

Towards a Multi-Subject Analysis of Neural Connectivity

C. J. Oates¹, L. Costa¹, S. Mukherjee², T. Nichols¹,
J. Q. Smith¹, J. Cussens³

¹ Department of Statistics, University of Warwick, CV4 7AL, UK.

² MRC Biostatistics Unit and CRUK Cambridge Institute, University of Cambridge, CB2 0SR, UK.

³ Department of Computer Science and York Centre for Complex Systems Analysis, University of York, YO10 5GE, UK.

3rd September 2014

Workshop on Statistical Systems Biology

9th -11th December 2014

warwick.ac.uk/SSB

Keynote Speakers:

Prof. David Balding, Chair of the UCL Genetics Institute.

Dr. Clive Bowsher, School of Mathematics, Bristol.

Prof. Mustafa Khammash, Control Theory and Systems Biology, ETH Zurich.

Prof. Walter Kolch, Director of the Conway Institute and Systems Biology Ireland, University College Dublin.

Prof. John Lygeros, Head of the Automatic Control Laboratory, ETH Zurich.

Dr. Sach Mukherjee, Group Leader at the Netherlands Cancer Institute.

Prof. Simon Tavaré FRS, Director of Cancer Research UK Cambridge Institute.

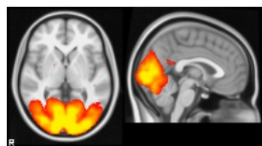
Prof. Darren Wilkinson, School of Mathematics and Statistics, Newcastle.

Call for Papers, Presentations and Posters

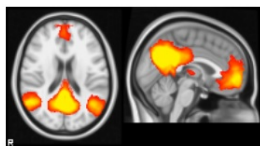
In partnership with *Statistical Applications in Genetics and Molecular Biology* (De Gruyter) we are soliciting high-quality research for joint journal submission and presentation at the workshop. In addition, we are encouraging the submission of contributed presentations and posters in relevant research areas.



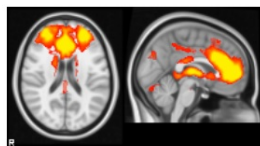
Motivation: Uncovering neural connectivity



(a) Visual



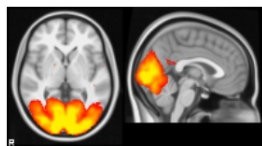
(b) DMN



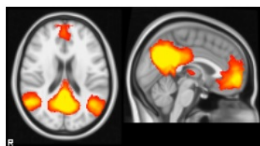
(c) Executive Control

How are these brain regions interacting?

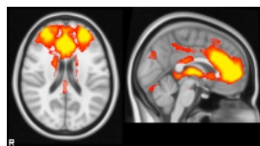
Motivation: Uncovering neural connectivity



(a) Visual



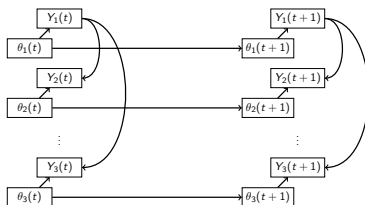
(b) DMN



(c) Executive Control

How are these brain regions interacting?

Multiregression
dynamical models
(MDMs; Catriona
Queen and Jim Smith,
1993):



Motivation: Uncovering neural connectivity

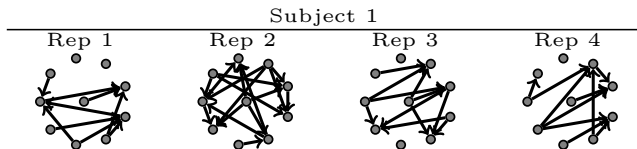


Figure: fMRI data; replicate data from the same subject. DAGs estimated from time series data using MDMs.

<u>Node Number</u>	<u>Symmetry</u>	<u>Summary</u>
1	Bilateral	Motor:hand/face
2	Bilateral	Sensory:All-but-face
3	Bilateral	Motor:All-but-face
4	Bilateral	UNKNOWN
5	Left Dominant	Sensorimotor: L Hand+Arms
6	Right Dominant	Sensorimotor: R Hand+Arms
7	Bilateral	Sensory: Trunk-to-feet
8	Bilateral	Sensory: Face
9	Bilateral	Auditory
10	Bilateral	Sensorimotor:All-but-face - Sensory:Face

An ideal algorithm

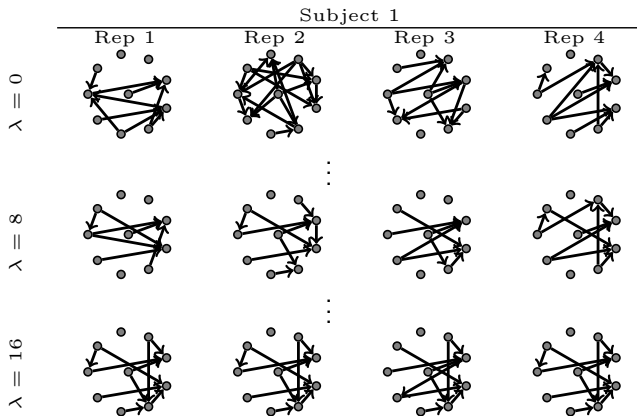


Figure: fMRI data; joint learning of all DAGs simultaneously. [λ is a “regularity” parameter.]

An ideal algorithm

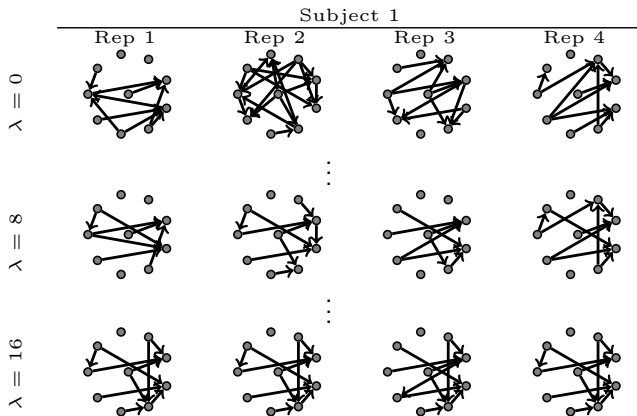


Figure: fMRI data; joint learning of all DAGs simultaneously. [λ is a “regularity” parameter.]

But how might this work? Seems challenging...

Joint statistical model for multiple DAGs

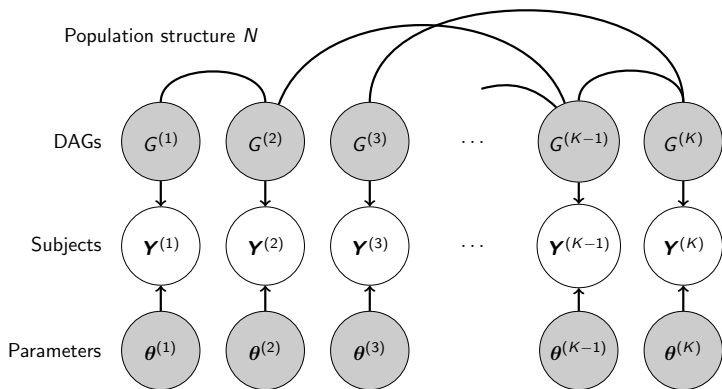


Figure: A Bayesian hierarchical model for multiple DAGs.

Joint statistical model for multiple DAGs

Joint prior over multiple DAGs:

$$p(G^{(1:K)}|N) \propto \underbrace{\left(\prod_{(k,l) \in N} r(G^{(k)}, G^{(l)}) \right)}_{\text{regularity}} \times \underbrace{\left(\prod_{k=1}^K m(G^{(k)}) \right)}_{\text{multiplicity correction}}$$

Joint statistical model for multiple DAGs

Joint prior over multiple DAGs:

$$p(G^{(1:K)} | N) \propto \underbrace{\left(\prod_{(k,l) \in N} r(G^{(k)}, G^{(l)}) \right)}_{\text{regularity}} \times \underbrace{\left(\prod_{k=1}^K m(G^{(k)}) \right)}_{\text{multiplicity correction}}$$

Structural Hamming distance:

$$\log(r(G^{(k)}, G^{(l)})) = -\lambda \sum_{(j,i)} \mathbb{I}\{j \in G_i^{(k)} \Delta G_i^{(l)}\}.$$

Joint statistical model for multiple DAGs

Joint prior over multiple DAGs:

$$p(G^{(1:K)} | N) \propto \underbrace{\left(\prod_{(k,l) \in N} r(G^{(k)}, G^{(l)}) \right)}_{\text{regularity}} \times \underbrace{\left(\prod_{k=1}^K m(G^{(k)}) \right)}_{\text{multiplicity correction}}$$

Structural Hamming distance:

$$\log(r(G^{(k)}, G^{(l)})) = -\lambda \sum_{(j,i)} \mathbb{I}\{j \in G_i^{(k)} \Delta G_i^{(l)}\}.$$

Binomial correction:

$$m(G^{(k)}) = \prod_{i=1}^P \binom{P}{|G_i^{(k)}|}^{-1} \mathbb{I}\{|G_i^{(k)}| \leq d_{\max}\}.$$

Joint statistical model for multiple DAGs

Joint prior over DAGs and the network N :

$$p(G^{(1:K)}, N) \propto p(G^{(1:K)}|N)p(N)$$

where η controls the density of the network N and

$$\log(p(N)) \stackrel{+C}{=} \eta \|N\|.$$

Joint statistical model for multiple DAGs

Joint prior over DAGs and the network N :

$$p(G^{(1:K)}, N) \propto p(G^{(1:K)}|N)p(N)$$

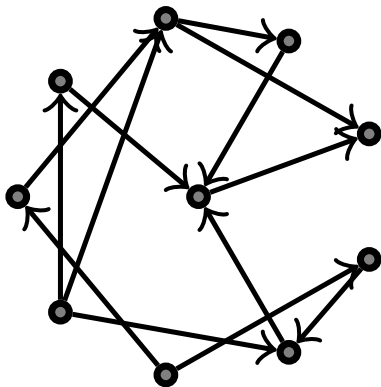
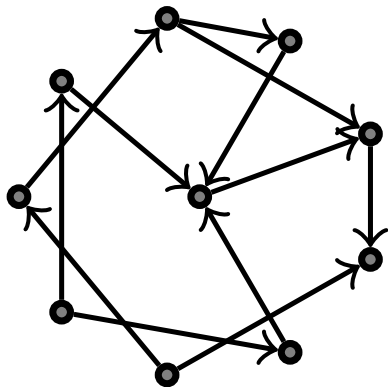
where η controls the density of the network N and

$$\log(p(N)) \stackrel{+C}{=} \eta \|N\|.$$

Then interest is in the “doubly joint” MAP

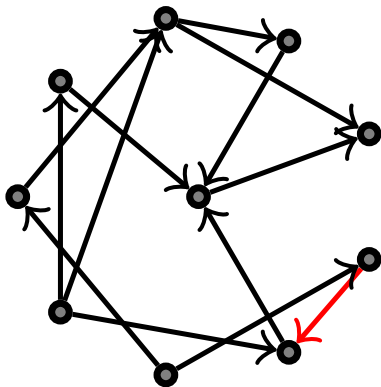
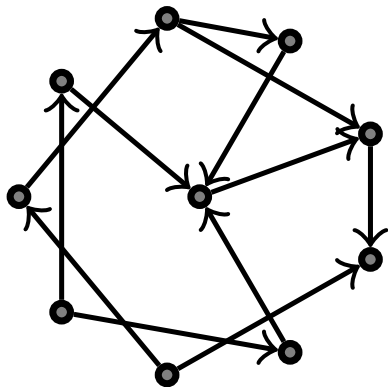
$$(\hat{G}^{(1:K)}, \hat{N}) := \arg \max_{G^{(1:K)}, N} p(\mathbf{Y}^{(1:K)}|G^{(1:K)}, N)p(G^{(1:K)}, N).$$

Why are multiple DAGs challenging? Acyclicity.



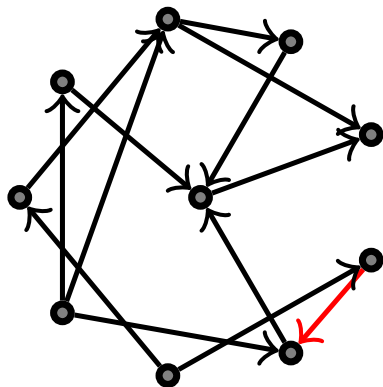
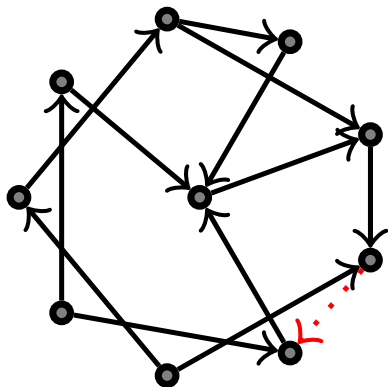
Design a local move that encourages more similar DAGs...

Why are multiple DAGs challenging? Acyclicity.



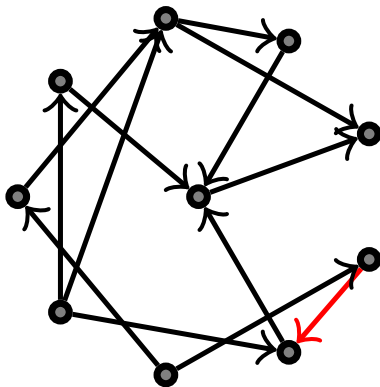
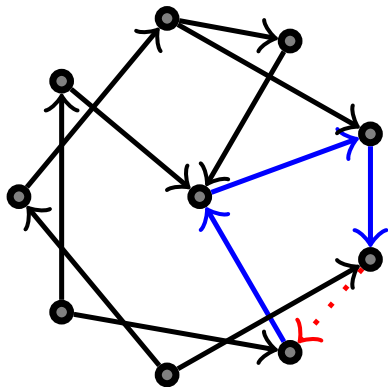
Pick an edge on which the two DAGs differ.

Why are multiple DAGs challenging? Acyclicity.



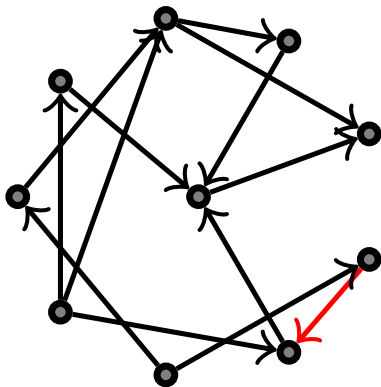
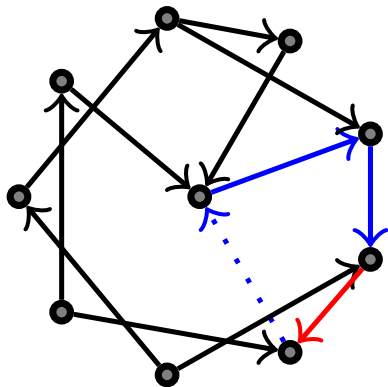
Propose to add/remove this edge to the other DAG.

Why are multiple DAGs challenging? Acyclicity.



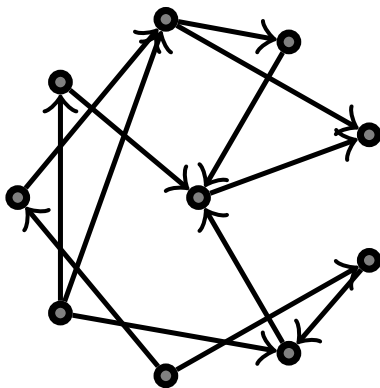
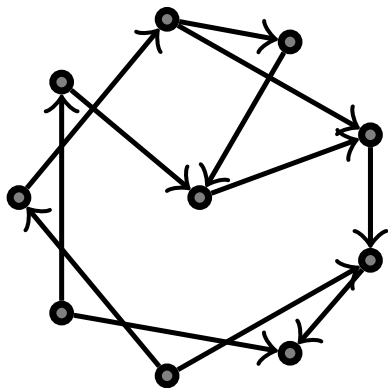
Check no cycles are created.

Why are multiple DAGs challenging? Acyclicity.



Delete an edge from each cycle.

Why are multiple DAGs challenging? Acyclicity.



But these new DAGs are as different as when we started!

Clearly a different approach is needed.

Integer linear programs for MAP DAGs

Inference for single DAGs is “easy”:

Integer linear programs for MAP DAGs

Inference for single DAGs is “easy”:

Consider a Bayesian network \mathbf{Y} with respect to a directed acyclic graph (DAG) model G . i.e.

$$p_{\mathbf{Y}}(\mathbf{y}|G) = \prod_{i=1}^P p(\mathbf{y}_p | \mathbf{y}_{G_i}, G_i).$$

Integer linear programs for MAP DAGs

Inference for single DAGs is “easy”:

Consider a Bayesian network \mathbf{Y} with respect to a directed acyclic graph (DAG) model G . i.e.

$$p_{\mathbf{Y}}(\mathbf{y}|G) = \prod_{i=1}^P p(\mathbf{y}_p | \mathbf{y}_{G_i}, G_i).$$

There is interest in the *maximum a posteriori* (MAP) estimate

$$\hat{G} := \arg \max_G p_{\mathbf{Y}}(\mathbf{y}|G)p(G).$$

Integer linear programs for MAP DAGs

Inference for single DAGs is “easy”:

Consider a Bayesian network \mathbf{Y} with respect to a directed acyclic graph (DAG) model G . i.e.

$$p_{\mathbf{Y}}(\mathbf{y}|G) = \prod_{i=1}^P p(\mathbf{y}_p | \mathbf{y}_{G_i}, G_i).$$

There is interest in the *maximum a posteriori* (MAP) estimate

$$\hat{G} := \arg \max_G p_{\mathbf{Y}}(\mathbf{y}|G)p(G).$$

Choose a “nice” $p(G) = \prod_{i=1}^P p_{G_i}(G_i)$.

Integer linear programs for MAP DAGs

Cussens '10 and Jaakola *et al.* '10 cast the MAP estimator in a DAG model as an integer linear program (ILP):

$$\max \mathbf{f}^T \mathbf{x} \quad \text{subject to} \quad \mathbf{A}\mathbf{x} \leq \mathbf{b}, \quad \mathbf{C}\mathbf{x} = \mathbf{d}, \quad \mathbf{x} \in \mathbb{Z}^d$$

Integer linear programs for MAP DAGs

Cussens '10 and Jaakola *et al.* '10 cast the MAP estimator in a DAG model as an integer linear program (ILP):

$$\max \mathbf{f}^T \mathbf{x} \quad \text{subject to} \quad \mathbf{A}\mathbf{x} \leq \mathbf{b}, \quad \mathbf{C}\mathbf{x} = \mathbf{d}, \quad \mathbf{x} \in \mathbb{Z}^d$$

Objective function:

$$\begin{aligned} \mathbf{f}^T \mathbf{x} = \log[p_{\mathbf{Y}}(\mathbf{y}|G)p(G)] &= \sum_{i=1}^P \log[p(\mathbf{y}_i|\mathbf{y}_{G_i}, G_i)p_{G_i}(G_i)] \\ &= \sum_{i=1}^P \sum_{\pi \subseteq \{1:P\}} \log[p(\mathbf{y}_i|\mathbf{y}_{\pi}, \pi)p_{G_i}(\pi)]x_{i,\pi} \end{aligned}$$

where $x_{i,\pi} = \mathbb{I}\{G_i = \pi\}$ and e.g. $\mathbf{x} = (0, 0, 1, 0, 0, 0, \dots, 1, 0, 0)$.

Integer linear programs for MAP DAGs

Cussens '10 and Jaakola *et al.* '10 cast the MAP estimator in a DAG model as an integer linear program (ILP):

$$\max \mathbf{f}^T \mathbf{x} \quad \text{subject to} \quad \mathbf{A}\mathbf{x} \leq \mathbf{b}, \quad \mathbf{C}\mathbf{x} = \mathbf{d}, \quad \mathbf{x} \in \mathbb{Z}^d$$

Objective function:

$$\begin{aligned} \mathbf{f}^T \mathbf{x} = \log[p_{\mathbf{Y}}(\mathbf{y}|G)p(G)] &= \sum_{i=1}^P \log[p(\mathbf{y}_i|\mathbf{y}_{G_i}, G_i)p_{G_i}(G_i)] \\ &= \sum_{i=1}^P \sum_{\pi \subseteq \{1:P\}} \log[p(\mathbf{y}_i|\mathbf{y}_{\pi}, \pi)p_{G_i}(\pi)]x_{i,\pi} \end{aligned}$$

where $x_{i,\pi} = \mathbb{I}\{G_i = \pi\}$ and e.g. $\mathbf{x} = (0, 0, 1, 0, 0, 0, \dots, 1, 0, 0)$.

Q: How to ensure \mathbf{x} corresponds to a well-defined DAG?

Integer linear programs for MAP DAGs

Convexity:

$$\sum_{\pi \subseteq \{1:P\}} x_{i,\pi} = 1 \quad \forall i \in \{1:P\}$$

No self-loops:

$$x_{i,\pi} = 0 \quad \forall \pi \in i$$

Acyclicity (version of Jaakola *et al.*, '10):

$$\sum_{i \in C} \sum_{\substack{\pi \subseteq \{1:P\} \\ \pi \cap C = \emptyset}} x_{i,\pi} \geq 1 \quad \forall \emptyset \neq C \subseteq \{1:P\}.$$

These constraints together exactly characterise the space of DAGs.

Integer linear programs for multiple DAGs

Claim: $(\hat{G}^{(1:K)}, \hat{N})$ is characterised by an extended ILP.

Integer linear programs for multiple DAGs

Claim: $(\hat{G}^{(1:K)}, \hat{N})$ is characterised by an extended ILP.

Sketch: Need to encode N in the state vector \mathbf{x} :

$$E^{(k,l)} := \mathbb{I}\{(k, l) \in N\} \quad \forall k, l \text{ with } k < l$$

Integer linear programs for multiple DAGs

Claim: $(\hat{G}^{(1:K)}, \hat{N})$ is characterised by an extended ILP.

Sketch: Need to encode N in the state vector \mathbf{x} :

$$E^{(k,l)} := \mathbb{I}\{(k, l) \in N\} \quad \forall k, l \text{ with } k < l$$

Now construct binary indicators of individual edges and disagreement between edges:

$$e_{j,i}^{(k)} = \sum_{\substack{\pi \subseteq \{1:P\} \\ j \in \pi}} x_{i,\pi}^{(k)} \quad \forall i, j, k.$$

$$d_{j,i}^{(k,l)} = \mathbb{I}\{j \in G_i^{(k)} \Delta G_i^{(l)}\} \quad \forall i, j, k, l \text{ with } k < l$$

$$D_{j,i}^{(k,l)} = \mathbb{I}\{j \in G_i^{(k)} \Delta G_i^{(l)} \text{ and } (k, l) \in N\} \quad \forall i, j, k, l.$$

Integer linear programs for multiple DAGs

Claim: $(\hat{G}^{(1:K)}, \hat{N})$ is characterised by an extended ILP.

Sketch: Need to encode N in the state vector \mathbf{x} :

$$E^{(k,l)} := \mathbb{I}\{(k, l) \in N\} \quad \forall k, l \text{ with } k < l$$

Now construct binary indicators of individual edges and disagreement between edges:

$$e_{j,i}^{(k)} = \sum_{\substack{\pi \subseteq \{1:P\} \\ j \in \pi}} x_{i,\pi}^{(k)} \quad \forall i, j, k.$$

$$d_{j,i}^{(k,l)} = \mathbb{I}\{j \in G_i^{(k)} \Delta G_i^{(l)}\} \quad \forall i, j, k, l \text{ with } k < l$$

$$D_{j,i}^{(k,l)} = \mathbb{I}\{j \in G_i^{(k)} \Delta G_i^{(l)} \text{ and } (k, l) \in N\} \quad \forall i, j, k, l.$$

XOR and AND constraints \equiv integer linear inequalities.

Example: AND constraint $A = \text{AND}(B, C)$

$$\begin{array}{rcccccc} +A & & -B & & \leq & 0 \\ +A & & & & -C & \leq & 0 \\ -A & & +B & & +C & \leq & 1 \end{array}$$

This linearisation of AND is **optimal**, in the sense that it describes all facets of the convex hull of feasible solutions for the AND constraint.

Some preliminary results...

Results: Simulation study

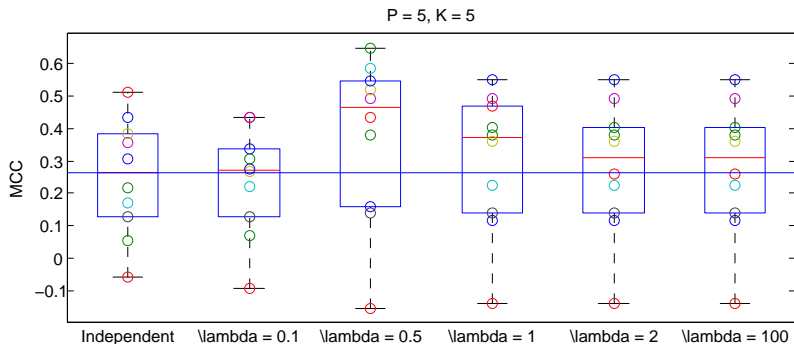


Figure: Simulated data; fixed $N = \text{complete}$, varying λ . [MCC = Matthews' correlation coefficient.]

Intuition: A modest amount of regularisation should help, but too much can lead to artefacts.

Results: Group analysis of fMRI data

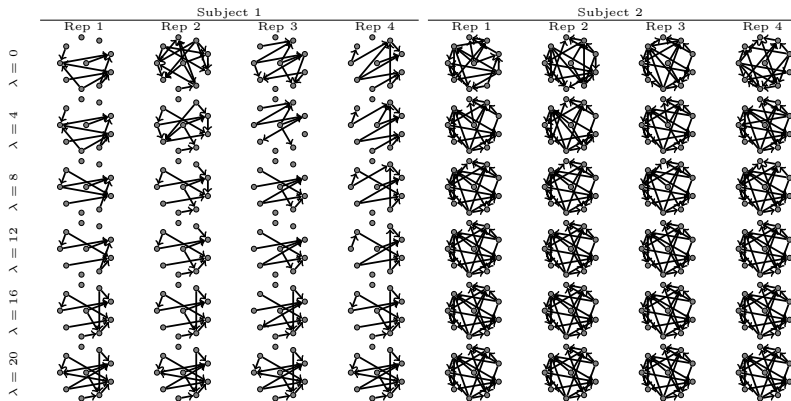


Figure: fMRI data on two subjects; learning λ .

Results: Group analysis of fMRI data

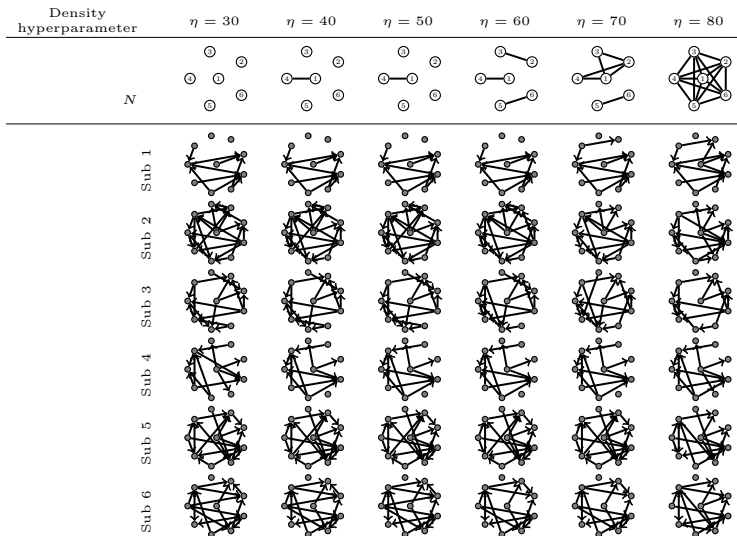


Figure: fMRI data on six subjects; learning N . [$\lambda = 4$]

Summary

To do:

- ▶ Large-scale empirical study
- ▶ Informative group priors (e.g. based on demographic covariates or genealogy)
- ▶ Causal semantics for transfer learning

References:

- ▶ CJO, Lilia Costa, Tom Nichols (2014) *Towards a Multi-Subject Analysis of Neural Connectivity*.

To appear in *Neural Computation*.

arXiv:1404.1239.

- ▶ CJO, Jim Smith, Sach Mukherjee, James Cussens (2014) *Exact Estimation of Multiple Directed Acyclic Graphs*.

arXiv:1404.1238.