# IARPA <br> Machine Intelligence from Cortical Networks (MICrONS) workshop 

# On Statistical Inference for the Cortical Column Conjecture 

Carey E. Priebe

Department of Applied Mathematics \& Statistics Johns Hopkins University, Baltimore, MD, USA

February 11-12, 2014

## Outline

Introduction
Cortical Column Conjecture Random Graph Model
Hypotheses
Biologically Motivated Example

Cortical Column Conjecture Test
Known Graph Alignment
Effect of Graph Matching
Identifying Large Scale Structures
Effect of Misspecified Subgraphs
Effect of Misspecified Subgraphs in Practice

Discussion

## 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.

## Introduction

## Cortical Column Conjecture

Neurons are connected in a graph that exhibits motifs representing repeated processing modules (Horton \& Adams, 2005; Mountcastle, 1997).

## Question

- Can we test whether an observed cortical graph satisfies the conjecture?
- Can we estimate relevant cortical computing parameters from an observed cortical graph?

Goal
To present a notional demonstration of how statistical inference on graphs can inform our understanding of cortical computing.
J. C. Horton and D. L. Adams. "The cortical column: a structrue 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.


Pipeline

Cortical Column Conjecture


## Cortical Column Graph



- cortical graph G: hierarchical block model on $n$ vertices (neurons).
- induced subgraph $H_{r}$ : stochastic block models on $n_{r}, r=1, \ldots, R$.

1. identify large-scale structures in $G \rightarrow H_{r}$,
2. test if $\hat{H}_{r}$ 's correspond to repeated motifs.

## Stochastic Block Model

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

$$
\mathcal{B}_{K}:=\left\{B \in[0,1]^{K \times K_{\mid}} \mid B^{T}=B\right\} .
$$

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

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

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 $G \sim S B M(V, \vec{n}, B)$ if the vertices $V$ are partitioned into subsets $V_{1}, \cdots, V_{K}$ with sizes given by $\vec{n}=\left(\vec{n}_{1}, \cdots, \vec{n}_{K}\right)$ and edges $\mathbb{1}\{u \sim v\}$ are independent Bernoulli random variables with $P[u \sim v]=B_{i j}$ for $u \in V_{i}$ and $v \in V_{j}$.

## Hierarchical Stochastic Block Model

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

- $\left\{V_{r}\right\}_{r=1}^{R}$ : a disjoint partition of the vertices $V$, and for each $r=1,2, \ldots, R$, write $n_{r}:=\left|V_{r}\right|$.
- $H_{r}=\Omega\left(V_{r}\right)=\left(V_{r}, E_{r}\right)$.
- $H_{r} \sim \operatorname{SBM}\left(V_{r}, \vec{m}_{r}, B_{r}\right)$, where $\vec{m}_{r} \in \mathcal{M}_{n_{r}, K}$ and $B_{r} \in \mathcal{B}_{K}$.
- for $u \in V_{r}$ and $v \in V_{r^{\prime}}$ with $r \neq r^{\prime}, \mathbb{1}\{u \sim v\} \sim \operatorname{Bernoulli}(p)$.
$G \sim \operatorname{HSBM}(V, \vec{m}, B)$, where $\vec{m}=\left[\vec{m}_{1}^{\top}\left|\vec{m}_{2}^{\top}\right| \cdots \mid \vec{m}_{R}^{\top}\right]^{\top}$ and $B$ is given by

$$
B=\left[\begin{array}{cccc}
B_{1} & p J_{K, K} & \cdots & p J_{K, K} \\
p J_{K, K} & B_{2} & \ddots & \vdots \\
\vdots & \ddots & \ddots & p J_{K, K} \\
p J_{K, K} & \cdots & p J_{K, K} & B_{R}
\end{array}\right]
$$

## Hypothesis

Given generating block-probability matrix $B^{*} \in \mathcal{B}_{K}$ and $\epsilon \in[0,1]$, we let $B_{r} \stackrel{\text { iid }}{\sim} F_{B^{*}, \epsilon}$ where $F_{B^{*}, \epsilon}$ specifies mutually independent entries with

$$
B_{r, i j} \sim \operatorname{Uniform}\left(B_{i j}^{*} \pm \epsilon \bar{u}_{i j}\right)
$$

for $i \leqslant j$ (recall symmetry of $B_{r}$ ). Here $\bar{u}_{i j}=\min \left\{B_{i j}^{*}, 1-B_{i j}^{*}\right\}$ is the maximum allowable range guaranteeing that the random variable $B_{r, i j}$ stays in $[0,1]$.
Hypotheses of interest are given by

$$
H_{0}: \epsilon \geqslant \epsilon_{0}
$$

versus

$$
H_{A}: \epsilon<\epsilon_{0}
$$

with $\epsilon_{0} \in(0,1]$. A smaller value of $\epsilon$ corresponds to motifs having more coherent structure.

## Biologically Motivated Model

Our cortical graph G:

- $n$ vertices, with $n \in\left\{10^{3}, 10^{4}, 10^{5}, 10^{6}\right\}$.
- $V=[n]$ is partitioned into $R$ subsets $\left\{V_{r}\right\}_{r=1}^{R}$.
- $H_{r}=\Omega\left(V_{r}\right)$ each have $\left|V_{r}\right|=n_{r}=m=100, K=5$ blocks.
- $R=n / m$.
- $\vec{m}_{r}=[2,50,15,8,25]^{\top} \forall r$.

$$
B^{*}=\left[\begin{array}{ccccc}
0.1 & 0.045 & 0.015 & 0.19 & 0 \\
0.045 & 0.05 & 0.035 & 0.14 & 0.03 \\
0.015 & 0.035 & 0.08 & 0.105 & 0.04 \\
0.19 & 0.14 & 0.105 & 0.29 & 0.13 \\
0 & 0.03 & 0.04 & 0.13 & 0.09
\end{array}\right]
$$

- $p \in\left\{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\right\}$.

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

# 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, colorcoded according to their destination.


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


Dimensions: $1000 \times 1000$
(c) unorganized alternative


Dimensions: $1000 \times 1000$
(b) organized null


Dimensions: $1000 \times 1000$
(d) organized alternative


## Cortical Column Conjecture Test

Let $A_{r}$ be the adjacency matrix associated with induced subgraph $H_{r}$.

We consider the test statistic

$$
T=\sum_{r=1}^{R}\left\|A_{r}-\bar{A}\right\|
$$

(Small values of $T$ provide evidence in favor of $H_{A}$.)

## If the $H_{r}$ are known and aligned



For example, the $p$-value for the alternative observation, with $T \approx 48.17$, is $p=0.00345$, justifying our claim that "the repeated motif structure is evident."

## The Effect of Graph Matching

For the test statistic

$$
T=\sum_{r=1}^{R}\left\|A_{r}-\bar{A}_{r}\right\|
$$

we use post-matching sample graph means

$$
\bar{A}_{r}=(1 / R) \sum_{r^{\prime}} \tilde{A}_{r, r^{\prime}}
$$

Here

$$
\tilde{A}_{r, r^{\prime}}=\underset{P \in P\left(n_{r} \vee n_{r^{\prime}}\right)}{\arg \min }\left\|A_{r}-P A_{r^{\prime}} P^{\top}\right\|_{F}
$$

is the adjacency matrix for the graph $H_{r^{\prime}}$ matched to the graph $H_{r}$. $P(n)$ is the set of $n \times n$ permutation matrices (padding adjacency matrices with zeros as needed to make the matrix multiplication consistent).

## The Effect of Graph Matching



Consider two cases:

1. identity: $\beta \approx 0.865$
2. barycenter: $\beta \approx 0.849$

Both are significantly superior to power without graph matching $(\beta \approx 0.739)$ !?

## Identifying Large Scale Structures

- So far, we have considered the situation in which we know the true subgraphs $\left\{H_{r}=\Omega\left(V_{r}\right)\right\}_{r=1}^{R}$.
- In practice, we must first identify these large-scale structures in G, e.g., community detection.
- We consider the case where we have imperfectly identified subgraphs $\left\{\hat{H}_{r}=\Omega\left(\hat{V}_{r}\right)\right\}_{r=1}^{\hat{R}}$, and it will be the collection $\left\{\hat{H}_{r}\right\}$ that we ultimately must use to test the cortical column conjecture.


## Robustness



Consider two cases:

1. blue: using true subgraphs $H_{r}=\Omega\left(V_{r}\right)$.
2. red: using imperfectly identified subgraphs $\left\{\hat{H}_{r}\right\}_{r=1}^{R}$.
NB: significant reduction of power!

## Community Detection

We now consider the effect of using community detection algorithms for (necessarily errorful) identification of large-scale structures (candidate motifs) in G.

- Louvain (Walktrap, Infomap, ...).
- clustering $\circ$ adjacency spectral embedding.
- Adjusted Rand Index as a performance measure?
- Subsequent power as the performance measure!
V. D. Blondel, J.L. Guillaume, R. Lambiotte, and E. Lefebvre. "Fast unfolding of communities in large networks," Journal of Statistical Mechanics: Theory and Experiment, 2008(10):P10008, 2008.


## Example

Consider $n=10^{3}, R=10$, and $\epsilon_{H_{0}}=1.0, \epsilon_{H_{A}}=0.3$.
With $p=0.001$ :

```
## Louvain
wcL <- multilevel.community(g)
membL <- membership(wcL)
table(membL)
# HO:
# 1 rrrrrrrrrrrrrrrrrrrrrrrra
# HA:
# 1
#101 1
(ariL <- adjustedRandIndex(labG, membL))
#[1] 0.8965267 : HO
#[1] 0.9319579 : HA
```

We truncate the collection of identified subgraphs $\left\{\hat{H}_{r}\right\}_{r=1}^{\hat{R}}$ by discarding all those "tiny" clusters, yielding $\underline{\hat{R}} \leqslant \hat{R}$. We proceed with the truncated collection $\left\{\hat{H}_{r}\right\}_{T=1}^{\hat{R}}$. (We use $\tau=10$.)

## Example

| algorithm | $\operatorname{ARI}\left(H_{0}, H_{A}\right)$ | power |
| :---: | :---: | :---: |
| Louvain | $(0.90,0.93)$ | 0.81 |
| Walktrap | $(0.88,0.92)$ | 0.82 |
| Infomap | $(0.86,0.92)$ | 0.86 |
|  |  |  |
| skmeans $\circ$ ase | $(0.85,0.91)$ | 0.76 |

## Example

"louvain"
$\mathrm{p}=0.001$, power $=0.8128$


$$
p=0.001, \beta=0.813
$$

"louvain"

$p=0.01, \beta=0.158$


## Discussion

- Ky Fan \& Schatten
- estimation: $\widehat{\epsilon}$
- $(p-c) J_{K, K}+q c I_{K}$
- directed, weighted, loopy, multi, ...
- soft seeds: $80 / 20$ vertex attribute split (excitatory/inhibitory)
- a limited set of computing primitives; that is, $\geqslant 1$ repeated motifs
- extension of this method to allow for overlapping subgraphs; that is, a given node can participate in multiple instances of a repeated subgraph
- graph-centric computer vision
- graph anova
- invariants
- scalability
- QQ / errroful / experimental design
- ...


## Matching C.elegans



Figure 4: Plotting the matched ratio $R_{m} \pm 2$ s.e. for matching the 253 vertex chemical and electrical C. elegans connectomes for seed values $m=0,50,100,150,200$. We show the performance of the JOFC and SGM algorithms for matching the graphs for all combinations of with/without directness and with/without edge weights. JOFC is plotted in green, SGM in red, and chance in black. Note that JOFC for each combination of with/without directedness and with/without edge weights significantly outperforms the best SGM combination (directed and unweighted). For each combination of $m$ and $\rho$, we ran 100 MC replicates.

Vince Lyzinski, Sancar Adali, Joshua T. Vogelstein, Youngser Park, Carey E. Priebe, "Seeded Graph Matching Via Joint Optimization of Fidelity and Commensurability," arXiv, 2014.

## Multiple Repeated Motifs

After performing SGM \& Louvain procedures as needed, provided that the number $k$ of repeated motifs is $\widehat{k}$,

$$
T=\sum_{r=1}^{R}\left\|A_{r}-\widehat{A}_{r}\right\|
$$

where $\widehat{A}_{r}:=\sum_{i=1}^{\widehat{k}} \widehat{\mathrm{G}}_{i} \widehat{H}_{i r}$, each $n \times n$ non-negative matrix $\widehat{\mathrm{G}}_{i}$ represents an (estimated) repeated motif, and each $\left(\widehat{H}_{1 r}, \ldots, \widehat{H}_{\widehat{k r}}\right)$ is a probability vector as estimated using the following procedures:

SVT+NMF @ rank $k$

$$
\begin{aligned}
X e_{r} & \leftarrow \operatorname{vec}(G(r)) \text { for } r=1, \ldots, R \\
\widehat{X} & \leftarrow\left(U \Sigma V^{\top} \text { of } X @ \operatorname{rank} k\right)^{+} \\
(\widehat{W}, \widehat{H}) & \leftarrow \arg \min \|\widehat{X}-\widehat{W} \hat{H}\|_{F}
\end{aligned}
$$

Model Selection

$$
\begin{aligned}
& \widehat{k} \leftarrow \underset{k=1, \ldots, R}{\arg \min } \operatorname{AIC}(k), \\
& \widehat{G}_{i} \leftarrow \operatorname{graph}\left(\widehat{W} e_{i}\right) \text { for } i=1, \ldots, \widehat{k}, \\
& \text { where } e_{i} \text { is a standard basis in } \mathbb{R}^{\widehat{r}} .
\end{aligned}
$$

## Quantity/Quality Trade-Off

$\beta$ and $\partial \beta$


Figure 1. Power $\beta$ and its derivative $\frac{\partial \beta}{\partial \varepsilon}$ as functions of the edge tracing error rate $\varepsilon$ for our example scenario (see text for details). (We plot $\left(\frac{1}{2}\right) \frac{\partial \beta}{\partial \varepsilon}(\varepsilon)$ so that the two curves are on approximately the same scale and can productively be presented on the same plot.)
C.E. Priebe, J.T. Vogelstein and D. Bock "Optimizing the quantity/quality trade-off in connectome inference," Communications in Statistics - Theory and Methods, Volume 42, Issue 19, pp. 3455-3462, 2013.

## Hierarchical Experimental Design

4) 



Sir Ronald A. Fisher
(Photograph courtesy of Antony Barrington-Brown)

## The <br> Design of Experiments

## By

Sir Ronald A. Fisher, Sc.D., F.R.S.
Honorary Research Fellow, Division of Mathematical Statistics, C.S.I.R.O., University of Adelaide; Foreign Associate, United States National Academy of Sciences, and Foreign Honorary Member, American Academy of Arts and Sciences; Foreign Member of the Swedish Royal Academy of Sciences, and the Royal Danish Academy of Sciences and Letters; Member of the Pontifal Academy; Galton Professor, University of London, and Arthur Balfour Professor of Genetics, University of Cambridge


HAFNER PUBLISHING COMPANY New York 1971

Leopold Kronecker to Hermann von Helmholtz:
"The wealth of your practical experience with sane and interesting problems
will give to mathematics
a new direction and a new impetus."


Kronecker


Helmholtz

