}$$

where $\eta^{(v)} \in [0, 1]$ is a learnable parameter.

Using the spatial graph $\mathbf{A}_1^{(v)}$, feature similarity graph $\mathbf{A}_2^{(v)}$, and the fused graph $\mathbf{A}_3^{(v)}$, we employ three graph convolutional encoders to extract view-specific latent representations:

$$\mathbf{Z}_k^{(v)} = F_{\theta_k^{(v)}}(\mathbf{X}^{(v)}, \mathbf{A}_k^{(v)}), \tag{3}$$

where $F_{\theta_k^{(v)}}(\cdot)$ denotes a graph convolutional operation.

Cross-graph Feature Integration

The latent representations obtained from the three graph encoders are first concatenated:

$$(\mathbf{Z}')^{(v)} = [\mathbf{Z}_1^{(v)} \| \mathbf{Z}_2^{(v)} \| \mathbf{Z}_3^{(v)}]. \tag{4}$$

To capture inter-feature correlations and adaptively integrate information from different graph views, we apply learnable importance weights followed by a projection layer:

$$\mathbf{Z}^{(v)} = F\left([u_1^{(v)} \odot \mathbf{Z}_1^{(v)} \| u_2^{(v)} \odot \mathbf{Z}_2^{(v)} \| u_3^{(v)} \odot \mathbf{Z}_3^{(v)}]\right), \tag{5}$$

where $u_k^{(v)}$ represents the learned importance coefficient for the representation derived from the k-th graph, and $F(\cdot)$ is a linear transformation.

Inter-omics Integration

To capture cross-modal dependencies across different omics views, we apply a multi-head self-attention mechanism over the view-specific representations. The outputs of all attention heads are concatenated:

$$\hat{\mathcal{Z}} = \mathrm{Concat}\left(\tilde{\mathcal{Z}}_1, \tilde{\mathcal{Z}}_2, \cdots, \tilde{\mathcal{Z}}_H\right), \tag{6}$$

where $\tilde{\mathcal{Z}}_h$ denotes the output of the h-th attention head. The final refined representation for each omics view is then obtained by splitting $\hat{\mathcal{Z}}$ along the modality dimension.

3.2 Dual-flow Latent Refinement

Normalizing flows offer an invertible and probabilistic framework for mapping a simple base distribution to a complex target distribution via a sequence of bijective transformations. We extend this paradigm to spatial multi-omics data by introducing a dual-flow architecture that explicitly disentangles relational (graph-structured) consistency from global distributional flexibility.

Graph Normalizing Flow for Structured Transformation

The structure-aware branch employs a Graph Normalizing Flow (GNF) [Liu *et al.*, 2019], which incorporates reversible message passing to enable exact likelihood computation on graph-structured data while preserving invertibility. Following the GRevNet design, at each layer t, the node feature matrix $\mathbf{Z}_t^{(v)}$ is split along the feature dimension into $\mathbf{Z}_t^{(0)}$ and $\mathbf{Z}_t^{(1)}$. A reversible transformation step consists of two half-steps:

$$\begin{aligned}
\mathbf{Z}_{t+\frac{1}{2}}^{(0)} &= \mathbf{Z}_t^{(0)} \odot \exp\left(F_1\left(\mathbf{Z}_t^{(1)}, \mathbf{G}^{(v)}\right)\right) + F_2\left(\mathbf{Z}_t^{(1)}, \mathbf{G}^{(v)}\right), \\
\mathbf{Z}_{t+1}^{(1)} &= \mathbf{Z}_{t+\frac{1}{2}}^{(1)} \\
&\quad \odot \exp\left(G_1\left(\mathbf{Z}_{t+\frac{1}{2}}^{(1)}, \mathbf{G}^{(v)}\right)\right) + G_2\left(\mathbf{Z}_{t+\frac{1}{2}}^{(1)}, \mathbf{G}^{(v)}\right) \\
\mathbf{Z}_{t+\frac{1}{2}}^{(1)} &= \mathbf{Z}_t^{(1)}, \mathbf{Z}_{t+1}^{(0)} = \mathbf{Z}_{t+\frac{1}{2}}^{(0)}
\end{aligned} \tag{7}$$

where F_1, F_2, G_1, G_2 are graph-conditioned transformation functions implemented as Graph Neural Networks (GNNs) that operate on the refined affinity matrix $\mathbf{G}^{(v)}$. This formulation ensures that the entire transformation is invertible:

$$\begin{aligned}
\mathbf{Z}_{t+1/2}^{(0)} &= \mathbf{Z}_{t+1}^{(0)}, \mathbf{Z}_t^{(1)} = \mathbf{Z}_{t+1/2}^{(1)} \\
\mathbf{Z}_{t+1/2}^{(1)} &= \frac{\left(\mathbf{Z}_{t+1}^{(1)} - G_2\left(\mathbf{Z}_{t+1/2}^{(1)}, \mathbf{G}^{(v)}\right)\right)}{\exp\left(G_1\left(\mathbf{Z}_{t+1/2}^{(1)}, \mathbf{G}^{(v)}\right)\right)} \\
\mathbf{Z}_t^{(0)} &= \frac{\left(\mathbf{Z}_{t+1/2}^{(0)} - F_2\left(\mathbf{Z}_t^{(1)}, \mathbf{G}^{(v)}\right)\right)}{\exp\left(F_1\left(\mathbf{Z}_t^{(1)}, \mathbf{G}^{(v)}\right)\right)}
\end{aligned} \tag{8}$$

The Jacobian of each step is block triangular, allowing efficient computation of the log-determinant required for likelihood evaluation.

Vanilla Normalizing Flow for Global Flexibility

In parallel, we employ a structure-free Vanilla Normalizing Flow (VNF) based on RealNVP-style affine coupling layers. This flow operates independently on each node's feature vector without considering graph structure, thereby preserving global distributional flexibility. Given a latent vector $\mathbf{z} \in \mathbb{R}^d$, we split it into two halves $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^{d/2}$. A coupling layer is defined as:

$$\begin{bmatrix} z_1^m \\ z_2^m \end{bmatrix} = \begin{cases} \begin{bmatrix} z_1^{m-1} \odot e^{s_{\text{mlp}}^m(z_2^{m-1})} + t_{\text{mlp}}^m(z_2^{m-1}) \\ z_2^{m-1} \end{bmatrix}, & m \text{ is odd}, \\ \begin{bmatrix} z_1^{m-1} \\ z_2^{m-1} \odot e^{s_{\text{mlp}}^m(z_1^{m-1})} + t_{\text{mlp}}^m(z_1^{m-1}) \end{bmatrix}, & m \text{ is even}. \end{cases} \tag{9}$$

where s_{mlp}^m and t_{mlp}^m are Multi-Layer Perceptrons (MLPs). The triangular Jacobian ensures tractable likelihood computation.

Dual-flow Architecture and Fusion

Given the refined multi-omics representation $\hat{\mathbf{Z}}^{(v)}$, we process it through both flows:

$$\hat{\mathbf{Z}}_{\text{graph}}^{(v)} = f_{\text{graph}}(\hat{\mathbf{Z}}^{(v)}, \mathbf{G}^{(v)}), \tag{10}$$

$$\hat{\mathbf{Z}}_{\text{plain}}^{(v)} = f_{\text{plain}}(\hat{\mathbf{Z}}^{(v)}), \tag{11}$$

The two complementary representations are then adaptively fused:

$$\tilde{\mathbf{Z}}^{(v)} = \alpha \hat{\mathbf{Z}}_{\text{graph}}^{(v)} + (1-\alpha)\hat{\mathbf{Z}}_{\text{plain}}^{(v)}, \tag{12}$$

where $\alpha \in [0, 1]$ is a learnable parameter. This dual-flow design explicitly separates structural constraints from global distribution modeling, allowing GBFlow to balance relational coherence with distributional expressiveness and produce robust latent embeddings for spatial domain identification.

3.3 Grouping Belief based Affinity Refinement

To suppress noisy graph connections, we introduce a Grouping Belief based Affinity Refinement (GBAR) strategy. Given similarity matrices $\mathbf{A}_h^{(v)}$ (from dynamic subspace) and

$\mathbf{A}_a^{(v)} = \mathbf{A}_3^{(v)}$, we first construct top-t neighbor masks $\mathbf{M}_h^{(v)}$ and $\mathbf{M}_a^{(v)}$. From these, soft relation matrices are computed as:

$$\mathbf{R}_h^{(v)} = \frac{\mathbf{M}_h^{(v)}(\mathbf{M}_h^{(v)})^\top}{t}, \quad \mathbf{R}_a^{(v)} = \frac{\mathbf{M}_a^{(v)}(\mathbf{M}_a^{(v)})^\top}{t}. \quad (13)$$

Next, hard relation matrices $\mathbf{G}_h^{(v)}$ and $\mathbf{G}_a^{(v)}$ are obtained by thresholding with confidence θ. The GBAR loss consists of two terms:

$$\mathcal{L}_{reg1} = -\sum_{v=1}^{V}\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_1((a_h^v)_{ij}\log((g_a^v)_{ij}+\varepsilon) + (a_a^v)_{ij}\log((g_h^v)_{ij}+\varepsilon)) \quad (14)$$

$$\mathcal{L}_{reg2} = -\sum_{v=1}^{V}\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_2(g_{ij}^v\log((r_a^v)_{ij}) + g_{ij}^v\log((r_h^v)_{ij})) \quad (15)$$

where $\mathbf{g}^{(v)} = (\mathbf{G}_h^{(v)} + \mathbf{G}_a^{(v)})/2$, and β_1, β_2 are hyperparameters. The total GBAR loss is:

$$\mathcal{L}_{\text{GBAR}} = \mathcal{L}_{\text{reg1}} + \mathcal{L}_{\text{reg2}}. \quad (16)$$

The refined graph $\mathbf{G}^{(v)}$ resulting from this procedure is subsequently used in GNF to guide structure-aware latent transformations.

3.4 Dynamic Graph Learning

Shared and Omics-specific Subspace Modeling

To simultaneously capture globally shared structures and omics-specific dependencies, we propose a dynamic subspace modeling strategy. We first construct a global self-expression matrix: $\mathbf{C} = \bar{\mathbf{c}}\bar{\mathbf{c}}^\top$, where $\bar{\mathbf{c}}$ denotes a low-dimensional latent embedding capturing global sample relationships. For each omics v, we additionally learn an omics-specific self-expression matrix $\mathbf{Q}^{(v)}$ to encode modality-dependent connectivity patterns. The adaptive graph structure for each omics is thus defined as:

$$\mathbf{H}^{(v)} = \bar{\mathbf{c}}\bar{\mathbf{c}}^\top + \mathbf{Q}^{(v)}. \quad (17)$$

The self-expression reconstruction objective is:

$$\mathcal{L}_{\text{self}} = \frac{\lambda}{V}\sum_{v=1}^{V}\|\tilde{\mathbf{Z}}^{(v)} - \mathbf{H}^{(v)}\tilde{\mathbf{Z}}^{(v)}\|_F^2 + \mathcal{L}_1, \quad (18)$$

where $\mathcal{L}_1$ enforces graph smoothness and omics disentanglement.

Structure-aware Contrastive Learning

We construct a unified affinity graph by aggregating the fused graphs learned from all omics:

$$\mathbf{W}^{ij} = \max_v(\mathbf{A}_3^{(v)})^{ij}. \quad (19)$$

Based on this structure-aware affinity, we design a contrastive objective:

$$\mathcal{L}_{\text{cont}} = -\sum_{v=1}^{V}\sum_{i=1}^{n}\log\frac{P_i^v}{N_i^v}, \quad (20)$$

where P_i^v and N_i^v denote aggregated similarities over positive and negative sample pairs determined by $\mathbf{W}$:

$$P_i^v = \sum_{j=1}^{n}W^{ij}g\left(\bar{\tilde{z}}_i^v, \tilde{z}_j^v\right) + \sum_{\substack{j=1\\ i\neq j}}^{n}W^{ij}g\left(\bar{\tilde{z}}_i^v, \bar{\tilde{z}}_j^v\right)$$
$$N_i^v = \sum_{j=1}^{n}\rho_{W^{ij}}g\left(\bar{\tilde{z}}_i^v, \tilde{z}_j^v\right) + \sum_{\substack{j=1\\ i\neq j}}^{n}\rho_{W^{ij}}g\left(\bar{\tilde{z}}_i^v, \bar{\tilde{z}}_j^v\right), \quad (21)$$

where $\tilde{z}_i^v \in \mathbb{R}^{n\times l}$ denotes the latent feature vector of sample i in view v before self-expression, obtained from the dual-flow latent refinement module, and $\bar{\tilde{z}}_i^v \in \mathbb{R}^{n\times l}$ represents the self-expressed feature after applying the dynamic self-expression model, i.e., $\bar{\tilde{Z}}^{(v)} = H^{(v)}\tilde{Z}^{(v)}$. Both are l-dimensional latent embeddings. $g(\cdot,\cdot)$ is the cosine similarity, and $\rho_{W^{ij}}$ is an indicator function defined as:

$$\rho_{W^{ij}} = \begin{cases} W^{ij}, & W^{ij} > 0 \\ 1, & W^{ij} = 0 \end{cases} \quad (22)$$

3.5 Overall Loss Function and Spatial Domain Identification

To ensure that the cyclic autoencoder preserves both structural and semantic information, we employ a reconstruction loss $\mathcal{L}_{\text{rec}}$ that penalizes discrepancies in both node features and self-representations. The overall training objective integrates all components of our framework:

$$\mathcal{L} = \mathcal{L}_{\text{rec}} + \mathcal{L}_{\text{self}} + \mathcal{L}_{\text{cont}} + \mathcal{L}_{\text{GBAR}}. \quad (23)$$

For spatial domain identification, we construct the final affinity matrix by combining the global and omics-specific self-expression components:

$$\mathbf{H} = \bar{\mathbf{c}}\bar{\mathbf{c}}^\top + \frac{1}{V}\sum_{v=1}^{V}\frac{\mathbf{Q}^{(v)} + (\mathbf{Q}^{(v)})^\top}{2}, \quad (24)$$

Spectral clustering is then applied on $\mathbf{H}$ to obtain the final spatial domain assignments, yielding coherent and biologically meaningful spatial partitions. The overall algorithm process is detailed in the supplementary materials.

4 Experiments

4.1 Datasets and Preprocessing

In this study, we utilize eight datasets: human breast cancer[Long *et al.*, 2024], mouse brain[Long *et al.*, 2024; Dong and Zhang, 2022], human lymph node[Long *et al.*, 2024], and mouse embryo[Jiang *et al.*, 2023]. Dataset-specific information and experimental settings are provided in the supplementary materials. For the breast cancer and mouse brain datasets, we select the top 3,000 highly variable genes,

apply normalization and log transformation, and reduce dimensionality to 40 via PCA. Whole-slide histology images are cropped into 50×50 patches based on spatial coordinates and flattened into 1D vectors. For the lymph node dataset, RNA is filtered to retain genes expressed in at least 10 cells, normalized, log-transformed, and reduced to 100 dimensions via PCA; paired ADT data are similarly normalized, scaled, and PCA-reduced. For the mouse embryo dataset, RNA undergoes filtering, highly variable gene selection, normalization, log transformation, and PCA reduction to 50 dimensions; ATAC features detected in fewer than 6% of cells are removed and projected to 21 dimensions via LSI.

4.2 Baseline and Metrics

We compare GBFlow with seven baseline methods, including GraphST[Long *et al.*, 2023], STAGATE[Dong and Zhang, 2022], PAST[Li *et al.*, 2023], SpatialGlue[Long *et al.*, 2024], PRAGA[Huang *et al.*, 2025], stDCL[Yu *et al.*, 2025], and DUSTED[Zhu *et al.*, 2025]. To comprehensively evaluate spatial domain delineation, we employ a set of standard metrics: Accuracy (ACC), F1-Score, Jaccard coefficient, Adjusted Rand Index (ARI), Mutual Information (MI), Normalized Mutual Information (NMI), Adjusted Mutual Information (AMI), V-Measure, and Completeness (Compl.). Across these metrics, GBFlow consistently achieves top or near-top performance.

4.3 Experimental Results

Qualitative Experimental Results

GBFlow consistently achieves state-of-the-art or highly competitive performance across diverse spatial multi-omics datasets. As shown in Table 1 and Supplementary Table 1 our method yields the highest ACC (76.65%) on the Human breast cancer dataset, outperforming strong baselines such as STAGATE and DUSTED. Notably, in the high-complexity Mouse brain anterior task, which involves 52 distinct clusters, GBFlow demonstrates superior precision by securing top scores in FMI (48.98%), ARI (44.65%), and ACC (47.38%). The model's robustness is further validated on the Human lymph node and Mouse embryo datasets, where it maintains stable and leading performance in terms of NMI and ARI, underscoring its ability to generalize effectively across varying degrees of tissue heterogeneity and cluster scales.

An interesting phenomenon observed in Table 1 is that the ranking of the same method can vary across different evaluation metrics. For example, on the Human lymph node D1 dataset, PRAGA achieves the highest MI (65.42) but a much lower ARI (24.94%), while GBFlow attains a slightly lower MI (65.37%) but leads in ARI (34.41%). This highlights the distinct emphases of different evaluation metrics: information-theoretic measures such as MI and NMI focus more on the overall alignment between feature distributions and label distributions, whereas ARI and FMI place greater emphasis on the pairwise consistency of clustering results. GBFlow's consistent superiority in the latter category directly reflects how its grouping belief module enhances the pairwise accuracy of clustering by reinforcing high-confidence neighborhood relationships.

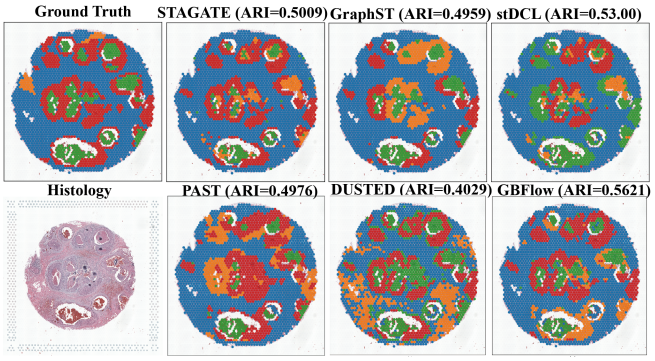

Figure 2: Spatial domain delineation on human breast cancer. The ground truth annotations (top row) correspond to histologically defined tissue structures: desmoplastic changes, lymphocytes, and necrosis/hemorrhage. The spatial domains identified by all methods are visualized in subsequent rows. GBFlow achieves superior structural consistency and biological interpretability, as reflected in its higher ARI score.

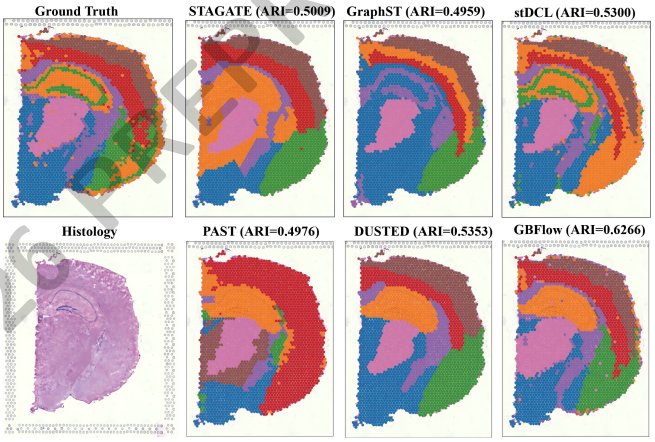

Figure 3: Spatial domain delineation on mouse brain coronal. The ground truth (top row) corresponds to anatomically defined cortical layers L1–L4. Subsequent rows visualize clustering results from all methods.

The performance gains of GBFlow stem from its dual-flow latent refinement strategy, which integrates a structure-aware Graph Normalizing Flow (GNF) with a structure-free Vanilla Normalizing Flow (VNF). This design enables the model to preserve local spatial coherence while capturing global molecular distribution patterns. In addition, the grouping belief-based affinity refinement module effectively suppresses noise in the initial graphs and reinforces high-confidence neighborhood relationships. By pruning spurious connections and enhancing discriminative features, GBFlow learns more separable latent representations, leading to improved spatial domain identification.

Quantitative Experimental Results

Figure 2 shows GBFlow's ability to delineate histologically meaningful regions in human breast cancer tissue, including desmoplastic changes, lymphocyte infiltration, and necrosis/hemorrhage. Compared with baseline methods,

Method	Metrics (%)									
	MI	NMI	AMI	FMI	ARI	V-Measure	F1-Score	Jaccard	Compl.	ACC
Human breast cancer section1 (n=2518, clusters=4)										
GraphST	38.72	36.38	36.27	69.71	49.59	36.28	69.70	53.50	37.11	49.80
STAGATE	50.25	44.87	44.78	68.60	50.09	**44.87**	68.27	51.82	41.99	71.45
stDCL	41.17	39.24	39.14	72.65	53.00	39.24	72.61	57.00	39.03	62.87
PAST	41.34	37.90	37.80	69.32	49.76	37.90	69.27	52.98	36.32	68.75
DUSTED	44.21	38.51	38.42	62.02	40.29	38.51	61.60	44.50	35.29	69.38
GBFlow	**47.72**	**44.02**	**43.93**	**73.72**	**56.21**	44.02	**73.71**	**58.37**	**45.75**	**76.65**
Mouse brain anterior (n=2695, clusters=52)										
GraphST	253.44	68.82	64.35	39.32	36.32	68.82	38.31	23.69	70.89	43.49
STAGATE	**265.82**	**71.71**	**67.60**	40.84	37.77	**71.71**	39.63	24.71	69.26	45.27
stDCL	249.93	70.95	67.47	45.54	43.29	70.95	45.54	29.48	**72.03**	41.69
PAST	228.24	63.80	58.96	33.13	30.48	63.76	33.10	19.83	63.76	37.81
DUSTED	263.58	71.37	67.21	42.48	39.63	71.37	41.51	26.19	69.17	46.49
GBFlow	224.51	66.80	62.90	**48.98**	**44.65**	66.80	**47.38**	**31.05**	62.79	**47.38**
Human lymph node D1 (n=3359, clusters=10)										
GraphST	62.69	34.05	33.60	37.33	21.26	34.05	37.19	22.84	36.05	44.66
STAGATE	17.82	12.25	11.65	33.61	5.61	12.25	32.31	19.27	15.22	34.59
PAST	59.21	33.25	32.77	42.66	26.05	33.25	42.66	27.11	32.48	42.33
SpatialGlue	63.95	39.01	38.67	**50.60**	35.05	39.01	50.47	33.76	41.54	**64.39**
PRAGA	**65.42**	34.67	34.22	39.44	24.94	34.67	38.94	24.17	32.15	47.51
DUSTED	51.04	28.40	27.91	36.26	19.92	28.40	36.13	22.05	29.35	41.95
GBFlow	65.37	**39.49**	**39.07**	50.18	**34.41**	**39.49**	**50.04**	**33.37**	**37.60**	61.95
Mouse embryo E18.5 (n=2129, clusters=14)										
GraphST	**118.25**	52.80	51.93	48.53	39.78	52.80	48.05	31.63	55.87	54.86
STAGATE	16.81	12.77	11.11	40.48	5.03	12.77	30.28	17.84	32.60	34.34
PAST	101.22	44.04	43.04	34.15	24.00	44.04	32.93	19.71	40.81	38.94
SpatialGlue	104.28	**55.95**	55.36	57.98	46.55	**55.95**	56.80	39.67	**64.75**	61.44
PRAGA	112.02	49.42	48.50	41.46	31.92	49.42	40.61	25.48	46.35	46.78
DUSTED	109.20	48.68	47.74	37.41	27.14	48.68	36.93	22.65	51.59	42.79
GBFlow	117.00	55.33	54.49	**58.55**	**50.39**	55.33	**58.53**	**41.37**	55.27	**63.03**

Table 1: Quantitative experimental results on multiple spatial transcriptomics datasets (Part 1)

which exhibit over-segmentation (GraphST), boundary blurring (stDCL), or internal noise (PAST, DUSTED), GBFlow produces sharper boundaries and more coherent intra-region structures, reflecting effective integration of multi-omics signals and the benefit of affinity refinement.

Figure 3 illustrates GBFlow's preservation of fine-grained anatomical organization in mouse brain coronal sections. While STAGATE and GraphST show layer fragmentation and stDCL blurs inter-layer transitions, GBFlow maintains clear laminar boundaries (L1–L4) and continuous intra-layer homogeneity. The dual-flow design ensures neighborhood consistency within layers (GNF) and captures global expression variability across layers (VNF), yielding spatially smooth and anatomically faithful partitions. Overall, GBFlow demonstrates higher structural consistency and robustness across diverse tissue types, with quantitative improvements in ACC, ARI, and FMI corroborating these observations.

5 Conclusion

We propose GBFlow, a novel grouping belief-guided dual normalizing flow framework for robust spatial domain identification. By integrating structure-aware and structure-free normalizing flows alongside a grouping belief-based affinity refinement strategy, GBFlow effectively balances relational consistency with global distributional flexibility. Experiments across eight spatial multi-omics datasets demonstrate that GBFlow consistently outperforms state-of-the-art methods, achieving more accurate and robust spatial delineation. Additional analyses including marker gene visualization, latent representation visualization, and scalability analysis are provided in the supplementary material. The framework provides a principled approach to multi-omics integration and is widely applicable to spatially resolved biomedical data analysis.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (62372279, 62322215, 62532017), the Natural Science Foundation of Shandong Province (ZR2025QB62, ZR2023MF119), and the Natural Science Foundation of Hunan Province (2026JJ30018). This study was also supported in part by the High-Performance Computing Center of Central South University.

References

[Cai and Li, 2025] Wenhao Cai and Weizhong Li. Spatialfuser: A unified deep learning framework for spatial multi-omics data integrative analysis. *bioRxiv*, page 2025.09.14.676067, 2025.

[Dong and Zhang, 2022] Kangning Dong and Shihua Zhang. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nature communications*, 13(1):1739, 2022.

[Duan *et al.*, 2023] Honglin Duan, Tao Cheng, and Hui Cheng. Spatially resolved transcriptomics: advances and applications. *Blood Science*, 5(1):1–14, 2023.

[Gong *et al.*, 2025] Qiyu Gong, Kai Cao, Jackson A Weir, Qitian Wu, Dongshunyi Li, and Fei Chen. Energy-regularized graph learning for multiscale spatial representation. *bioRxiv*, pages 2025–08, 2025.

[Han *et al.*, 2025] Shen Han, Zhiyao Zhou, Jiawei Chen, Zhezheng Hao, Sheng Zhou, Gang Wang, Yan Feng, Chun Chen, and Can Wang. Uncertainty-aware graph structure learning. In *Proceedings of the ACM on Web Conference 2025*, pages 4863–4874, 2025.

[Hu *et al.*, 2021] Jian Hu, Xiangjie Li, Kyle Coleman, Amelia Schroeder, Nan Ma, David J Irwin, Edward B Lee, Russell T Shinohara, and Mingyao Li. Spagcn: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nature methods*, 18(11):1342–1351, 2021.

[Huang *et al.*, 2025] Xinlei Huang, Zhiqi Ma, Dian Meng, Yanran Liu, Shiwei Ruan, Qingqiang Sun, Xubin Zheng, and Ziyue Qiao. Praga: prototype-aware graph adaptive aggregation for spatial multi-modal omics analysis. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 326–333, 2025.

[Jiang *et al.*, 2023] Fuqing Jiang, Xin Zhou, Yingying Qian, Miao Zhu, Li Wang, Zhuxia Li, Qingmei Shen, Minhan Wang, Fangfang Qu, and Guizhong Cui. Simultaneous profiling of spatial gene expression and chromatin accessibility during mouse brain development. *Nature Methods*, 20(7):1048–1057, 2023.

[Jiang *et al.*, 2025a] Bingbing Jiang, Chenglong Zhang, Zhongli Wang, Xinyan Liang, Peng Zhou, Liang Du, Qinghua Zhang, Weiping Ding, and Yi Liu. Scalable fuzzy clustering with collaborative structure learning and preservation. *IEEE Transactions on Fuzzy Systems*, 33(9):3047–3060, 2025.

[Jiang *et al.*, 2025b] Linling Jiang, Wenchang Zhang, Hua Wang, and Fan Zhang. Robust memory-based graph neural networks for noisy and sparse graphs. *Engineering Applications of Artificial Intelligence*, 160:111874, 2025.

[Kang *et al.*, 2025] Liping Kang, Qinglong Zhang, Fan Qian, Junyao Liang, and Xiaohui Wu. Benchmarking computational methods for detecting spatial domains and domain-specific spatially variable genes from spatial transcriptomics data. *Nucleic Acids Research*, 53(7):gkaf303, 2025.

[Li *et al.*, 2022] Jiachen Li, Siheng Chen, Xiaoyong Pan, Ye Yuan, and Hong-Bin Shen. Cell clustering for spatial transcriptomics data with graph neural networks. *Nature Computational Science*, 2(6):399–408, 2022.

[Li *et al.*, 2023] Zhen Li, Xiaoyang Chen, Xuegong Zhang, Rui Jiang, and Shengquan Chen. Latent feature extraction with a prior-based self-attention framework for spatial transcriptomics. *Genome Research*, 33(10):1757–1773, 2023.

[Liu *et al.*, 2019] Jenny Liu, Aviral Kumar, Jimmy Ba, Jamie Kiros, and Kevin Swersky. Graph normalizing flows. *Advances in Neural Information Processing Systems*, 32, 2019.

[Long *et al.*, 2023] Yahui Long, Kok Siong Ang, Mengwei Li, Kian Long Kelvin Chong, Raman Sethi, Chengwei Zhong, Hang Xu, Zhiwei Ong, Karishma Sachaphibulkij, and Ao Chen. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with graphst. *Nature Communications*, 14(1):1155, 2023.

[Long *et al.*, 2024] Yahui Long, Kok Siong Ang, Raman Sethi, Sha Liao, Yang Heng, Lynn van Olst, Shuchen Ye, Chengwei Zhong, Hang Xu, and Di Zhang. Deciphering spatial domains from spatial multi-omics with spatialglue. *Nature Methods*, 21(9):1658–1667, 2024.

[Lu and Lin, 2025] Yong Lu and Zhengguo Lin. Aro-gnn: Adaptive relation-optimized graph neural networks. *Journal of King Saud University Computer and Information Sciences*, 37(7):174, 2025.

[Lu *et al.*, 2025] Qiaolin Lu, Jiayuan Ding, Lingxiao Li, and Yi Chang. Graph contrastive learning of subcellular-resolution spatial transcriptomics improves cell type annotation and reveals critical molecular pathways. *Briefings in Bioinformatics*, 26(1):bbaf020, 2025.

[Ma *et al.*, 2024] Yuanyuan Ma, Lifang Liu, Yongbiao Zhao, Bo Hang, and Yanduo Zhang. Hypergcn: an effective deep representation learning framework for the integrative analysis of spatial transcriptomics data. *BMC genomics*, 25(1):566, 2024.

[Moses and Pachter, 2022] Lambda Moses and Lior Pachter. Museum of spatial transcriptomics. *Nature methods*, 19(5):534–546, 2022.

[Pham *et al.*, 2023] Duy Pham, Xiao Tan, Brad Balderson, Jun Xu, Laura F. Grice, Sohye Yoon, Emily F. Willis, Minh Tran, Pui Yeng Lam, Arti Raghubar, Priyakshi Kalita-de Croft, Sunil Lakhani, Jana Vukovic, Marc J. Ruitenberg,

and Quan H. Nguyen. Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues. *Nature Communications*, 14(1):7739, 2023.

[Sun *et al.*, 2025] Shuli Sun, Jixin Liu, Guojun Li, and Bingqiang Liu. Deepgft: identifying spatial domains in spatial transcriptomics of complex and 3d tissue using deep learning and graph fourier transform. *Genome Biology*, 26(1):153, 2025.

[Tang *et al.*, 2023] Ziyang Tang, Zuotian Li, Tieying Hou, Tonglin Zhang, Baijian Yang, Jing Su, and Qianqian Song. Sigra: single-cell spatial elucidation through an image-augmented graph transformer. *Nature Communications*, 14(1):5618, 2023.

[Wang *et al.*, 2024] Lequn Wang, Yaofeng Hu, Kai Xiao, Chuanchao Zhang, Qianqian Shi, and Luonan Chen. Multi-modal domain adaptation for revealing spatial functional landscape from spatially resolved transcriptomics. *Briefings in Bioinformatics*, 25(4):bbae257, 2024.

[Xu *et al.*, 2025] Ke Xu, Yu Xu, Zirui Wang, Xin Maizie Zhou, and Lu Zhang. stdyer enables spatial domain clustering with dynamic graph embedding. *Genome Biology*, 26(1):34, 2025.

[Yu *et al.*, 2025] Zhuohan Yu, Yuning Yang, Xingjian Chen, Ka-Chun Wong, Zhaolei Zhang, Yuming Zhao, and Xiangtao Li. Accurate spatial heterogeneity dissection and gene regulation interpretation for spatial transcriptomics using dual graph contrastive learning. *Advanced Science*, 12(3):2410081, 2025.

[Zhang *et al.*, 2025a] Jingyun Zhang, Hao Peng, Li Sun, Guanlin Wu, Chunyang Liu, and Zhengtao Yu. Unsupervised graph clustering with deep structural entropy. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V. 2*, pages 3752–3763, 2025.

[Zhang *et al.*, 2025b] Tianjiao Zhang, Hongfei Zhang, Zhongqian Zhao, Saihong Shao, Yucai Jiang, Xiang Zhang, and Guohua Wang. spamgcn: a graph convolutional network with autoencoder for spatial domain identification using multi-scale adaptation. *Genome Biology*, 26(1):159, 2025.

[Zhu *et al.*, 2025] Jun Zhu, Yifu Li, Zhenchao Tang, and Cheng Chang. Dusted: Dual-attention enhanced spatial transcriptomics denoiser. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 1219–1227, 2025.