Neuronal Classification from Network Connectivity

*Rebecca F. Goldin[1,2], Paul M. Salomonsky[3], Carey E. Priebe[4], David J. Marchette[3], Giorgio A. Ascoli[1]**

[1]Center for Neural Informatics, Structure, & Plasticity; and Molecular Neuroscience Dept; Krasnow Institute for Advanced Study; George Mason Univ. Fairfax, VA (USA) *ascoli@gmu.edu [2]Mathematical Sciences, George Mason Univ., Fairfax, VA (USA) [3]Naval Surface Warfare Center, Code Q33, Dahlgren, VA (USA) [4]Applied Mathematics and Statistics, Johns Hopkins Univ. Baltimore, MD (USA)

Introduction: Identifying Neuronal Classes

The mammalian brain is a large network of neurons (~10^8 in rodents, up to 10^{11} in humans), sparsely interconnected by synapses (~10^4 per neuron). Most synapses are directional contacts from the axon of the sending neuron to the dendrite of the receiving neuron. Although no two neurons have the exact same synaptic connectivity, clear similarity patterns support the fundamental working assumption of identifiable neuronal classes with distinct connectivity and function.

Figure 1. Model of data and classification

Given a large network of neurons with their connections ("data"), we use statistical techniques such as cluster analysis and a Bayesian Information Criterion[1] (BIC) function to identify groups of neurons (estimated neuronal classes) with similar connectivity. To check the robustness of this technique, we simulate a network with a biologically plausible set of neuronal classes and connections. We then measure how closely the estimated classes mimic the original "true" neuronal classes.

Simulation Approach

Mathematical Terminology

Directed Graph: A collection of ***Vertices*** (neurons) and ***Directed Edges*** (connections from anaxon of one neuron to a dendrite of another).
Classes of Vertices: Groups of neurons that have similar connectivity properties. Each neuron is in exactly one group.
True Classes: Classes of vertices pre-defined in the simulation.
Estimated Classes: Classes of vertices obtained by the model.
Probability Matrix: A table of probabilities; the entry in the a^{th} row and b^{th} column is the probability that a neuron in class *a* connects with one in class *b*.
Adjacency Matrix: An *n* x *n* matrix (*n* large) of 0s and 1s indicating the absence or presence of a connection, respectively, from one neuron to another.

We assume k=8 classes of neurons and n=32,768 neurons divided into classes specified by color.

	15768 CA1 Pyramidal Cells: Principal output neurons of the hippocampus. Excitatory. Constitute one of the most studied and best characterized neuron types in the brain.
	4000 CA1 Oriens/Lacunosum-Moleculare Cells: local inhibitory neurons. Dendrites are in the oriens layer and their axons start in the oriens and go up to Lacunosum-Moleculare .
	1000 CA1 Basket Cells: local peri-somatic inhibitory interneurons. Axons target pyramidal and basket cells. Their dendrites span all layers of CA1.
	3000 CA1 Perforant Pathway-Associated Cells: local inhibitory interneurons with dendrites and axons confined to the Lacunosum-Moleculare layer.
	2000 CA1 Oriens Cells: Local inhibitory interneurons. Dendrites and axons confined to the oriens layer.
	2500 Entorhinal Cortex Layer 5 Pyramidal Cells: play the role of deep layer 'input' neurons. They are excitatory and have dendrites and axons through the deep and superficial layers of the entorhinal cortex.
	2500 Entorhinal Cortex Layer 3 Pyramidal Cells: One of the superficial excitatory layer 'output' neurons. Dendrites through the deep and superficial layers of the EC. Axons starting in layer 3,projecting to CA1LM.
	2000 Entorhinal Cortex GABAergic Cells: Inhibitory local interneurons of the EC, with axons and dendrites through the deep and superficial layers of the entorhinal cortex.

Figure 2. Numbers of neurons, names, and brief descriptions of the neuronal classes used to simulate a biologically plausible network inspired by a a simplified and scaled-down mouse hippocampus

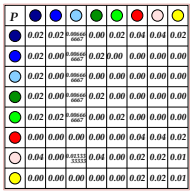


Figure 3. Probability matrix *P*

Cluster analysis using BIC

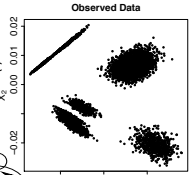


Figure 7. Graph of first coordinates of data

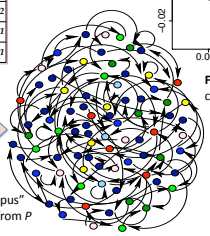
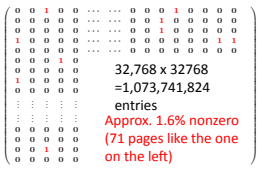


Figure 4. A "hippocampus" with classes sampled from *P*

Forget the classes

SVD top dimension

- ≈1 page: number of neurons in a mouse hippocampus slice
- ≈10 pages: number of neurons in a mouse whole hippocampus
- ≈1,000 pages: number of neurons in a mouse brain
- ≈100,000 pages: number of neurons in a human brain
- ≈100,000 pages: number of connections in a mouse hippocampus
- ≈10,000,000 pages: number of connections in a mouse brain
- ≈1,000,000,000 (1 billion) pages: number of connections in a human brain



32,768 x 32768 =1,073,741,824 entries

Approx. 1.6% nonzero (71 pages like the one on the left)

Figure 6. Adjacency matrix (the pixels on the left represent only the 1s, not the 0s)

Figure 5. Squint! Can you see 250,000 pixels? (lift for explanation)

Methods and Experimental Design

For each *n* x *n* adjacency matrix *A* and each dimension *d*=1,2,..., 10, we use Singular Value Decomposition (SVD) to obtain an approximation of *A*:

$$\begin{pmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ u_{21} & u_{22} & u_{23} & u_{24} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ u_{n-1,1} & u_{n-1,2} & u_{n-1,3} & u_{n-1,4} \\ u_{n1} & u_{n2} & u_{n3} & u_{n4} \end{pmatrix} \begin{pmatrix} s_1 & 0 & 0 & 0 \\ 0 & s_2 & 0 & 0 \\ 0 & 0 & s_3 & 0 \\ 0 & 0 & 0 & s_4 \end{pmatrix} \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1,n-1} & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2,n-1} & v_{2n} \\ v_{31} & v_{32} & \cdots & v_{3,n-1} & v_{3n} \\ v_{41} & v_{42} & \cdots & v_{4,n-1} & v_{4n} \end{pmatrix}$$

d=4 *d*<<*n*

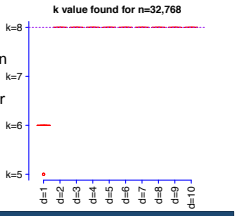
Color-consistent columns and rows correspond to singular values, and multi-colored columns and rows correspond to individual neurons/vertices. For each vertex, we have *d* parameters, *d*<<*n*, from which we can estimate the connectivity of the graph. The row $(u_{i1}\ u_{i2}\ u_{i3}\ u_{i4})$ records the vectors connecting to the i^{th} neuron while the column $(v_{1i}\ v_{2i}\ v_{3i}\ v_{4i})$ records those vertices to which the i^{th} neuron connects.
We apply the *kmeans* clustering technique[2] to these *n* rows and *n* columns (*d*-vectors) to find clusters of vertices; these represent estimated classes of neurons with similar connectivity statistics. The estimated number of classes is determined using BIC. We estimate *P* by assigning each entry[3]

$$p_{ab} = \overline{u}_{a1}s_1\overline{v}_{1b} + \overline{u}_{a2}s_2\overline{v}_{2b} + \overline{u}_{a3}s_3\overline{v}_{3b} + \overline{u}_{a4}s_4\overline{v}_{4b}$$

These probabilities estimate the original probabilities; error obtained by comparing this matrix to the true probabilities are represented in **Figure 10.** We correct the clustering by 10 percent of the minimal distance among rows and columns in *P*.

Results: Estimating Class Assignments

We ran 50 simulations (adjacency matrices) with different random seeds for 32,678 vertices; this results in 500 experiments as we change the SVD analysis (*d*=1,...,10.)
We analyzed the data for different values of *d*. After cluster analysis, the correct number of classes was found 100 percent of the cases for *d*>1. The probabilities differed from true probability (in Euclidean norm, square root of sum of squares of differences in matrices) on the order of 10^{-7} for *d*>2. Classes were assigned correctly 100 percent of the time for *d*>1. For smaller *n* and/or for *d*=1, class assignment was not perfect.



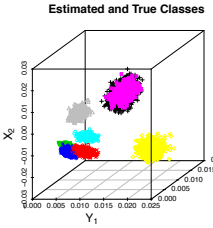


Figure 9. 8 true classes (by color) and 8 estimated class (by shape)

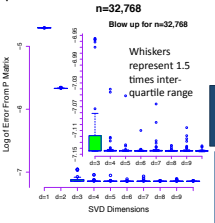


Figure 10. Error in probability table found compared to true probability

Conclusions

- Large quantities of connectivity data can be analyzed to obtain a meaningful interpretation of neuronal class. Suitable data include dense electron microscopy reconstructions (identified synapses) and light microscopy (potential synapses).
- This work may direct data collection methods, indicating whether experimental data from a large number of preparations, each with small samples, or a large sample of from fewer preparation would yield better statistical results.
- Determination of neuronal classes has the prospect of becoming rigorous, verifiable and reproducible.

Future Directions

- Using model selection techniques we may find other methods more effective than *kmeans* in finding classes.
- Robustness tests need to be implemented.
- Spatial position (or other properties) of the neurons could be considered to affect connection probabilities.
- In the simulations, connections are formed by sampling a binomial distribution for each probability (Erdös-Renyí graphs). Other distributions may be more appropriate to describe brain networks.
- Dependence among connections could be considered.

Select References

[1]Schwarz, G. Estimating the Dimension of a Model, *Annals of Statistics,* 6:461-464, 1978.
[2]Lloyd., S. P. Least Squares Quantization in PCM. *IEEE Trans. on Information Theory* 28(2):129-137, 1982.
[3]Scheinerman E. and Tucker, K. Modeling graphs using dot product representations. *Computational Statistics,* 25:1–16, 2010.

Support: NIH R01 NS39600 & ONR MURI 141010198 to GAA Partial Support, Office of Naval Research In-House Laboratory Independent Research Program, to DM