CELL-TYPE PREDICTION IN SPATIAL TRANSCRIPTOMICS DATA USING GRAPH NEURAL NETWORKS

Moritz Lampert
Center for AI and Data Science (CAIDAS)
Chair of Machine Learning for Complex Networks
Julius-Maximilians-Universiät Würzburg, DE
`moritz.lampert@uni-wuerzburg.de`

Christopher Blöcker
Data Analytics Group, Department of Informatics
University of Zurich, CH
Center for AI and Data Science (CAIDAS)
Chair of Machine Learning for Complex Networks
Julius-Maximilians-Universiät Würzburg, DE

Ingo Scholtes
Center for AI and Data Science (CAIDAS)
Chair of Machine Learning for Complex Networks
Julius-Maximilians-Universiät Würzburg, DE
Data Analytics Group, Department of Informatics
University of Zurich, CH

Dominic Grün
Center for AI and Data Science (CAIDAS)
Würzburg Institute of Systems Immunology
Max Planck Research Group
Julius-Maximilians-Universiät Würzburg, DE

ABSTRACT

Recent advances in spatial transcriptomics enable the exploration of biological processes between cells at an unprecedented resolution. Leveraging spatial information allows us to construct cell-to-cell interaction graphs that describe possible communication between cells. Combining spatial interactions through graph neural networks (GNNs) with cells' gene expressions is a promising avenue for uncovering the underlying mechanisms behind, for example, cell differentiation. However, how to best construct a meaningful graph that captures relevant spatial information remains an open question. Moreover, what GNN architectures perform well on typical prediction tasks, such as cell-type prediction, is unclear. We address these questions by systematically evaluating several graph construction methods with common GNNs on four publicly available spatial transcriptomics datasets. Our results show that the spatial cell-to-cell interaction graphs contain relevant information for predicting cell types. Despite differences in graph topology, the choice of graph construction method affects cell-type prediction performance only minimally. Common GNN models do not perform better than a simpler multi-layer perceptron that does not have access to spatial information.

1 INTRODUCTION

Recent breakthroughs in sequencing technology enable us to measure gene expression at the cell level while simultaneously capturing each cell's location inside a given tissue sample. Such spatial transcriptomics data opens new opportunities to understand how the spatial organization of cells influences processes like gene expression or differentiation, thus facilitating the development of targeted treatments against diseases like cancer. Cell-to-cell interactions between nearby cells are an important mechanism through which the spatial organization of tissue can influence the function of individual cells (Armingol et al., 2021). Graphs that connect nodes via edges are a natural fit to model possible cell-to-cell interactions. The rich toolkit of measures originally developed in graph theory and social network analysis enables us to investigate how the interactions between cells influence dynamical processes like cell differentiation based on the graph topology. The development of graph neural networks (GNNs) further facilitates the application of deep learning techniques to spatial transcriptomics data, assuming that we can infer meaningful cell-to-cell interaction graphs. In particular, recent works in this area have investigated different GNN architectures for tasks like spatial domain clustering (Liu et al., 2024b) or prediction (Yasin & Ahmed, 2023).

While these works show the potential of GNNs, the application of graph-based deep learning to spatial transcriptomics data also raises questions that are currently poorly understood. Despite recent advances in sequencing and imaging technology, spatial transcriptomics data typically does not include information on actual cell-to-cell interactions. This necessitates inferring interaction graphs based on other information, e.g., the spatial organisation of cells. However, it is not clear how those methods influence the results. Additionally, the different recently developed GNN architectures have been shown to differ in terms of performance and expressivity. To the best of our knowledge, it has not been systematically evaluated (i) how the performance of GNN architectures differs in the context of cell-type prediction using spatial transcriptomics data, (ii) to what extent it is influenced by the underlying graph construction and (iii) what advantages graph-based deep learning provides over methods that do not account for the topology of potential cell-to-cell interactions.

Addressing these issues, this work investigates the application of different graph construction methods and GNN architectures to cell-type prediction using spatial transcriptomics data. Our key contributions are as follows:

- Using four publicly available spatial transcriptomics datasets, we apply four commonly used methods to construct cell-to-cell interaction graphs.
- We evaluate four GNN models and compare them to a baseline that only uses a cell's gene expressions to investigate the influence of spatial information captured by different graph construction methods.
- By randomizing both gene expression and the interaction graph, we isolate the predictive power that is due to the interaction graph alone.

Our work builds on the hypothesis that graph models that perform better in cell-type prediction are likely to better model the actual (unknown) topology of cell-to-cell interactions. Our systematic evaluation aims to aid the development of GNN-based methods for spatial transcriptomics data to understand the role of cell-to-cell interactions better and thus aid the design of targeted treatments.

2 BACKGROUND AND RELATED WORK

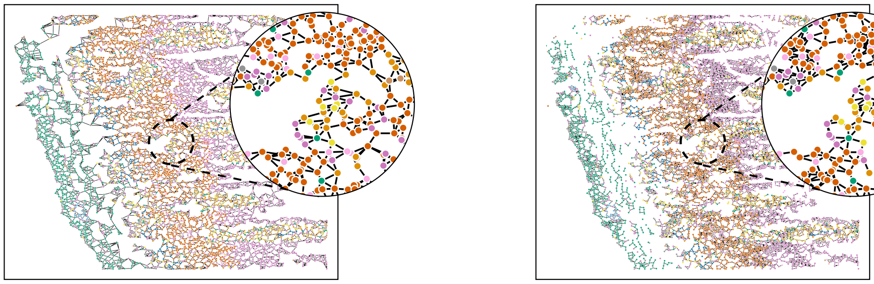
Providing the background of our work, we briefly summarise key methods to construct graphs from spatial data in the following. We further introduce popular graph neural network architectures and briefly review recent applications of deep graph learning in spatial transcriptomics data.

Graph Construction Applying graph-based machine learning to spatial transcriptomics data requires using the cells' position to construct a graph that captures possible cell-to-cell interactions. Given a set of n cells $i = 1, \dots, n$ with coordinates $c_i \in \mathbb{R}^2$, we create a graph $G = (V, E)$ with nodes $v_i \in V$ and undirected edges $(v_i, v_j) \in E$ whenever cells i and j are likely to interact based on their coordinates c_i and c_j. We consider the following construction methods:

- **k-nearest-neighbour (k-NN) graph** : For each cell i we create undirected edges (v_i, v_j) for those k cells c_j that are closest in terms of their Euclidean distance.
- **ϵ-radius graph**: For two cells i and j we include an undirected edge (v_i, v_j) iff their Euclidean distance $dist(c_i, c_j) < \epsilon$.
- **Delaunay graph**: We perform a Delaunay triangulation (Delaunay, 1934) of cell coordinates and use edges of the triangles as edges of the graph.
- **δ-radius-Delaunay graph**: We compute the Delaunay graph and remove all edges (v_i, v_j) where $dist(c_i, c_j) \geq \delta$, i.e. the resulting edges correspond to the intersection of edges in the Delaunay and the ϵ-radius-graph.

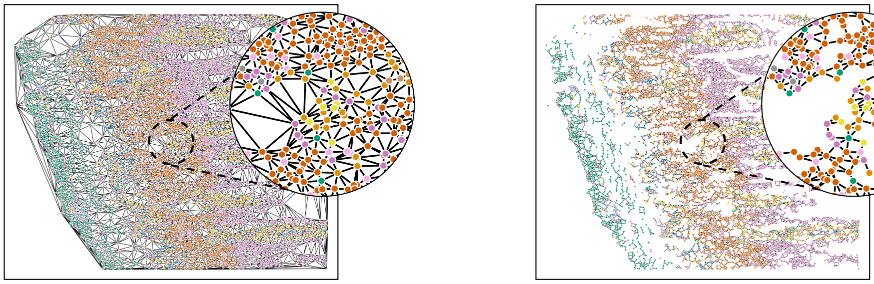
The average node degree for the first two methods largely depends on parameter choices. Since the Delaunay graph does not have any parameters, we choose the parameters k and ϵ for our evaluations to match the average degree of the Delaunay graph approximately and set $\delta = \epsilon$. Properties other than the average degree can vary significantly for differently constructed graphs from the same dataset (for more details see appendix C).

In figure 1, we visualise the resulting graph topologies for an example dataset that captures cell coordinates of an intestine tissue sample, where colours indicate cell types (Petukhov et al., 2022). The included magnifications highlight some key differences: The Delaunay graph contains many long-distance edges connecting cells to closest neighbours in all directions due to the underlying triangulation. In contrast, the ϵ-radius graph imposes a hard upper distance limit, which leads to high connectivity in dense areas and sparse connectivity in regions with low cell density. The k-NN graph balances between the purely distance-based construction of the ϵ-radius graph and the triangulation-based Delaunay graph. In the following, we investigate whether these differences in graph topology translate to differences in the predictive power of graph-based deep learning methods.



(a) k-NN graph

(b) ϵ-radius graph



(c) Delaunay graph

(d) δ-radius-Delaunay graph

Figure 1: Each visualization shows one of the graph construction methods for the intestine dataset (Petukhov et al., 2022). Each color denotes a different cell type.

Graph Neural Networks The use of standard machine learning techniques, such as logistic regression, support vector machines or multi-layer perceptrons (MLPs), requires an additional preprocessing step when applied to graph-based data. In this step nodes in the graph are embedded in Euclidean space such that the embeddings "encode" the node's position in the graph. More recent graph-based deep learning techniques address the limitations of this approach by learning the node representations end-to-end during the optimization of the machine learning model that is used for the actual prediction task. Graph neural networks implement this with a deep neural network architecture that includes at least one *neural message passing layer* (Gilmer et al., 2017), in which nodes exchange features with their neighbours. The feature aggregation ψ from the neighbours $N(v)$ (and v's own feature) and the subsequent application of a transformation function ϕ with learnable parameters allow GNNs to learn internal node representations that capture patterns in the topology of the graph and in the distribution of additional node features. Formally, a message-passing layer can be defined as

$$\boldsymbol{h}'_v = \phi\big(\boldsymbol{h}_v, \psi\big(\boldsymbol{h}_v, \{\boldsymbol{h}_u | u \in N(v)\}\big)\big),$$

where $\boldsymbol{h}_v$ is node v's feature vector. After one or more message-passing layers, a GNN typically uses one final transformation to generate the model's output.

We select four popular GNN variants that differ in expressive power measured by the Weisfeiler-Leman graph isomorphism test (Li & Leskovec, 2022) to investigate their performance on cell-type prediction in section 3: (i) The graph convolutional network (GCN) (Kipf & Welling, 2017) that uses a mean aggregation allowing the model to learn from the distribution of elements in the neighbourhood. (ii) SAGE (Hamilton et al., 2017) that aggregates the neighbours using max-pooling which means it only learns to distinguish unique elements from the neighbourhood. (iii) The graph isomorphism network (GIN) (Xu et al., 2019) that sums up all representations followed by an MLP which leads to the most expressive power since it recognises repetitive elements in the neighbourhood multiset. Lastly, (iv) the graph attention network (GAT) (Velickovic et al., 2018) which employs self-attention to learn how to aggregate the node's neighbourhood to maximize the performance. GAT is similarly expressive as the GCN since it also averages the neighbouring features. Formal definitions of all GNN models are included in appendix A.2.

Application of GNNs to Omics Data In recent years, many approaches have been proposed to solve a variety of tasks by applying GNN models to spatial transcriptomics data (Hetzel et al., 2021; Danishuddin et al., 2024). These include cell-type annotation via label transfer or clustering (Liu et al., 2024a), latent representation learning or the inference of communications between genes or cells (Almet et al., 2021). Most of these methods (systematically summarized in table 1) use one of the aforementioned graph construction methods and GNN models but have not been compared to many of the other approaches highlighting the necessity of a systematic evaluation.

Before the advent of spatial transcriptomics, GNNs had already been applied in a wide range of tasks using a variety of constructed graphs based on, e.g., similarities in gene expression or gene regulation (Hetzel et al., 2021). For the latter case, benchmark studies on phenotype prediction have shown that simpler machine learning models often outperform more sophisticated deep learning approaches (Smith et al., 2020) like GNNs (Brouard et al., 2024) which questions the necessity for more complex models. A similar question was recently raised by Yuan (2024) for spatial domain detection – one of the tasks for spatial transcriptomics where GNNs are commonly applied. This emphasises the need to compare typically used models against simple baselines to measure the influence of spatial information as well as the impact of model complexity on the task.

Table 1: Overview of GNN approaches that solve different tasks in spatial transcriptomics.

Reference	Goal	Space	Graph Constr.	Model	Features
STELLAR (Brbić et al., 2022)	label transfer	cell coord.	ϵ-radius graph	GCN/SAGE	gene expr.
GraphST (Long et al., 2023)	clustering	cell coord.	k-NN graph	GCN	gene expr.
SpaGCN (Hu et al., 2021)	clustering	cell coord. + histology	fully connected and weighted	GCN	PCA of gene expr.
SCGDL (Liu et al., 2023)	clustering	cell coord.	"neighbor graph"	(Bresson & Laurent, 2017)	gene expr.
STAGATE (Dong & Zhang, 2022)	clustering	cell coord.	(pruned) ϵ-radius graph	GAT	gene expr.
SpaceFlow (Ren et al., 2022)	clustering	cell coord.	δ-radius-Del. graph	GCN	gene expr.
SCAN-IT (Cang et al., 2021)	clustering	cell coord.	δ-radius-Del. graph	GCN	gene expr.
CytoCommunity (Hu et al., 2024)	clustering	cell coord.	k-NN graph	(Morris et al., 2019)	cell type
DeepST (Xu et al., 2022)	clustering	cell coord.	k-NN graph	GCN or GAT	gene expr. + hist.
CCSt (Li et al., 2022)	clustering	cell coord.	ϵ-radius graph	GCN	gene expr.
stAA (Fang et al., 2024)	clustering	cell coord.	ϵ-radius graph	custom GNN	gene expr.
Spatial-MGCN (Wang et al., 2023)	clustering	cell coord. + gene expr.	multi-view k-NN graph	GCN	gene expr.
GRAPHDeep (Liu et al., 2024b)	clustering	cell coord.	ϵ-radius or k-NN graph	20 GNN models	gene expr.
SEDR (Xu et al., 2024)	latent repr.	cell coord.	k-NN graph	GCN	gene expr.
stMVC (Zuo et al., 2022)	latent repr.	cell coord. + histology	k-NN graph	GAT	gene expr.
Spage2Vec (Partel & Wählby, 2021)	gene signat. clust.	RNA coord.	ϵ-radius graph	SAGE	gene expr.
GCNG (Yuan & Bar-Joseph, 2020)	gene interaction	cell coord.	ϵ-radius graph	GCN or GAT (graph classification)	gene expr. of ligand-receptor
NCEM (Fischer et al., 2023)	intercell. commun.	cell coord.	ϵ-radius graph	GCN	cell type
CLARIFY (Bafna et al., 2023)	cell + gene interaction	cell coord.	k-NN + gene regul. net.	GCN	gene expr.

3 EXPERIMENTAL EVALUATION

To investigate the influence of different graph construction methods and GNN architectures for cell-type prediction, we perform the following experimental evaluation: Using four datasets detailed in appendix A.1, we use the cell coordinates c_i to construct an undirected graph G using the graph construction methods introduced in section 2. On each of the resulting graphs, we then train a GNN with a dropout and a linear output layer using the gene expressions as node features $\boldsymbol{h}_v$ to predict the cell type for each node. We systematically evaluate the GNN models outlined in section 2 and optimise hyperparameters using a grid search with cross-validation (cf. appendix A.2). In figure 2 we illustrate the architecture used in our experiments. Additionally, we train an MLP on the node features as a baseline that does not use any spatial information. Our code is available on GitHub.

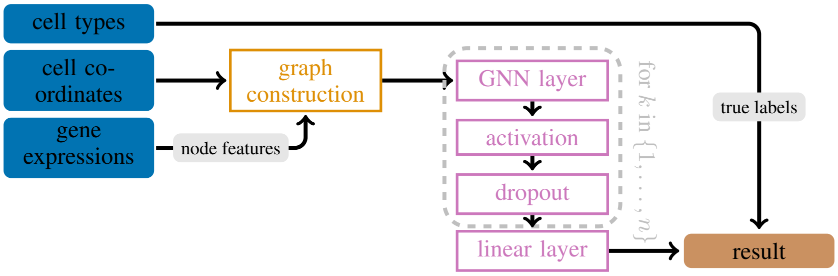


Figure 2: Overview of the architecture that is used for the experiments.

Results Table 2 shows the average weighted F1 scores over 10 runs on a test set for each combination of dataset, graph construction method and GNN architecture. We find that the MLP baseline performs best on all four datasets, closely followed by SAGE. Other models perform substantially worse than the baseline, with GIN performing the worst in our experiments. The performance of all models is similar across graph construction methods, with GIN showing the largest variation.

Table 2: Average weighted F1 scores and standard deviations over 10 runs on a test set. For dataset, graph and model, the best hyperparameters listed in table 5 are chosen based on a validation set, respectively. Note that the MLP does not receive the graph as input.

Dataset	**Graph**	**MLP**	**GCN**	**SAGE**	**GIN**	**GAT**
intestine	k-NN	0.91 ± 0.01	0.80 ± 0.01	0.89 ± 0.01	0.71 ± 0.01	0.78 ± 0.01
	radius	-	0.79 ± 0.01	0.89 ± 0.01	0.70 ± 0.01	0.78 ± 0.02
	Delaunay	-	0.78 ± 0.01	0.90 ± 0.01	0.68 ± 0.01	0.79 ± 0.02
	radius-Del.	-	0.81 ± 0.01	0.90 ± 0.01	0.71 ± 0.01	0.78 ± 0.01
embryo	k-NN	0.85 ± 0.01	0.78 ± 0.01	0.84 ± 0.00	0.76 ± 0.01	0.76 ± 0.01
	radius	-	0.78 ± 0.01	0.84 ± 0.01	0.76 ± 0.01	0.76 ± 0.01
	Delaunay	-	0.72 ± 0.01	0.84 ± 0.01	0.72 ± 0.01	0.76 ± 0.01
	radius-Del.	-	0.78 ± 0.01	0.84 ± 0.01	0.76 ± 0.01	0.76 ± 0.01
hypoth.	k-NN	0.86 ± 0.01	0.68 ± 0.01	0.85 ± 0.01	0.60 ± 0.01	0.74 ± 0.02
	radius	-	0.65 ± 0.01	0.85 ± 0.01	0.57 ± 0.01	0.74 ± 0.02
	Delaunay	-	0.65 ± 0.01	0.85 ± 0.02	0.57 ± 0.02	0.74 ± 0.02
	radius-Del.	-	0.68 ± 0.01	0.85 ± 0.01	0.61 ± 0.01	0.73 ± 0.03
brain	k-NN	0.96 ± 0.01	0.87 ± 0.01	0.95 ± 0.01	0.70 ± 0.01	0.82 ± 0.02
	radius	-	0.86 ± 0.01	0.95 ± 0.01	0.68 ± 0.02	0.83 ± 0.01
	Delaunay	-	0.85 ± 0.01	0.95 ± 0.01	0.67 ± 0.01	0.84 ± 0.02
	radius-Del.	-	0.88 ± 0.01	0.95 ± 0.01	0.73 ± 0.01	0.82 ± 0.01

The results show that using node features (i.e. gene expression) without spatial information is sufficient to make good cell-type predictions. This is expected since the assumed ground truth labels of the datasets were inferred based on the gene expressions. Surprisingly, the best GNN model, namely SAGE, does not perform better but slightly worse than the baseline on all datasets, which indicates that the graph topology – irrespective of the graph construction method – does not contribute any additional information that improves the performance of the model over the baseline. We note that the max-pool aggregator used in SAGE is less expressive compared to the other aggregation steps. The most expressive model (GIN) surprisingly performs worst (Xu et al., 2019). This suggests that SAGE may perform this well – not despite – but *because* it uses the least amount of topological information from the graph, which increases the importance of the cells' own gene expressions. This would mean that SAGE's prediction is mostly based on the node's gene expression, while GIN is likely to assign more weight to the gene expression of other nodes as compared to the gene expression of a cell itself. A possible interpretation of this result is that – at least in the data used in our study – spatial information provides no additional benefit for predicting a cell's type, regardless of the graph construction method. In the following, we test this assumption by isolating spatial information to predict the cell type solely based on the graph topology.

Influence of Graph Topology To exclusively study the predictive power of the cell interaction graph for different graph construction methods and GNN architectures, we perform additional experiments where we randomise the data. We replace the gene expressions with random node features sampled from a normal distribution with the same number of dimensions while leaving the underlying graph untouched. We then compare the predictive power of GNN-based models in the randomised data with *intact* graph topology to a further randomised version where we additionally randomise the graph's topology. Using a Molloy-Reed configuration model, we preserve node degrees but randomly permute edges (Newman, 2018). With this setup, a difference in predictive power between the data with and without randomly shuffled graph is due to the underlying topology, which is constructed from the cells' spatial organisation.

Table 3 shows the results of the GNN models with randomised node features where the value in parentheses gives the baseline performance of a model with both the graph and the node features randomised. Although we observe a considerable drop in predictive performance compared to the results with gene expressions, the models that have access to information on the graph topology perform substantially better than the baseline without topological information. In contrast to our findings above, this shows that the graph topology provides useful information to predict cell types, although it does not seem to provide information that is useful *in addition* to gene expression. Nevertheless, the performance remains similar across graph construction methods despite the differences in graph topology (cf. appendix C). Referring to our assumption that better performance is associated with a more biologically meaningful graph model of cell-to-cell interactions, all tested graph construction methods seem to be equally suitable.

Table 3: Average weighted F1 scores over 10 runs on a test set with random node features. The scores with the shuffled graphs as input are printed in parentheses. Standard deviations are reported in appendix B and hyperparameters are chosen as in table 2.

Dataset	**Graph**	**MLP**	**GCN**	**SAGE**	**GIN**	**GAT**
intestine	k-NN	0.12 (0.12)	0.52 (0.10)	0.43 (0.10)	0.52 (0.12)	0.53 (0.12)
	radius	-	0.51 (0.10)	0.45 (0.11)	0.51 (0.12)	0.53 (0.13)
	Delaunay	-	0.53 (0.10)	0.44 (0.11)	0.52 (0.12)	0.53 (0.12)
	radius-Del.	-	0.51 (0.11)	0.42 (0.10)	0.50 (0.12)	0.52 (0.12)
embryo	k-NN	0.10 (0.10)	0.66 (0.09)	0.53 (0.09)	0.65 (0.09)	0.67 (0.09)
	radius	-	0.66 (0.10)	0.53 (0.10)	0.65 (0.11)	0.67 (0.11)
	Delaunay	-	0.68 (0.08)	0.59 (0.09)	0.67 (0.09)	0.68 (0.09)
	radius-Del.	-	0.66 (0.09)	0.52 (0.10)	0.64 (0.10)	0.67 (0.10)
hypoth.	k-NN	0.22 (0.22)	0.32 (0.20)	0.26 (0.21)	0.32 (0.21)	0.33 (0.22)
	radius	-	0.33 (0.20)	0.27 (0.20)	0.32 (0.22)	0.33 (0.21)
	Delaunay	-	0.33 (0.20)	0.26 (0.20)	0.32 (0.21)	0.33 (0.21)
	radius-Del.	-	0.32 (0.20)	0.26 (0.21)	0.30 (0.21)	0.33 (0.22)
brain	k-NN	0.08 (0.08)	0.39 (0.08)	0.29 (0.07)	0.38 (0.09)	0.40 (0.09)
	radius	-	0.38 (0.08)	0.29 (0.07)	0.38 (0.08)	0.40 (0.09)
	Delaunay	-	0.39 (0.07)	0.28 (0.07)	0.38 (0.09)	0.40 (0.08)
	radius-Del.	-	0.37 (0.08)	0.27 (0.07)	0.36 (0.09)	0.39 (0.09)

4 CONCLUSION

In summary, we investigated approaches to model the spatial organisation of cells using different graph construction methods. We further compared the predictive power of common GNN-based approaches for cell type prediction with a simpler baseline model that does not explicitly utilise spatial information. The results show that – at least for the datasets investigated in our work – including spatial information with GNNs does not increase predictive performance. A possible reason is that the cells' gene expression profiles already carry high predictive power for determining cell types. This is in line with other recent works (Brouard et al., 2024; Yuan, 2024) that questioned the benefit of GNN models for related tasks. Nevertheless, our experiments show that the graph *does* contain useful information for cell-type prediction. At least in our data, the specific method used to construct the graph did not considerably influence the performance of GNNs.

Our work acts as a reminder to carefully evaluate the performance of new machine learning techniques to simpler baseline methods. At the same time, by uncovering intricacies of spatial transcriptomics data, we hope to aid the future development of specialised GNN approaches for this domain and, eventually, the design of novel targeted treatments against diseases.

ACKNOWLEDGEMENTS

Moritz Lampert and Christopher Blöcker acknowledge funding from the German Federal Ministry of Education and Research, Grant No. 100582863 (TissueNet). Ingo Scholtes and Christopher Blöcker acknowledge funding through the Swiss National Science Foundation, Grant No. 176938. Dominic Grün acknowledges funding from the German Federal Ministry of Education and Research (TissueNet - 031L0311A).

REFERENCES

Axel A. Almet, Zixuan Cang, Suoqin Jin, and Qing Nie. The landscape of cell-cell communication through single-cell transcriptomics. *Curr. Opin. Syst. Biol.*, 26:12–23, June 2021. doi: 10.1016/j.coisb.2021.03.007.

Erick Armingol, Adam Officer, Olivier Harismendy, and Nathan E Lewis. Deciphering cell–cell interactions and communication from gene expression. *Nature Reviews Genetics*, 22(2):71–88, 2021.

Mihir Bafna, Hechen Li, and Xiuwei Zhang. CLARIFY: Cell-cell interaction and gene regulatory network refinement from spatially resolved transcriptomics. *Bioinformatics*, 39(39 Suppl 1):i484–i493, June 2023. doi: 10.1093/bioinformatics/btad269.

Thorsten Beier, Constantin Pape, Nasim Rahaman, Timo Prange, Stuart Berg, Davi D. Bock, Albert Cardona, Graham W. Knott, Stephen M. Plaza, Louis K. Scheffer, Ullrich Koethe, Anna Kreshuk, and Fred A. Hamprecht. Multicut brings automated neurite segmentation closer to human performance. *Nat. Methods*, 14(2):101–102, January 2017. doi: 10.1038/nmeth.4151.

Stuart Berg, Dominik Kutra, Thorben Kroeger, Christoph N. Straehle, Bernhard X. Kausler, Carsten Haubold, Martin Schiegg, Janez Ales, Thorsten Beier, Markus Rudy, Kemal Eren, Jaime I. Cervantes, Buote Xu, Fynn Beuttenmueller, Adrian Wolny, Chong Zhang, Ullrich Koethe, Fred A. Hamprecht, and Anna Kreshuk. Ilastik: Interactive machine learning for (bio)image analysis. *Nat. Methods*, 16(12):1226–1232, December 2019. doi: 10.1038/s41592-019-0582-9.

Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *J. Stat. Mech.*, 2008(10):P10008, October 2008. doi: 10.1088/1742-5468/2008/10/P10008.

Maria Brbić, Kaidi Cao, John W. Hickey, Yuqi Tan, Michael P. Snyder, Garry P. Nolan, and Jure Leskovec. Annotation of spatially resolved single-cell data with STELLAR. *Nat. Methods*, 19 (11):1411–1418, November 2022. doi: 10.1038/s41592-022-01651-8.

Xavier Bresson and Thomas Laurent. Residual Gated Graph ConvNets. *CoRR*, abs/1711.07553, 2017.

Céline Brouard, Raphaël Mourad, and Nathalie Vialaneix. Should we really use graph neural networks for transcriptomic prediction? `http://nathalievialaneix.eu/doc/pdf/brouard_etal_BB2024.pdf`, 2024. Accessed: 2024-01-29.

Zixuan Cang, Xinyi Ning, Annika Nie, Min Xu, and Jing Zhang. SCAN-IT: Domain segmentation of spatial transcriptomics images by graph neural network. *BMVC*, 32:406, November 2021.

Danishuddin, Shawez Khan, and Jong Joo Kim. Spatial transcriptomics data and analytical methods: An updated perspective. *Drug Discovery Today*, pp. 103889, 2024. doi: 10.1016/j.drudis.2024.103889.

Boris Delaunay. Sur la sphère vide. *Bull. de l'Académie des Sciences de l'URSS*, (6):793–800, 1934.

Kangning Dong and Shihua Zhang. Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. *Nat. Commun.*, 13(1):1739, April 2022. doi: 10.1038/s41467-022-29439-6.

Zhaoyu Fang, Teng Liu, Ruiqing Zheng, Jin A, Mingzhu Yin, and Min Li. stAA: Adversarial graph autoencoder for spatial clustering task of spatially resolved transcriptomics. *Brief. Bioinform.*, 25 (1):bbad500, January 2024. doi: 10.1093/bib/bbad500.

Matthias Fey and Jan Eric Lenssen. Fast Graph Representation Learning with PyTorch Geometric. *CoRR*, abs/1903.02428, 2019.

David S. Fischer, Anna C. Schaar, and Fabian J. Theis. Modeling intercellular communication in tissues using spatial graphs of cells. *Nat. Biotechnol.*, 41(3):332–336, March 2023. doi: 10.1038/s41587-022-01467-z.

Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural Message Passing for Quantum Chemistry. In *PMLR*, volume 70, pp. 1263–1272, 2017.

Laleh Haghverdi, Aaron T. L. Lun, Michael D. Morgan, and John C. Marioni. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nat. Biotechnol.*, 36(5):421–427, June 2018. doi: 10.1038/nbt.4091.

William L. Hamilton, Zhitao Ying, and Jure Leskovec. Inductive Representation Learning on Large Graphs. In *NIPS*, pp. 1024–1034, 2017.

Leon Hetzel, David S. Fischer, Stephan Günnemann, and Fabian J. Theis. Graph representation learning for single-cell biology. *Curr. Opin. Syst. Biol.*, 28:100347, December 2021. doi: 10.1016/j.coisb.2021.05.008.

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. *Nat. Methods*, 18(11):1342–1351, November 2021. doi: 10.1038/s41592-021-01255-8.

Yuxuan Hu, Jiazhen Rong, Yafei Xu, Runzhi Xie, Jacqueline Peng, Lin Gao, and Kai Tan. Unsupervised and supervised discovery of tissue cellular neighborhoods from cell phenotypes. *Nat. Methods*, pp. 1–12, January 2024. doi: 10.1038/s41592-023-02124-2.

Thomas N. Kipf and Max Welling. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR (Poster)*. OpenReview.net, 2017.

Jiachen Li, Siheng Chen, Xiaoyong Pan, Ye Yuan, and Hong-Bin Shen. Cell clustering for spatial transcriptomics data with graph neural networks. *Nat. Comput. Sci.*, 2(6):399–408, June 2022. doi: 10.1038/s43588-022-00266-5.

Pan Li and Jure Leskovec. The Expressive Power of Graph Neural Networks. In *Graph Neural Networks: Foundations, Frontiers, and Applications*, pp. 63–98. Springer Nature, Singapore, 2022. doi: 10.1007/978-981-16-6054-2_5.

Teng Liu, Zhao-Yu Fang, Xin Li, Li-Ning Zhang, Dong-Sheng Cao, and Ming-Zhu Yin. Graph deep learning enabled spatial domains identification for spatial transcriptomics. *Brief. Bioinform.*, pp. bbad146, April 2023. doi: 10.1093/bib/bbad146.

Teng Liu, Zhao-Yu Fang, Zongbo Zhang, Yongxiang Yu, Min Li, and Ming-Zhu Yin. A comprehensive overview of graph neural network-based approaches to clustering for spatial transcriptomics. *Comput. Struct. Biotech. J.*, 23:106–128, December 2024a. doi: 10.1016/j.csbj.2023.11.055.

Teng Liu, Zhaoyu Fang, Xin Li, Lining Zhang, Dong-Sheng Cao, Min Li, and Mingzhu Yin. Assembling spatial clustering framework for heterogeneous spatial transcriptomics data with GRAPHDeep. *Bioinformatics*, pp. btae023, January 2024b. doi: 10.1093/bioinformatics/btae023.

T. Lohoff, S. Ghazanfar, A. Missarova, N. Koulena, N. Pierson, J. A. Griffiths, E. S. Bardot, C.-H. L. Eng, R. C. V. Tyser, R. Argelaguet, C. Guibentif, S. Srinivas, J. Briscoe, B. D. Simons, A.-K. Hadjantonakis, B. Göttgens, W. Reik, J. Nichols, L. Cai, and J. C. Marioni. Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nat. Biotechnol.*, 40 (1):74–85, January 2022. doi: 10.1038/s41587-021-01006-2.

Yahui Long, Kok Siong Ang, Mengwei Li, Kian Long Kelvin Chong, Raman Sethi, Chengwei Zhong, Hang Xu, Zhiwei Ong, Karishma Sachaphibulkij, Ao Chen, Li Zeng, Huazhu Fu, Min Wu, Lina Hsiu Kim Lim, Longqi Liu, and Jinmiao Chen. Spatially informed clustering, integration, and deconvolution of spatial transcriptomics with GraphST. *Nat. Commun.*, 14(1):1155, March 2023. doi: 10.1038/s41467-023-36796-3.

Aaron T. L. Lun, Davis J. McCarthy, and John C. Marioni. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Res*, 5:2122, 2016. doi: 10.12688/f1000research.9501.2.

Leland McInnes and John Healy. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *CoRR*, abs/1802.03426, 2018.

Jeffrey R. Moffitt, Dhananjay Bambah-Mukku, Stephen W. Eichhorn, Eric Vaughn, Karthik Shekhar, Julio D. Perez, Nimrod D. Rubinstein, Junjie Hao, Aviv Regev, Catherine Dulac, and Xiaowei Zhuang. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science*, 362(6416):eaau5324, November 2018. doi: 10.1126/science.aau5324.

Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. Weisfeiler and Leman Go Neural: Higher-Order Graph Neural Networks. In *AAAI*, pp. 4602–4609. AAAI Press, 2019. doi: 10.1609/aaai.v33i01.33014602.

Mark E. J. Newman. *Networks*. Oxford University Press, 2nd edition edition, 2018. doi: 10.1093/oso/9780198805090.001.0001.

Gabriele Partel and Carolina Wählby. Spage2vec: Unsupervised representation of localized spatial gene expression signatures. *FEBS J.*, 288(6):1859–1870, March 2021. doi: 10.1111/febs.15572.

Viktor Petukhov, Rosalind J. Xu, Ruslan A. Soldatov, Paolo Cadinu, Konstantin Khodosevich, Jeffrey R. Moffitt, and Peter V. Kharchenko. Cell segmentation in imaging-based spatial transcriptomics. *Nat. Biotechnol.*, 40(3):345–354, March 2022. doi: 10.1038/s41587-021-01044-w.

Blanca Pijuan-Sala, Jonathan A. Griffiths, Carolina Guibentif, Tom W. Hiscock, Wajid Jawaid, Fernando J. Calero-Nieto, Carla Mulas, Ximena Ibarra-Soria, Richard C. V. Tyser, Debbie Lee Lian Ho, Wolf Reik, Shankar Srinivas, Benjamin D. Simons, Jennifer Nichols, John C. Marioni, and Berthold Göttgens. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature*, 566(7745):490–495, February 2019. doi: 10.1038/s41586-019-0933-9.

Anjali Rao, Dalia Barkley, Gustavo S. França, and Itai Yanai. Exploring tissue architecture using spatial transcriptomics. *Nature*, 596(7871):211–220, August 2021. doi: 10.1038/s41586-021-03634-9.

Honglei Ren, Benjamin L. Walker, Zixuan Cang, and Qing Nie. Identifying multicellular spatiotemporal organization of cells with SpaceFlow. *Nat. Commun.*, 13(1):4076, July 2022. doi: 10.1038/s41467-022-31739-w.

Aaron M. Smith, Jonathan R. Walsh, John Long, Craig B. Davis, Peter Henstock, Martin R. Hodge, Mateusz Maciejewski, Xinmeng Jasmine Mu, Stephen Ra, Shanrong Zhao, Daniel Ziemek, and Charles K. Fisher. Standard machine learning approaches outperform deep representation learning on phenotype prediction from transcriptomics data. *BMC Bioinformatics*, 21(1):119, March 2020. doi: 10.1186/s12859-020-3427-8.

V. A. Traag, L. Waltman, and N. J. van Eck. From Louvain to Leiden: Guaranteeing well-connected communities. *Sci. Rep.*, 9(1):5233, March 2019. doi: 10.1038/s41598-019-41695-z.

Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph Attention Networks. In *ICLR (Poster)*. OpenReview.net, 2018.

Bo Wang, Jiawei Luo, Ying Liu, Wanwan Shi, Zehao Xiong, Cong Shen, and Yahui Long. Spatial-MGCN: A novel multi-view graph convolutional network for identifying spatial domains with attention mechanism. *Brief. Bioinform.*, 24(5):bbad262, September 2023. doi: 10.1093/bib/bbad262.

Chang Xu, Xiyun Jin, Songren Wei, Pingping Wang, Meng Luo, Zhaochun Xu, Wenyi Yang, Yideng Cai, Lixing Xiao, Xiaoyu Lin, Hongxin Liu, Rui Cheng, Fenglan Pang, Rui Chen, Xi Su, Ying Hu, Guohua Wang, and Qinghua Jiang. DeepST: Identifying spatial domains in spatial transcriptomics by deep learning. *Nucleic Acids Res.*, 50(22):e131, December 2022. doi: 10.1093/nar/gkac901.

Hang Xu, Huazhu Fu, Yahui Long, Kok Siong Ang, Raman Sethi, Kelvin Chong, Mengwei Li, Rom Uddamvathanak, Hong Kai Lee, Jingjing Ling, Ao Chen, Ling Shao, Longqi Liu, and Jinmiao Chen. Unsupervised spatially embedded deep representation of spatial transcriptomics. *Genome Med.*, 16(1):12, January 2024. doi: 10.1186/s13073-024-01283-x.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How Powerful are Graph Neural Networks? In *ICLR*. OpenReview.net, 2019.

Alhassan Yasin and Junaid Ahmed. Graph Neural Operators for Learning on Spatial Transcriptomics Data. May 2023. doi: 10.21203/rs.3.rs-2912958/v1.

Ye Yuan and Ziv Bar-Joseph. GCNG: Graph convolutional networks for inferring gene interaction from spatial transcriptomics data. *Genome Biol.*, 21(1):300, December 2020. doi: 10.1186/s13059-020-02214-w.

Zhiyuan Yuan. MENDER: Fast and scalable tissue structure identification in spatial omics data. *Nat. Commun.*, 15(1):207, January 2024. doi: 10.1038/s41467-023-44367-9.

Meng Zhang, Stephen W. Eichhorn, Brian Zingg, Zizhen Yao, Kaelan Cotter, Hongkui Zeng, Hongwei Dong, and Xiaowei Zhuang. Spatially resolved cell atlas of the mouse primary motor cortex by MERFISH. *Nature*, 598(7879):137–143, October 2021. doi: 10.1038/s41586-021-03705-x.

Chunman Zuo, Yijian Zhang, Chen Cao, Jinwang Feng, Mingqi Jiao, and Luonan Chen. Elucidating tumor heterogeneity from spatially resolved transcriptomics data by multi-view graph collaborative learning. *Nat. Commun.*, 13(1):5962, October 2022. doi: 10.1038/s41467-022-33619-9.

A EXPERIMENTAL SETUP

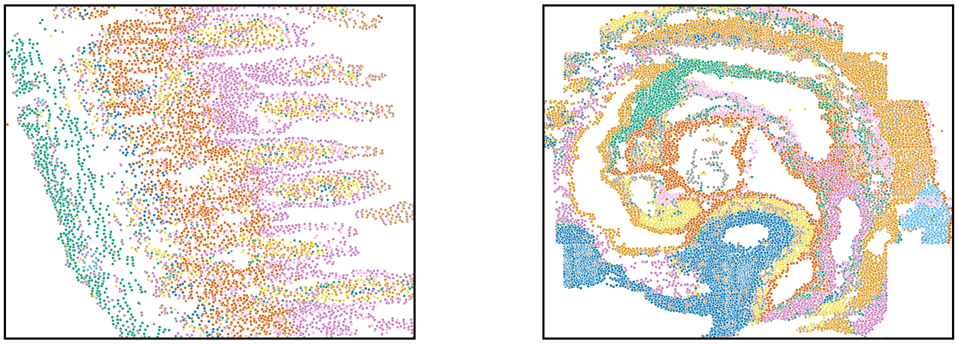
The following provides further details on the experimental setup of this work. We start by describing the datasets and then specify the hyperparameters that are used in the performed grid search.

A.1 DATASETS

The input datasets consist of three parts as visualized in figure 2: cell coordinates, gene expressions and cell types. Cell coordinates and gene expressions were measured using MERFISH and SeqFish (Rao et al., 2021). For the cell types, there are many different established ways to infer them based on clustering or label-transfer approaches and manual preparation. The datasets use different inference methods but all of them only use the gene expression profiles and are independent of the cell coordinates. Table 4 summarises general properties and figure 3 visualises the datasets by coloring each cell type differently.

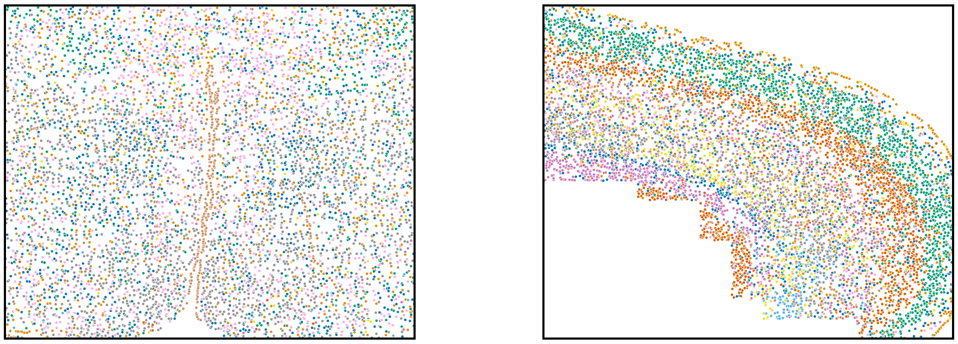
Table 4: Summary of the used datasets.

Name	**Number of Cells**	**Number of Genes**	**Number of Cell Types**
intestine	7416	241	19
embryo	19 451	351	24
hypothalamus	6412	155	16
brain	7626	254	24



(a) Intestine

(b) Embryo



(c) Hypothalamus

(d) Brain

Figure 3: All cells are visualised at the measured coordinates where each colour denotes a different cell type in each tissue. To enhance the contrast between different cell types, we use a colour scheme that contains 10 distinct colours. In datasets with more than ten cell types, multiple cell types are represented with the same colour.

All datasets contain cells from different body parts which means that each dataset can contain different cell types. All of them have an unbalanced distribution of cell types which is visualised in figure 4. The following provides detailed information about the used measurement technology, preprocessing procedures and cell type inference method for each of the datasets.

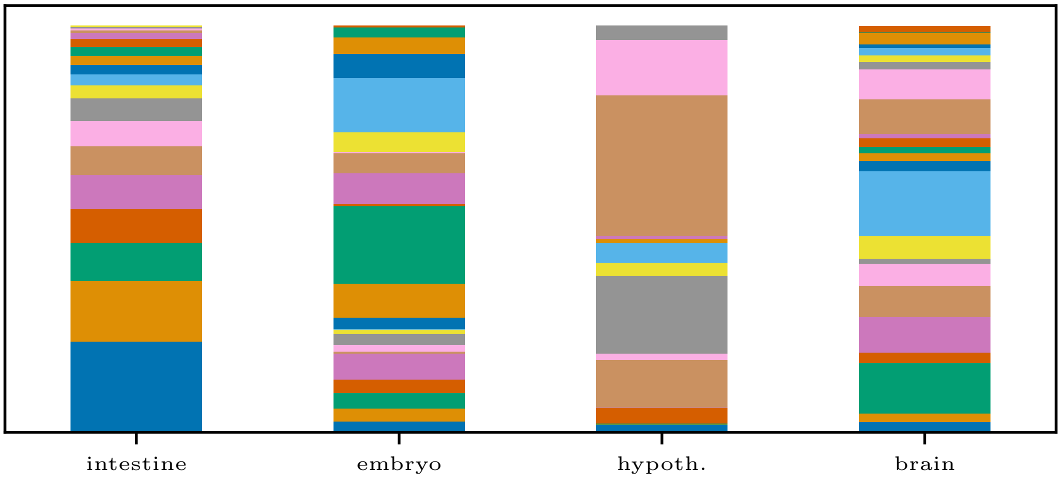


Figure 4: The class imbalance is shown by colouring the bar with the cell-type proportions.

Mouse Intestine The first dataset (Petukhov et al., 2022) contains genes measured in the ileum – the final section of the small intestine – of a mouse using MERFISH. The protocol measures molecule coordinates that correspond to a gene transcript. Petukhov et al. (2022) performed the segmentation using a novel method called Baysor with some priors of membrane stains to assign the gene transcripts to cells. Afterwards, all cells with ten gene transcripts or less and all cells with 900 or more were discarded as a preprocessing step. This work uses the transcript counts and positions of each cell as provided by the authors.

For the cell-type annotation, Petukhov et al. (2022) normalised the gene transcript counts using the cell area obtained via the segmentation. 200 genes were considered highly variable and used for PCA. They projected the first 50 principal component using UMAP (McInnes & Healy, 2018). The authors obtained the final cell types that are also used in this work using Leiden clustering (Traag et al., 2019).

Mouse Embryo The second dataset (Lohoff et al., 2022) contains cells from a mouse embryo recorded with SeqFISH. The original dataset has three samples and only the first is used in this work. Measurements from two different tissue slices are available. This is represented in the data as a third dimension for the cell coordinates. Lohoff et al. (2022) segmented the data into cells using the tools ilastik (Berg et al., 2019) and Multicut (Beier et al., 2017).

The authors transferred the cell types from another dataset (Pijuan-Sala et al., 2019) using PCA (Lun et al., 2016) and a mutual-nearest-neighbours (Haghverdi et al., 2018) approach to improve the data quality. Lohoff et al. (2022) assigned each cell the type of the majority of the cell's labelled neighbours in a joint k-NN graph. This work uses the cell coordinates, transcript counts and cell-type annotation as provided by the authors.

Mouse Hypothalamus The next dataset (Moffitt et al., 2018) contains cells from the hypothalamus of a mouse. It was captured using MERFISH. The original dataset contains samples from multiple mice at different positions. This work only uses the data of the first female mouse that is located $-0.24\,\mathrm{mm}$ away from the bregma – the intersection point of the different skull bones. The cells were segmented using a seeded watershed approach. As gene expressions, this dataset provides a normalised count that incorporates the cell area and the number of gene transcripts per cell.

Moffitt et al. (2018) identified the cell-types with Louvain clustering (Blondel et al., 2008) on the k-NN graph constructed from the data after applying PCA. In this work, we use the dataset as provided.

Mouse Brain This dataset (Zhang et al., 2021) is comprised of samples from a mouse's primary motor cortex We refer to it as the brain dataset for simplicity. The original data consists of slices from multiple samples captured with MERFISH. This work only uses the largest slice – namely slice 153 from Mouse 1 Sample 3. For segmentation, a seeded watershed was used similar to the processing pipeline of the previous dataset. The gene expressions that we utilize in this work were normalized using the cell area.

We further use the cell-type annotation from Zhang et al. (2021). They applied common preprocessing steps like outlier removal, normalization and PCA and constructed a k-NN graph using the 35 first principal components. The authors obtained the cell types with Louvain clustering.

A.2 HYPERPARAMETERS

We train the four GNN models using the methods explained above to construct the input graphs $G = (V, E)$ with nodes $v \in V$ and corresponding node features $\boldsymbol{h}_v$. Given a nodes neighbourhood $N(v) = \{u : (v, u) \in E\}$ and degree $d_v = |N(v)|$, we define a single GNN layer of each selected model formally as follows:

- GCN$(\boldsymbol{h}_v) = \sigma\Big(\Big(\sum_{u \in N(v) \cup \{v\}} \frac{1}{\sqrt{d_v d_u}} \boldsymbol{h}_u\Big) W\Big)$ (Kipf & Welling, 2017)
- SAGE$(\boldsymbol{h}_v) = \sigma\Big(\max_{u \in N(v)}\Big(\sigma\Big(\boldsymbol{h}_u W_{pool}\Big)\Big) W_n + \boldsymbol{h}_v W\Big)$ (Hamilton et al., 2017)
- GIN$(\boldsymbol{h}_v) = \sigma\Big(\sigma\Big(\Big(\boldsymbol{h}_v + \sum_{u \in N(v)} \boldsymbol{h}_u\Big) W_1\Big) W_2\Big)$ (Xu et al., 2019)
- GAT$(\boldsymbol{h}_v) = \sigma\Big(\sum_{u \in N(v) \cup \{v\}} \boldsymbol{a}_{vu} \boldsymbol{h}_u W\Big)$ with attention weights normalized across $N(v)$
  $\boldsymbol{a}_{vu} = \text{Softmax}_{u \in N(v) \cup \{v\}}\Big(\text{LeakyReLU}\Big(\boldsymbol{h}_v W A + \boldsymbol{h}_u W A_n\Big)\Big)$ (Velickovic et al., 2018)

The matrices W and A denote learnable weights and function σ a non-linear activation where we use ReLU throughout this work. For the graph attention network (Velickovic et al., 2018), a common approach is to use multiple heads per layer. This means that each layer can consist of multiple GAT$(\boldsymbol{h}_v)$ that are concatenated to get the final output representation.

For all dataset, graph construction method and GNN model combinations, the hyperparameters listed in table 5 are tested using grid search and we select the best configuration based on the average weighted F1 score over ten runs each on a random validation set. The models are optimized with Adam using a batch size of 512 and the training stops early if there is no improvement after 5 epochs to avoid overfitting. As a baseline, we train MLP models that receive only a node's gene expressions to predict its cell type using the same hyperparameter combinations.

Table 5: Overview of hyperparameters that are used.

Hyperparameters	**Options**
number of layers	1; 2; 3
number of neurons	64; 128
learning rate	10^{-2}; 10^{-3}
dropout rate	0.0; 0.2
number of heads (only GAT)	1; 8

The evaluations are run on a machine with an AMD Ryzen™ 9 7900X 12-Core Processor as CPU with 64 GB memory. As GPU, an NVIDIA® GeForce RTX® 4080 with 32 GB RAM is used. We use PyTorch Geometric (Fey & Lenssen, 2019) for the GNN implementations.

B STANDARD DEVIATIONS FOR TABLE 3

We report the standard deviations that correspond to the average values of table 3 in table 6.

Table 6: Standard deviations corresponding to the mean values reported in table 3.

Dataset	**Graph**	**MLP**	**GCN**	**SAGE**	**GIN**	**GAT**
intestine	k-NN	0.01 (0.01)	0.01 (0.01)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)
	radius	-	0.01 (0.01)	0.02 (0.01)	0.01 (0.01)	0.01 (0.01)
	Delaunay	-	0.01 (0.01)	0.02 (0.01)	0.01 (0.00)	0.02 (0.01)
	radius-Del.	-	0.02 (0.01)	0.02 (0.01)	0.01 (0.01)	0.01 (0.01)
embryo	k-NN	0.01 (0.00)	0.01 (0.00)	0.02 (0.01)	0.01 (0.00)	0.01 (0.00)
	radius	-	0.01 (0.01)	0.02 (0.00)	0.01 (0.01)	0.01 (0.00)
	Delaunay	-	0.01 (0.00)	0.01 (0.00)	0.01 (0.01)	0.01 (0.01)
	radius-Del.	-	0.01 (0.00)	0.01 (0.01)	0.01 (0.00)	0.01 (0.00)
hypoth.	k-NN	0.01 (0.01)	0.02 (0.02)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)
	radius	-	0.01 (0.02)	0.02 (0.02)	0.02 (0.01)	0.02 (0.01)
	Delaunay	-	0.01 (0.01)	0.02 (0.02)	0.01 (0.01)	0.01 (0.01)
	radius-Del.	-	0.03 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.02)
brain	k-NN	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.02 (0.00)	0.01 (0.01)
	radius	-	0.01 (0.00)	0.02 (0.01)	0.02 (0.01)	0.01 (0.01)
	Delaunay	-	0.01 (0.01)	0.02 (0.01)	0.01 (0.01)	0.01 (0.01)
	radius-Del.	-	0.01 (0.01)	0.01 (0.01)	0.02 (0.01)	0.01 (0.00)

C GRAPH CONSTRUCTION METHODS

We provide a detailed analysis of the graph construction methods evaluated in this work and their differences in the following. Table 7 shows the mean with the corresponding standard deviation of the degrees. As mentioned above, the parameters k and ϵ for the k-NN and the ϵ-radius graph are chosen to approximate the degree of the parameterless Delaunay construction which is approximately eight for all datasets except the embryo dataset. This is due to the additional third dimension of the cell coordinates (see appendix A.1). Since the nearest neighbour relation between nodes is not symmetric, the mean degree is expected to be larger than k for the undirected k-NN graph that results from removing the direction of all directed edges to the nearest neighbours. We choose $k = 5$ for all datasets to get an average degree that is close to the average degree of the Delaunay graph. The distance threshold of the ϵ-radius graph is selected using a binary-search-like approach that searches for an ϵ producing a graph with an average degree of eight. The algorithm selects an ϵ with a fault tolerance of 0.2 for each dataset. The δ-radius-Delaunay graph uses the same value, i.e. $\delta = \epsilon$. All of the resulting graphs contain a self-loop for each node. Table 7 further includes baseline statistics for each method that are based on $10\,000$ nodes placed uniformly at random inside a unit square.

As mentioned above, the magnifications in figure 1 visualise the key differences between the evaluated construction methods: The graph resulting from the Delaunay triangulation contains many long-distance edges connecting cells to neighbours in all directions. Because of this, the graph exhibits a similar connectivity in densely populated areas as in regions with only a few cells. In contrast, the ϵ-radius graph contains edges based on a hard upper limit on the distance between nodes, which leads to high connectivity in dense areas and very sparse connectivity in regions with low cell density. The k-NN graph balances the purely distance-based ϵ-radius approach and the triangulation-based Delaunay graph construction resulting in a well-connected graph with similar connectivity in all regions. The δ-radius-Delaunay graph maintains similar connectivity in areas with high cell density as the k-NN graph but is less connected over larger distances due to the upper limit of the ϵ-radius graph.

We can see in table 7 that the properties described above also manifest in some of the measures that are commonly used to quantify the graph topology (Newman, 2018). In particular, the diameter – i.e. the longest path out of all shortest paths between all node pairs – and the average shortest path are substantially smaller for the Delaunay graph that contains edges over long distances. Note that across datasets, the path length measures differ according to the number of cells that are used to construct the graph. The number of components $\#C$ and the size of the largest component $\mathcal{S}(G_1, G)$ highlights the good connectivity of the Delaunay graph for all regions but also how the k-NN graph can be seen as a balance between ϵ-radius and Delaunay graph in terms of connectivity.

The average clustering coefficient measures how often two neighbours of a node are connected and is similar across the investigated approaches and datasets. The degree assortativity coefficient describes the tendency of nodes to connect to similar nodes. The values differ for all construction methods but are similar across datasets. The small degree assortativity coefficients in combination with the degree's small standard deviations of the k-NN graph and Delaunay graph show that the nodes in all regions have a similar degree. The higher values for ϵ-radius and δ-radius-Delaunay graph emphasize that the node degrees vary but are similar in the respective regions of cell density.

Generally, we observe that the investigated properties are different across construction methods but similar for each method across datasets. Furthermore, the baselines using random coordinates for the graph constructions also exhibit similar properties as the other graphs respectively. Thus, regardless of the cells' locations used for the graph construction, all graphs of the evaluated construction methods exhibit substantially different topological properties motivating the above evaluation to investigate if these differences translate to varying GNN performance for cell type prediction.

Table 7: Properties of the graphs constructed from real-world datasets. The average degree $\langle d\rangle$ and standard deviation, the diameter diam(G), the average shortest path length $\langle l\rangle$, the number of components #C., the relative size of the largest component $\mathcal{S}(G_1, G)$, the average clustering coefficient $\mathcal{C}(G)$ and the degree assortativity coefficient r are listed (Newman, 2018). Note that the path measures are calculated for the largest connected component which can result in differences due to the component size.

Dataset	**Constr.**	$\langle d\rangle$	**diam**(G)	$\langle l\rangle$	**# C.**	$\mathcal{S}(G_1, G)$	$\mathcal{C}(G)$	r
intestine	k-NN	6.77 ± 0.86	164	54.33	2	0.99	0.45	0.29
	ϵ-radius	8.10 ± 2.81	141	46.88	108	0.79	0.50	0.79
	Delaunay	7.99 ± 1.22	54	24.08	1	1.00	0.43	0.15
	δ-radius-Del.	6.40 ± 1.51	161	53.64	108	0.79	0.41	0.58
embryo	k-NN	6.64 ± 0.76	220	71.84	9	0.57	0.38	0.29
	ϵ-radius	8.12 ± 2.31	242	77.69	148	0.56	0.46	0.78
	Delaunay	16.49 ± 6.34	11	6.10	1	1.00	0.48	0.28
	δ-radius-Del.	7.32 ± 1.69	255	80.91	148	0.56	0.44	0.60
hypothalamus	k-NN	6.71 ± 0.83	114	44.36	1	1.00	0.44	0.28
	ϵ-radius	8.09 ± 2.24	111	40.83	17	0.99	0.51	0.72
	Delaunay	7.99 ± 1.24	44	22.89	1	1.00	0.43	0.14
	δ-radius-Del.	6.54 ± 1.32	119	46.00	17	0.99	0.42	0.48
brain	k-NN	6.79 ± 0.89	154	53.07	2	1.00	0.48	0.28
	ϵ-radius	8.00 ± 2.43	167	53.26	52	0.98	0.54	0.73
	Delaunay	7.99 ± 1.28	53	23.69	1	1.00	0.43	0.14
	δ-radius-Del.	6.31 ± 1.37	188	59.37	52	0.98	0.43	0.52
random	k-NN	6.85 ± 0.95	148	57.61	2	1.00	0.53	0.28
	ϵ-radius	8.02 ± 2.54	164	62.29	62	0.97	0.60	0.72
	Delaunay	7.99 ± 1.38	52	26.14	1	1.00	0.44	0.15
	δ-radius-Del.	6.13 ± 1.34	186	71.42	62	0.97	0.47	0.49