Resting-State Functional MRI Analysis by Graphical Model

With Applications on Functional Network Estimation

Wei Liu

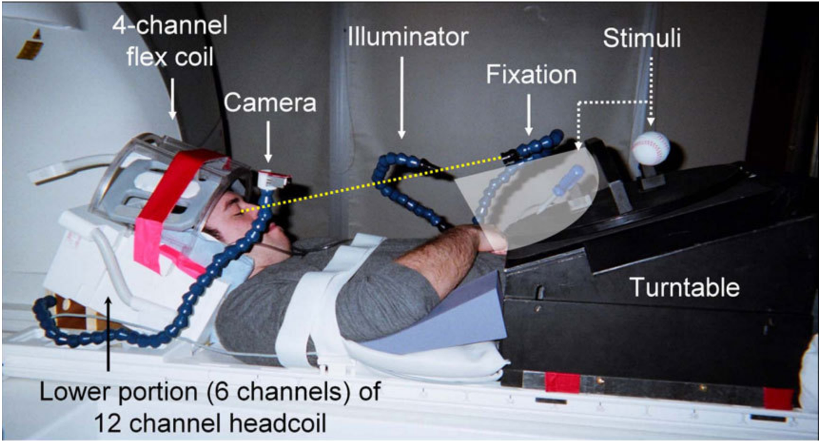
Scientific Computing and Imaging Institute
University of Utah
Advisor: Tom Fletcher
Co-advisor: Suyash Awate

November 25, 2013

www.sci.utah.edu

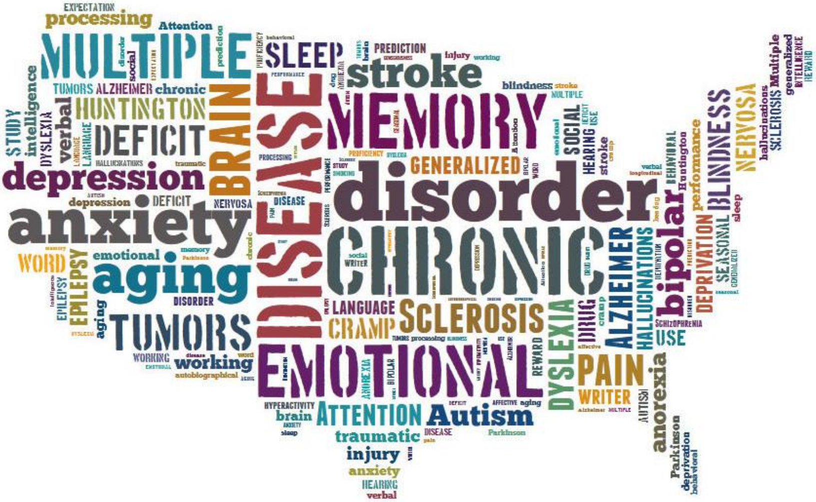
THE UNIVERSITY OF UTAH

fMRI Experiments

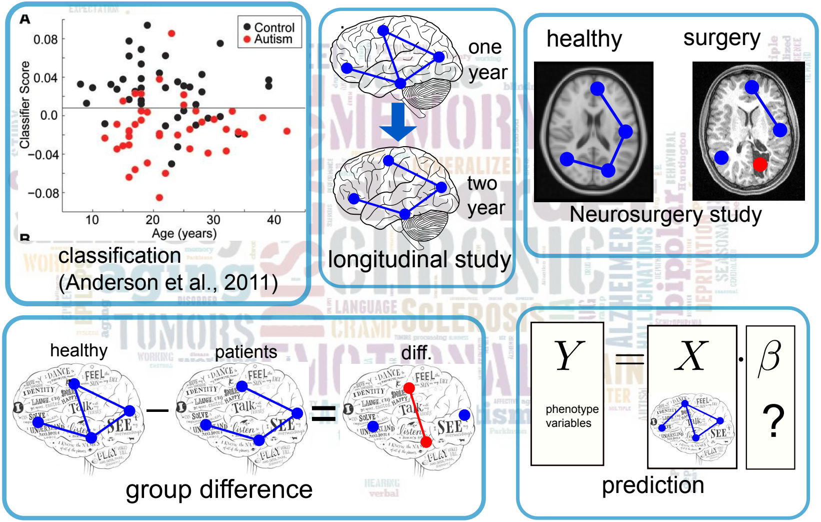


J C. Snow, Nature 2011

Clinical Applications of rs-fMRI



Clinical Applications of rs-fMRI



Outline

Thesis Statement

Statement

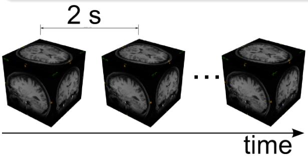
A multilevel Markov Random Field model improves the reliability of the functional network estimation in rs-fMRI group study by taking into account context information as a prior. The data-driven Bayesian model can jointly estimate the functional networks of both the group and the subjects, as well as drawing inference on the variability of the network maps across subjects.

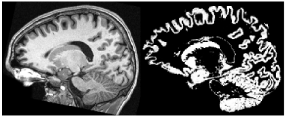
Contributions

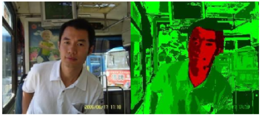
- Full pairwise connectivity with spatial coherence.
- Identify consistent, spatially coherent multiple functional networks.
- Hierarchical model for joint estimation of group and subject networks.
- Consistency analysis of the hierarchical model.

Functional MRI Data

- Blood oxygen level dependent (BOLD) indirectly measures neuronal activity.
- 3D volumes sampled at each time point.
- Fast scan, but noisy.
- Spatio-temporal dependency.




seg. of T1 MRI


seg. of natural images

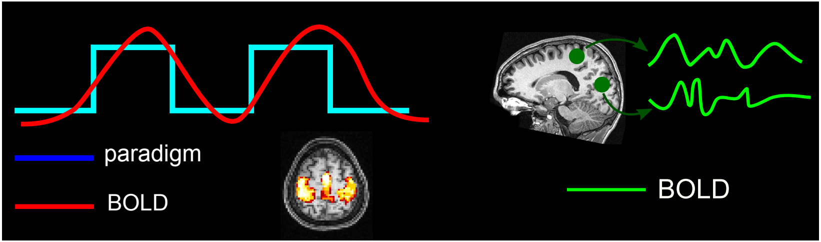
Task-based v.s. rs-fMRI

Task-based

- Experiment stimulus signal.
- Subjects undertake cognitive tasks.
- Linear regression between stimulus and BOLD.

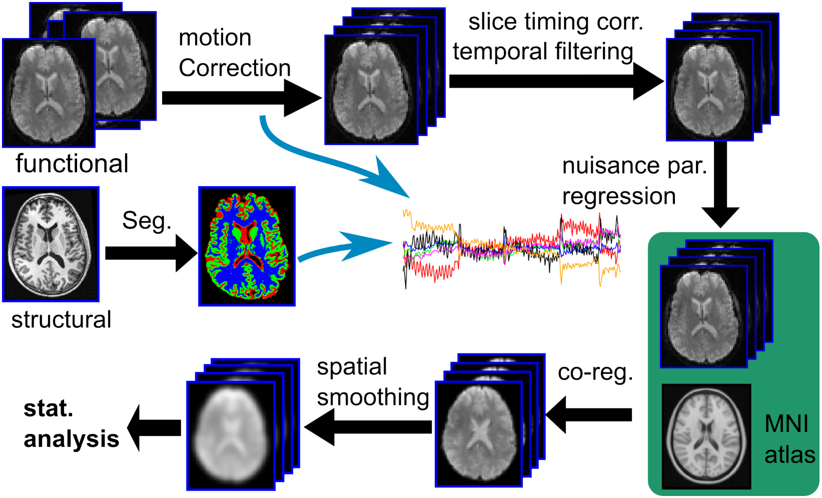
Resting-State

- No experiment paradigm signal.
- Subject stays in scanner. Eyes closed/open to a fixation cross.
- Correlation between two voxels.

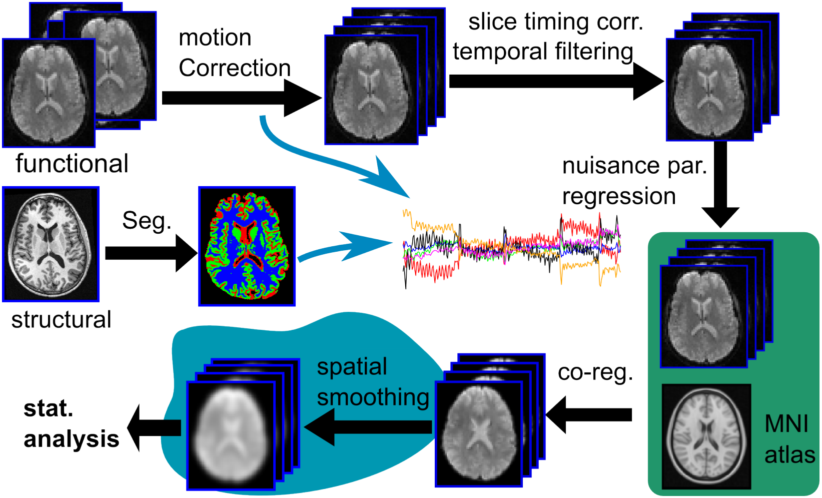


task resting-state

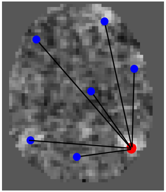
fMRI Processing Pipeline



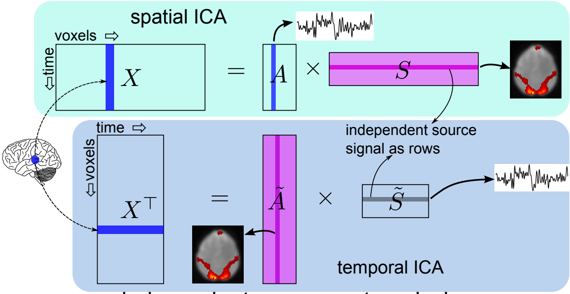
fMRI Processing Pipeline



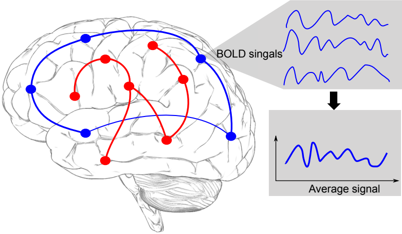
Network Analysis Methods



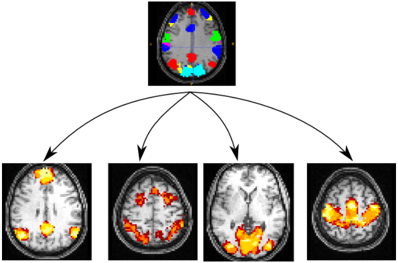
seed-based methods



independent component analysis

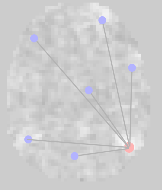


graph-based methods

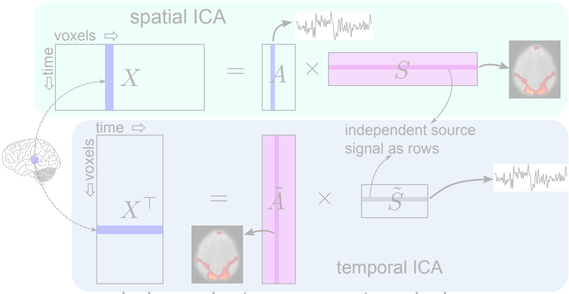


segmentation/clustering

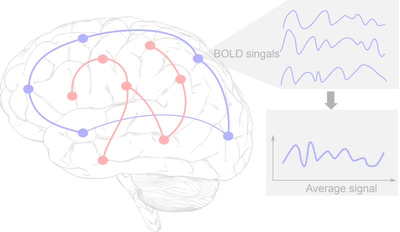
Network Analysis Methods



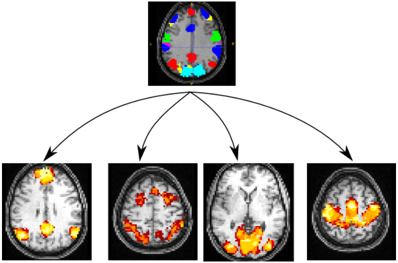
seed-based methods



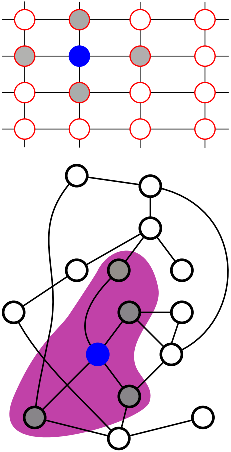
independent component analysis



graph-based methods



segmentation/clustering



Definition

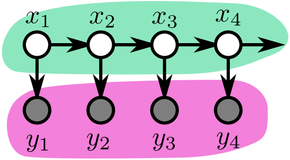
$\mathcal{G} = (\mathcal{V}, \mathcal{E})$: undirected graph.
$s \in \mathcal{V}$: a node/site in $\mathcal{V}$.
$X = \{x_1, \dots x_N\}$: Set of random variables defined on $\mathcal{G}$.
$\mathcal{N}_s$: Set of nodes neighboring s.
$(r, s) \in \mathcal{E} \Leftrightarrow r \in \mathcal{N}_s$.

Definition

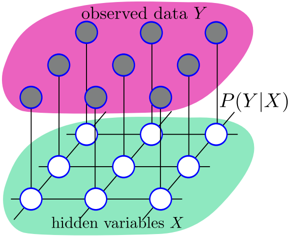
A Markov Random Field (MRF) is a collection of variables X defined on graph $\mathcal{G}$ if for all $s \in \mathcal{V}$

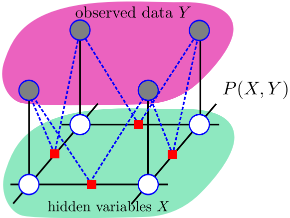
$$P(s_s | X_{\mathcal{V}-s}) = P(s_s | x_{\mathcal{N}_s})$$

Hidden Markov Model: A Generative Model



- X is defined on MRF.
- Y is assumed to be generated from X.
- Inverse problem: Given Y, estimate X.





Other forms exist: conditional random field, but no Bayesian interpretation.

Outline for section 2

2. Pairwise Connectivity with Six Dimensional MRF
   - Problem Statement
   - A Six Dimensional MRF model
   - Statistical Inference
   - Experiments on Synthetic and Real Data

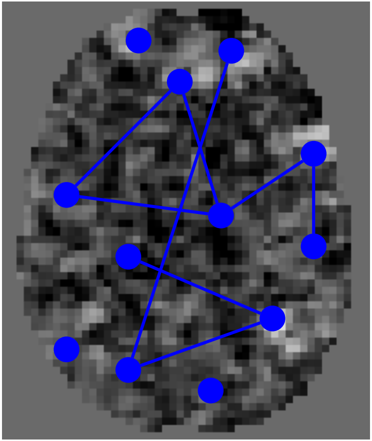
Contributions

- Full pairwise connectivity with spatial coherence.
- Identify consistent, spatially coherent multiple functional networks.
- Hierarchical model for jointly estimation of group and subject networks.
- Consistency analysis of the hierarchical model.

Pairwise Connectivity With Spatial Coherence [Liu2010]

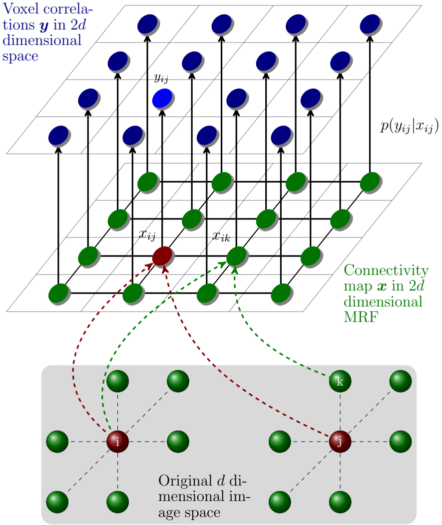
The Goal

- The connectivity between each pair of voxels in single subject rs-fMRI.
- Spatial regularization by MRF, without changing signals.
- Learn the strength of the smoothness from the data.



Model Definition

- MRF defined on 6D graph.
- Pairwise connectivity variable $X \in \{0,1\}^N$, sample correlation Y.
- $(x_{ij}, x_{ik}) \in \mathcal{E} \Leftrightarrow j \in \mathcal{N}(k)$.
- $P(F(y_{ij})|x_{ij}) \sim \mathcal{N}(\mu, \sigma^2)$.



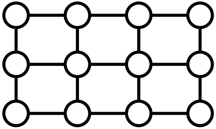
Statistical Inference

Questions

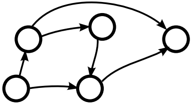
- Given x_{-s} and Y, what is $P(x_s)$.
- $x_s = \operatorname{argmax} P(X|Y)$
- $X^* = \operatorname{argmax} P(X|Y)$.

Algorithms

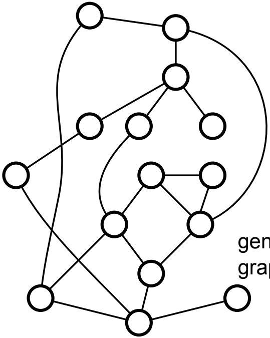
- Exact solutions for simple graphs (trees, chains): sum-product, max-sum, belief propagation.
- No exact solution for general graphs.



regular lattice

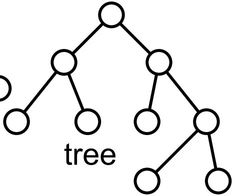


directed graph



general graph

chain



tree

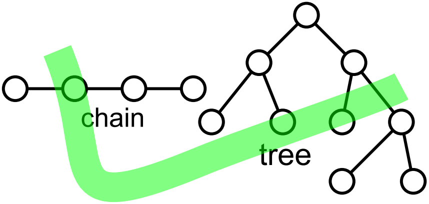
Statistical Inference

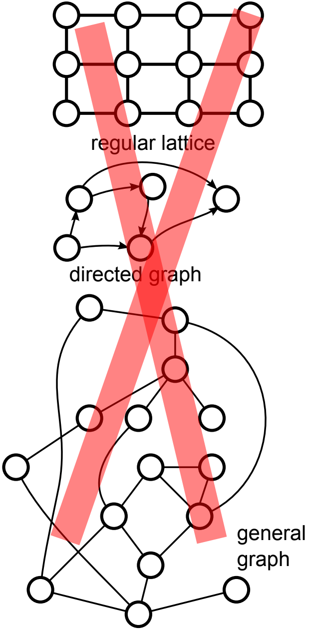
Questions

- Given x_{-s} and Y, what is $P(x_s)$.
- $x_s = \operatorname{argmax} P(X|Y)$
- $X^* = \operatorname{argmax} P(X|Y)$.

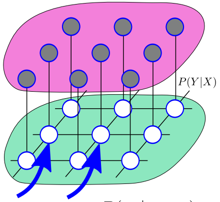
Algorithms

- Exact solutions for simple graphs (trees, chains): sum-product, max-sum, belief propagation.
- No exact solution for general graphs.



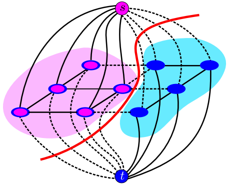


Approximate Inference

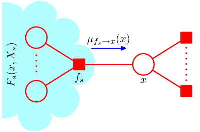


$x_s = \mathrm{argmax} P(x_s|x_{\mathcal{N}(s)})$

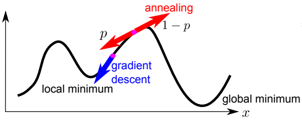
iterated conditional modes
= coordinate gradient descent



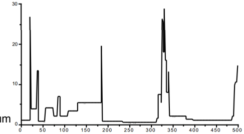
graph cuts: min (E) = min(Cuts)
= maxflow global optimum for
2 classes



And many other methods:
belief/expectation propagation,
Viterbi,...



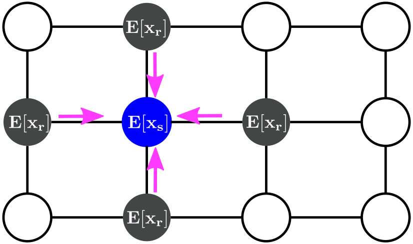
simulated annealing: non-zero probability of going uphill



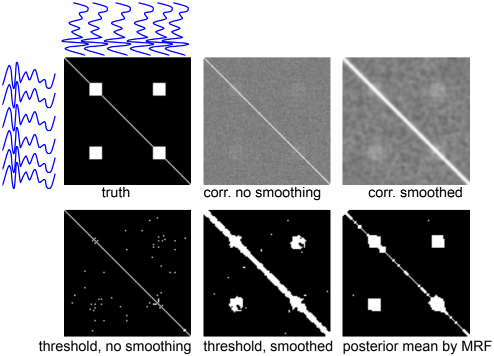
sampling from P(X|Y) to approximate posterior

Variational Inference

- Assuming $Q(X) = \prod_s q_s(x_s)$.
- Search $Q(x)$ in a smaller space.
- $\log Q_s(x_s) = E_{r \neq s}[\log P(X, Y)] + \text{const}$



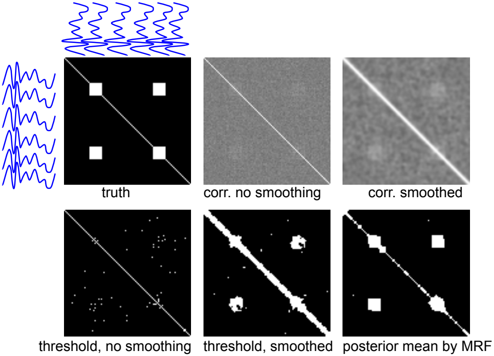
Experiments



Simulated data

- Construct 1D image. Connected voxels are added with sine wave signal.
- Spatial smoothing improves results, but increased false positive.

Experiments

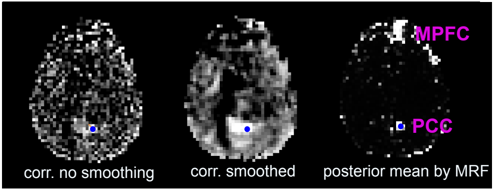


Simulated data

- Construct 1D image. Connected voxels are added with sine wave signal.
- Spatial smoothing improves results, but increased false positive.

Real data

- connectivity between a voxel in PCC and current slice.
- Detecting PCC-MPFC links in default mode network.



Outline for section 3

3 Consistent, Spatially Coherent Multiple Functional Networks
- Problem Statement
- A MAP framework with MRF prior
- MCEM for Statistical Inference
- Experiments on Synthetic Data and Real Data

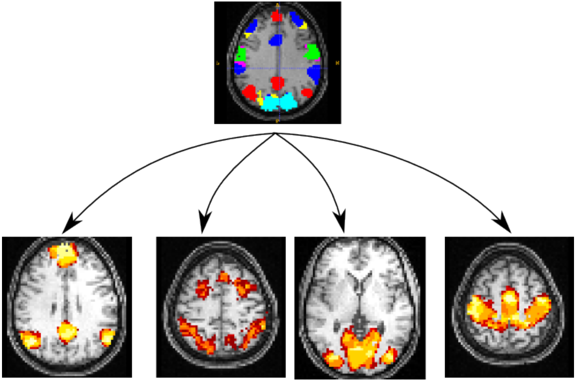
Contributions

- Full pairwise connectivity with spatial coherence.
- Identify consistent, spatially coherent multiple functional networks.
- Hierarchical model for jointly estimation of group and subject networks.
- Consistency analysis of the hierarchical model.

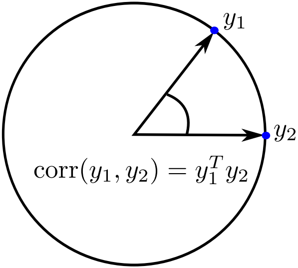
Consistent, Spatially Coherent Multiple Functional Networks [Liu2011]

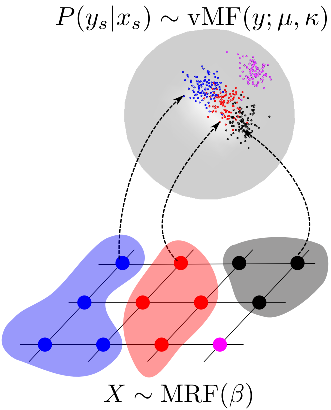
The Goal

- Partition the brain into multiple functional networks.
- Spatial coherence is respected.
- Parameter estimation.



MRF and vMF Models





Model

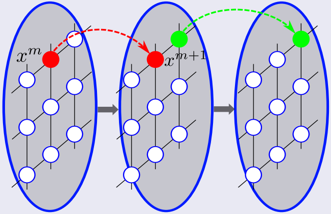
- $P(X) = (1/Z) \exp\left(-\beta \sum_{(r,s)\in\mathcal{E}} \psi(x_s, x_r)\right)$.
- $x_s \in \{1, \ldots, L\}$
- $P(y_s|x_s) = C_p(\kappa_l) \exp(\kappa_l \mu_l^\top y_s)$, $y_s \in S^{p-1}$ (von Mises-Fisher distribution).
- Solve $P(X|Y) \propto P(X) \cdot P(Y|X)$

Expectation Maximization

$$Q(\theta) = \mathbb{E}_{X|Y}[\log P(X, Y; \theta)]$$

Monte Carlo Expectation Maximization (MCEM)

$$\mathbb{E}_{X|Y}[\log P(X, Y; \theta)] \approx \frac{1}{M} \sum_m \log P(X^m, Y; \theta)$$



Statistical Inference by MCEM

Expectation Maximization

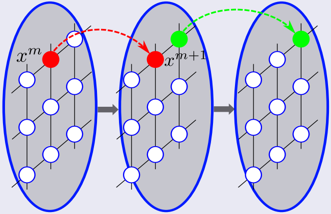
$Q(\theta) = \mathbb{E}_{X|Y}[\log P(X, Y; \theta)]$

Pseudo Likelihood

$\log P(X^m; \theta) \approx \sum_{s \in \mathcal{V}} \log P(x_s | x_{\mathcal{N}_s}; \theta)$

Monte Carlo Expectation Maximization (MCEM)

$$\mathbb{E}_{X|Y}[\log P(X, Y; \theta)]$$
$$\approx \frac{1}{M} \sum_m \log P(X^m, Y; \theta)$$



Statistical Inference by MCEM

Expectation Maximization

$Q(\theta) = \mathbb{E}_{X|Y}[\log P(X, Y; \theta)]$

Pseudo Likelihood

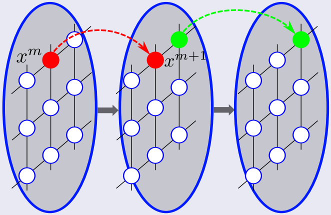
$\log P(X^m; \theta) \approx \sum_{s \in \mathcal{V}} \log P(x_s | x_{\mathcal{N}_s}; \theta)$

Parameter Estimation

$\hat{\mu}_l = \| R_l \|, R_l = \sum_{s \in \mathcal{V}_l} y_s$
$\hat{\kappa}_l \approx (pR_l - R^3)/(1 - R^2)$
$\hat{\beta} = \mathrm{argmax}_\beta \log P(X; \theta)$ by Newton's method.

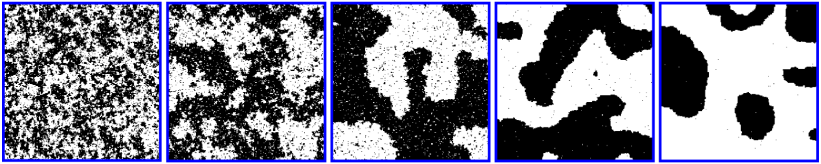
Monte Carlo Expectation Maximization (MCEM)

$$\mathbb{E}_{X|Y}[\log P(X, Y; \theta)] \approx \frac{1}{M} \sum_m \log P(X^m, Y; \theta)$$



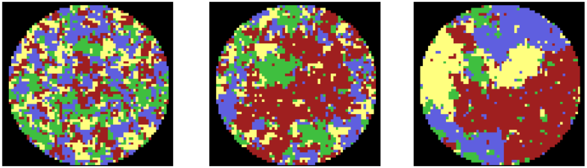
Simulation of Ising and Potts Models

Ising model

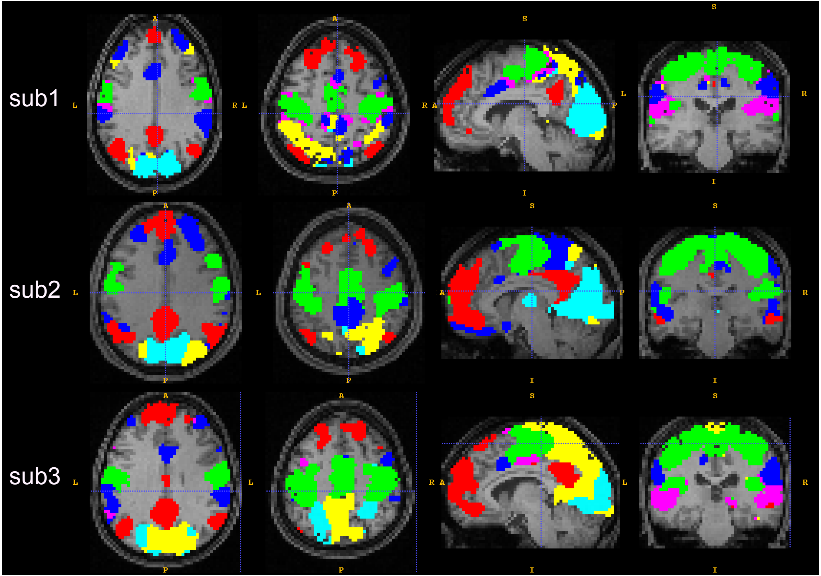


$\beta = 0.8$ $\beta = 0.88$ $\beta = 1.0$ $\beta = 1.5$ $\beta = 2.0$

Potts model



$\beta = 1.0$ $\beta = 1.1$ $\beta = 1.2$



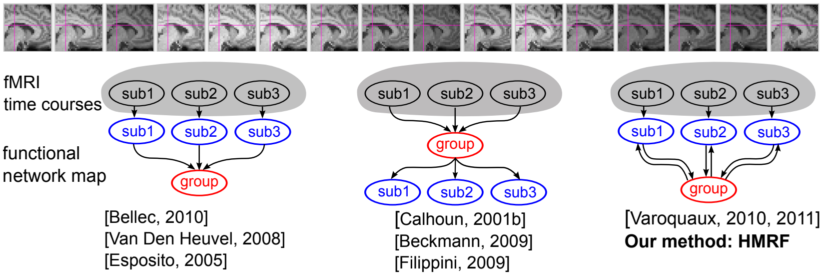
visual (cyan), motor (green), executive control (blue), salience (magenta), dorsal attention (yellow), and default mode (red) networks

Outline for section 4

Contributions

- Full pairwise connectivity with spatial coherence.
- Identify consistent, spatially coherent multiple functional networks.
- Hierarchical model for jointly estimation of group and subject networks.
- Consistency analysis of the hierarchical model.

Hierarchical Model: a Better Approach [Liu2012]



Existing Methods

- Bottom-up or top-down.
- Subject is estimated independently.
- Estimation is one way.

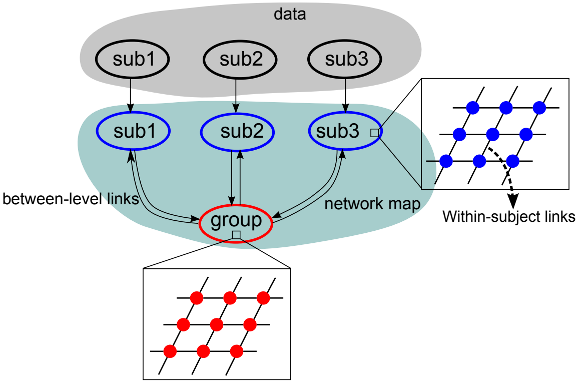
We Propose

- A hierarchical structure including group and subject.
- Jointly estimate both levels iteratively.
- Bayesian framework. Data driven. Parameter estimation.

Joint Estimation with MRF

Build a graph

- within-subject piecewise constant constraints.
- Between-subject (between-level) dependency.

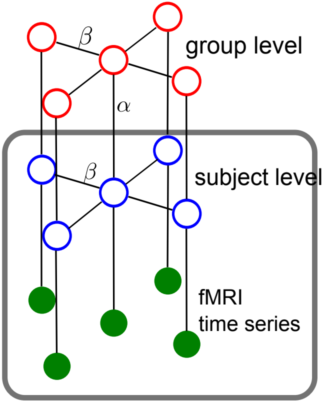


A Graphical View of the Hierarchical Model

$$\mathcal{G} = (\mathcal{V}_G, \mathcal{V}_1, \ldots, \mathcal{V}_J)$$

$$\mathcal{E} = \{(r,s) | (r,s) \in \mathcal{E}_G \text{ or } s \in \mathcal{V}_G, r \in \mathcal{V}_j \text{ or } (r,s) \in \mathcal{V}_j\}$$

$$U(X) = \sum_{(s,r)\in\mathcal{V}_G} \beta\psi(x_s, x_r) + \sum_{j=1}^{J} \left(\sum_{(s,\tilde{s})} \alpha\psi(x_s, x_{\tilde{s}}) + \sum_{(s,r)\in\mathcal{V}_j} \beta\psi(x_s, x_r) \right).$$



Likelihood

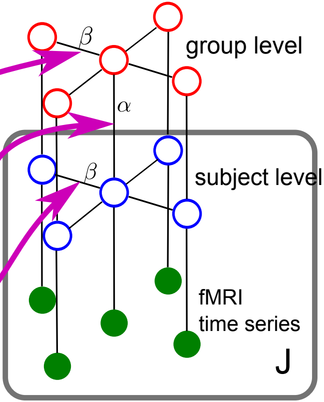
$P(y_s | x_s) \sim vMF(\mu, \kappa)$

A Graphical View of the Hierarchical Model

$$\mathcal{G} = (\mathcal{V}_G, \mathcal{V}_1, \dots, \mathcal{V}_J)$$

$$\mathcal{E} = \{(r,s) | (r,s) \in \mathcal{E}_G \text{ or } s \in \mathcal{V}_G, r \in \mathcal{V}_j \text{ or } (r,s) \in \mathcal{V}_j\}$$

$$U(X) = \sum_{(s,r)\in\mathcal{V}_G} \beta\psi(x_s, x_r) + \sum_{j=1}^{J} \left(\sum_{(s,\tilde{s})} \alpha\psi(x_s, x_{\tilde{s}}) + \sum_{(s,r)\in\mathcal{V}_j} \beta\psi(x_s, x_r) \right).$$

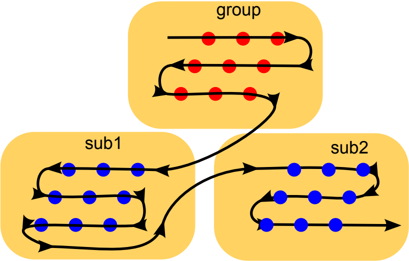


Likelihood

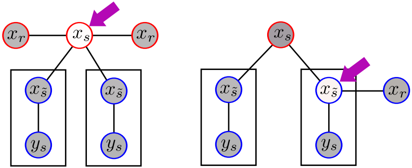
$P(y_s | x_s) \sim vMF(\mu, \kappa)$

Bayesian Inference: Gibbs Sampling

- Monte Carlo Sampling used to approximate $\mathbb{E}_{X|Y}[\log P(X, Y; \theta)]$.
- Gibbs sampling also in a multi-level fashion.



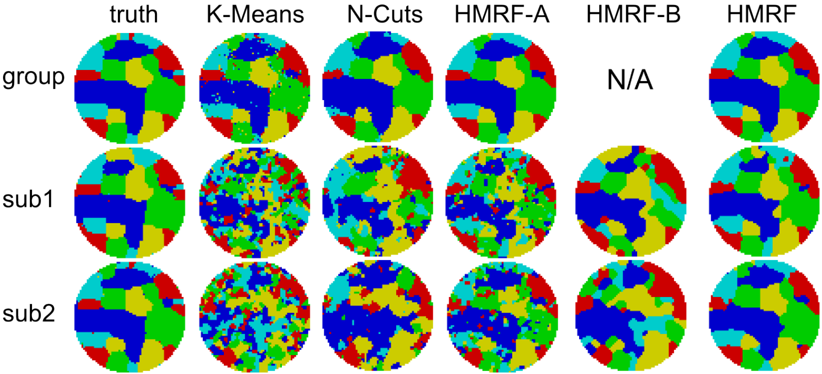
Gibbs sampling Schedule



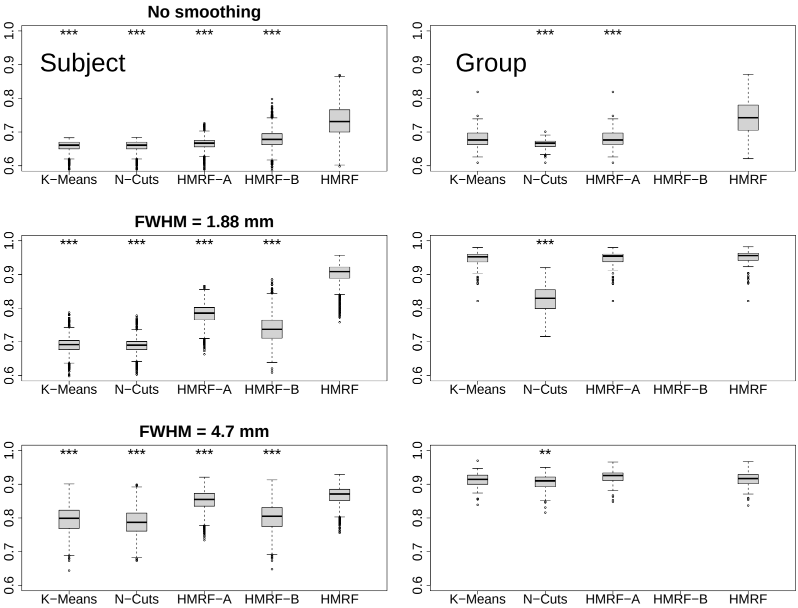
Sampling voxels in group map

Sampling voxels in subjects

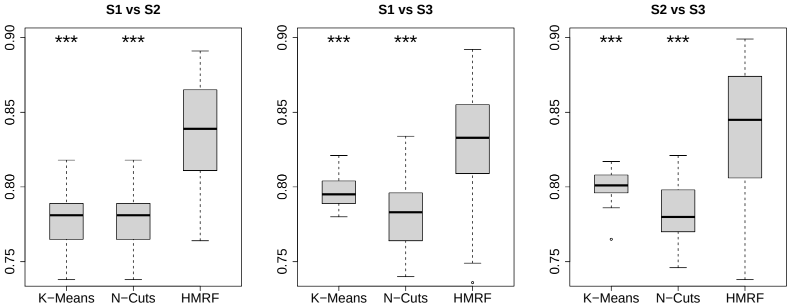
Synthetic Data: Estimation [Liu2013]



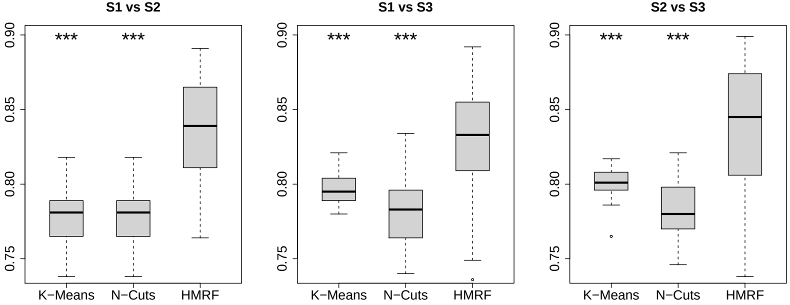
Synthetic Data: Monte Carlo Test

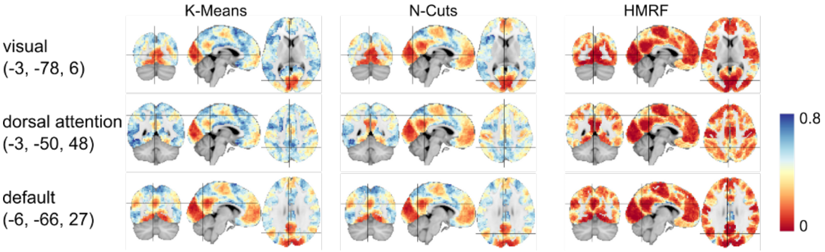


Cross-Session consistency

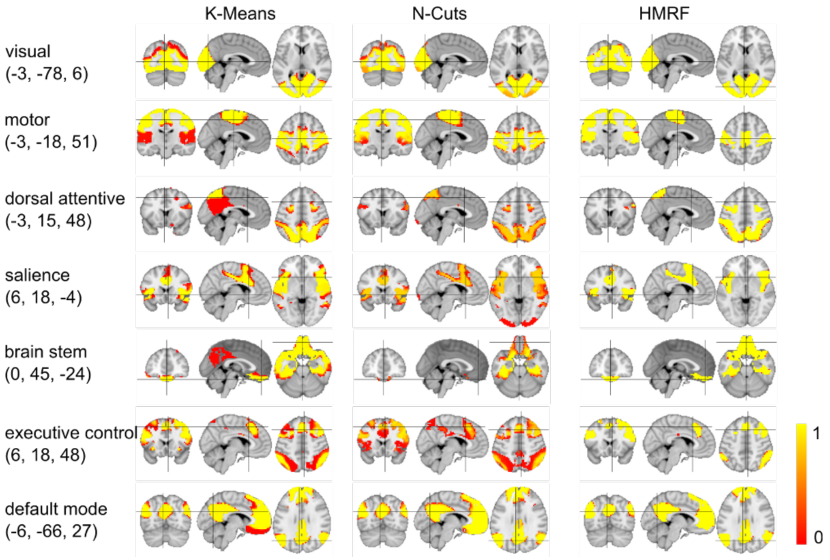


Cross-Session consistency

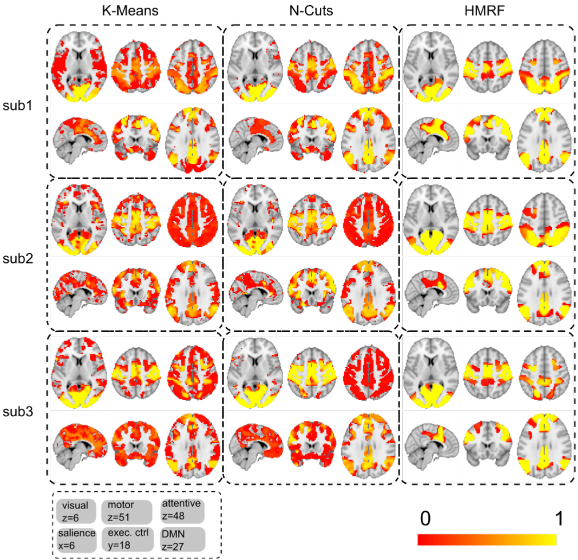




Bootstrapping: Group Mean Maps



Bootstrapping: Subject Mean Maps



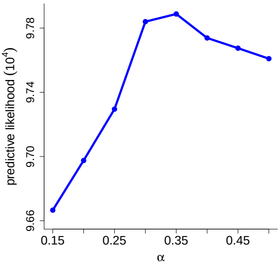
Inter-level Links Estimation

Bayesian Cross-Validation

$$\alpha = \operatorname{argmax} P(Y_t | Y; \alpha, \theta_t)$$

$$\begin{aligned} & P(Y_t | Y; \alpha, \theta_t) \\ & = \int P(Y_t | X_t; \theta_t) P(X_t | Y; \alpha) \, \mathrm{d}X_t \\ & \approx (1/M) \sum_m P(Y_t | X_t^m; \alpha, \theta_t), \\ & \quad X_t^m \sim P(X_t | Y; \alpha). \end{aligned}$$

X_t's are generated within MCEM.



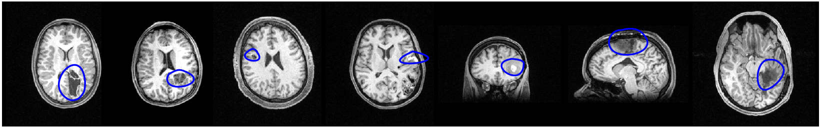
Conclusions

Summary of Contributions

- Pairwise connectivity estimation with MRF regularization.
- Segmentation of fMRI images for multiple network estimation.
- Group analysis by hierarchical MRF with consistency.
- Consistency of the hierarchical model.

Future Works

- Group analysis on subjects with large variation (neurosurgery).
- The correlation between functional connectivities (networks) and clinical variables.
- Dynamics of the functional networks.
- Spatio-temporal Gaussian-Markov model.



Publications

[Liu2010] **W. Liu**, P. Zhu, J. Anderson, D. Yurgelun-Todd, and P.T. Fletcher.
Spatial Regularization of Functional Connectivity using high-dimensional Markov random fields.
Medical Image Computing and Computer-Assisted Intervention–MICCAI 2010, pages 363–370, 2010.

[Liu2011] **W. Liu**, S. Awate, J. Anderson, D. Yurgelun-Todd, and P.T. Fletcher.
Monte Carlo expectation maximization with hidden Markov models to detect functional networks in resting-state fMRI.
Machine Learning in Medical Imaging, pages 59–66, 2011.

[Liu2012] **W. Liu**, S. Awate, and P.T. Fletcher.
Group analysis of resting-state fMRI by hierarchical Markov random fields.
Medical Image Computing and Computer-Assisted Intervention–MICCAI 2012, pages 189-196, 2012.

Under Revision/Review:

[Liu2013] **W. Liu**, S. Awate, J. Anderson, and P.T. Fletcher.
A Functional Networks Estimation Method of Resting-State fMRI Using a Hierarchical Markov Random Field.
NeuroImage, Under revision.

[Wang2013] Bo Wang, **Wei Liu**, Marcel Prastawa, Andrei Irimia, Paul M. Vespa, Johh D. Van Horn, P. Thomas Fletcher, Guido Gerig.
4D Active Cut: An Interactive Tool for Pathological Anatomy Modeling
IEEE International Symposium on Biomedical Imaging (2014), under review.

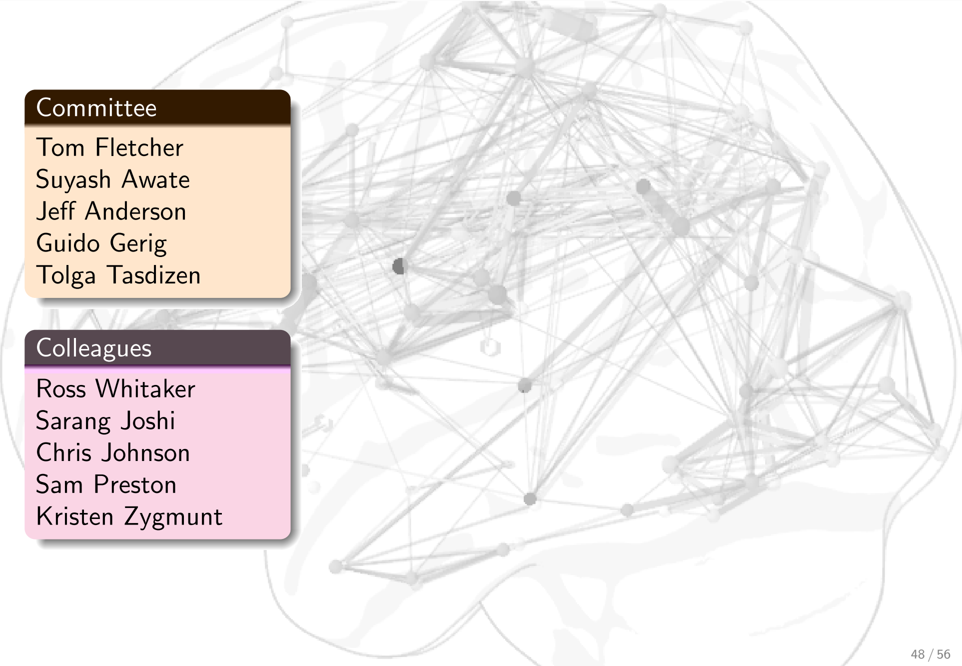
Acknowledgment

Committee

Tom Fletcher
Suyash Awate
Jeff Anderson
Guido Gerig
Tolga Tasdizen

Colleagues

Ross Whitaker
Sarang Joshi
Chris Johnson
Sam Preston
Kristen Zygmunt



Acknowledgment

Committee

Tom Fletcher
Suyash Awate
Jeff Anderson
Guido Gerig
Tolga Tasdizen

Colleagues

Ross Whitaker
Sarang Joshi
Chris Johnson
Sam Preston
Kristen Zygmunt

Fellow Students

Nikhil Singh
Gopal Veni
Bo Wang
Xiang Hao
Liang Zhou
Prasanna Venkatesh
Yen-Yun Yu

Acknowledgment

Committee

Tom Fletcher
Suyash Awate
Jeff Anderson
Guido Gerig
Tolga Tasdizen

Colleagues

Ross Whitaker
Sarang Joshi
Chris Johnson
Sam Preston
Kristen Zygmunt

Fellow Students

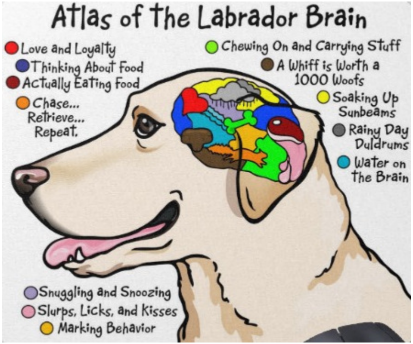
Nikhil Singh
Gopal Veni
Bo Wang
Xiang Hao
Liang Zhou
Prasanna Venkatesh
Yen-Yun Yu

Media, Admin. and Facility

Nathan Galli
Chris Pickett
Deb Zemek
Magali Coburn
Tony Portilo
Chris Bright
Ali Moeinvaziri

Dedication

Peihong Zhu



Thank you.
This is the end of the talk.

MRF and Gibbs Field

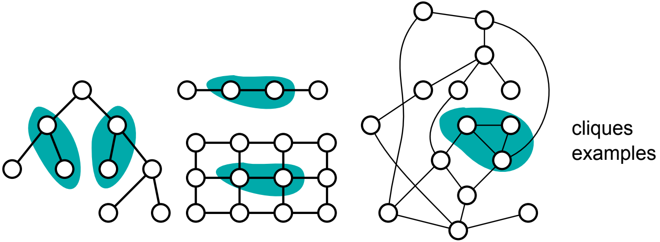
Theorem (Hammersley-Clifford, 1971)

X is an MRF on $\mathcal{G}$ if and only if X obeys Gibbs distribution in the following form

$$P(X) = \frac{1}{Z} \exp\left(-\frac{1}{T} U(X)\right),$$

$$U(X) = \sum_{c \in \mathcal{C}} V_c(X_c).$$

$Vc(X_c) = \beta \sum_{(r,s) \in \mathcal{V}} \psi(x_r, x_s)$: Ising, Potts model.



Inter-Subject Consistency

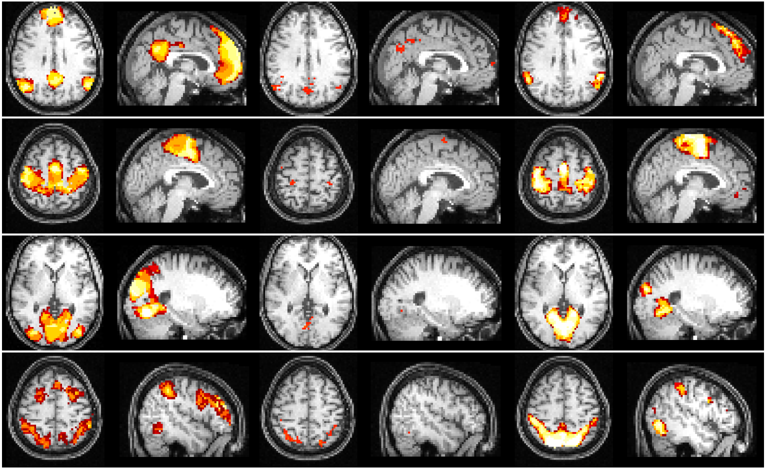


Figure : Comparison of the overlap of the label maps estimated by our MCEM approach, group ICA and single subject ICA on 16 subjects. Left: MCEM methods. Middle: single subject ICA. Right: group ICA. Color map ranges from 8 (red) 16 (yellow).

The Algorithm: MCEM Sampling on HMRF Model

Data: Normalized rs-fMRI, initial group label map
Result: MC samples of label maps $\{X^m, m = 1, \dots, M\}$, parameters $\{\beta, \mu, \sigma\}$

while $\mathbb{E}_{P(Y|X)}[\log P(Y, X; \theta)]$ *not converged* **do**
  repeat
    foreach $s \in \mathcal{V}_G$ **do** Draw sample of x_s from $P(x_s | x_{-s}, y_s; \theta)$;
    foreach $j = 1 \dots J$ **do**
      foreach $s \in \mathcal{V}_j$ **do** Draw sample of x_s from $P(x_s | x_{-s}, y_s; \theta)$;
    end
    Save sample X^m after B burn-ins;
  until $B + M$ *times*;
  foreach $l = 1 \cdots L$ **do**
    Estimate $\{\mu_l, \kappa_l\}$ by maximizing $(1/M) \sum_{m=1}^{M} \log P(Y | X^m; \theta)$;
  end
  Estimate β by maximizing $(1/M) \sum_{m=1}^{M} \log P(X^m; \theta)$;
end

Bootstrapping: Subject Variance Maps