

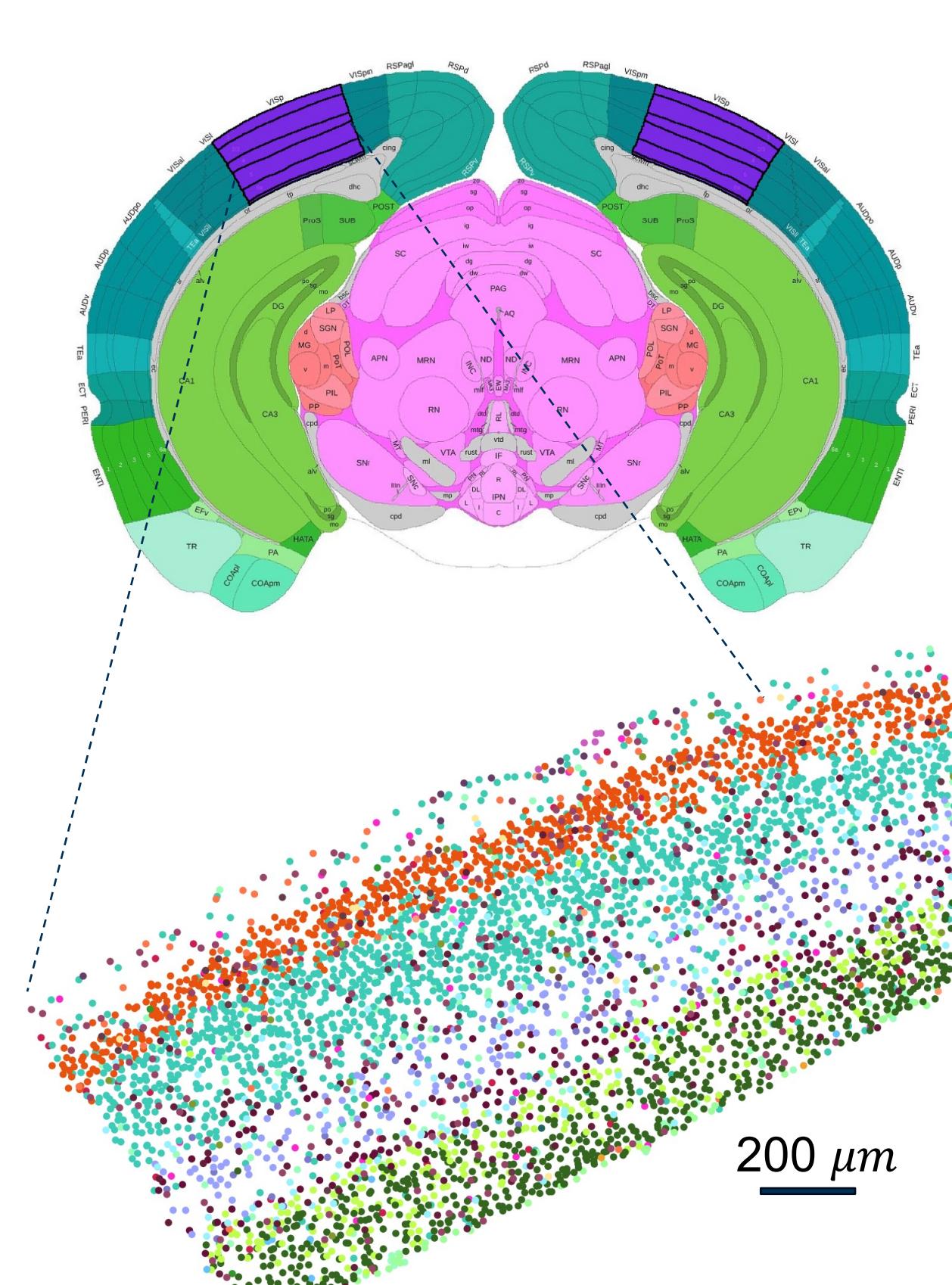
1. Introduction

Peptides released by one neuron can bind to G-protein coupled receptors (GPCRs) on other neurons and modulate their activity. Many such cognate neuropeptide-GPCR pairs have been identified in the literature. These constitute multiple molecularly distinct communication networks between neurons. Single cell transcriptomic data suggests that a vast majority of cortical neurons participate in such communication in a cell-type dependent manner (Smith et al. 2019).

Spatial transcriptomic experiments measure gene expression while preserving context within the tissue. We consider a recent whole mouse brain dataset obtained with an *in-situ* hybridization-based technique called MERFISH (Yao et al. 2023) and develop graph-based methods to study the organization of putative peptidergic communication networks.

We explore regimes in which neighborhood information is helpful for cell type classification with graph neural networks and adapt Node2Vec (Grover et al. 2016) to obtain multilayer graph embeddings. We speculate that cell clusters obtained in this way correspond to spatial domains relevant from a functional point of view.

2. Dataset overview



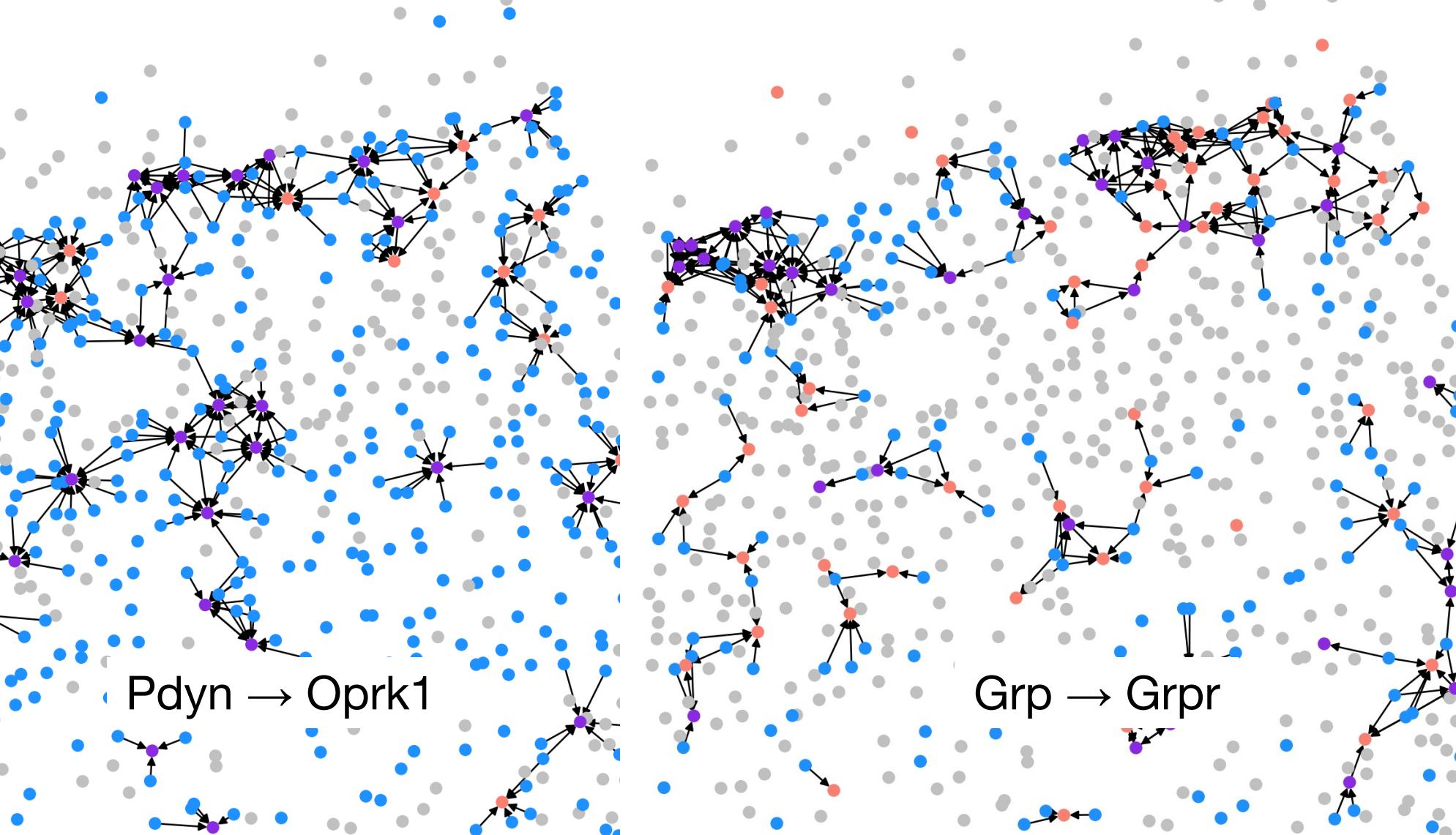
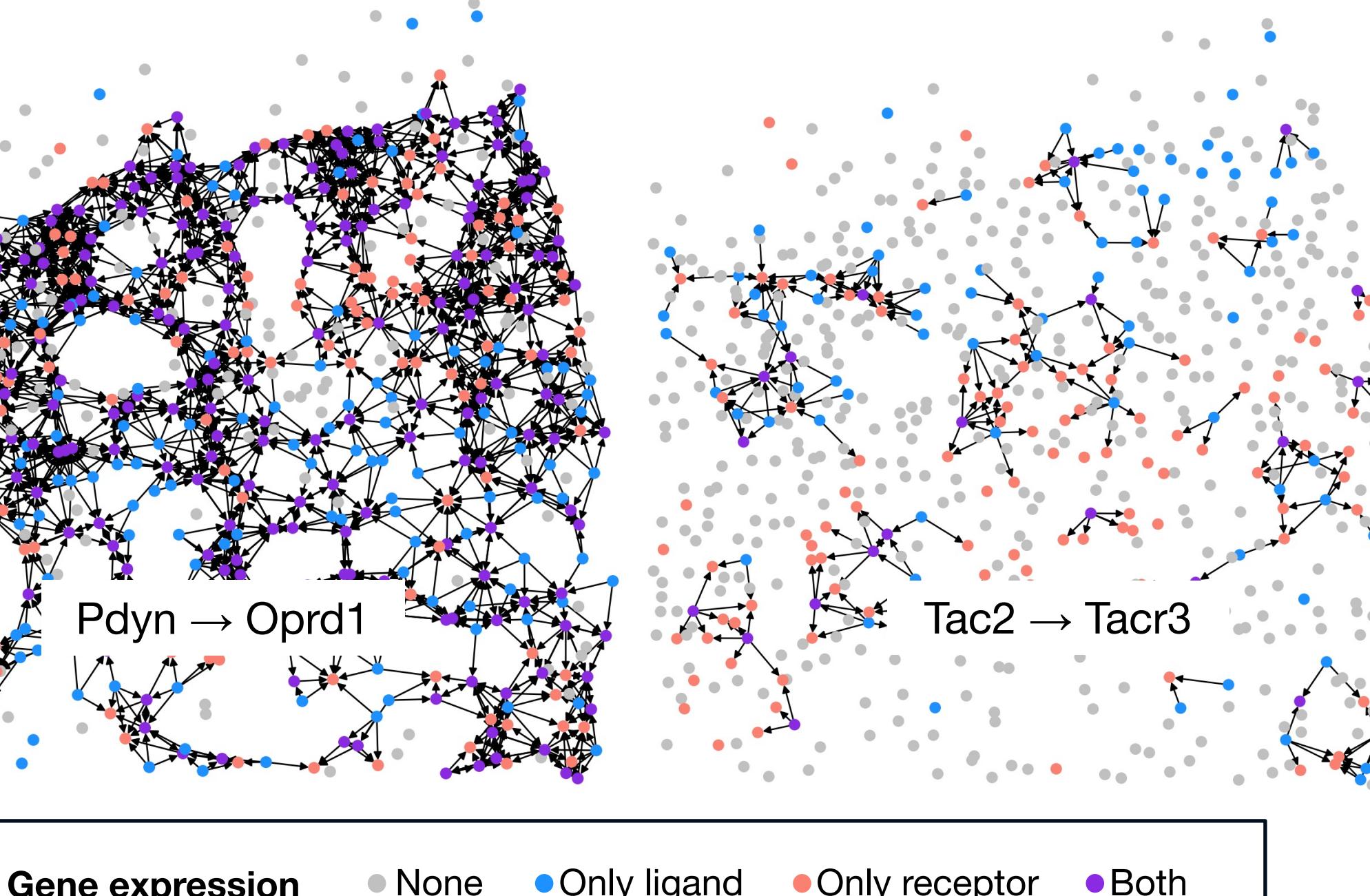
Peptidergic Communication networks are directed multilayer graphs.

The multilayer graph G is the set $\{V, \{E_l\}\}$ where:
 V is the set of vertices (cells)
 E_l is the set of directed edges for the l^{th} network.

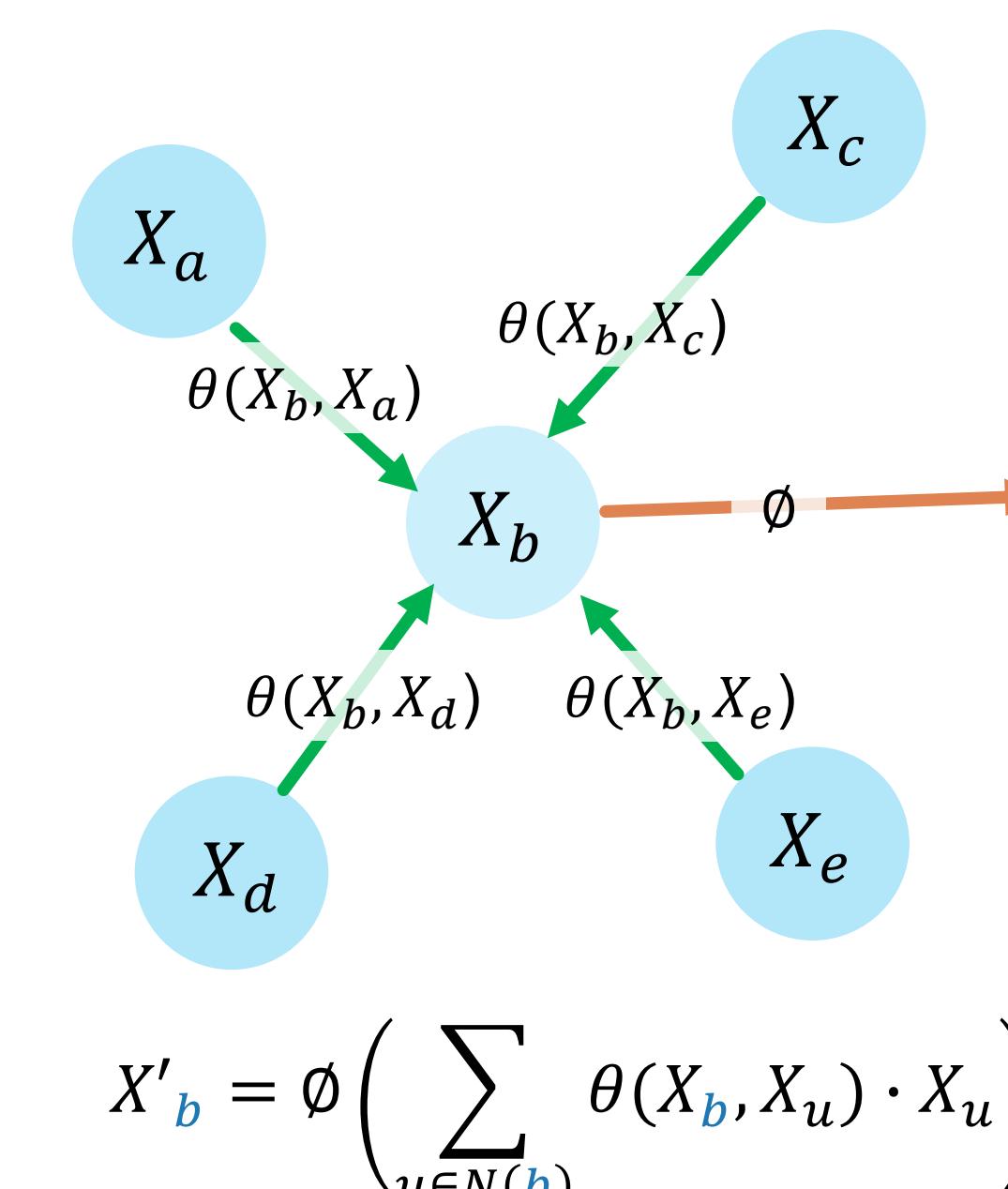
For the l^{th} network, an edge exists from vertex i to vertex j if all the following conditions are met:

1. distance $d[i, j] \leq d_{threshold}$
2. vertex i expresses the peptide precursor gene
3. vertex j expresses the GPCR gene

We set $d_{threshold} = 40 \mu m$ for all experiments unless mentioned otherwise.



3. Node classification with graph neural networks

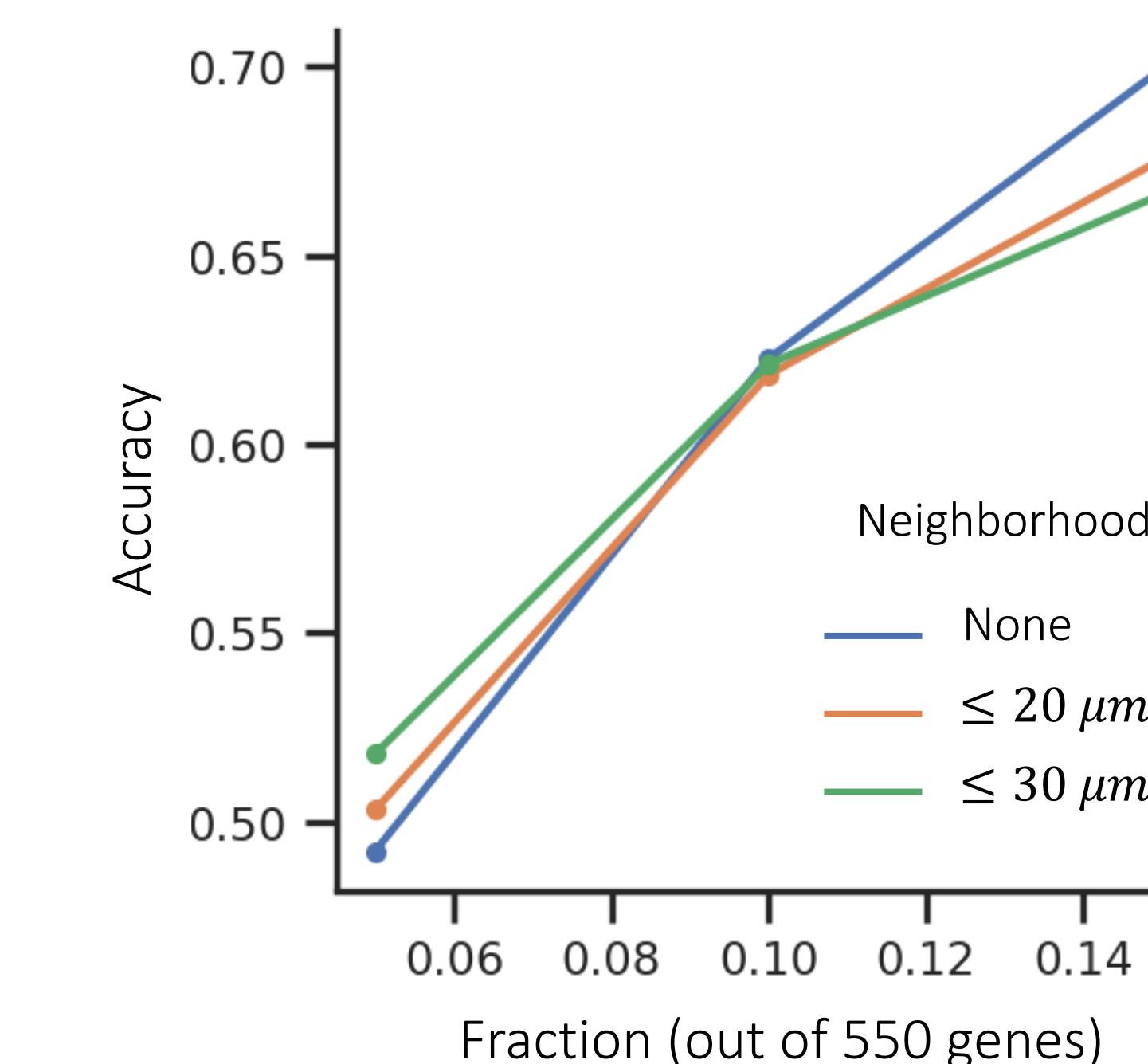


Graph neural networks combine features of nodes that are adjacent according to the underlying graph.

A nonlinear function ϕ of aggregated (sum here) weighted node features determines updated node representation.

In **GATv2**, the node features are weighted using a function of node pairs, θ .

Functions ϕ and θ are learned to optimize a per-node classification objective.

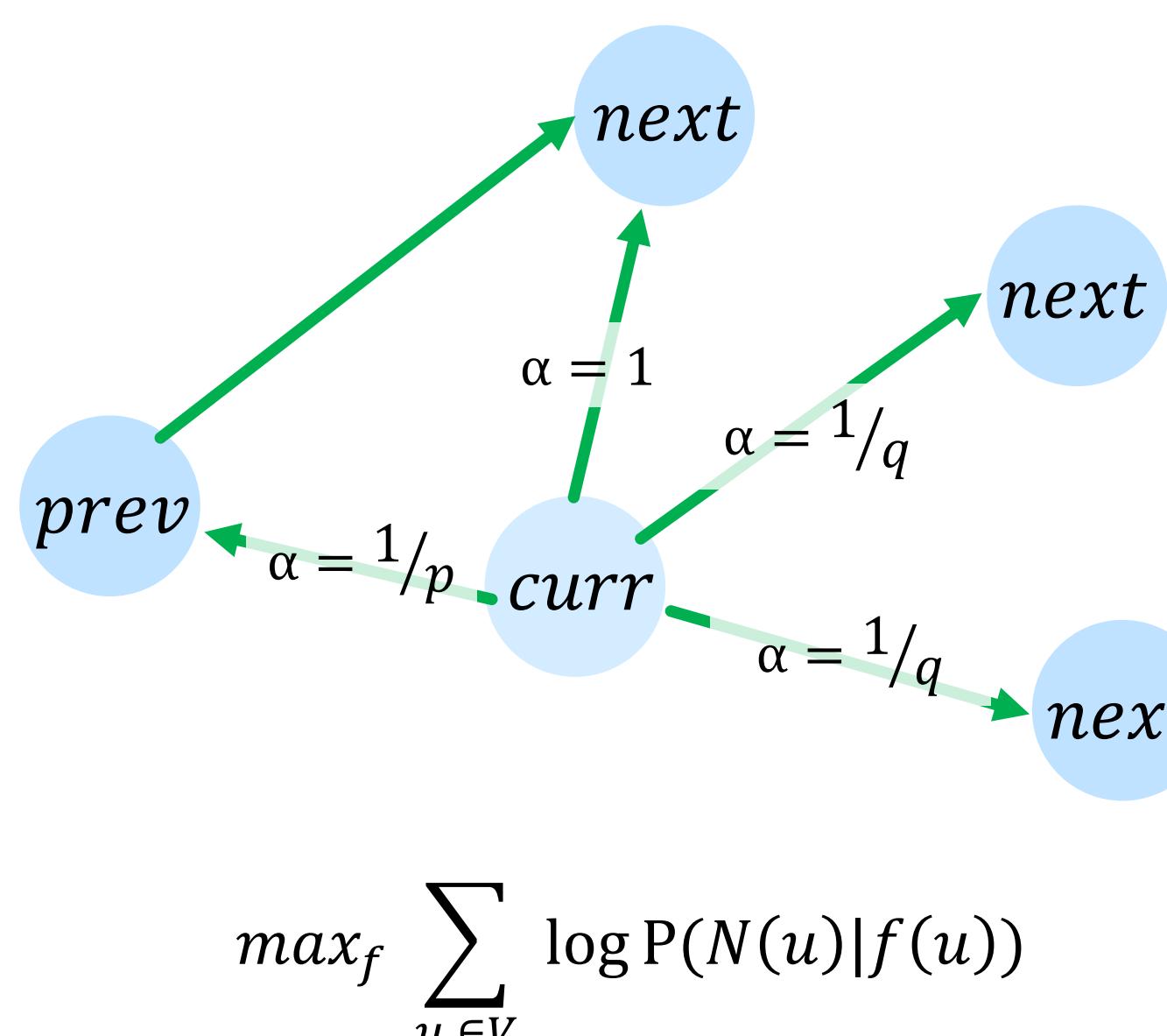


Here, we consider cells as nodes, and their gene expression as initial node features.

Supertype classification is challenging when few genes are available.

We find that gene expression in neighboring cells can improve classification in this regime.

4. Node embeddings

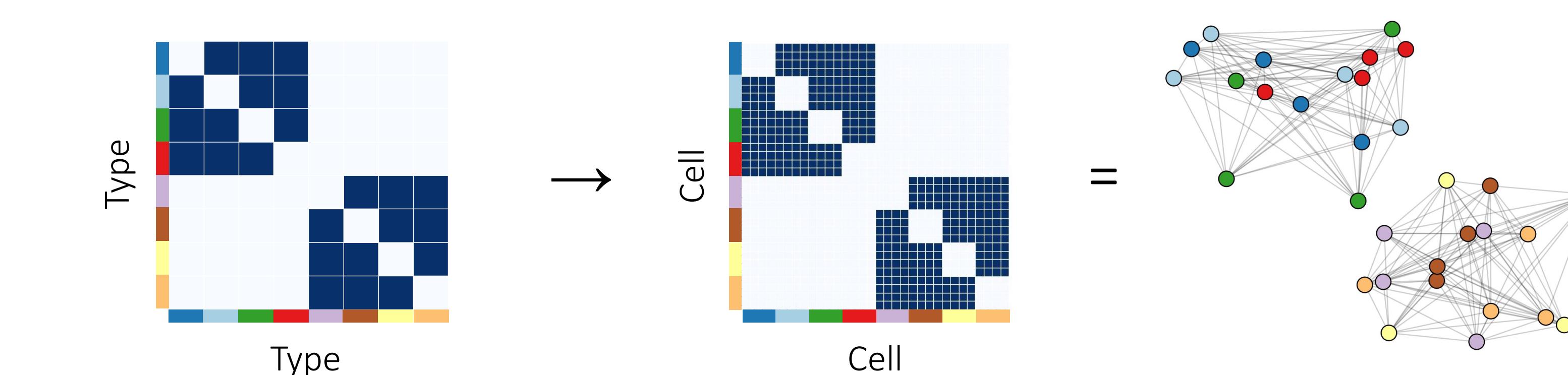


Node2Vec learns an embedding $f(u)$ that is predictive of the neighborhood $N(u)$ of each node u .

The neighborhood for each node is determined through fixed length, 2nd order random walks starting from each node.

Parameters p and q determine the transition probabilities for the random walk.

We build intuition for how peptidergic connectivity could partition cells into meaningful groups even in the absence of explicit gene expression information with a toy problem.

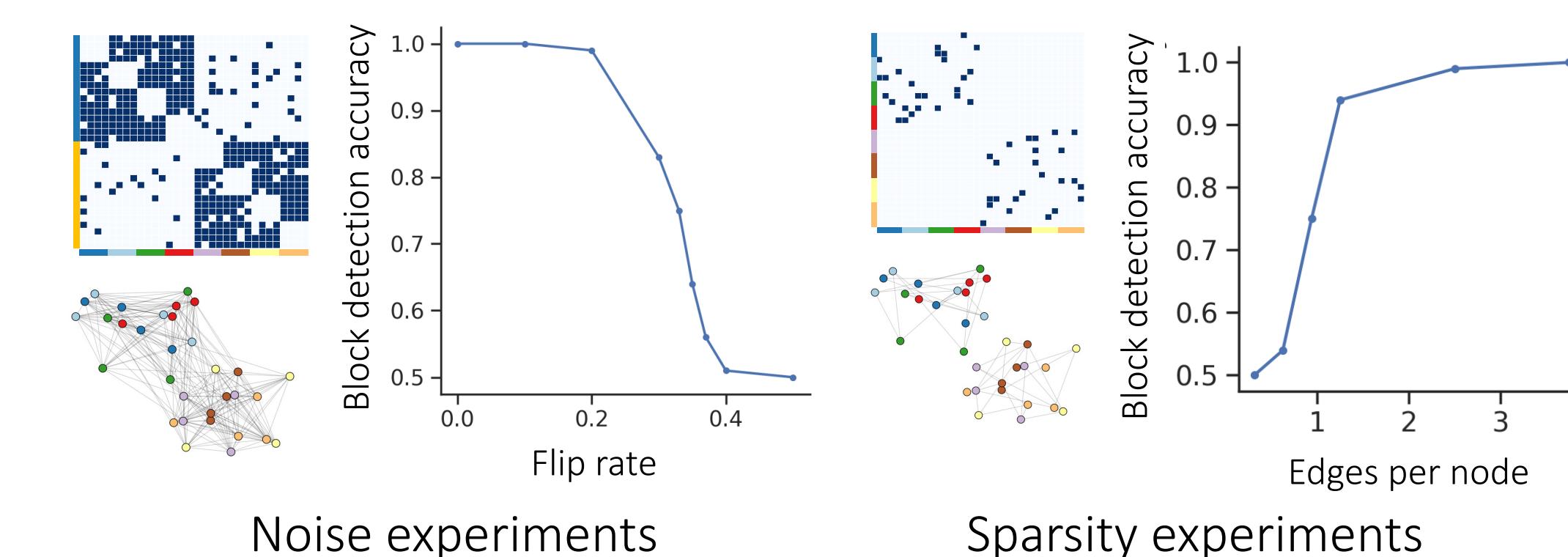


The type x type adjacency matrix can be sampled to obtain a cell x cell adjacency matrix. The block structure is equivalent to the graph having distinct connected components.

Clustering the embeddings obtained with Node2Vec is a scalable way to identify such components.

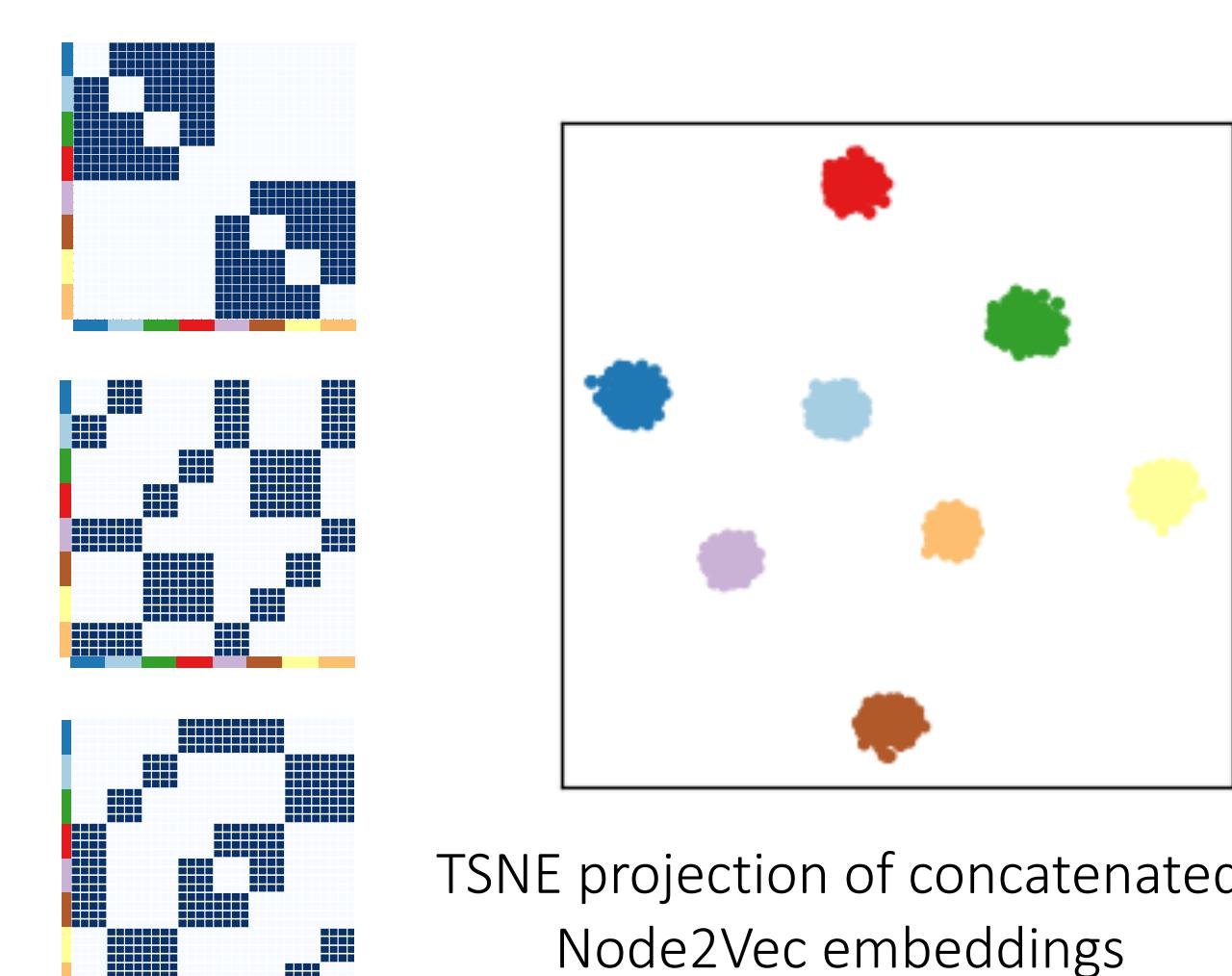
5. Multilayer graph clustering

Node2Vec embeddings cluster according to the underlying block structure in a manner that is robust to sparsity as well as noise in the cell x cell graph.



For multilayer graphs, one approach is to obtain independent embeddings for each layer in the multigraph. This approach can identify all the ‘types’ when the blocks are chosen carefully across layers

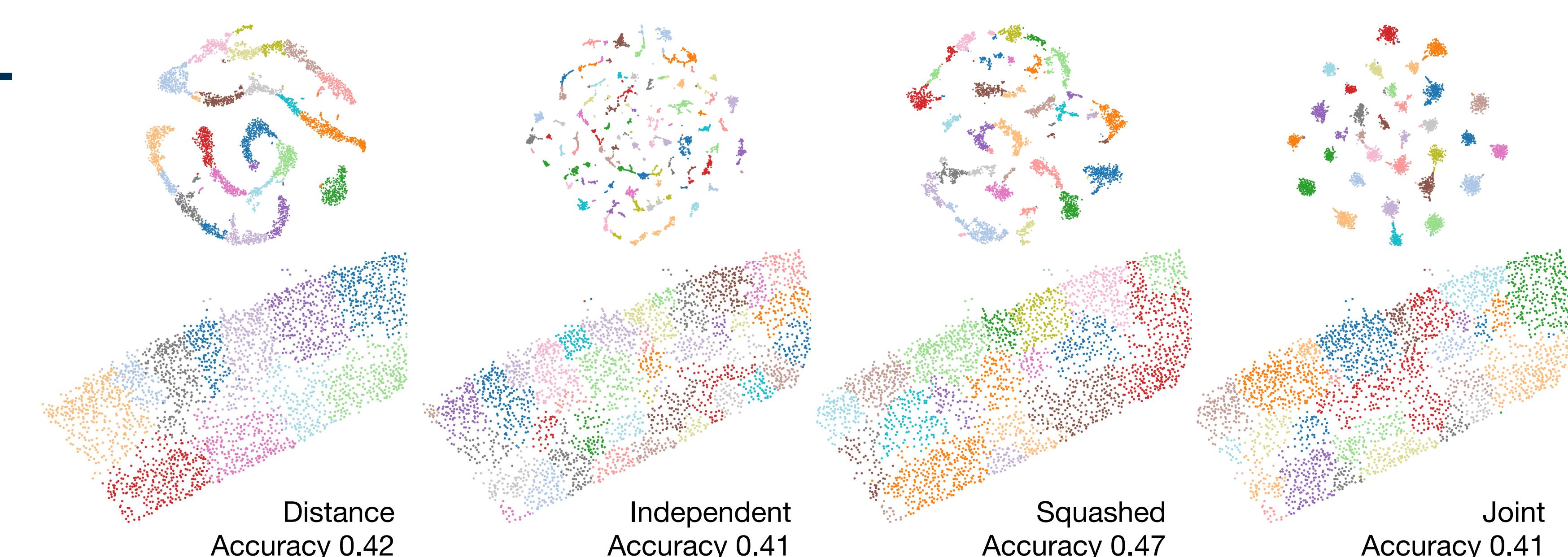
The peptidergic multilayer graph does not contain any edges between layers. We can exploit this fact to jointly consider all the layers of the graph, and efficiently construct random walks from each layer.



6. Application to MERFISH VI Sp data

We inspect the utility of graph embeddings in identifying celltypes (subclass) and spatial groupings in a few different settings:

1. considering only proximity of cells in the graph (**distance**)
2. concatenated embeddings of individual peptidergic graphs (**independent**)
3. collapsing the multi-layer graphs into a single layer graph (**squashed**)
4. jointly considering the multilayer graphs with an adapted Node2Vec model (**joint**)



7. Limitations and future work

Only 5 out of the 37 cognate neuropeptide precursor and GPCR pairs were experimentally measured in the MERFISH dataset, which precludes an assessment at more detailed levels of the taxonomy

Neuron morphology may be an important consideration relevant for the full spatial extent of peptidergic communication. Arbor density representations derived from patch-seq datasets could offer a way forward.

8. References and Acknowledgments

Smith, S. J., Sümbül, U., Graybuck, L. T., Collman, F., Seshamani, S., Gala, R., ... & Hawrylycz, M. (2019). Single-cell transcriptomic evidence for dense intracortical neuropeptide networks. *elife*, 8, e47889.

Yao, Z., van Velthoven, C. T., Kunst, M., Zhang, M., McMillen, D., Lee, C., ... & Zeng, H. (2023). A high-resolution transcriptomic and spatial atlas of cell types in the whole mouse brain. *Nature*, 624(7991), 317-332.

Grover, A., & Leskovec, J. (2016, August). node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 855-864).

Code repository: <https://github.com/donghyun-ethan-kim/cciGraphs>

We thank Tom Chartrand, Ian Convy, Olga Gliko, Anna Grim, Yeganeh Marghi, Uygar Sümbül, Meghan Turner, and Kasey Zhang for helpful feedback and discussions.