A Graph Attention Network Approach to Super-Resolution Spatial Transcriptomic Data

Luis Alonso, Mikel Hernáez and Idoia Ochoa

Graph Signal Processing workshop 2026

10th June 2026

Tecnun
Universidad
de Navarra

Cima
Universidad
de Navarra

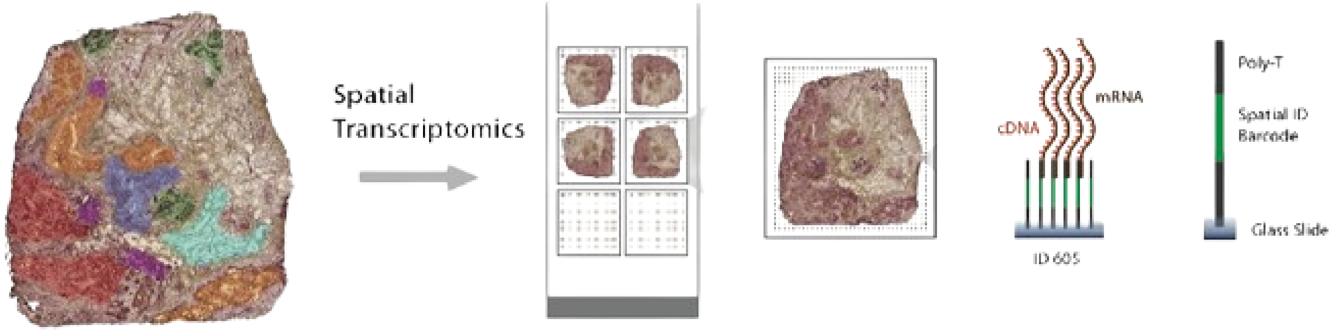
Content

1. Introduction to ST
2. Problem statement
3. Contribution
4. Results
5. Conclusions

Introduction

Spatial transcriptomics (ST)

1. A ST experiment generates:
    1. A gene expression matrix $\mathbf{X} \in \mathbb{N}_0^{G\times S}$.
        1. G genes, up to 30,000.
        2. S spots (spatial unit), up to a few thousand.
    2. A coordinate matrix $\mathbf{Z} \in \mathbb{N}^{S\times 2}$, where $Z_i = (x_i, y_i)^T$ are the coordinates of spot i.



Introduction
Spatial transcriptomics (ST)

1. Low spatial resolution.
    1. More than one cell per spot.
    2. Mixed signals.
2. The spots’ signal has:
    1. High overdispersion.
    2. High levels of sparsity (up to ~60%).

$$X_{gs} \sim ZINB\left(r_{gs}, p_{gs}, \pi_{gs}\right)$$

$$P(X_{gs} = x) = \begin{cases} \pi_{gs} + {p_{gs}}^{r_{gs}} & \text{if } x = 0 \\ \left(1 - \pi_{gs}\right)\binom{r_{gs}+x+1}{x}{p_{gs}}^{r_{gs}}\left(1 - p_{gs}\right)^{x} & \text{if } x > 0 \end{cases}$$

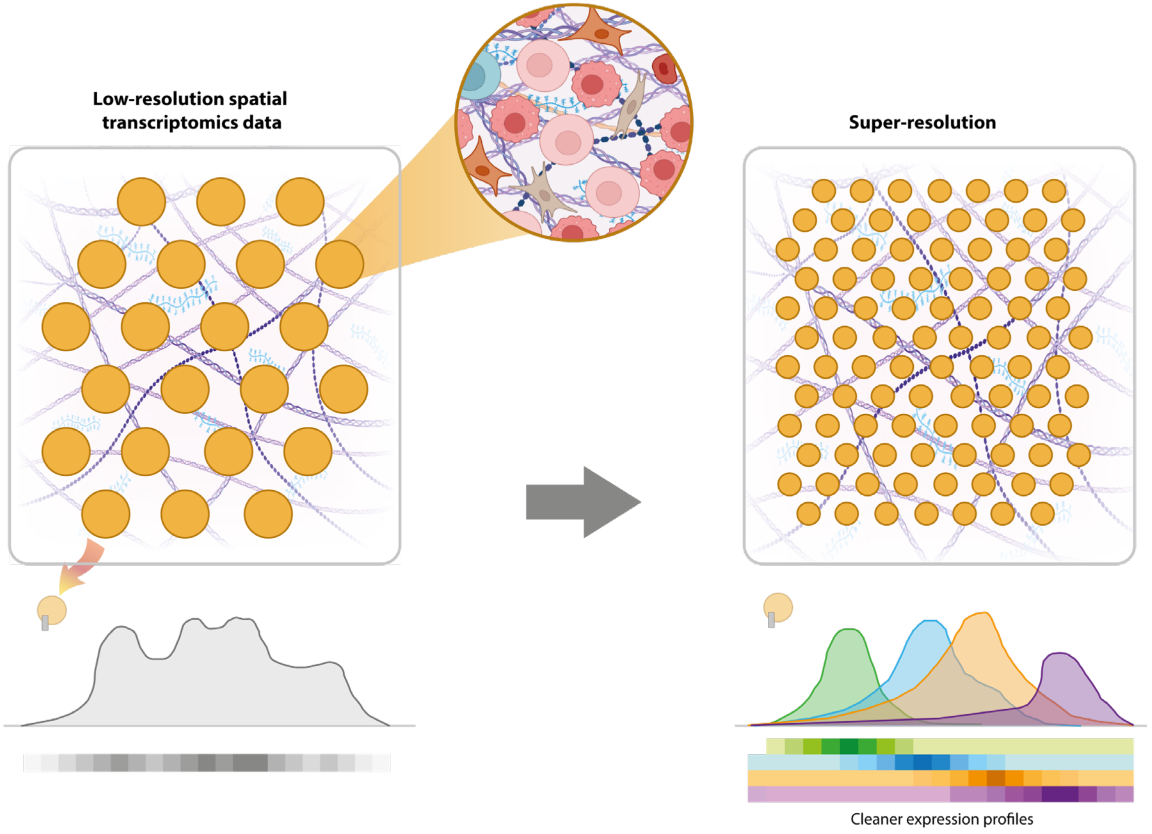
Problem statement

$$(\mathbf{X}, \mathbf{Z}) \to (\mathbf{X}_e, \mathbf{Z}_e)$$

$$\mathbf{X}_e \in \mathbb{N}_0{}^{G\times kS}$$
$$\mathbf{Z}_e \in \mathbb{N}^{2\times kS}$$

$k = 9$ (default)

Recover latent subspot signals from aggregated spot measurements.



Hypothesis: $Cov(x_i, x_j) \propto d(i,j)^{-1}$

Contribution

1. Self-supervised training approach:
    - Variational Autoencoder (VAE) to mitigate training-inference distribution shift.
2. Graph-based spatial representation.
    - Virtual nodes to represent subspots.
3. Attention learning on graphs.

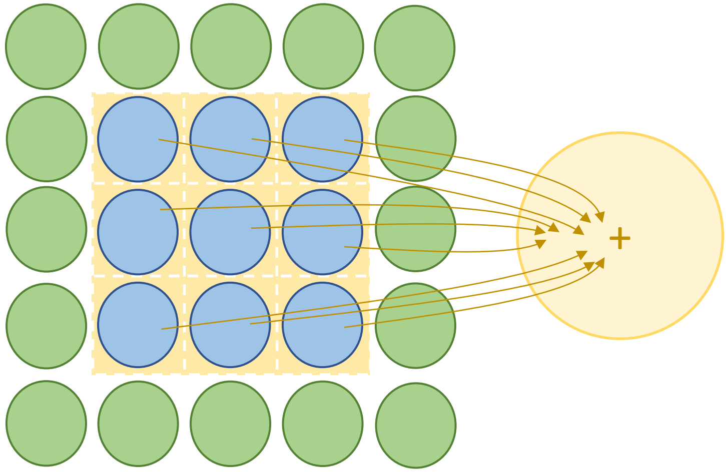
Contribution

1. Self-supervised training approach:
   - Variational Autoencoder (VAE) to mitigate training-inference distribution shift.
2. Graph-based spatial representation
   - Virtual nodes to represent subspots.
3. Attention learning on graphs.

Contribution

- No ground-truth available a priori.
- Pooling mask M_r to predict initial spots that originated the central **superspot**.
- Superspot's neighbors $\mathcal{N}_r$ to have spatial context.
- Training dataset:

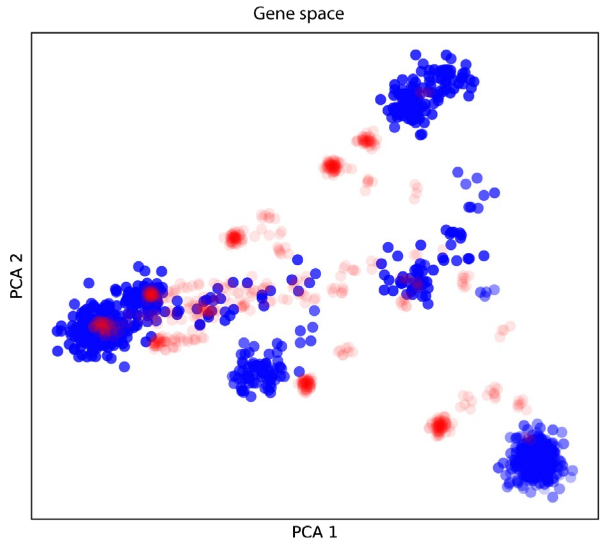
$$\mathcal{D} = \left\{ \left(\left\{ \sum_{j \in M_r} x_j \right\} \cup \{x_j\}_{j \in \mathcal{N}_r}, \{x_1, \dots, x_k\}_{M_r} \right) \right\}_{r=1}^{R}$$

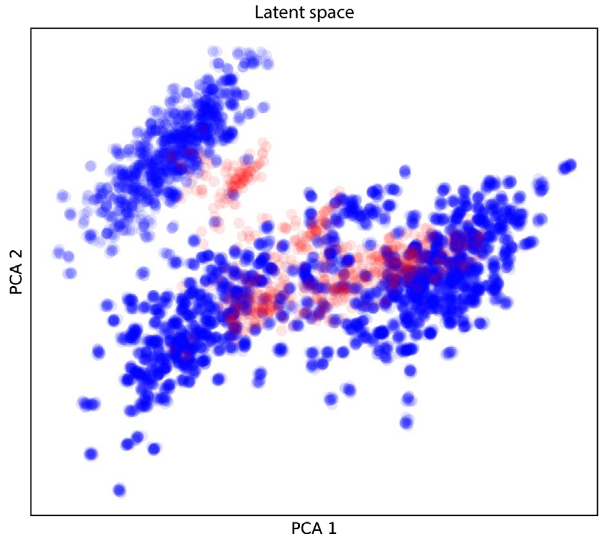


Contribution

- **Distribution shift** between training and inference data.
  - Sparsity differences.

- Train a Variational Autoencoder (VAE) to learn a **resolution-agnostic latent space** where biological signal can be encoded.
  - Encoder $q_\phi(\boldsymbol{z}|\boldsymbol{x}) = \mathcal{N}\left(\mu_\phi(\boldsymbol{x}), \Sigma_\phi(\boldsymbol{x})\right)$.
  - Gaussian prior on the latent variables, $p(\boldsymbol{z}) \sim \mathcal{N}(\mathbf{0}, \mathbb{I})$.





Contribution

1. Self-supervised training approach:
    - Variational Autoencoder (VAE) to mitigate training-inference distribution shift.
2. Graph-based spatial representation
    - Virtual nodes to represent subspots.
3. Attention learning on graphs.

Contribution

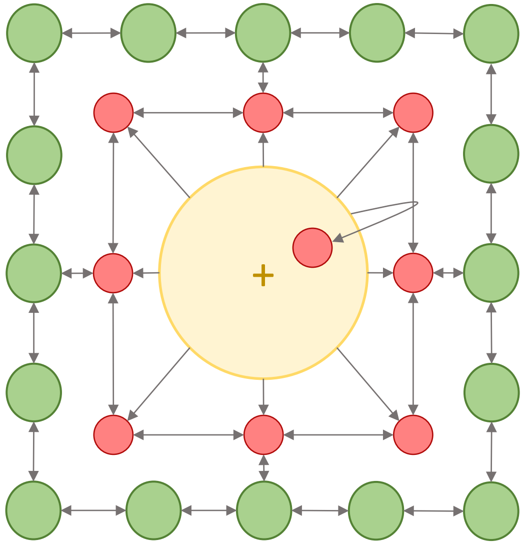
- Build a directed graph to **encode spatial dependencies** in our data.
- Represent predicted spots as **virtual nodes**.

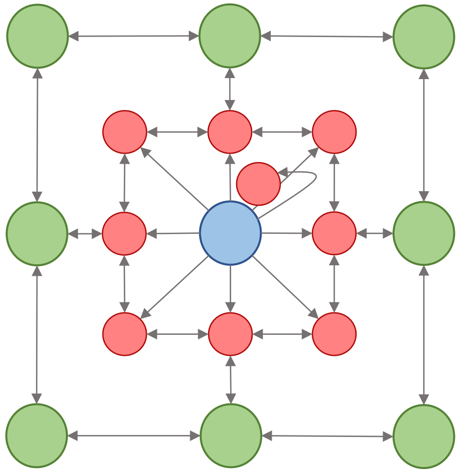
$$\mathcal{V} = \mathcal{V}_{\text{obs}} \cup \mathcal{V}_{\text{virt}}$$

- $\mathcal{G} = (\mathcal{V}, \mathcal{E}, A)$

$$A_{ij} = \begin{cases} d(s_i, s_j)^{-1} & \text{same node type} \\ t & \text{central} \rightarrow \text{virtual} \\ 1 & \text{neighbor} \rightarrow \text{virtual} \\ 0 & \text{otherwise} \end{cases}, \qquad t > 1$$

- During inference, place an initial spot itself at the center.





Contribution

1. Self-supervised training approach:
    - Variational Autoencoder (VAE) to mitigate training-inference distribution shift.
2. Graph-based spatial representation
    - Virtual nodes to represent subspots.
3. Attention learning on graphs.

Contribution

1. Graph Atention Network (GAT) [1] to learn **anisotropic diffusion**.
    1. Edge scoring function e: $\mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$

$$e(\boldsymbol{h}_i, \boldsymbol{h}_j) = \mathrm{LeakyReLU}\left(\boldsymbol{a}^T \cdot \left[\boldsymbol{W}_\delta \boldsymbol{h}_i \,||\, \boldsymbol{W}_\delta \boldsymbol{h}_j\right]\right)$$

2. Message passing:

$$h_i^{(l+1)} = \sigma\left(\max\left(\mathrm{softmax}_j\left(e\left(\boldsymbol{h}_i^{(l)}, \boldsymbol{h}_j^{(l)}\right)\right)\right) \cdot \boldsymbol{W}_\delta\, \boldsymbol{h}_j^{(l)}\right)$$

Learned graph filter instead of fixed Laplacian smoothing.

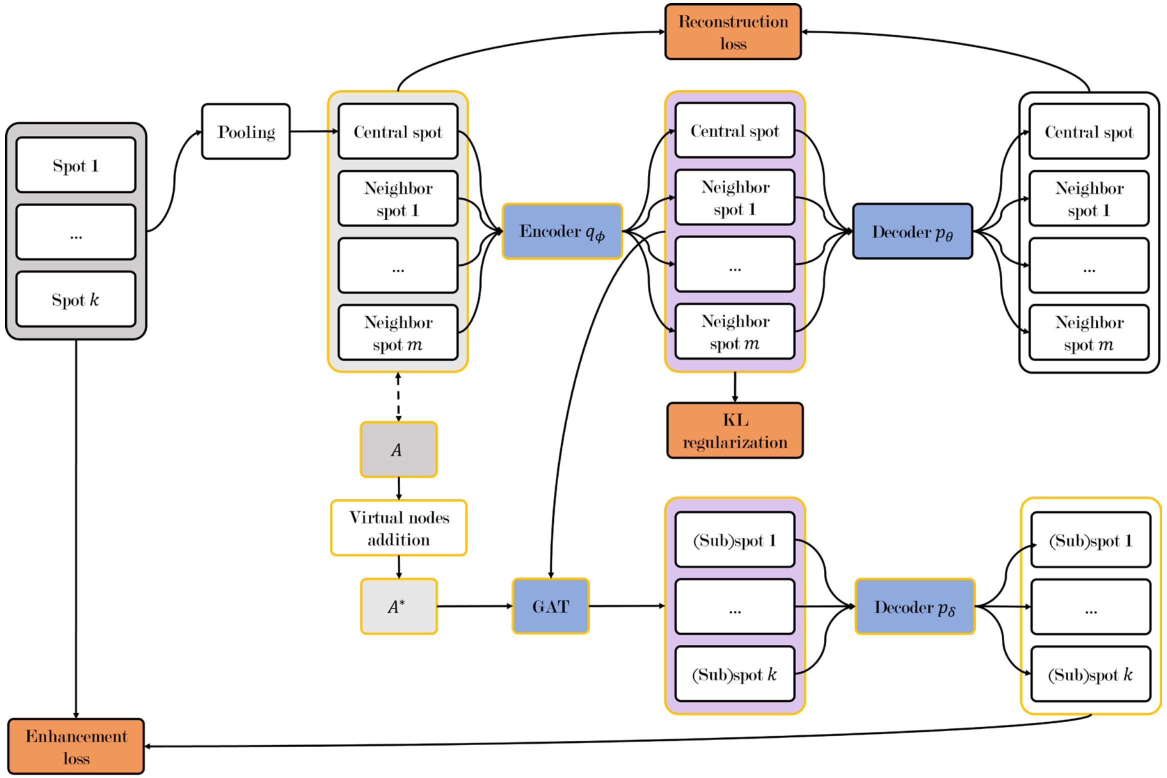
[1] S. Brody, U. Alon, and E. Yahav. “How attentive are graph attention networks?” *arXiv preprint arXiv*:2105.14491, 2021.

Contribution

$$\max_{\phi,\theta,\delta} \mathrm{ELBO}(\boldsymbol{x};\phi,\theta) + \mathcal{J}^{\delta}_{\mathrm{enhance}}$$

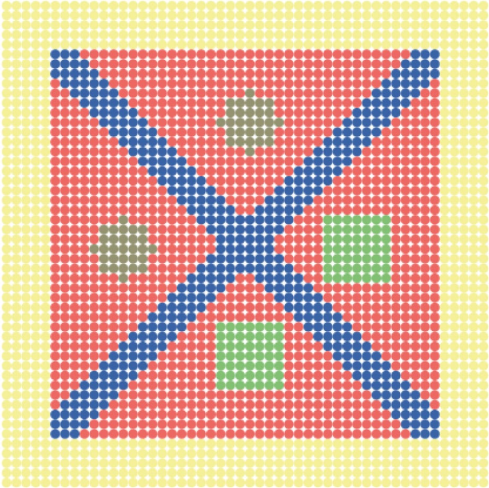
$$\mathrm{ELBO}(\boldsymbol{x};\phi,\theta) = \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}[\log p_\theta(\boldsymbol{x}|\boldsymbol{z})] - D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x}) \,||\, p(\boldsymbol{z}))$$

$$\mathcal{J}^{\delta}_{\mathrm{enhance}} = \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}[\log p_\delta(\boldsymbol{x}|\boldsymbol{z};\mathrm{A})]$$



Results I
Simulated dataset

- Ablated models:
  - *w/o* virtual nodes: subsampling neighbors' embeddings.
  - *w/o* VAE: message passing on gene space.
  - *w/o* attention: vanilla GNN.



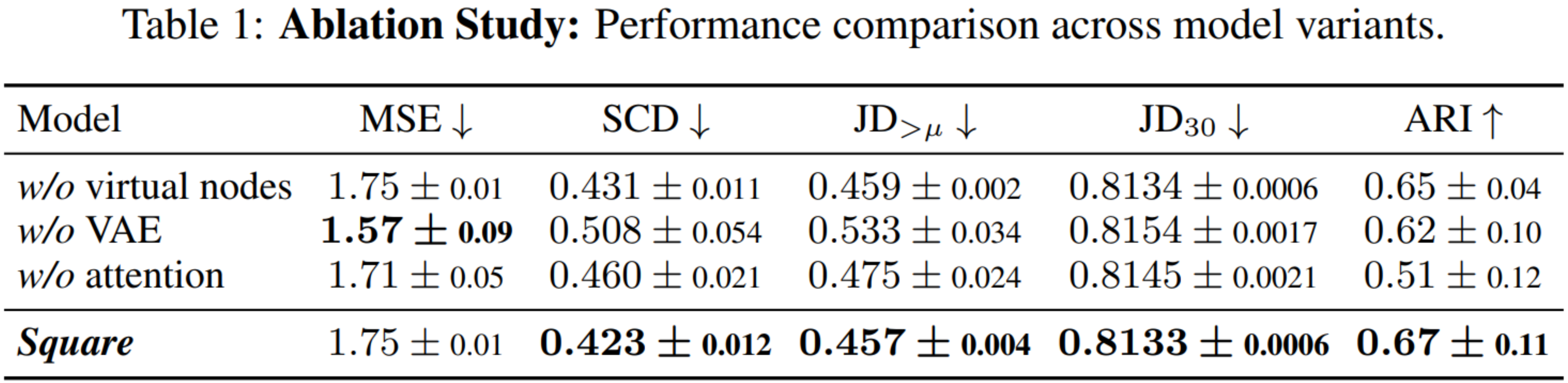
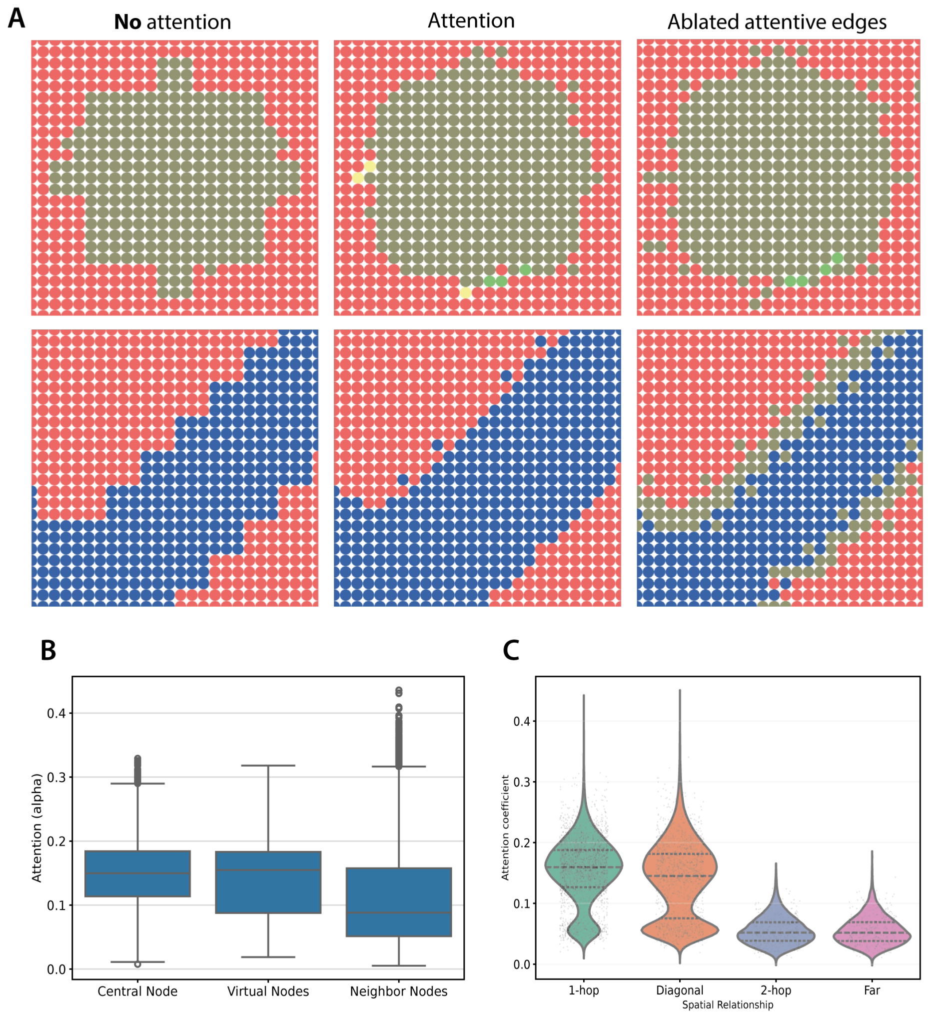


Table 1: **Ablation Study:** Performance comparison across model variants.

Model	MSE ↓	SCD ↓	$\mathrm{JD}_{>\mu}$ ↓	JD_{30} ↓	ARI ↑
w/o virtual nodes	1.75 ± 0.01	0.431 ± 0.011	0.459 ± 0.002	0.8134 ± 0.0006	0.65 ± 0.04
w/o VAE	**1.57 ± 0.09**	0.508 ± 0.054	0.533 ± 0.034	0.8154 ± 0.0017	0.62 ± 0.10
w/o attention	1.71 ± 0.05	0.460 ± 0.021	0.475 ± 0.024	0.8145 ± 0.0021	0.51 ± 0.12
Square	1.75 ± 0.01	**0.423 ± 0.012**	**0.457 ± 0.004**	**0.8133 ± 0.0006**	**0.67 ± 0.11**

Results I
Simulated dataset

- Artifact using vanilla GNN, due to homogeneous difussion within sets of 9 subspots.
- Sanity check: ablated most attentive edge decreases clustering performance.
- Central node must be the most attended.
- Closer nodes must be more attended.



Results II
Simulated dataset

- Benchmarked methods:
  - BayesSpace [1]
  - CARD [2]
  - STAGE [3]

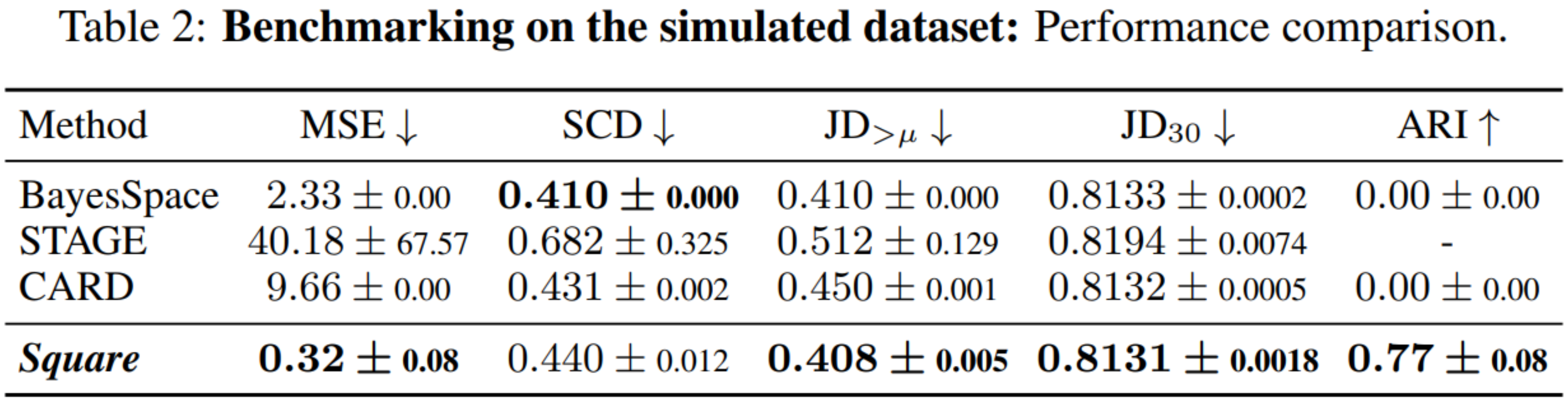


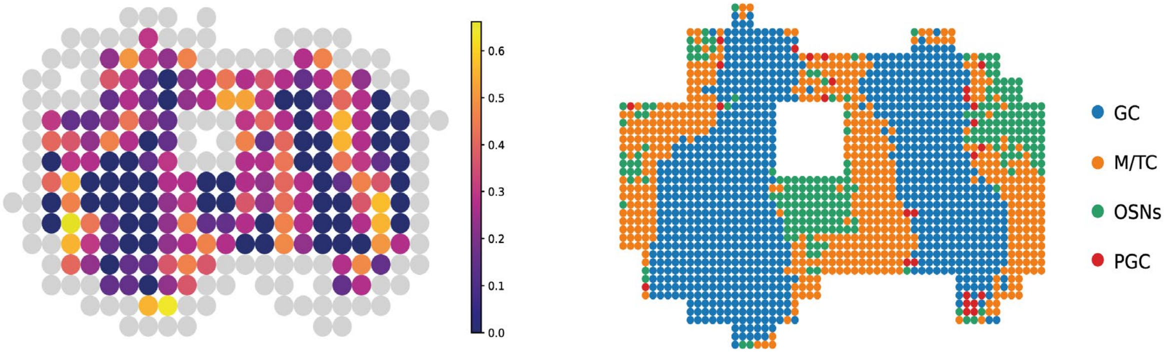
Table 2: **Benchmarking on the simulated dataset:** Performance comparison.

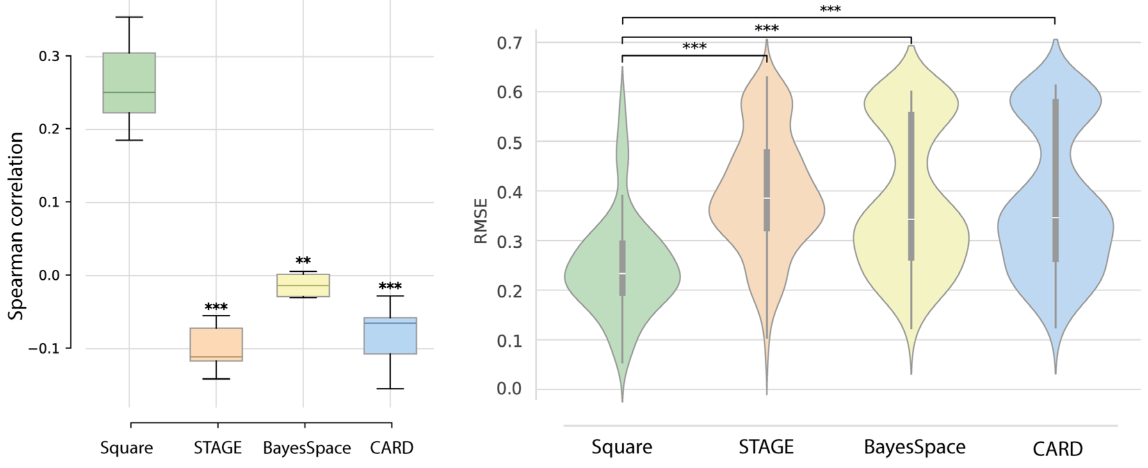
Method	MSE ↓	SCD ↓	JD$_{>\mu}$ ↓	JD$_{30}$ ↓	ARI ↑
BayesSpace	2.33 ± 0.00	**0.410 ± 0.000**	0.410 ± 0.000	0.8133 ± 0.0002	0.00 ± 0.00
STAGE	40.18 ± 67.57	0.682 ± 0.325	0.512 ± 0.129	0.8194 ± 0.0074	-
CARD	9.66 ± 0.00	0.431 ± 0.002	0.450 ± 0.001	0.8132 ± 0.0005	0.00 ± 0.00
Square	**0.32 ± 0.08**	0.440 ± 0.012	**0.408 ± 0.005**	**0.8131 ± 0.0018**	**0.77 ± 0.08**

[1] Zhao, Edward, et al. "Spatial transcriptomics at subspot resolution with BayesSpace." *Nature biotechnology* 39.11 (2021): 1375-1384.
[2] Ma, Ying, and Xiang Zhou. "Spatially informed cell-type deconvolution for spatial transcriptomics." *Nature biotechnology* 40.9 (2022): 1349-1359.
[3] Li, Shang, et al. "High-density generation of spatial transcriptomics with STAGE." *Nucleic Acids Research* 52.9 (2024): 4843-4856.

Results III
Synthetic dataset

- GT cell-type proportions available.
- Signal purity proxy: entropy.
- Check correlation between entropy and cluster frequency for differents levels of coarseness.
- Annotation of those clusters and assingment of one of the initial cell types.





Conclusions

- **Decoupling** the learning processes of **node embeddings** and **spatial dependencies** outperforms doing so jointly.

- Graphs can be used to encode **spatial dependencies** by representing the context with Euclidean **distances** as edge weights.

- **Attention** mechanism on regular graphs can help in scenarios where **anisotropic difussion** is desired, at least on synthetic data.

- Our method **outperforms the SOTA**, even with biological data.

GSP 2026
Madrid, Spain

A Graph Attention Network Approach t -Resolution Spatial Tr omic Data

Luis Alonso Idoia Ochoa

Graph Signal Processing workshop 2026

SCAN ME

Tecnun
Universidad
de Navarra

Cima
Universidad
de Navarra