

Integrating NetMotifs with FlyWire Connectome and Simulated Neural Dynamics

Lindsey Ma

May 18, 2025

Abstract

This project extends the NetMotifs framework to analyze structural connectomes and simulated functional dynamics, building on the foundational framework proposed by Jonas and Kording (2015) [1]. We implemented a Logistic-Distance Bayesian block model with a CRP prior to infer latent cell types and distance-dependent wiring rules from both synthetic graphs and the FlyWire connectome. Additionally, we developed a burst-based Poisson generator to simulate functional activity, enabling a joint structural-functional likelihood that sharpens cluster inference. Results demonstrate improved block recovery and reduced classification errors on synthetic data, while highlighting challenges in modeling real connectome data. Code at [Lindsey-cyber/mynetmotif](#) (currently a private repository).

1 Fitting Class-Dependent Connectivity in Bayesian Models

How good is Bayesian inference at fitting latent block structures from synthetic connectomes? To answer this, we simulate structural connectivity using class-dependent connection probabilities with and without spatial constraints by stochastic block model (see figure below).

We infer the joint posterior over cell-type assignments (c), interaction parameters (η), and global hyperparameters (θ) given the observed connectivity matrix R , where $R(e_i, e_j)$ denotes the connectivity between cells, $d(e_i, e_j)$ is their distance, and $\eta_{c_i c_j}$ governs the connection probability between cell types. This captures how connection probabilities depend jointly on cell identity and spatial proximity:

$$p(c, \eta, \theta \mid R) \propto \prod_{i,j} p(R(e_i, e_j) \mid f(d(e_i, e_j) \mid \eta_{c_i c_j}, \theta)) \prod_{m,n} p(\eta_{mn} \mid \theta) p(\theta) p(c \mid \alpha) p(\alpha)$$

(For the distance-dependent case, we use the Logistic Distance Model:

$$p^* = \frac{1}{1 + \exp\left(\frac{d(e_i, e_j) - \mu_{mn}}{\lambda_{mn}}\right)}, \quad p = p^* \cdot (p_{\max} - p_{\min}) + p_{\min}$$

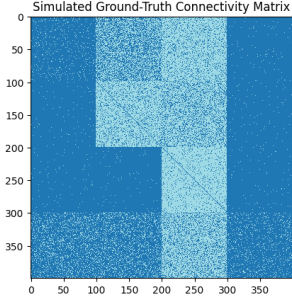


Figure 1: Ground-Truth Adjacency (Stochastic Block Model). Four distinct cell types exhibit strong intra-class connectivity (bright diagonal blocks) and weak inter-class connections (dark off-diagonal blocks), reflecting a clear modular organization.

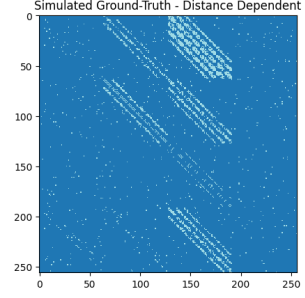


Figure 2: Ground-Truth Adjacency (Distance-Dependent Model). Connectivity depends purely on spatial proximity, producing diagonal bands instead of distinct blocks. This pattern reflects a smooth, continuous organization driven by distance constraints.

under the parameter condition $\mu_{mn} \sim \text{Exp}(\mu_{hp})$, $\lambda_{mn} \sim \text{Exp}(\lambda_{hp})$.)

We simulate structural connectivity by placing nodes in a planar grid with class labels and adding spatial jitter to their positions. For each pair of cell types, we define a distance threshold and connection probability; nodes closer than the threshold connect with higher probability. In the distance-dependent case, we set $\mu_{mn} \sim \text{Exp}(1.0)$, $\lambda_{mn} \sim \text{Exp}(1.0)$, $p_{\max} = 0.95$, and $p_{\min} = 0.01$, applying the Logistic Distance Model to compute connection probabilities based on spatial proximity.

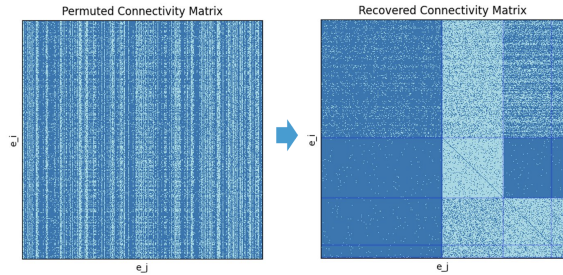


Figure 3: Recovered Connectivity Map (Beta-Bernoulli Model). The recovered matrix clearly restores the ground-truth block structure. Rows and columns align well with true classes, indicating precise recovery of class boundaries and connectivity patterns.

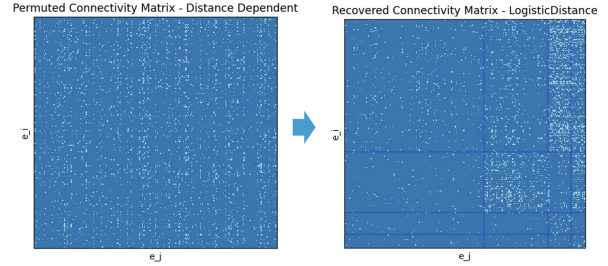


Figure 4: Recovered Connectivity Map (Logistic-Distance Model). Although modular boundaries are less distinct, meaningful clusters emerge. As the Logistic-Distance Model collapses distance dependencies, the matrix primarily visualizes whether interpretable spatial clusters are successfully inferred. This highlights the need for quantitative evaluation, as shown below.

The result matrices illustrate model recovery performance. Ideally, each row aligns with one column, indicating perfect clustering. In Figure 3, the Beta-Bernoulli model recovers four classes but introduces one extra inferred class, reflecting over-categorization. In Figure 4, the Logistic-Distance model further fragments the data, recovering six inferred classes with higher error but improved spatial coherence (baseline error is 75% under four classes).

$$\begin{bmatrix} 3 & 0 & 21 & 0 & 1 \\ 0 & 25 & 0 & 0 & 0 \\ 0 & 0 & 0 & 20 & 5 \\ 25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 5: Recovery Matrix under Beta-Bernoulli Model (Error Rate = 9%)

$$\begin{bmatrix} 5 & 0 & 8 & 73 & 14 & 0 \\ 98 & 0 & 2 & 0 & 0 & 0 \\ 0 & 95 & 0 & 0 & 0 & 5 \\ 14 & 0 & 78 & 8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 6: Recovery Matrix under Logistic-Distance Model (Error Rate = 14%)

2 Fitting Class-Dependent Models in FlyWire Connectomes

The next step is to validate our model on real connectome data using a subset of the FlyWire [2] dataset. We construct binary chemical synapse matrices by thresholding synapse counts (≥ 3 synapses define an edge). The first row of data includes neurons from the “proximal medulla”, “serpentine medulla”, and “inner lobe” regions, where no evident distance-dependent wiring patterns are observed. For this case, we apply the Beta-Bernoulli Non-Conjugate model, achieving a moderate recovery with an error rate of 19.3%.

In contrast, the second row contains neurons from the “photoreceptor”, “serpentine medulla”, and “inner lobe” regions, where photoreceptors exhibit strong spatially localized connectivity. To capture this, we apply the Logistic-Distance model with hyperparameters $\mu_{mn} \sim \text{Exp}(1.0)$, $\lambda_{mn} \sim \text{Exp}(1.0)$, $p_{\max} = 0.95$, and $p_{\min} = 0.01$. The overall error rate remains over 50%, suggesting that a simple distance-dependent model is insufficient to fully explain the connectivity patterns in these regions. This is because only the photoreceptors exhibit clear distance dependence, while the “serpentine medulla” and “inner lobe” regions do not follow this assumption. As a result, directly applying a distance-dependent model leads to systematic errors.

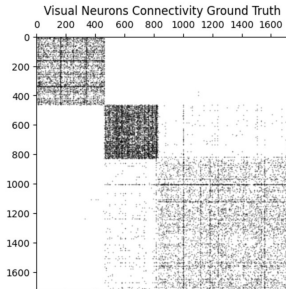


Figure 7: Connectivity among neurons from the serpentine medulla, proximal medulla, and inner lobe regions.

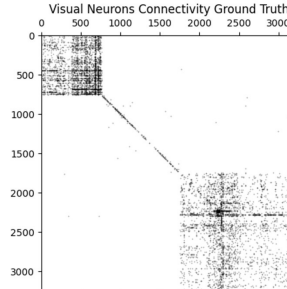


Figure 8: Connectivity among neurons from the serpentine medulla, photoreceptor, and inner lobe regions.

The training loss curve shows rapid initial improvement but quickly saturates, suggesting

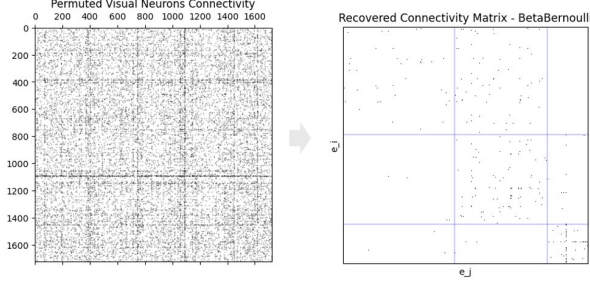


Figure 9: Recovered Connectivity Map (Beta-Bernoulli Model) reveals clear patterns of three categories, indicated by the blue lines, demonstrating effective class recovery.

$$\begin{bmatrix} 4 & 5 & 218 \\ 12 & 132 & 31 \\ 295 & 14 & 89 \end{bmatrix}$$

Figure 11: Recovery Matrix under Beta-Bernoulli Model (Error Rate = 19.3%)

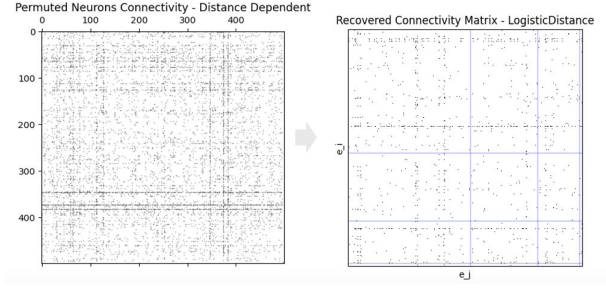
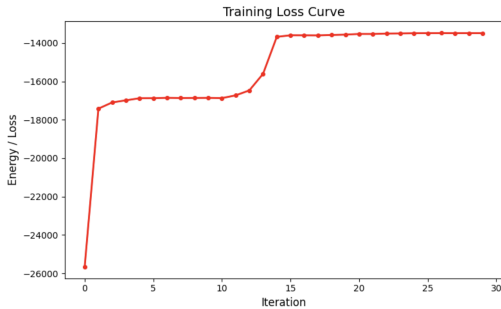


Figure 10: Recovered Connectivity Map (Logistic-Distance Model). No clear class boundaries are recovered, and the matrix lacks distinct structural patterns, reflecting poor clustering performance.

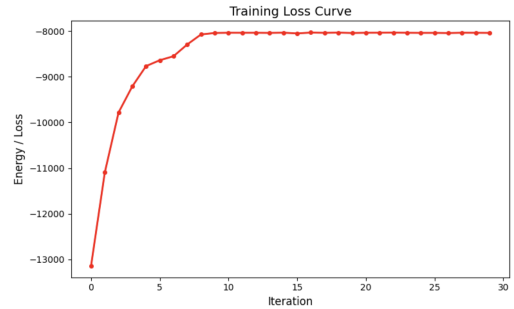
$$\begin{bmatrix} 65 & 30 & 21 & 1 \\ 82 & 51 & 32 & 3 \\ 115 & 62 & 37 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Figure 12: Recovery Matrix under Logistic-Distance Model (Error Rate = 60.8%)

the need for more expressive models or improved hyperparameter tuning. Future work will incorporate depth-aware priors and variational inference to better model these complex biological structures.



Training Loss Curve under Beta-Bernoulli Model



Training Loss Curve under Logistic-Distance Model

Figure 13: Training Loss Curves. Rapid early improvement followed by saturation indicates potential model capacity limitations or suboptimal hyperparameter settings.

3 Functional-Structural Data Simulation

The next step is to build a functional simulation model by generating spike trains using a burst-based Poisson generator. At each time step $\Delta t = 1$ ms, neurons fire independently with baseline probability $p_{\text{base}} = r_{\text{base}} \cdot \Delta t$, where $r_{\text{base}} = 0.2$ Hz. Additionally, bursts are triggered per class with probability $p_{\text{burst}} = r_{\text{burst}} \cdot \Delta t$, where $r_{\text{burst}} = 1.5$ Hz. If a burst occurs, each neuron in the class fires with probability $p_{\text{within burst}} = 0.9$. The total firing probability for neuron i is given by $P(\text{neuron } i \text{ fires}) = 1 - (1 - p_{\text{base}}) \cdot (1 - I[\text{burst for } c_i] \cdot p_{\text{within burst}})$, combining baseline and burst-driven firing.

Next, the model was extended to include functional data by introducing an additional functional loss term and set $w = 0.5$ to control its influence.

$$\mathcal{L} = \mathcal{L}_{\text{struct}}(R \mid c, \eta) + w \mathcal{L}_{\text{func}}(F \mid c)$$

where

$$\mathcal{L}_{\text{func}} = - \sum_{i < j} [F_{ij} \log p_{c_i c_j} + (1 - F_{ij}) \log (1 - p_{c_i c_j})]$$

In this formulation, F_{ij} is the binary functional adjacency matrix derived from co-firing events. The term $p_{c_i c_j}$ represents the model’s current belief (or estimated probability) that cells i and j belong to the same functional module, which can be directly computed from class assignments. Finally, w is a weighting parameter that controls the influence of functional connectivity information on the overall objective. Here we fix $w = 0.5$ (equal structural/functional weight) after coarse grid search.

Next, we simulate a total duration of 20 seconds with 1 ms resolution and compute co-firing matrices using 50 ms time bins. Functional connectivity is thresholded by requiring at least 4 co-firing events within each bin. This threshold ensures reliable detection of synchronous activity while avoiding spurious connections. These simulated functional adjacency matrices are then integrated with structural connectivity to jointly infer latent blocks.

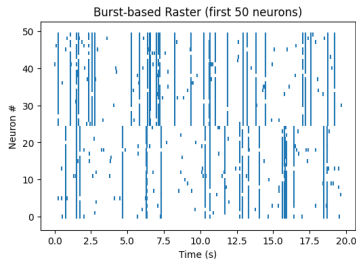


Figure 14: Example raster plot of simulated burst-driven activity for 50 neurons. Distinct periods of synchronized firing within cell types highlight functional assembly structures.

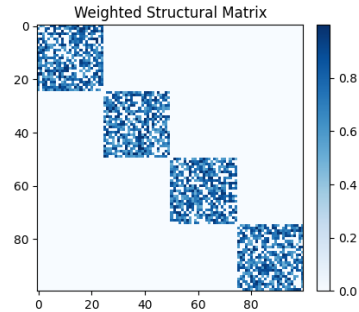


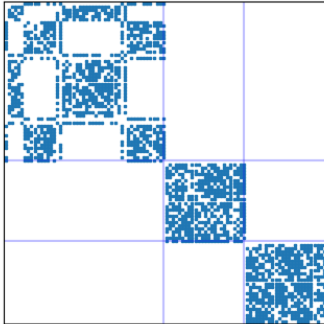
Figure 15: Simulated structural adjacency matrix with weighted connections. High intra-class connectivity and minimal inter-class links demonstrate clear modular organization.

Here, the simulated spike raster (Figure 14) shows within-type synchrony during bursts,

replicating structured co-firing patterns typical of functional assemblies. The co-firing matrix further confirms strong within-type correlations and sparse between-type interactions.

Integrating these functional observations with the structural connectivity (Figure 15), the joint inference sharpens cluster assignments, effectively reducing uncertainty in block structure recovery.

Recovered Connectivity (Beta-Bernoulli IRM)



$$\begin{bmatrix} 0 & 25 & 0 & 0 \\ 25 & 0 & 0 & 0 \\ 0 & 0 & 25 & 0 \\ 25 & 0 & 0 & 0 \end{bmatrix}$$

Recovery Matrix under New Model (Error Rate = 25%). Partial class separation achieved, indicating underfitting due to unresolved group distinctions.

Figure 16: Recovered Connectivity Map and Recovered Class Connectivity Matrix.

In Figure 16, the model successfully identifies clear boundaries for three classes, demonstrating strong recovery of dominant structures. While two classes remain merged, this suggests that enhancing the model’s sensitivity to finer distinctions may further improve class separation. Incorporating richer priors or more expressive models may further improve its ability to more completely resolve latent class structures.

4 Future Goals

In future work, we aim to extend our analysis to real structural and functional datasets, specifically leveraging the MICrONS v3 dataset, which includes approximately 150,000 synapses and 75 minutes of two-photon calcium imaging. This will enable us to replicate and validate the joint inference framework under biologically realistic conditions. Additionally, we plan to incorporate depth-aware connectivity priors by extending the logistic likelihood model with per-type Gaussian depth priors, following the approach outlined in Equation (5) of the referenced paper. To improve scalability for large-scale graphs with over 10,000 nodes, we also intend to accelerate inference by porting the current slice-sampling kernels to a stochastic variational inference (SVI) framework, as proposed by Hoffman *et al.* (2013)[3].

References

- [1] Jonas, E., & Kording, K. P. (2015). Automatic Discovery of Cell Types and Microcircuitry from Neural Connectomics. *eLife*, 4:e04250.
- [2] Dorkenwald, S., Turner, N. L., Macrina, T., Lee, K., Lu, R., Wu, J., Bodor, A. L., Bleckert, A., Brittain, D., Kemnitz, N., *et al.* (2022). FlyWire: Online Community for Whole-Brain Connectomics. *Nature Methods*, 19(1), 119–128.

- [3] Hoffman, M. D., Blei, D. M., Wang, C., & Paisley, J. (2013). Stochastic Variational Inference. *Journal of Machine Learning Research*, 14(1), 1303–1347.