

GraSPy: an Open Source Python Package for Statistical Connectomics

Benjamin D. Pedigo¹, Jaewon Chung¹, Eric W. Bridgeford², Bijan Varjavand³, Carey E. Priebe³, Joshua T. Vogelstein¹

¹Department of Biomedical Engineering, Johns Hopkins University ²Department of Biostatistics, Johns Hopkins University ³Department of Applied Mathematics and Statistics, Johns Hopkins University

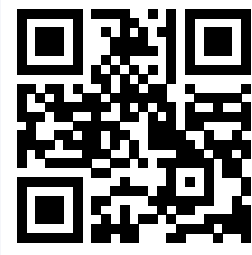
Summary

- Connectome datasets are growing in size
- Analysis is a current bottleneck

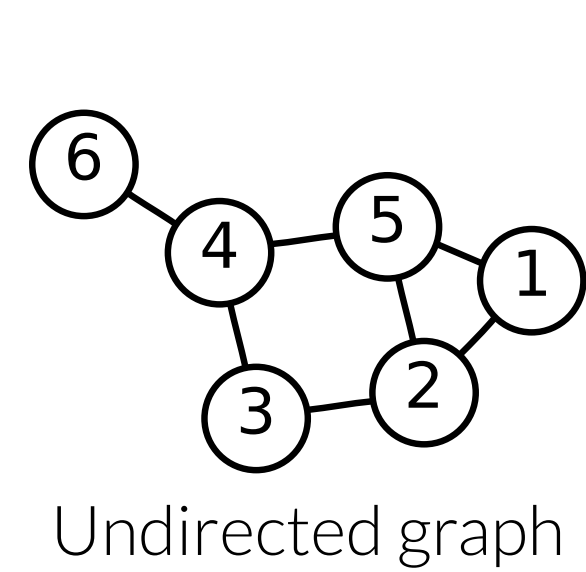
- Graphs (networks) are natural models
- Require specific statistical tools and implementations [1]

- GraSPy**: open source python toolkit [2]
- Graph sampling, estimation, embedding, testing

- Accelerate understanding of connectomes with valid graph inference
- Available at neurodata.io/graspy



Graphs



$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Adjacency matrix A

- Model adjacency matrices as samples from a matrix of probabilities P :
- Latent position random graphs: latent vector for each node determining probability of connections
- Find P using the dot product of the latent positions

$$P = XX^T, \quad X \in \mathbb{R}^{n \times d}$$

where row i of X is a latent vector for node j [1]

Random Graph Models

Graph model	Latent representation	Realization (sample)
Ernyos-Reyni (ER): $p_{ij} = p$ p is overall connection probability <i>Latent representation:</i> all nodes share same latent position		
Stochastic block model (SBM): $p_{ij} = B_{\tau_i, \tau_j}$ $B \in [0, 1]^{K \times K}, \tau \in \mathbb{R}^n$ B contains block-block connection probabilities, τ block assignments <i>Latent representation:</i> nodes in a block share same latent position.		
Degree-corrected SBM (DCSBM): $p_{ij} = B_{\tau_i, \tau_j} d_i d_j$ $d \in \mathbb{R}^n$ represents expected degree for each node <i>Latent representation:</i> nodes in block live on a line, expected degree determines their position on line		
Random dot product graph (RDPG): $p_{ij} = \langle x_i, y_j \rangle$ $x_i, y_i \in \mathbb{R}^d$ x_i, y_j are latent positions for node i, j <i>Latent representation</i> point in \mathbb{R}^d for each node		

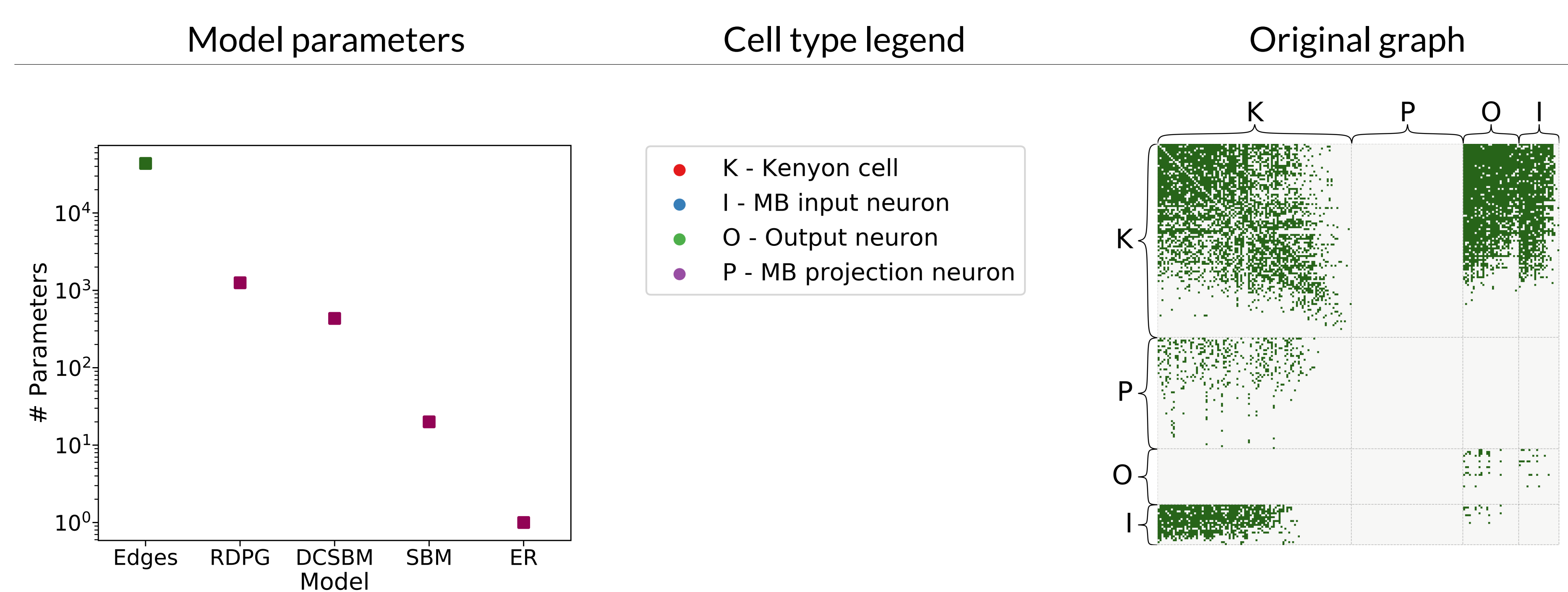


Figure 1. Four random graph models displayed with their corresponding representations in latent space and a sample from the graph model. All graph models are fit to the *Drosophila* left mushroom body from Eichner et al [4]. The latent space representations are calculated by computing an adjacency spectral embedding of the estimated probability matrix from each model.

Graph embeddings

- Embeddings convert graphs into Euclidian representations, allowing subsequent inference and estimation
- Multigraph embeddings can place a population of graphs in the same latent dimensions
- Embeddings such as adjacency spectral embedding (ASE) can be generative models for random graphs (Figure 1).

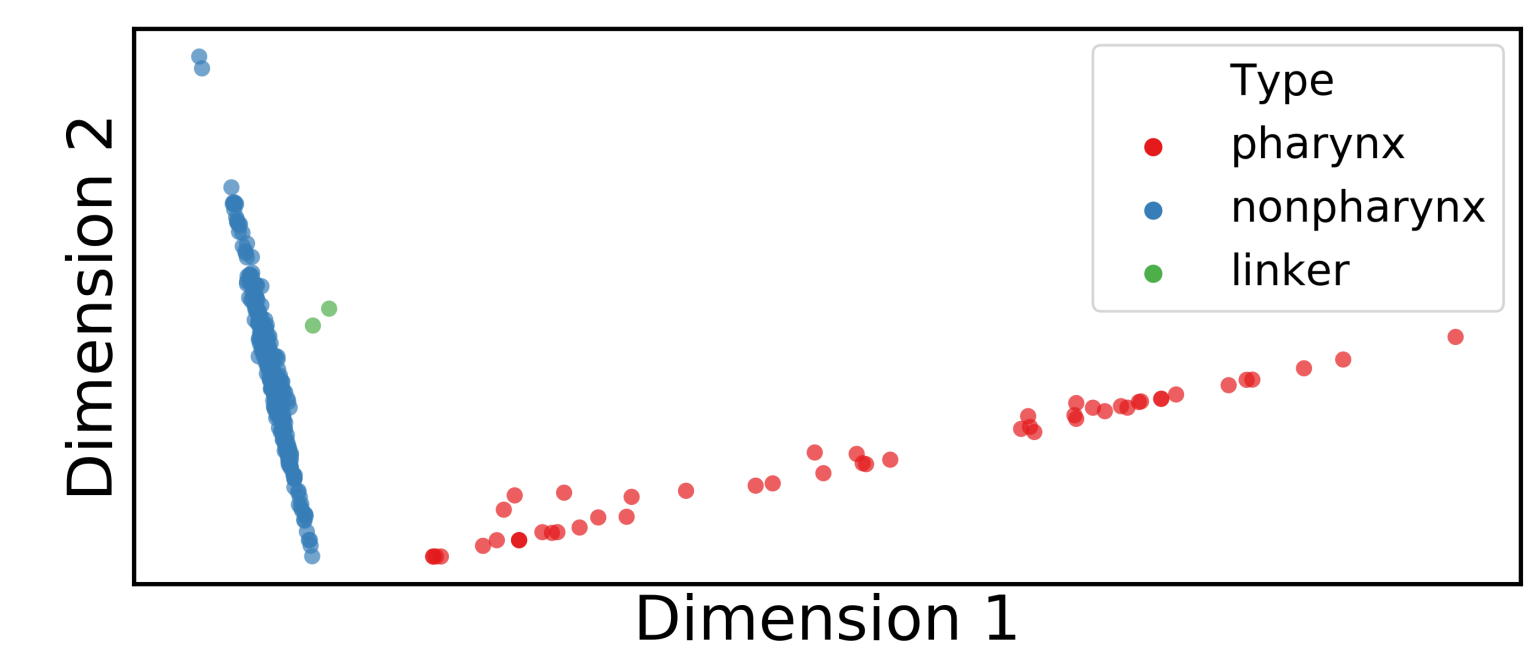


Figure 2. Laplacian spectral embedding (LSE) on the hermaphrodite *C. elegans* connectome [3], showing pharynx/nonpharynx division

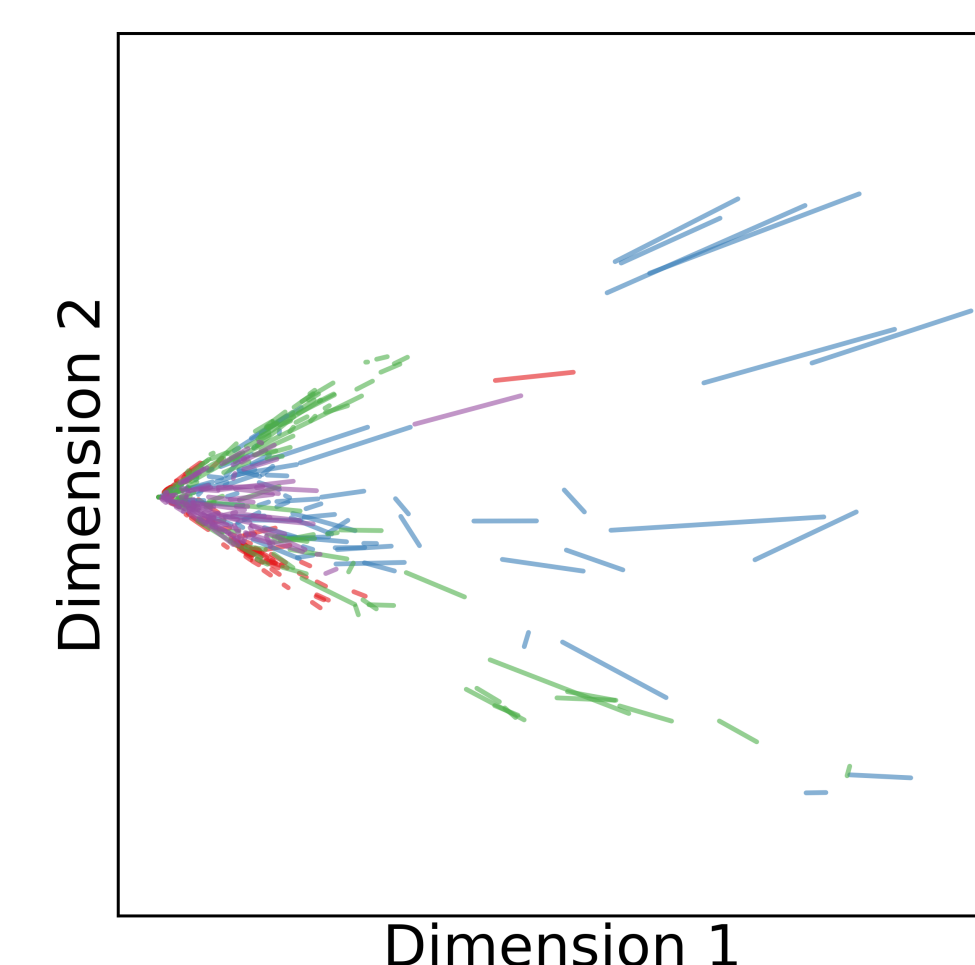


Figure 3. Omnibus embedding of the male and hermaphrodite *C. elegans* connectome, lines show disparity between sex for each node

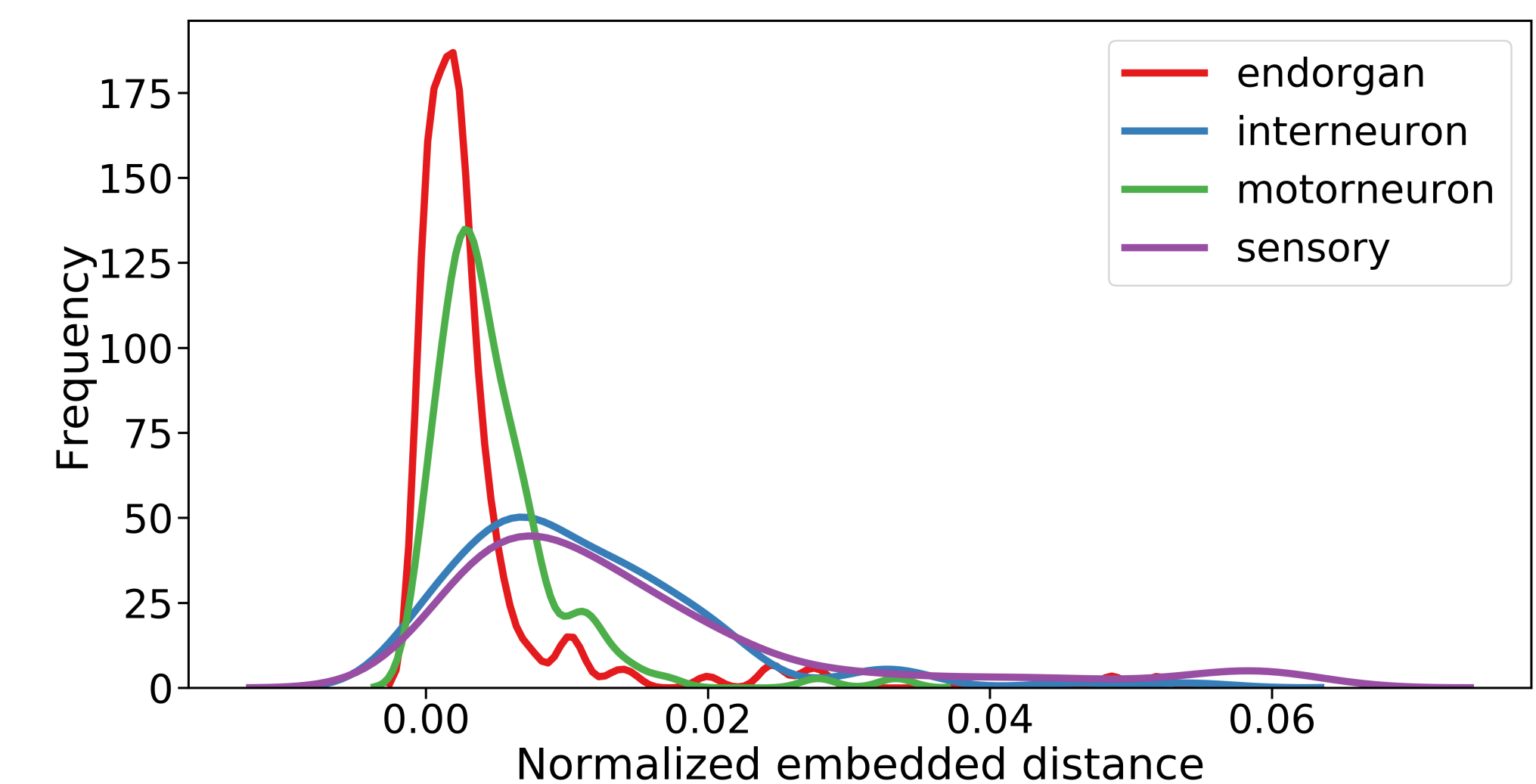


Figure 4. Kernel density estimates (KDEs) of distances in the embedding plotted by cell type, color coded as in Figure 3

Graph hypothesis testing

- How to test if two graphs (G_1 and G_2) were generated from same distribution?
 $G_1 \sim P_1, G_2 \sim P_2$
 $H_0: P_1 = P_2, H_a: P_1 \neq P_2$
- GraSPy** implements two such hypothesis tests
 - Matched test**: correspondence between node identities is known
 - Unmatched test**: correspondence between nodes is unknown or does not exist

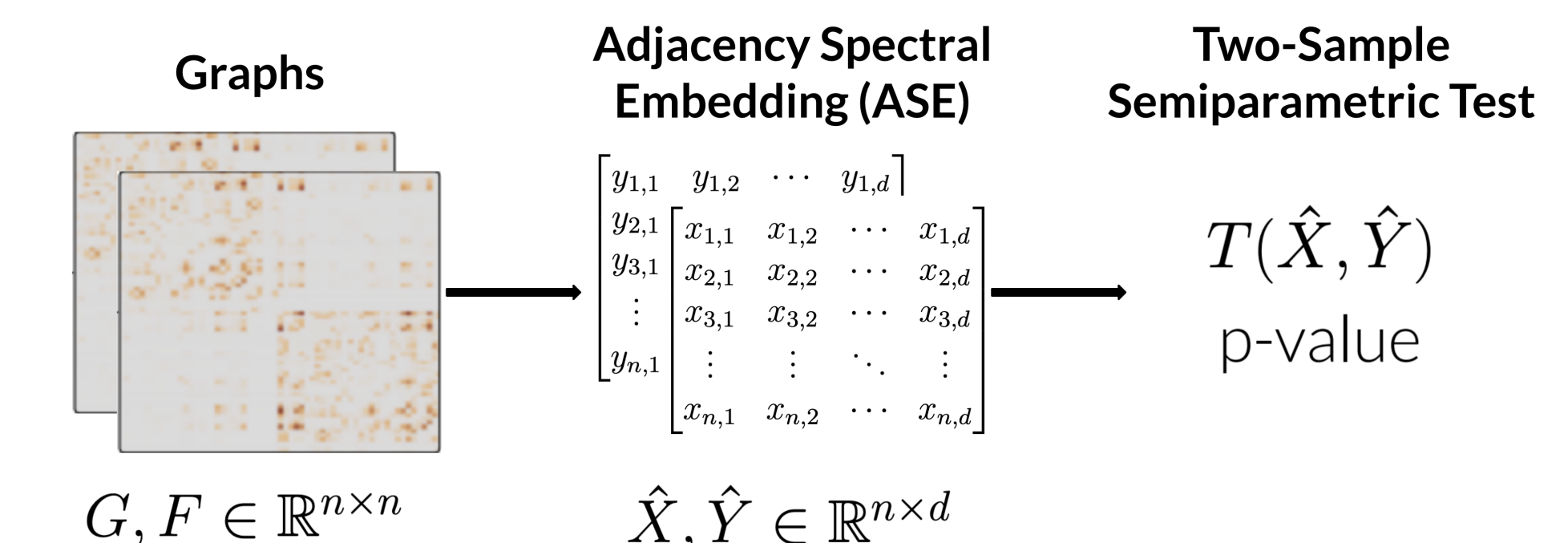


Figure 5. Procedure for the matched test of Tang et al. [5]. Two graphs are embedded using ASE, embeddings are aligned, test statistic is computed as distance between the aligned embeddings.

- Different ways to compare latent positions:
 - Exact: $H_0: X = YW$, $H_a: X \neq YW$, $W \in \mathbb{R}^{d \times d}$ and $WW^T = I$
 - Scaling: $H_0: X = cYW$, $H_a: X \neq cYW$, for some $c > 0$
 - Diagonal: $H_0: X = DYW$, $H_a: X \neq DYW$, for some diagonal $D \in \mathbb{R}^{n \times n}$

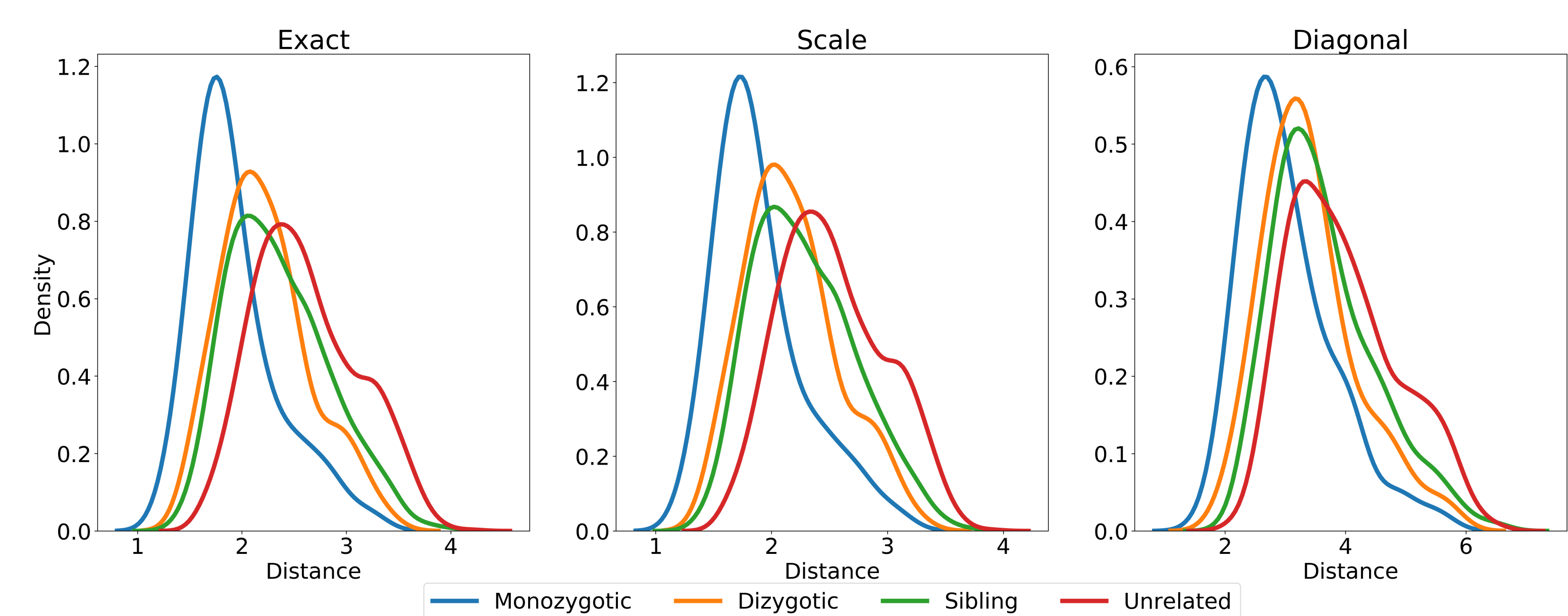


Figure 6. KDEs of matched test statistics for dMRI connectomes from Human Connectome Project Young Adult study [6], compared for different levels of relatedness

Conclusion

- GraSPy** provides tools for graph hypothesis testing and estimation
- Tools will allow neuroscientists to make claims about graph-value data
- GraSPy** will continue to grow and add functionality - in particular, we aim to consider node and edge attributes

References

- Avanti Athreya, Donnell E. Fishkind, Minh Tang, Carey E. Priebe, Youngser Park, Joshua T. Vogelstein, Keith Levin, Vince Lyzinski, Yichen Qin, and Daniel L. Sussman. Statistical inference on random dot product graphs: a survey. *Journal of Machine Learning Research*, 18(226):1-92, 2018.
- Jaewon Chung*, Benjamin D Pedigo*, Eric W Bridgeford, Bijan K Varjavand, and Joshua T Vogelstein. GraSPy: Graph statistics in python. *Nature*, 2019.
- Cook, S. J., Jarrell, T. A., Brittin, C., Wang, Y., Bloniarz, A. E., Yakovlev, M. A., Nguyen, K. C. Q., Tang, L. T.-H., Bayer, E. A., Duerri, J. S., Buelow, H., Hobert, O., Hall, D. H., and Emmons, S. W. Whole animal connectomes of both *C. elegans* sexes. *Nature*, 2019.
- Katharina Eichler, Feng Li, Ashok Litwin-Kumar, Youngser Park, Ingrid Andrade, Casey M Schneider-Mizell, Timo Saumweber, Annina Huser, Claire Eschbach, Bertram Gerber, et al. The complete connectome of a learning and memory centre in an insect brain. *Nature*, 548(7666):175, 2017.
- Minh Tang, Avanti Athreya, Daniel L. Sussman, Vince Lyzinski, Youngser Park, and Carey E. Priebe. A semiparametric two-sample hypothesis testing problem for random graphs. *Journal of Computational and Graphical Statistics*, 26(2):344-354, 2017.
- David C Van Essen, Stephen M Smith, Deanna M Barch, Timothy EJ Behrens, Essa Yacoub, Kamil Ugurbil, Wu-Minn HCP Consortium, et al. The wu-minn human connectome project: an overview. *Neuroimage*, 80:62-79, 2013.