The Gordon and Betty Moore Foundation
Data-Driven Discovery Symposium

Data-Driven Discovery for the Cortical Column Conjecture

Carey E. Priebe
http://www.ams.jhu.edu/~priebe

Department of Applied Mathematics & Statistics
Center for Imaging Science
Department of Computer Science
Department of Electrical and Computer Engineering
Johns Hopkins University
Baltimore, MD, USA

July 28-29, 2014
Palo Alto, California

Outline

Introduction

Motivation

Many contemporary theories of neural information processing suggest that the neocortex employs algorithms composed of repeated instances of a limited set of computing primitives. There is a recognized need for tools for interrogating the structure of the cortical microcircuits believed to embody these primitives.

Cortical Column Conjecture

Neurons are connected in a graph that exhibits motifs representing repeated *processing modules.*

J. C. Horton and D. L. Adams. "The cortical column: a structure without a function," *Philosophical Transactions of the Royal Society B*, 360(1456):837-862, Apr. 2005.

V. Mountcastle. "The columnar organization of the neocortex," *Brain*, 120(4):701-722, Apr. 1997.

Introduction

Fundamental Question

What are the relevant cortical computing structures associated with recurring circuit motifs?

Goals

1. To obtain, and to understand the error characteristics of, an observed cortical graph.
2. To develop, and to understand the theoretical & practical properties of, general methodologies for end-to-end cortical graph inference.

We aim to advance theoretical neuroscience, and impact algorithms for machine intelligence, through insights derived from high-fidelity reconstructions of cortical microcircuits; in particular, we will estimate relevant cortical computing parameters from observed cortical graphs.

Notable Findings I

Science, April 25, 2014

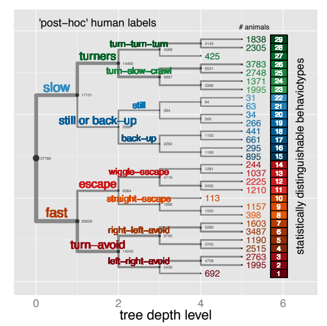
Discovery of Brainwide Neural-Behavioral Maps via Multiscale Unsupervised Structure Learning

Joshua T. Vogelstein,[1,2]* Youngser Park,[1]* Tomoko Ohyama,[3]* Rex A. Kerr,[3] James W. Truman,[3] Carey E. Priebe,[1]†‡ Marta Zlatic[3]†‡

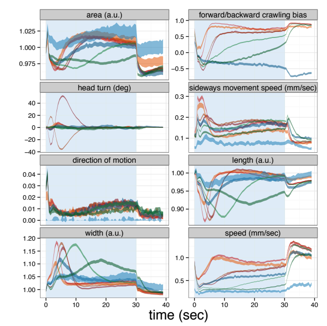
A single nervous system can generate many distinct motor patterns. Identifying which neurons and circuits control which behaviors has been a laborious piecemeal process, usually for one observer-defined behavior at a time. We present a fundamentally different approach to neuron-behavior mapping. We optogenetically activated 1054 identified neuron lines in *Drosophila* larvae and tracked the behavioral responses from 37,780 animals. Application of multiscale unsupervised structure learning methods to the behavioral data enabled us to identify 29 discrete, statistically distinguishable, observer-unbiased behavioral phenotypes. Mapping the neural lines to the behavior(s) they evoke provides a behavioral reference atlas for neuron subsets covering a large fraction of larval neurons. This atlas is a starting point for connectivity- and activity-mapping studies to further investigate the mechanisms by which neurons mediate diverse behaviors.

Timothy O'Leary and Eve Marder, "Mapping Neural Activation onto Behavior in an Entire Animal," *Science*, April 25, 2014.
"(Authors) usher in a new era of integrated methods for deciphering how an entire nervous system generates behavior... (They) have thus achieved a technical, multidisciplinary tour de force that will provide a rich source of research questions."

Notable Findings I



(f) Cluster Tree



(e) Cluster Response

Figure 2. (f) Post-hoc human labels assigned to the automatically detected behaviotype families and subfamilies. (e) Mean and standard error of the responses of the eight behaviotype subfamilies.

Vogelstein et al, "Discovery of Brainwide Neural-Behavioral Maps via Multiscale Unsupervised Structure Learning," *Science*, vol. 344 no. 6182, 386-392, April 25, 2014.

Notable Findings I

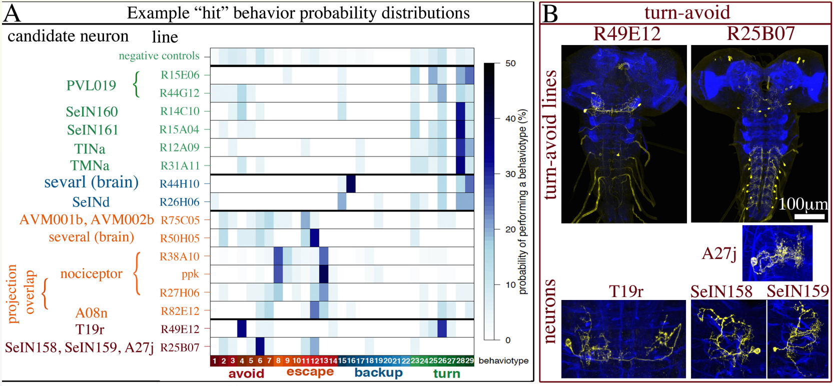


Figure 4. Examples of significant neuron lines and candidate neurons involved in distinct behaviotypes.

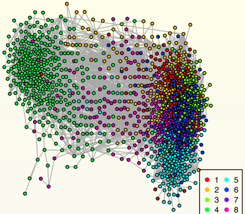
Vogelstein et al, "Discovery of Brainwide Neural-Behavioral Maps via Multiscale Unsupervised Structure Learning," *Science*, vol. 344 no. 6182, 386-392, April 25, 2014.

Notable Findings II

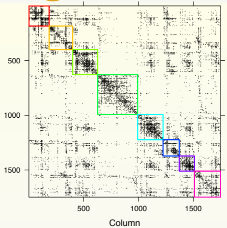
Adjacency Spectral Graph Embedding & Subsequent Inference

The eigendecomposition of an adjacency matrix provides a way to embed a graph as points in finite dimensional Euclidean space. This embedding allows the full arsenal of statistical and machine learning methodology for multivariate Euclidean data to be deployed for graph inference. Our work analyzes this embedding in the context of various random graph models with a focus on the impact for subsequent inference. In summary, this body of work demonstrates that for a broad class of graph models and inference tasks, adjacency spectral embedding allows for accurate graph inference via standard multivariate methodology.

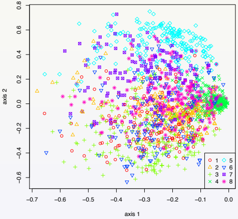
Notable Findings II



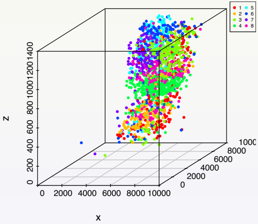
(a) graph



(b) adjacency matrix



(c) spectral embedding



(d) spatial embedding

Takemura et al, "A visual motion detection circuit suggested by Drosophila connectomics," *Nature*, 500, 175-181, August 8, 2013.

Notable Findings II

- D.L. Sussman, M. Tang, D.E. Fishkind, and C.E. Priebe, "A consistent adjacency spectral embedding for stochastic blockmodel graphs," *Journal of the American Statistical Association*, Vol. 107, No. 499, pp. 1119-1128, 2012.
- D.E. Fishkind, D.L. Sussman, M. Tang, J.T. Vogelstein, and C.E. Priebe, "Consistent adjacency-spectral partitioning for the stochastic block model when the model parameters are unknown," *SIAM JMAA*, Vol. 34, No. 1, pp. 23-39, 2013.
- V. Lyzinski, D.L. Sussman, M. Tang, A. Athreya, and C.E. Priebe, "Perfect Clustering for Stochastic Blockmodel Graphs via Adjacency Spectral Embedding," submitted for publication, 2013.
- D.L. Sussman, M. Tang, and C.E. Priebe, "Consistent latent position estimation and vertex classification for random dot product graphs," *IEEE TPAMI*, Vol. 36, No. 1, pp. 48-57, 2014.
- M. Tang, D.L. Sussman, and C.E. Priebe, "Universally consistent vertex classification for latent positions graphs," *Annals of Statistics*, Vol. 41, No. 3, pp. 1406-1430, 2013.
- M. Tang, Y. Park, and C.E. Priebe, "Out-of-sample extension for latent position graphs," submitted for publication, 2013.
- A. Athreya, V. Lyzinski, D.J. Marchette, C.E. Priebe, D.L. Sussman, and M. Tang, "A limit theorem for scaled eigenvectors of random dot product graphs," submitted for publication, 2013.
- S. Suwan, D.S. Lee, R. Tang, D.L. Sussman, M. Tang, and C.E. Priebe, "Empirical Bayes Estimation for the Stochastic Blockmodel," submitted for publication, 2014.
- M. Tang, A. Athreya, D.L. Sussman, V. Lyzinski, and C.E. Priebe, "Two-sample Hypothesis Testing for Random Dot Product Graphs via Adjacency Spectral Embedding," submitted for publication, 2014.

Cortical Column Conjecture

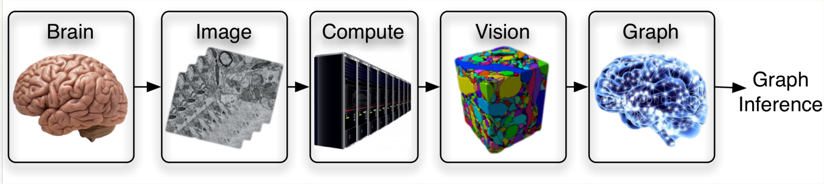
Cortical Column Conjecture

The cortical column conjecture suggests that neurons in the neocortex are connected in a graph that exhibits motifs representing repeated processing modules.

Our focus is on *extracting*, and then *estimating* the structure of, the cortical graph, for the purpose of subsequent modeling and algorithm development.

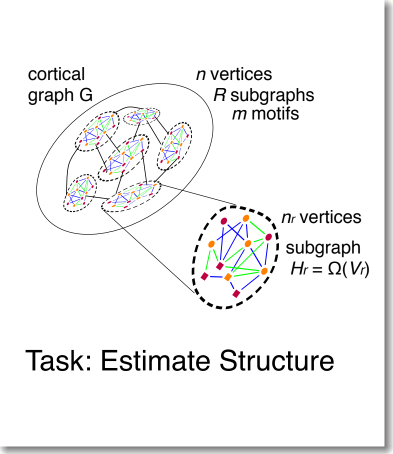
Vision & Plans

End-to-End Pipeline
for estimating cortical graph structure...



... for the purpose of
subsequent modeling and algorithm development

Cortical Graph



- Cortical graph G: hierarchical block model on n vertices (neurons).
- Induced subgraphs H_r: stochastic block models on n_r vertices, $r = 1, \ldots, R$.

1. Identify large-scale structures in $G \to H_r$.
2. Identify clusters of the $\hat{H}_r$'s corresponding to repeated motifs.
3. Estimate structural parameters.

Vision & Plans

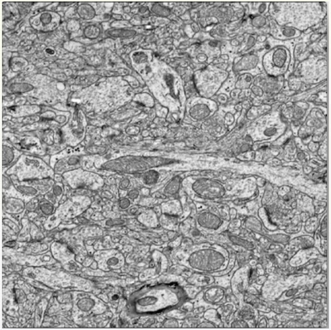
The Big Five Graph Inference Challenges

1. Clustering the vertices in a graph
2. Clustering a collection of graphs
3. Graph matching
4. Testing and estimation
5. Robustness to errorfully observed graphs

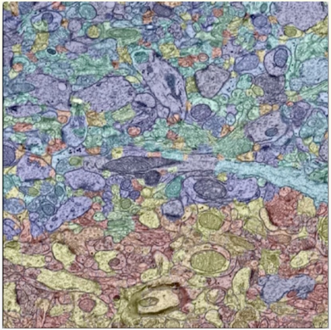
All of these challenges must be addressed in the context of available vertex- and edge-attributes – we will have neuron types (excitatory, inhibitory), edge weights (synapse strengths), spatial information, etc. The development of optimal robust methodologies for attributed graphs presents significant additional challenges.

From Images to Graphs to Inference

"... a goal of the computer vision is to reconstruct a graph with error properties that the subsequent inference can tolerate ..."



(a) EM data
(Davi Bock, HHMI Janelia)



(b) Segmentation
(Hanspeter Pfister, Harvard)

C.E. Priebe, D.L. Sussman, M. Tang, J.T. Vogelstein, "Statistical inference on errorfully observed graphs," *Journal of Computational and Graphical Statistics*, accepted for publication, 2014.

Stochastic Block Model

Let $\mathcal{B}_K$ be the collection of symmetric $K \times K$ block-probability matrices;

$$\mathcal{B}_K := \{B \in [0,1]^{K \times K} | B^T = B\}.$$

Let $\mathcal{M}_{n,K}$ be the collection of length K non-negative integer-valued vectors $\vec{n}$ with entries summing to n;

$$\mathcal{M}_{n,K} := \{\vec{n} \in \mathbb{N}^K | \sum_{k=1}^{K} \vec{n}_k = n\}.$$

Given positive integers n and K, consider $\vec{n} \in \mathcal{M}_{n,K}$ and $B \in \mathcal{B}_K$. We say a random graph G on vertex set V is a stochastic block model graph $G \sim SBM(V, \vec{n}, B)$ if the vertices V are partitioned into subsets $V_1, \cdots, V_K$ with sizes given by $\vec{n} = (\vec{n}_1, \cdots, \vec{n}_K)$ and edges $\mathbb{1}\{u \sim v\}$ are independent Bernoulli random variables with $P[u \sim v] = B_{ij}$ for $u \in V_i$ and $v \in V_j$.

Hierarchical Stochastic Block Model

Given $G = (V, E)$, where $V = [n]$.

- $\{V_r\}_{r=1}^R$: a disjoint partition of the vertices V, and for each $r = 1, 2, \ldots, R$, write $n_r := |V_r|$.
- $H_r = \Omega(V_r) = (V_r, E_r)$.
- $H_r \sim SBM(V_r, \vec{m}_r, B_r)$, where $\vec{m}_r \in \mathcal{M}_{n_r,K}$ and $B_r \in \mathcal{B}_K$.
- for $u \in V_r$ and $v \in V_{r'}$ with $r \neq r'$, $\mathbb{1}\{u \sim v\} \sim Bernoulli(p)$.

$G \sim HSBM(V, \vec{m}, B)$, where $\vec{m} = [\vec{m}_1^\top \mid \vec{m}_2^\top \mid \cdots \mid \vec{m}_R^\top]^\top$ and B is given by

$$B = \begin{bmatrix} B_1 & pJ_{K,K} & \cdots & pJ_{K,K} \\ pJ_{K,K} & B_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & pJ_{K,K} \\ pJ_{K,K} & \cdots & pJ_{K,K} & B_R \end{bmatrix}.$$

Biologically Motivated Model

Our cortical graph G:

- n vertices, with $n \in \{10^3, 10^4, 10^5, 10^6\}$.
- $V = [n]$ is partitioned into R subsets $\{V_r\}_{r=1}^R$.
- $H_r = \Omega(V_r)$ each have $|V_r| = n_r = m = 100$, $K = 5$ blocks.
- $R = n/m$.
- $\vec{m}_r = [2, 50, 15, 8, 25]^\top \ \forall r$.
- $$B_r = \begin{bmatrix} 0.1 & 0.045 & 0.015 & 0.19 & 0 \\ 0.045 & 0.05 & 0.035 & 0.14 & 0.03 \\ 0.015 & 0.035 & 0.08 & 0.105 & 0.04 \\ 0.19 & 0.14 & 0.105 & 0.29 & 0.13 \\ 0 & 0.03 & 0.04 & 0.13 & 0.09 \end{bmatrix}.$$
- $p \in \{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$.

Izhikevich E. M. and Edelman G. M. (2008) "Large-Scale Model of Mammalian Thalamocortical Systems," *PNAS*, 105:3593-3598

Biologically Motivated Model

Large-scale model of mammalian thalamocortical systems

Eugene M. Izhikevich and Gerald M. Edelman*

The Neurosciences Institute, 10640 John Jay Hopkins Drive, San Diego, CA 92121

Contributed by Gerald M. Edelman, December 27, 2007 (sent for review December 21, 2007)

The understanding of the structural and dynamic complexity of mammalian brains is greatly facilitated by computer simulations. We present here a detailed large-scale thalamocortical model based on experimental measures in several mammalian species. The model spans three anatomical scales. (*i*) It is based on global (white-matter) thalamocortical anatomy obtained by means of diffusion tensor imaging (DTI) of a human brain. (*ii*) It includes multiple thalamic nuclei and six-layered cortical microcircuitry based on *in vitro* labeling and three-dimensional reconstruction of single neurons of cat visual cortex. (*iii*) It has 22 basic types of neurons with appropriate laminar distribution of their branching dendritic trees. The model simulates one million multicompartmental spiking neurons calibrated to reproduce known types of responses recorded *in vitro* in rats. It has almost half a billion synapses with appropriate receptor kinetics, short-term plasticity, and long-term dendritic spike-timing-dependent synaptic plasticity (dendritic STDP). The model exhibits behavioral regimes of normal brain activity that were not explicitly built-in but emerged spontaneously as the result of interactions among anatomical and dynamic processes. We describe spontaneous activity, sensitivity to changes in individual neurons, emergence of waves and rhythms, and functional connectivity on different scales.

brain models | cerebral cortex | diffusion tensor imaging | oscillations | spike-timing-dependent synaptic plasticity

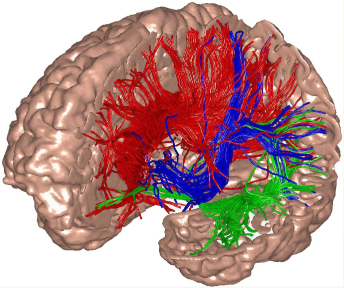


Fig. 1. The model's global thalamocortical geometry and white matter anatomy was obtained by means of diffusion tensor imaging (DTI) of a normal human brain. In the illustration, left frontal, parietal, and a part of temporal cortex have been cut to show a small fraction of white-matter fibers, color-coded according to their destination.

Izhikevich E. M. and Edelman G. M. (2008) "Large-Scale Model of Mammalian Thalamocortical Systems," *PNAS*, 105:3593-3598

Biologically Motivated Model

presynaptic neurons

postsynaptic neurons		percent of cells	number of synapses	nb1	p2/3	b2/3	nb2/3	ss4(L4)	ss4(L2/3)	p4	b4	nb4	p5(L2/3)	p5(L5/6)	b5	nb5	p6(L4)	p6(L5/6)	b6	nb6	corticocortical	TCs	TCn	TIs	TIn	TRN
nb1		1.5	8890	10.1	6.3	0.6	1.1	.	.	0.1	.	.	0.1	.	.	.	.	.	.	.	77.6	.	4.1	.	.	.
p2/3	L2/3	26	5800	.	59.9	9.1	4.4	0.6	6.9	7.7	.	0.8	7.4	.	.	.	2.3	.	.	0.8	.	.	.	.	.	.
	L1		1306	10.2	6.3	0.1	1.1	.	.	0.1	.	.	0.1	.	.	.	.	.	.	.	78	.	4.1	.	.	.
b2/3		3.1	3854	1.3	51.6	10.6	3.4	0.5	5.8	6.6	.	0.8	6.3	.	.	.	2.1	.	.	0.7	9.8	.	0.5	.	.	.
nb2/3		4.2	3307	1.7	48.6	11.4	3.3	0.5	5.5	6.2	.	0.8	5.9	.	.	.	1.8	.	.	0.6	13	.	0.7	.	.	.
ss4(L4)		9.2	5792	.	2.7	0.2	0.6	11.9	3.7	4.1	7.1	2	0.8	0.1	.	.	32.7	.	.	5.8	25.3	1.7	1.3	.	.	.
ss4(L2/3)		9.2	4989	.	5.6	0.4	0.8	11.3	3.8	4.3	7.2	2.1	1.1	0.1	.	.	31.1	.	.	5.5	23.9	1.7	1.3	.	.	.
p4	L4	9.2	5031	.	4.3	0.2	0.6	11.5	3.6	4.2	7.2	2.1	1.2	0.1	.	.	31.4	0.1	.	5.9	24.5	1.7	1.3	.	.	.
	L2/3		866	.	63.1	5.1	4.1	0.6	7.2	8.1	.	0.6	7.8	.	.	.	2.5	.	.	0.8	.	.	.	.	.	.
	L1		806	10.2	6.3	0.1	1.1	.	.	0.1	.	.	0.1	.	.	.	.	.	.	.	78	.	4.1	.	.	.
b4		5.4	3230	.	5.8	0.5	0.8	11	3.8	4.2	8.4	2.4	1.1	.	.	.	30.3	.	.	5.4	23.3	1.6	1.2	.	.	.
nb4		1.5	3688	.	2.7	0.2	0.6	11.7	3.6	4	8.2	2.3	0.8	0.1	.	.	32.2	.	.	5.7	24.9	1.7	1.3	.	.	.
p5(L2/3)	L5	4.8	4316	.	45.9	1.8	0.3	3.3	2	7.5	.	0.9	11.7	1	0.8	1.1	2.3	2.1	.	11.5	7.2	0.1	0.4	.	.	.
	L4		283	.	2.8	0.1	0.7	12.2	3.8	4.2	5.2	1.5	0.8	0.1	.	.	33.7	.	.	5.9	26	1.8	1.4	.	.	.
	L2/3		412	.	63.1	5.1	4.1	0.6	7.2	8.1	.	0.6	7.8	.	.	.	2.5	.	.	0.8	.	.	.	.	.	.
	L1		185	10.2	6.3	0.1	1.1	.	.	0.1	.	.	0.1	.	.	.	.	.	.	.	78	.	4.1	.	.	.
p5(L5/6)	L5	1.3	5101	.	44.3	1.7	0.2	3.2	2	7.3	.	0.8	11.3	1.2	0.8	1.1	2.3	2.5	0.3	11.3	9.2	0.2	0.5	.	.	.
	L4		949	.	2.8	0.1	0.7	12.2	3.8	4.2	5.2	1.5	0.8	0.1	.	.	33.7	.	.	5.9	26	1.8	1.4	.	.	.
	L2/3		1367	.	63.1	5.1	4.1	0.6	7.2	8.1	.	0.6	7.8	.	.	.	2.5	.	.	0.8	.	.	.	.	.	.
	L1		5658	10.2	6.3	0.1	1.1	.	.	0.1	.	.	0.1	.	.	.	.	.	.	.	78	.	4.1	.	.	.
b5		0.6	2981	.	45.5	2.3	0.2	3.3	2	7.5	.	1.1	11.6	1	0.9	1.3	2.3	2	.	11.4	7.2	0.1	0.4	.	.	.
nb5		0.8	2981	.	45.5	2.3	0.2	3.3	2	7.5	.	1.1	11.6	1	0.9	1.3	2.3	2	.	11.4	7.2	0.1	0.4	.	.	.
p6(L4)	L6	13.6	3261	.	2.5	0.1	0.1	0.7	0.9	1.3	.	0.1	0.1	4.9	.	0.3	1.2	13.2	7.7	7.7	55.7	0.6	2.9	.	.	.
	L5		1066	.	46.8	0.8	0.3	3.4	2.1	7.7	.	0.6	11.9	1	0.6	0.8	2.3	2.1	.	11.7	7.4	0.1	0.4	.	.	.
	L4		1915	.	2.8	0.1	0.7	12.2	3.8	4.2	5.2	1.5	0.8	0.1	.	.	33.7	.	.	5.9	26	1.8	1.4	.	.	.
	L2/3		121	.	63.1	5.1	4.1	0.6	7.2	8.1	.	0.6	7.8	.	.	.	2.5	.	.	0.8	.	.	.	.	.	.
p6(L5/6)	L6	4.5	5573	.	2.5	0.1	0.1	0.7	0.9	1.3	.	0.1	0.1	4.9	.	0.3	1.2	13.2	7.8	7.8	55.7	0.6	2.9	.	.	.
	L5		257	.	46.8	0.8	0.3	3.4	2.1	7.7	.	0.6	11.9	1	0.6	0.8	2.3	2.1	.	11.7	7.4	0.1	0.4	.	.	.
	L4		243	.	2.8	0.1	0.7	12.2	3.8	4.2	5.2	1.5	0.8	0.1	.	.	33.7	.	.	5.9	26	1.8	1.4	.	.	.
	L2/3		286	.	63.1	5.1	4.1	0.6	7.2	8.1	.	0.6	7.8	.	.	.	2.5	.	.	0.8	.	.	.	.	.	.
	L1		62	10.2	6.3	0.1	1.1	.	.	0.1	.	.	0.1	.	.	.	.	.	.	.	78	.	4.1	.	.	.
b6		2	3220	.	2.5	0.1	0.1	0.7	0.9	1.3	.	0.1	0.1	4.9	.	0.4	1.2	13.2	7.7	7.7	55.7	0.6	2.9	.	.	.
nb6		2	3220	.	2.5	0.1	0.1	0.7	0.9	1.3	.	0.1	0.1	4.9	.	0.4	1.2	13.2	7.7	7.7	55.7	0.6	2.9	.	.	.
				brainstem		sensory																				
TCs		0.5	4000	31	.	7.1	.	.	.	.	.	.	.	.	.	.	23	8	.	.	.	.	.	5	.	25.9
TCn		0.5	4000	31	.	7.1	.	.	.	.	.	.	14	3.8	.	.	.	13.2	.	.	.	.	.	.	5	25.9
TIs		0.1	3000	13.5	.	48.7	.	.	.	.	.	.	.	.	.	.	9.8	3.3	.	.	.	0.4	.	24.4	.	.
TIn		0.1	3000	13.4	.	48.7	.	.	.	.	.	.	5.8	1.6	.	.	.	5.4	.	.	.	.	0.6	.	24.4	.
TRN		0.5	4000	40	.	.	.	.	.	.	.	.	.	.	.	.	30	.	.	.	.	10	10	.	.	10

Izhikevich E. M. and Edelman G. M. (2008) "Large-Scale Model of Mammalian Thalamocortical Systems," *PNAS*, 105:3593-3598

Theorem

Let $G_n \sim HSBM(\theta)$.
Our goal is to consistently estimate θ.

Theorem

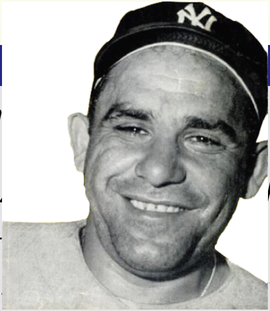
Under suitable eigenvalue assumptions,
a three-step algorithm given by

1. *cluster vertices* $\circ$ $ASE(G)$ *to estimate the* H_r
2. *cluster the collection* $\{ASE(\widehat{H}_r)\}_{r=1}^{\widehat{R}}$
3. *estimate motif parameters for each cluster of subgraphs*

yields $\widehat{\theta}_n \to \theta$ *as* $n \to \infty$.

Theorem

Let $G_n \sim HSBM(\theta)$.
Our goal is to consistently estimate θ.



Theorem

Under suitable eigenva
a three-step algorithm

1. *cluster vertices* $\circ$ *he* H_r
2. *cluster the collect*
3. *estimate motif pa ter of subgraphs*

yields $\widehat{\theta}_n \to \theta$ *as* $n \to \infty$.

"In theory there is no difference between theory and practice. In practice, there is." – Yogi Berra

Open Source

OPEN CONNECTOME PROJECT

- connectome data repository
- to make state-of-the-art neuroscience open and to facilitate the analysis of connectome data
- collaboration with Harvard, HHMI's Janelia Farm, etc.
- http://www.openconnectomeproject.org

igraph

- subcontract to Harvard as a part of DARPA XDATA project
- distributing our graph inference methodologies via "the network analysis package"
- http://igraph.org

Vision & Plans

Impact

Our work will provide pioneering robust inference capabilities for cortical graphs which will in turn

1. impact the future of neuroscience through fundamental insights into recurring motifs in the neocortex, and
2. impact the future of machine intelligence through estimates of the structure of cortical computing primitives.

Furthermore, we expect our work to have major impact on the burgeoning field of robust graph inference, and to unite the nascent field of data science for graphs through a compelling and high-profile application in the natural sciences.

Leopold Kronecker to Hermann von Helmholtz (1888):

"The wealth of your practical experience
with sane and interesting problems
will give to mathematics
a new direction and a new impetus."

Kronecker

Helmholtz