Lecture 26: Time-Varying Dynamic Bayesian Network Learning for fMRI Data

Based on the joint work: L. Sun, A. Zhang, and F. Liang (2024), published in *Statistics in Medicine*.

Outline

fMRI and BOLD Signals

- Functional MRI (fMRI) measures neuronal activity indirectly via blood oxygenation-level dependent (BOLD) signal.
- Typical fMRI study:
  - Many time points, many regions of interest (ROIs).
  - Massive multivariate time series.

Brain Connectivity from fMRI

- **Nodes**: ROIs (brain regions).
- **Edges**: statistical dependence or causal influence.
- Two main types of connectivity:
  - **Functional connectivity**: undirected; temporal correlation/dependence.
  - **Effective connectivity**: directed; information flow, causal influence.
- Goal: learn effective connectivity networks from task-based fMRI data.

Challenges

- High dimensionality:
  - Hundreds of ROIs $\Rightarrow$ tens of thousands of potential edges.
- Time-varying structure:
  - Brain connectivity may change over the course of a task.
- Multi-subject data:
  - Many subjects, possibly missing time points.
- Existing approaches:
  - Vector Autoregressive Model, dynamic Bayesian networks, hidden Markov models.
  - Often assume stationarity or are not scalable in p or T.

Aim of the Paper

- Develop a method to learn **time-varying dynamic Bayesian networks** (DBNs) from high-dimensional, multi-subject fMRI data.
- Requirements:
  - Scalable in the number of ROIs p.
  - Allow time-varying connectivity.
  - Naturally handle multi-subject data.
  - Provide statistical guarantees (consistency).
- Application: emotion-processing task fMRI (Human Connectome Project).

Main Contributions

- Propose a two-stage method:
  1. Joint Gaussian graphical model (JGGM) estimation across time.
  2. Markov Neighborhood Regression (MNR) for dynamic DBN learning.
- Break DBN learning into a sequence of regression problems.
- Use graphical structure (Markov neighborhoods) to reduce dimension.
- Show consistency and good empirical performance.
- Apply to large-scale fMRI emotion-processing data; identify key roles for subcortical-cerebellum.

Task-Based fMRI Model

For subject i at time t:

$$\boldsymbol{Y}_t^{(i)} = \boldsymbol{\mu}^{(i)} + \sum_{k=1}^{K} \boldsymbol{W}_k(t) \circ \gamma_k^{(i)} + \boldsymbol{X}_t^{(i)},$$

- $\boldsymbol{Y}_t^{(i)} \in \mathbb{R}^p$: BOLD signal at p ROIs.
- $\boldsymbol{\mu}^{(i)}$: baseline mean.
- $\boldsymbol{W}_k(t) \in \mathbb{R}^p$: design vector for stimulus k (HRF-convolved).
- $\gamma_k^{(i)} \in \mathbb{R}^p$: subject-specific activation coefficients.
- $\boldsymbol{X}_t^{(i)} \in \mathbb{R}^p$: residual process containing effective connectivity.

Stimulus Modeling via HRF

- Each ROI-specific design component:

$$W_{rk}(t) = \int_0^t w_k(\tau) h_r(t-\tau)\, d\tau,$$

  where
  - $w_k(\tau)$: external stimulus function.
  - $h_r(\cdot)$: hemodynamic response function (HRF).
- Use canonical HRF, common in motor, visual, and emotion tasks.
- In practice:
  - First estimate activation $\gamma_k^{(i)}$ using GLM (Friston et al.).
  - Then regress out activation, and analyze residual $\boldsymbol{X}_t^{(i)}$.

Time-Varying VAR for Effective Connectivity

Residual process:

$$\boldsymbol{X}_t^{(i)} = \sum_{l=1}^{L} \boldsymbol{A}_{t,l} \boldsymbol{X}_{t-l}^{(i)} + \boldsymbol{e}_i(t),$$

- $\boldsymbol{A}_{t,l} \in \mathbb{R}^{p \times p}$: time-varying transition matrices.
- L: lag order (typically small, e.g., $L = 1$ or 2).
- $\boldsymbol{e}_i(t) \sim N_p(0, \Sigma)$, often Σ diagonal.
- Different components of $\boldsymbol{X}_t$ are conditionally independent given past, but generally dependent marginally.

Goal: infer time-varying directed edges encoded in $\boldsymbol{A}_{t,l}$.

Dynamic Bayesian Network Interpretation

- The time-varying VAR defines a dynamic directed acyclic graph (DAG) over time.
- Directed Markov property:

$$X_{t,v} \perp \boldsymbol{X}_{t,an(v)\setminus pa(v)} \mid \boldsymbol{X}_{t,pa(v)}$$

- Assume:
  - Directed Markov property.
  - Faithfulness
- Resulting dynamic DAG encodes effective connectivity between ROIs.

High-Dimensional Regression Setup

- Classical model:

$$Y = \beta_0 + X_1\beta_1 + \cdots + X_p\beta_p + \varepsilon,$$

  with $\varepsilon \sim N(0, \sigma^2)$, $\boldsymbol{X} \sim N_p(0, \Sigma)$.
- $p \gg n$: direct inference on β_j is challenging.
- MNR idea:
  - Use the **Markov blanket** (neighborhood) in the GGM for $\boldsymbol{X}$.
  - Reduce to a low-dimensional subset regression.

Markov Neighborhoods

- Let GGM for $\boldsymbol{X}$ be $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.
- Markov neighborhood of X_j:

$$\xi_j = \{k : e_{jk} = 1\},$$

such that

$$X_j \perp X_i \mid \boldsymbol{X}_{\xi_j}, \quad \forall i \notin \xi_j \cup \{j\}.$$

- Any superset containing ξ_j is a valid Markov neighborhood.
- Key fact:
  - Inference for β_j can be done using regression on $\{j\} \cup \xi_j \cup$ active set.

Subset Regression for β_j

- Under suitable screening conditions:

$$D_j = \{j\} \cup \hat{\xi}_j \cup \hat{\varpi}^*,$$

  where
  - $\hat{\varpi}^*$: estimated set of active variables.
  - $\hat{\xi}_j$: estimated Markov neighborhood of X_j.
- Run OLS on subset:

$$Y = \beta_0 + \boldsymbol{X}_{D_j}\boldsymbol{\beta}_{D_j} + \varepsilon.$$

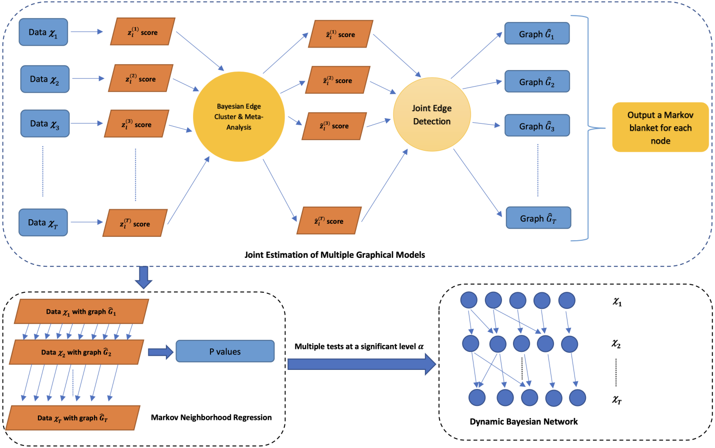
- Result: $\hat{\beta}_j$ has (approximate) standard low-dimensional asymptotics.

MNR Algorithm (Summary)

1. **Variable selection:**
   - Use SCAD, MCP, Lasso, or SIS to obtain $\hat{\varpi}^*$ containing all truly active variables.
2. **Graph estimation:**
   - Fit a GGM for $\boldsymbol{X}$ (e.g., ψ-learning, nodewise regression, graphical Lasso).
   - Extract $\hat{\xi}_j$ for each node j.
3. **Subset regressions:**
   - For each j, regress Y on $\boldsymbol{X}_{D_j}$.
   - Obtain estimates, standard errors, p-values, confidence intervals for β_j.

MNR turns one high-dimensional problem into many low-dimensional ones.

Overall Pipeline



- Stage 1: Joint estimation of multiple GGMs across time.
- Stage 2: MNR-based regression to construct dynamic DBN.

Data Structure

- Observations:

$$\{\boldsymbol{X}_t^{(i)} : t = 1, \ldots, T,\ i = 1, \ldots, n_t\},$$

  with possible missingness across t for each subject.
- At each time t:
  - $\mathcal{X}_t$: $n_t \times p$ matrix of ROIs.
  - Assume $\boldsymbol{X}_t^{(i)} \sim N_p(0, \Sigma_t)$.
- Goal Stage 1:
  - Jointly estimate $\mathcal{G}_t$ (GGM) for $t = 1, \ldots, T$.

Stage 1 (i): Edgewise Score Evaluation

- For each time t and node pair (i, j):
  1. Construct super-Markov blankets via SIS:
     - Reduce conditioning set size to $O(n_t / \log n_t)$.
  2. Conditional independence test:

$$X_{t,i} \perp X_{t,j} \mid \tilde{S}_{t,ij} \setminus \{i, j\}.$$

  3. Obtain p-values $p_{ij}^{(t)}$ and transform to z-scores:

$$z_{ij}^{(t)} = \Phi^{-1}(1 - p_{ij}^{(t)}).$$

- Complexity per time: $O(p^2)$.

Stage 1 (ii): Bayesian Data Integration

- For each edge l (pair of ROIs):
  - Edge status over time: $\boldsymbol{e}_l = (e_l^{(1)}, \ldots, e_l^{(T)})$.
  - Temporal prior:

$$P(\boldsymbol{e}_l \mid q) \propto q^{\#\text{no-change}}(1-q)^{\#\text{changes}},$$

    with $q \sim \text{Beta}(a_1, b_1)$.
  - z-scores modeled as mixture:

$$z_l^{(t)} \sim \begin{cases} N(\mu_{l0}, \sigma_{l0}^2), & e_l^{(t)} = 0, \\ N(\mu_{l1}, \sigma_{l1}^2), & e_l^{(t)} = 1. \end{cases}$$

- Use stochastic EM / IRO-type algorithm for clustering.

Stage 1 (iii): Integrated z-Scores and Edge Detection

- Given posterior samples of $\boldsymbol{e}_l$:
  - For each time t, compute Bayesian Stouffer integrated z-score $\hat{z}_l^{(t)}$ by meta-analysis across times in which $e_l^{(i)} = 0$ or 1.
- Multiple testing on $\hat{z}_l^{(t)}$:
  - Empirical Bayes mixture modeling of z-scores.
  - Control FDR via q-values (Storey, 2002).
- Output:
  - GGMs $\mathcal{G}_1, \ldots, \mathcal{G}_T$.
  - Markov blankets (neighborhoods) for each ROI at each time.

Stage 2: Dynamic DBN via MNR

For $t = 2, \ldots, T$ and each ROI j:

$$X_{t,j} = \beta_{t,j}^{(0)} + \sum_{k=1}^{p} \beta_{t,j}^{(k)} X_{t-1,k} + \varepsilon_{t,j}.$$

- Treat $X_{t,j}$ as response; predictors: $\boldsymbol{X}_{t-1}$.
- Use MNR:
  - Variable selection for relevant lagged ROIs.
  - Markov blankets from $\mathcal{G}_{t-1}$.
  - Subset regression to get p-values $q_{k,j}^{(t)}$.
- Transform $q_{k,j}^{(t)}$ to z-scores and apply joint multiple testing to decide directed edges from $X_{t-1,k}$ to $X_{t,j}$.

Tuning Parameters α_1 and α_2

- α_1:
  - Significance level in joint edge detection for GGMs.
  - Controls size of Markov neighborhoods.
  - Recommend relatively large (e.g., 0.1–0.2) to avoid missing neighbors.
- α_2:
  - Significance level in joint link detection for DBN edges.
  - Controls sparsity of dynamic DBNs.
  - Choose based on desired network density.
- α_1 and α_2 play roles analogous to regularization parameters.

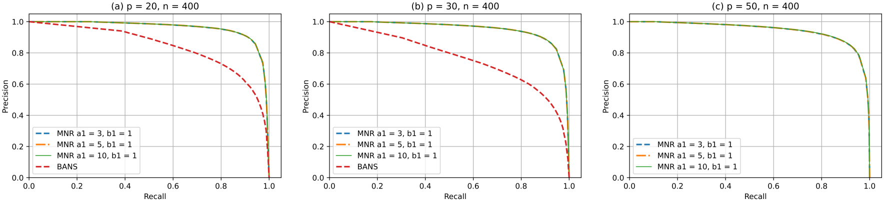
Computational Complexity and Parallelism

- Complexity roughly $O(p^2)$ (with $p \gg n$, $p \gg T$).
- Highly parallelizable:
  - Edgewise score evaluation: parallel over edges.
  - Bayesian edge clustering: parallel over edges.
  - MNR regressions: parallel over ROIs and times.
- Empirical Bayes mixture estimation can use stochastic gradient / minibatch.
- Practical implementation in R, parallelized on multi-core CPU.

Simulation Setup: Comparison with BANS

- Data mimic task-based fMRI:
  - $n = 400$ subjects, $T = 60$ time points.
  - $p = 20, 30, 50$ ROIs.
- Competing method:
  - BANS (Bayesian nodewise selection), designed for multi-layer GGM / DBN.
  - Regressions with spike-and-slab priors.
- Performance metrics:
  - Precision–recall curves.
  - AUC.
  - Computational time.

Precision–Recall Performance



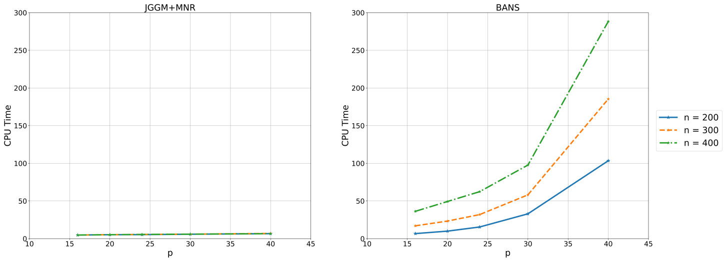
- JGGM+MNR consistently achieves higher AUC than BANS.
- For $p = 50$, BANS becomes prohibitively slow; omitted in plot.

AUC and Time Summary

p	Method	AUC	Time (min)
20	JGGM+MNR	≈ 0.959	≈ 5.1
	BANS	≈ 0.84	≈ 48
30	JGGM+MNR	≈ 0.952	≈ 5.8
	BANS	≈ 0.77	≈ 96
50	JGGM+MNR	≈ 0.945	≈ 7.7
	BANS	– (¿12 hrs)	$>$12h

- New method scales well in p; BANS does not.

Time Complexity vs. p and n



- JGGM+MNR: time grows mildly with p and almost flat in n.
- BANS: exponential-like growth in both p and n.

High-Dimensional and Large-Scale Cases

- Two scenarios:
  1. $n = 800,\ p = 300,\ T = 60.$
  2. $n = 400,\ p = 500,\ T = 60$ (small-n, large-p).
- Compare JGGM+MNR with:
  - Lasso.
  - Elastic Net.
  - MCP.
- Strategy:
  - Fit separate high-dimensional regressions for each ROI and time with regularization.
  - Compare edge recovery to JGGM+MNR.

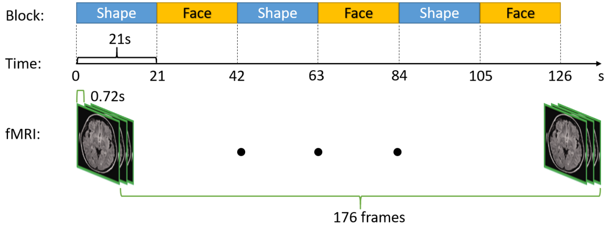
High-Dimensional Results (Summary)

- JGGM+MNR outperforms regularization methods in edge recovery (higher AUC / F1).
- Regularization methods struggle to capture subtle structure and often select too many or too few edges.
- MNR benefits from:
  - Graph-based dimension reduction.
  - Low-dimensional inference with proper uncertainty quantification.
- Supports feasibility of whole-brain time-varying DBN learning.

Human Connectome Project Emotion Task

- Data from HCP S1200 release.
- 867 subjects (22–35 years; 409 males, 458 females).
- Use task fMRI (emotion processing) with left-to-right encoding.
- Acquisition parameters:
  - TR/TE $= 720/33.1$ ms, multiband factor 8, voxel size $2 \times 2 \times 2$ mm.

Emotion Processing Task Design



- 6 blocks: 3 face-matching, 3 shape-matching.
- Each block: 21 s (6 trials, 2 s stimulus + 1 s ITI).
- Total duration: 2:16, 176 frames per subject.
- Faces show angry or fearful expressions.

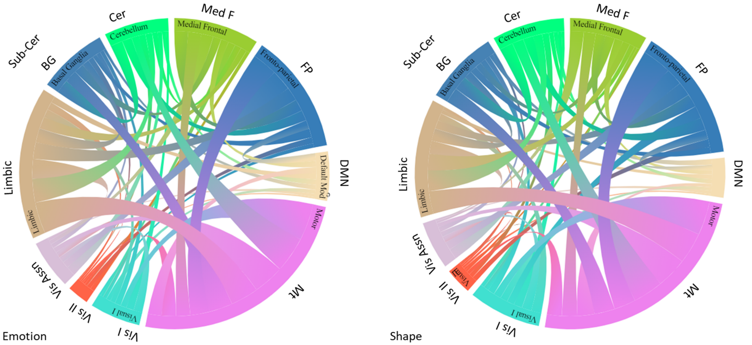
Preprocessing and ROI Definition

- HCP minimal preprocessing pipeline:
  - Distortion correction, motion correction, normalization to MNI, intensity normalization, etc.
- Additional steps:
  - Regress out 12 motion parameters (and derivatives).
  - Remove linear trends.
  - Band-pass filter (0.01–0.25 Hz).
- ROIs:
  - 268 ROIs from functional atlas (Finn et al., 2015).
  - Averaged voxel signals within each ROI.
  - Organized into 8 functional modules: Med F, FP, DMN, Sub-Cer, Motor, Vis I, Vis II, Vis Assn.

Parameter Settings for Analysis

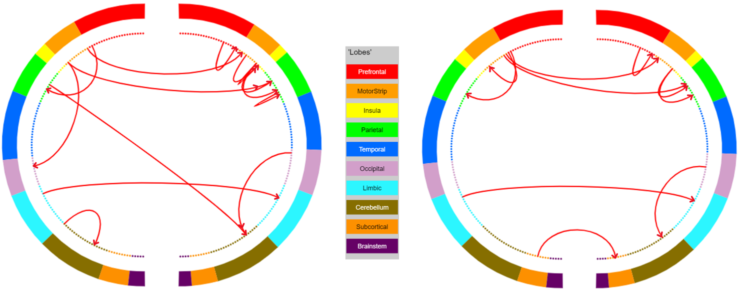
- JGGM stage:
  - Significance level $\alpha_1 = 0.2$ for multiple testing.
- DBN stage:
  - Main results: $\alpha_2 = 0.1$, lag order $L = 1$.
  - Sensitivity analysis: other α_1, α_2, and $L = 2$ in supplement.
- Output:
  - 175 dynamic DAGs (one per time transition) partitioned into emotion vs shape blocks.

Task-Related Networks: Chord Plots



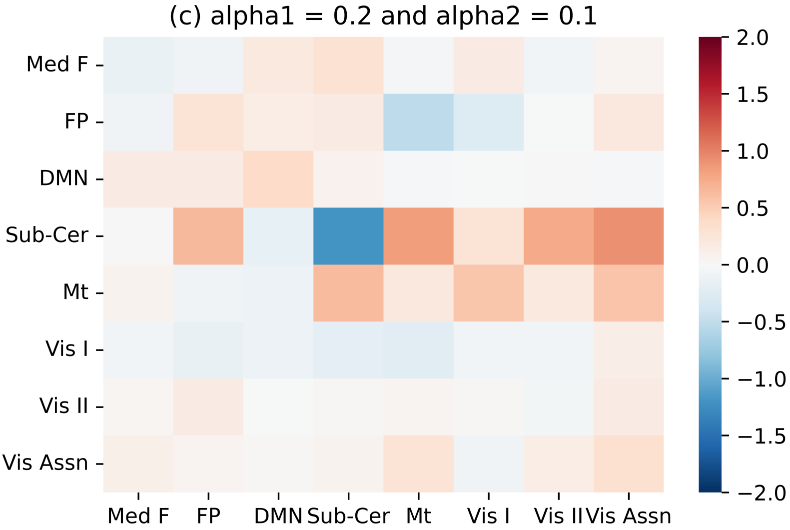
- Module-level chord plots for emotion and shape tasks.
- Overall similar structure, but different inter-module connectivity patterns.

Characteristic Edges



- Characteristic edges: edges appearing at least 2 SD above mean frequency.
- Emotion task: 14 characteristic edges.
- Shape task: 10 characteristic edges.
- Overlap: 6 common edges (e.g., bilateral motor strip, limbic connections).

Module-Level Differences



- Heat map: difference in mean edge degree (module-wise)

Role of Subcortical-Cerebellum

- Sub-Cer module includes subcortical structures and cerebellum.
- Emotion task:
  - Increased connectivity along:
    - Sub-Cer $\leftrightarrow$ Motor.
    - Sub-Cer $\rightarrow$ Vis II.
    - Sub-Cer $\rightarrow$ Vis Assn.
- Interpretation:
  - Consistent with known roles in emotion, autonomic regulation, and complex network processing.
  - Supports cerebellum's involvement in emotional face processing and recognition.

Identification of Emotion-Related Hubs

- Examine node degrees over time:
  - Row degree: number of outgoing edges from each ROI at each time.
- Compare emotion vs shape using Wilcoxon signed-rank test.
- ROI 32 stands out:
  - Significantly higher outgoing degree in emotion vs shape.
  - Belongs to Brodmann area 6 (BA6).
- Literature: BA6 implicated in facial emotion processing and evaluation.

Network Dynamics and Time Variation

- Use ψ-learning to test for differential edges across time:
  - Within emotion task (between consecutive transitions).
  - Within shape task.
  - Between emotion and shape tasks.
- At 10% significance:
  - Many differential edges within each task $\Rightarrow$ strong time variation.
  - Fewer differences detected for $L = 2$ than $L = 1$, suggesting higher-order Markov structure.
- Supports assumption of time-varying connectivity rather than stationarity.

Summary of fMRI Findings

- Emotion processing engages:
  - Stronger inter-module connectivity involving Sub-Cer, Motor, Vis II, Vis Assn.
  - Distinct dynamic patterns compared to shape processing.
- Subcortical-cerebellum:
  - Central hub coordinating motor and visual modules in emotion task.
- Hub ROI 32 (BA6):
  - Differentially connected in emotion vs shape, consistent with emotion-related function.
- Overall:
  - Method recovers biologically plausible and interpretable dynamic networks.

Methodological Advantages

- Scalable:
  - $O(p^2)$ complexity, parallelizable.
- Flexible:
  - Handles time-varying structure and multi-subject data.
  - Allows different n_t at different times.
- Statistically grounded:
  - Uses Markov blankets and MNR to leverage graph structure.
  - Consistency results under mild conditions.
- Uncertainty quantification:
  - z-scores for all possible edges.
  - Empirical Bayes FDR control and q-values.

Limitations and Extensions

- Current implementation assumes:
  - Gaussian data for GGMs and MNR.
  - Diagonal noise covariance Σ (with extension in remark).
- Extensions:
  - Mixed data: joint GGMs and MNR for mixed types (Sun & Liang, 2022).
  - Incorporate task information directly into graphical modeling and regressions.
  - Subject-specific weights via weighted MNR.

Conclusion

- Proposed a new framework (JGGM + MNR) for learning time-varying dynamic Bayesian networks from high-dimensional, multi-subject fMRI data.
- Demonstrated:
  - Superior performance to BANS and regularization methods in simulations.
  - Scalability to whole-brain networks.
  - Meaningful neuroscientific findings in an emotion processing task.
- Provides a general toolbox for spatio-temporal causal graph learning beyond fMRI.

Thank You

Questions?

- Paper: Sun, L., Zhang, A., and Liang, F. (2024). Time-varying dynamic Bayesian network learning for an fMRI study of emotion processing. *Statistics in Medicine*, 43 (14), 2713-2733.
- https://doi.org/10.1002/sim.10096