

EN.580.694
Statistical Connectomics
Final Project Proposal

William Hockeimer

April 2, 2015

1 Opportunity

The mammalian hippocampus is a subcortical structure heavily implicated in processing both spatial and episodic information. The Cornu Ammonis 3 (CA3) region of the hippocampus is hypothesized to serve as a content-addressible memory system, crucial to the overall processing algorithm of the hippocampus, thanks to its extensive recurrent connections. In previous theoretical studies, CA3 has been conceived as a single, homogeneous network. Under this hypothesis, information is stored as patterns of neural activity. Given an incomplete portion of the input associated with a given pattern, CA3 is thought to retrieve the full, stored pattern. This process, known as pattern completion, is thought to rely on the recurrent connectivity of CA3. However, recent genetic evidence shows the existence of at least nine distinct genetic expression regions of CA3. This calls into questions whether or not the connectivity of CA3 is truly homogeneous and what effects, if any, this would have on the pattern completion abilities of CA3.

2 Challenge

The neuroscience community faces a challenge concerning how to best integrate this new genetic information and assess what effect, if any, this higher- resolution picture of CA3 affects the predictions made by modeling studies concerning content-addressible memory.

3 Action

We will develop an Erdos-Reyni mixture model to explore how differential genetic expression profiles across CA3 affect its connectivity and, therefore, its computational capabilities. Each component of the Erdos-Reyni mixture model, denoted as ER_i will correspond to one of nine empirically determined expression

regions of CA3. Each block of the Erdos-Reyni model will be associated with a certain node attribute, a K-bit bitstring. Each bit of the string corresponds to a given gene and the value whether it's expressed (1) or not (0). Empirical studies will be used to determine 1) what subregion corresponds to what bitstring and 2) how the bitstrings interact, i.e. how the distribution of bitstring values affects the probability of connection between two mixtures. In short, a rulebook will be defined based on empirically known associations between cues and receptors and their effects on connectivity. For instance, if one bit corresponds to Netrin and another corresponds to Unc then the rule for each node is that a node expressing each (i.e. has a "1" in the column associated with each) will have a higher likelihood of connectivity. The magnitude of this likelihood is a parameter that can be experimentally tuned. Therefore there will exist two Erdos-Reyni models: 1) under the null hypothesis the model (ER_1) will correspond to the classical, fully-recurrent model of CA3 familiar to the computational literature 2) an Erdos-Reyni mixture model (ER_9) with one component corresponding to each of the 9 genetically distinct subregions of CA3. Each network will be used to define a 1-layer neural network which will subsequently be trained on a pattern completion task. The performance of each network can be assessed along multiple dimensions, including 1) maximum number of patterns that can be stored 2) maximum overlap between stored patterns 3) speed at pattern completion 4) minimum size of input pattern needed to retrieve whole pattern. The performance of each network on some given set of pattern completion tasks, for example some of those just mentioned, will be the output used to judge how changing the connectivity

4 Resolution

We will run each network - the ER_1 and ER_9 network on a pattern completion and assess how each network performs. This will help us understand how the connectivity within a recurrent network affects its computational abilities and, specifically, how geneic subdomains of CA3 affect its ability to perform the pattern completion tasks so commonly associated with it.

5 Future

Future experiments will determine 1) the parameters of the pattern needed to achieve successful completion, 2) the degree of reconnection between sub-regions of CA3 that still allows a basal level of pattern completion to take place.

6 Statistic Decision Theoretic

6.1 Sample Space

$G(V, E, Y)$ where $V = 300,000$ (empirically determined number of CA3) neurons
 $E, Y \in B^{n \times 1}$ where B is distribution of bit strings

6.2 Model

$SBM_9^N(\rho, \beta) : \rho = \delta_9, \beta \in (0, 1)^{(9 \times 9)}$

6.3 Action Space

$A_0, 1^{(n \times 1)}$ where n is number of neurons, i.e., 300,000. In English, this is the output pattern of the pattern completion task.

6.4 Decision Rule Class

$f - \mathcal{L} G_p x A_p$ where subscript p indicates the graph activation and output pattern associated with input pattern p . f indicates whether G is a successful representation of A , i.e. whether the pattern successfully retrieves p .

6.5 Loss Function

The MSE of the retrieved pattern from the null hypothesis network (ER_1) and the test network (Er_9) is computed across n patterns.

$$l = \text{MSE} = \frac{1}{n} * \sum_{k=1}^n (y_i - \hat{y})^2$$

6.6 Risk Function

$$R = E[L]$$