Towards a Multi-Subject Analysis of Neural Connectivity

C. J. Oates[1,*], **L. Costa**[1,†], and **T. E. Nichols**[1,2,‡]

[1]Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

[2]Warwick Manufacturing Group, University of Warwick, Coventry, CV4 7AL, UK

Abstract

Directed acyclic graphs (DAGs) and associated probability models are widely used to model neural connectivity and communication channels. In many experiments, data are collected from multiple subjects whose connectivities may differ but are likely to share many features. In such circumstances it is natural to leverage similarity between subjects to improve statistical efficiency. The first exact algorithm for estimation of multiple related DAGs was recently proposed by Oates *et al.* (2014); in this letter we present examples and discuss implications of the methodology as applied to the analysis of fMRI data from a multi-subject experiment. Elicitation of tuning parameters requires care and we illustrate how this may proceed retrospectively based on technical replicate data. In addition to joint learning of subject-specific connectivity, we allow for heterogeneous collections of subjects and simultaneously estimate relationships between the subjects themselves. This letter aims to highlight the potential for exact estimation in the multi-subject setting.

1 Introduction

Probabilistic graphical models are widely used to model neural connectivity and the transfer of information between regions of the brain (Poldrack *et al.*, 2011). In brief, vertices indexed by 1, …, P in a directed acyclic graph (DAG) G are identified with random variables Y_i that represent neural activity at a particular region and edges between the vertices describe conditional independence statements, whose interpretation depends on both the underlying statistical model for the data and the context in which data are obtained. In many neuroscience applications, subject-specific connectivity (i.e. the set of edges) itself is uncertain and an important challenge is to infer this structure from experimental data (Friston, 2011). There has been considerable statistical research into inference for graphical models in general over the last decade, with particular emphasis on Bayesian networks (BNs; Chickering, 2003; Friedman and Koller, 2003; Ellis and Wong, 2008), Gaussian graphical models (GGMs; Meinshausen and Bühlmann, 2006; Chandrasekaran *et al.*, 2012) and discrete graphical models (Loh and Wainwright, 2013). Nevertheless there remain two substantive barriers to the inference of graphical models from data: Firstly, inferred graphical structure is often not robust to reasonable perturbation of the underlying data

*c.oates@warwick.ac.uk
†l.c.carneiro-da-costa@warwick.ac.uk
‡t.e.nichols@warwick.ac.uk

(Claassen and Heskes, 2012). This is due to a combination of the high variance of graphical estimators themselves and any additional variance that is introduced if the structure learning algorithm returns only an approximation to the intended estimator. Secondly, conventional model selection criteria for graphical models are often biased towards selecting more complex models (i.e. more edges), since there are typically very many models in which the data-generating model is nested; these models are also able to fit the data well (albeit with some coefficients close or equal to zero; Consonni and La Rocca, 2010). Consequently many more data are required to exclude more complex alternatives. Taken together, these factors limit the extent to which neural connectivity can be accurately recovered from data.

Many experimental designs in neuroscience involve data collected on multiple subjects, indexed by $1, \dots, K$, that may differ with respect to neural connectivity, such that corresponding graphs $G^{(k)}$ may be subject-specific (Sugihara *et al.*, 2006; Li *et al.*, 2008). Efforts to analyse multi-subject experimental data have previously focussed on hierarchical models and imaging data, rather then connectivity *per se* (Mumford and Nichols, 2009; Sanyal and Ferreira, 2012; Badillo *et al.*, 2013; Marquand *et al.*, 2014). Given that elements of neural architecture are largely conserved between subjects, it is natural to leverage this similarity in order to improve statistical efficiency, by addressing both the robustness of inferred graphical structure and reducing small sample bias (Mechelli *et al.*, 2002). The statistical challenge of estimating multiple related graphical models has recently received much attention: For GGMs, Danaher *et al.* (2014) and others exploited L_1 penalties, such as the fused graphical LASSO, to couple together inference for multiple related subjects. Such penalised likelihood methods are computationally tractable and scale well to high dimensions. These studies demonstrate that it is possible to increase statistical efficiency, often considerably, by formulating an appropriate joint model that couples together multiple graphs. Likewise, the methodology improves robustness by requiring that graphical structure is approximately invariant to perturbations of the data that are, in effect, provided by the subjects themselves.

Whilst useful in many applications, GGMs are undirected graphs and hence cannot not represent the direction of information flow between neural regions. More fundamentally, GGMs do permit causal inference that is typically the scientific objective (Valdes-Sosa *et al.*, 2011). For this reason we focus attention on graphical models, such as BNs, that are based on DAGs and have an associated theory of inferred causation (Pearl, 2009). Research focussing on DAGs in this setting includes Ramsey *et al.* (2009), who constructed a hierarchical model in which graph structure was conserved between subjects but the parameters that describe the data-generating process were subject-specific. Waldorp *et al.* (2011) went further by permitting subject-specific graph structure and parameters in the context of Gaussian ancestral graphs whose parameters are constrained by a hierarchical model. This latter work is closest in spirit to the methodology that we discuss below, but we do not restrict attention to either stationary data or Gaussian data, rendering our approach considerably more flexible.

Until very recently, estimation of more general DAGs required either the strong assumption that an ordering of the variables $1, \dots, P$ applied equally to all subjects (Oyen and Lane, 2013), or the use of expensive computational approximations such as Markov chain Monte

Carlo that scale extremely poorly as either the number *P* of variables or number *K* of subjects grows (Werhli and Husmeier, 2008). An exact algorithm that facilitates the joint estimation of multiple DAGs was recently developed in the sister paper Oates *et al.* (2014), viewing the estimation problem within a hierarchical Bayesian framework (somewhat similar to a random effects model for the graph structure) and applying advanced techniques from integer linear programming to obtain a *maximum a posteriori* estimate of all DAGs simultaneously. The availability of exact algorithms offers the opportunity to analyse multi-subject neural connectivity using causal DAG models, whilst leveraging the similarity between subjects in order to improve statistical efficiency and robustness. This letter illustrates the scope and applicability of these exact algorithms within neuroscience, using a small functional magnetic resonance imaging (fMRI) time course dataset obtained on six subjects, coupled with multiregression dynamical models (MDMs; Queen and Smith, 1993) that permit statistically rigorous causal inference (Queen and Albers, 2009). It is envisaged that exact algorithms will play an important rôle in future studies of neural connectivity and this letter serves to illustrate their application by example.

2 Results

2.1 fMRI data and experimental setup

Exact algorithms are illustrated here with a small fMRI dataset consisting of six subjects from the Human Connectome Project (Van Essen *et al.*, 2013). Scans were acquired on each subject while they were in a state of quiet repose; data from one 15 minute session were used, with a spatial resolution of $2 \times 2 \times 2$ mm and a temporal resolution of 0.7 secs; see Smith *et al.* (2013) for full details. After correcting for head motion, all data was registered to a common reference atlas space and 100-dimensional independent component analysis (ICA) was conducted on the temporally concatenated data. The result of this ICA was 100 spatial modes (common to all subjects) and 100 corresponding temporal modes (subject-specific); at this high dimension, the 100 spatial modes are sparse and spatially compact (though possibly bilaterally symmetric) and so essentially provide a data-driven parcellation of the brain. Hierarchical clustering was used on the ICA temporal modes following Beckmann and Smith (2002), and the 10-mode cluster corresponding to motor cortex was selected for study here. Thus our data consists of 10 nodes, with a time series for each node for each subject. Figure 1a displays the neural regions that we consider and Figure 1b shows the approximate description of each region; note that region 4 was spatially diffuse and difficult to characterise, and thus is likely to be an artefactual component.

The goal here is to understand neural information transfer at resting state and establish subject-specific connectivity. By its very nature, estimation of resting state connectivity is challenging due to limited information content in the fMRI time series. Indeed, Smith *et al.* (2011) reported that whilst the presence or absence of connections can sometimes be estimated from fMRI time series data, estimating the direction of edges from data remains extremely challenging. The integration of data from multiple related subjects offers one route to increased statistical power and this is the approach that we pursue here.

2.2 MDMs for fMRI time series data

Following data preprocessing we are left with a collection of random variables $Y_i^{(k)}(n)$ representing the observed activity in subject k at region i and time point n. Following recent research by Costa *et al.* (2013a) into causal inference based on such fMRI time course data, we model the $Y_i^{(k)}(n)$ as arising from a causal MDM. Specifically, an MDM is defined on a multivariate time series $\boldsymbol{Y}^{(k)}$ is characterised by a contemporaneous DAG $G^{(k)}$, with information shared across time through evolution of the model parameters $\boldsymbol{\theta}_i^{(k)}(n)$. We consider the case where $\boldsymbol{Y}^{(k)}(n)|\boldsymbol{\theta}^{(k)}(n)$ satisfies linear Gaussian structural equations, though any formulation would be compatible with the methodology that we present. Write $G_i^{(k)} \subseteq \{1, \ldots, P\}\backslash\{i\}$ for the parents of vertex i in the DAG $G^{(k)}$ and write $\boldsymbol{Y}_S^{(k)}$ for the collection of univariate time series corresponding to the variables $\{Y_i^{(k)}: i \in S\}$. This MDM is described by the following observation equations

$$Y_i^{(k)}(n) = \boldsymbol{Y}_{G_i^{(k)}}^{(k)}(n)^T \boldsymbol{\theta}_i^{(k)}(n) + \varepsilon_i^{(k)}(n) \quad (1)$$

where $\varepsilon_i^{(k)}(n) \sim N(0, V_i^{(k)}(n))$, together with the system equations

$$\boldsymbol{\theta}^{(k)}(n) = \boldsymbol{\Gamma}^{(k)}(n)\boldsymbol{\theta}^{(k)}(n-1) + \mathbf{w}^{(k)}(n) \quad (2)$$

where $\boldsymbol{\Gamma}^{(k)}(n)$ is a matrix of autoregressive coefficients and $\mathbf{w}^{(k)}(n) \sim N(\mathbf{0}, \boldsymbol{W}^{(k)}(n))$. Default choices for $V_i^{(k)}(n)$, $\boldsymbol{\Gamma}^{(k)}(n)$, $\boldsymbol{W}^{(k)}(n)$ were assumed following Costa *et al.* (2013a). Model selection for MDMs is based on Bayes factors (see e.g. West and Harrison, 1997). The evidence in favour of the DAG $G^{(k)}$ under the MDM likelihood can be calculated as

$$p(\boldsymbol{Y}^{(k)}|G^{(k)}) = \prod_{i=1}^{P}\prod_{n=1}^{N} p(Y_i^{(k)}(n)|\boldsymbol{Y}_{G_i^{(k)}}^{(k)}(n), \boldsymbol{Y}^{(k)}(1:n-1), G_i^{(k)}). \quad (3)$$

In practice Eqn. 3 is evaluated using simple Kalman filter recurrences and we refer the reader to Costa *et al.* (2013a) for further details. Costa *et al.* (2013b) reports that the MDMs above are well-suited to the analysis of resting-state fMRI data, outperforming the methods surveyed by Smith *et al.* (2011) in both the detection of edges and also the orientation of edges. This promising performance appears to be driven by the information present in temporal spike patterns, as exploited directly in recent work by Diekman *et al.* (2014).

The MDMs here are reified with the interpretation that edges correspond to neural connectivity (Dawid, 2010). Independent estimation for the subject-specific DAGs $G^{(k)}$ based on the MDM score (Eqn. 3) yields graphs that display high between-subject variability (Fig. 2a). Thus the causal semantics that are associated with MDMs imply that neural connectivity is highly variable between subjects. This is unreasonable on neuroscientific grounds and likely reflects the lack-of-robustness and small sample bias that are often associated with graphical analyses. This motivates a hierarchical statistical model and exact estimation, as we describe below.

2.3 A hierarchical model for multi-subject analysis

Following unsatisfactory independent estimation, we now proceed to explore exact joint estimation as enabled by the recent methodological advances of Oates *et al.* (2014). We note that, providing that the quantities used to compute Eqn. 3 above have been cached, the joint analysis below does not require any further computation involving the MDM model equations. Write $\mathscr{G}$ for the collection of all DAGs on the vertices 1, …, P and write $G^{(1:K)} \in \mathscr{G}^K$ for the collection of all the DAGs $G^{(1)}$, …, $G^{(K)}$. Joint estimation proceeds within a hierarchical Bayesian framework that is specified by the "multiple DAG prior"

$$p(G^{(1:K)}|A) \propto \left(\prod_{(k,l) \in A} r(G^{(k)}, G^{(l)}) \right) \times \left(\prod_{k=1}^{K} m(G^{(k)}) \right). \quad (4)$$

The functions r and m are defined below. Here A denotes an undirected network on the indices 1, …, K that will be used to encode a similarity structure between subjects; the first product factorises along the edges of A. When A is complete, Eqn. 4 encodes an exchangeability assumption that any DAG $G^{(k)}$ is equally likely *a priori* to be similar to any other DAG $G^{(l)}$ ($k \neq l$). Such an exchangeability assumption is implicit in much of the recent literature on multiple graphical models (Werhli and Husmeier, 2008; Oyen and Lane, 2013; Danaher *et al.*, 2014). However, exchangeability will be inappropriate when the collection of subjects is heterogeneous, for instance containing groups or subgroups that correspond to differential neural connectivities. The methodology that we present below allows for arbitrary (and even uncertain) A, relaxing this exchangeability assumption and permitting more flexible estimation.

The function $r : \mathscr{G} \times \mathscr{G} \to [0, \infty)$ is used to encode regularity between pairs of DAGs, with larger values corresponding to *a priori* more similar DAG structures. Oates *et al.* (2014) showed that a particularly convenient form of regularity function is obtained by considering hyper-Markov properties (Dawid and Lauritzen, 1993):

$$\log(r(G^{(k)}, G^{(l)})) = - \sum_{i=1}^{P} \sum_{j=1}^{P} \lambda_{j,i}^{(k,l)} [(j \in G_i^{(k)}) \oplus (j \in G_i^{(l)})]. \quad (5)$$

Here $\oplus$ is the logical XOR operator and $[E]$ is used to denote an indicator function for the event E. The constants $\lambda_{j,i}^{(k,l)}$ can be used to encode which aspects of structure are more likely to be conserved across subjects, based on subjective prior information, or indeed to encode which subjects are more likely to share similar connectivity, based on ancillary covariates such as age, gender, disease status etc. For example one could exploit a penalty $\lambda_{j,i}^{(k,l)} = \lambda^{-|age(k)-age(l)|}$ for some $\lambda > 0$ that encourages sharing of graph structure among subjects of similar ages, but treats edges (j, i) as exchangeable. The function $m(G)$ in the multiple graphical model prior (Eqn. 4) provides an adjustment for the fact that the size of the space $\mathscr{G}$ grows super-exponentially with the number P of vertices (Consonni and La Rocca, 2010). In this paper we follow Chen and Chen (2008); Scott and Berger (2010); Foygel and Drton (2013) and control multiplicity using the binomial correction

$$m(G) = \prod_{i=1}^{P} \binom{P}{|G_i|}^{-1} [|G_i| \leq d_{\max}]. \quad (6)$$

Here $d_{\max}$ is a fixed upper bound on the in-degree of vertices in G that encodes prior knowledge on the support of the graphical models (e.g. Hill *et al.*, 2012). For all examples in this paper we made the subjective choice $d_{\max} = 3$ that reflects the degree of connectivity observed in previous literature (e.g. Ramsey *et al.*, 2009).

2.4 Exact estimation of graphical structure

Bayesian estimation of graphical structure is based on the *maximum a posteriori* (MAP) estimate that is obtained jointly over all DAGs as

$$\hat{G}^{(1:K)}|A := \underset{G^{(1:K)} \in \mathscr{G}^K}{\arg\max}\, p(G^{(1:K)}|\boldsymbol{Y}^{(1:K)}, A). \quad (7)$$

More generally, the network A that expresses similarity between the subjects may be subject to uncertainty. Write $\mathscr{A}$ for the set of undirected networks on the vertices $1, \ldots, K$. In this setting we impose a hyperprior distribution over $\mathscr{A}$ given by

$$\log(p(A)) \overset{+C}{=} \sum_{k=1}^{K} \sum_{l=k+1}^{K} \eta^{(k,l)}[(k,l) \in A] \quad (8)$$

where $\overset{+C}{=}$ denotes equality up to an unspecified additive constant. Here constants $\eta^{(k, l)}$ can again be used to encode prior similarity between subjects on the basis of ancillary covariates. The hyperprior distribution in Eqn. 8 has the effect of detering sparsity in the network A, leading to increased regularisation between DAGs and a more conservative estimate of between-subject variability. In this extended setting, our focus is now an extended MAP estimator

$$(\hat{G}^{(1:K)}, \hat{A}) := \underset{\substack{G^{(1:K)} \in \mathscr{G}^K \\ A \in \mathscr{A}}}{\arg\max}\, p(G^{(1:K)}, A|\boldsymbol{Y}^{(1:K)}) \quad (9)$$

that simultaneously estimates subject-specific DAGs $G^{(k)}$ and the network A that relates subjects.

For each of the optimisation problems in Eqns. 7, 9, Oates *et al.* (2014) describes how techniques from integer linear programming, including constraint propagation and cutting planes, may be used to exactly locate the MAP estimate with a minimal computational burden. Define the "local scores"

$$s^{(k)}(i, G_i^{(k)}) := \begin{cases} \log(p(\boldsymbol{Y}_i^{(k)}|\boldsymbol{Y}_{G_i^{(k)}}^{(k)}, G_i^{(k)})) - \log\binom{P}{|G_i^{(k)}|}) & \text{if}\, |G_i^{(k)}| \leq d_{\max} \\ -\infty & \text{otherwise} \end{cases} \quad (10)$$

that are sufficient statistics for structure learning in the MDM model. Specifically, $s^{(k)}(i, \pi)$ is equal (up to an additive constant) to the log-posterior probability associated with the parent set configuration $G_i^{(k)} = \pi$ for node i in subject k, based only on the subject's own data $Y^{(k)}$. The joint MAP estimators introduced above seek to maximise the sum of these terms subject to the hierarchical penalty (Eqn. 5) and the requirement that $G^{(1)}, \ldots, G^{(K)}$ are each well-defined DAGs.

In brief, the computational methodology exploits the fact that Eqns. 7, 9 can be encoded as an integer linear program of the form

$$\text{maximise } \boldsymbol{f}^T \boldsymbol{x} \text{ subject to } \boldsymbol{Ax} \leq \boldsymbol{b}, \boldsymbol{Cx} = \boldsymbol{d}, \boldsymbol{x} \geq \boldsymbol{0}, x_i \in \mathbb{Z} \quad (11)$$

through careful choices of the integer-valued matrices $\boldsymbol{A}$, $\boldsymbol{C}$ and integer-valued vectors $\boldsymbol{b}$, $\boldsymbol{d}$. Specifically, the entries of the vector $\boldsymbol{x}$ are taken to be binary indicators corresponding to events that include $[G_i^{(k)} = \pi]$, $[(k, l) \in A]$ and $[(j \in G_i^{(k)}) \oplus (j \in G_i^{(l)})] \cap [(k, l) \in A]$, whilst the vector $\boldsymbol{f}$ contains the local scores $s^{(k)}(i, \pi)$ and the constants $\lambda_{j,i}^{(k,l)}$ and $\eta^{(k, l)}$. By inspection of Eqns. 3, 5 and 8 we see that the posterior log-probability

$$\log(p(G^{(1:K)}, A | \boldsymbol{Y}^{(1:K)})) \overset{+C}{=} \sum_{k=1}^{K} \sum_{i=1}^{P} \sum_{\pi \subseteq \{1:P\} \setminus \{i\}} s^{(k)}(i, \pi)[G_i^{(k)}$$

$$= \pi] + \sum_{k=1}^{K} \sum_{l=k+1}^{K} \eta^{(k,l)}[(k, l) \in A] - \sum_{k=1}^{K} \sum_{l=k+1}^{K} \sum_{i,j=1}^{P} \lambda_{j,i}^{(k,l)}[(j \in G_i^{(k)}) \oplus (j \in G_i^{(l)})] \cap [(k, l) \in A]$$

(12)

can be written as an inner-product $\boldsymbol{f}^T \boldsymbol{x}$. The inequality constraints $\boldsymbol{Ax} \quad \boldsymbol{b}$ and equality constraints $\boldsymbol{Cx} = \boldsymbol{d}$ are carefully chosen to ensure that the feasible region for $\boldsymbol{x}$ consists of precisely those vectors that correspond to well-defined (multiple) DAG models. This final point is somewhat technical and we refer the reader to Oates *et al.* (2014) for full details.

2.5 Elicitation of tuning parameters; theory

The class of statistical models that is amenable to exact inference is substantial, but here we focus on particularly tractable prior specifications that allows us to clearly illustrate the methodology. Specifically, we reduce the number of hyperparameters to two by making the assumption that all edges are *a priori* equally likely to be shared between all pairs of subjects ($\lambda_{j,i}^{(k,l)} = \lambda$ for all i, j, k, l) and that all pairs of subjects are *a priori* equally likely to share similar graph structure ($\eta^{(k, l)} = \eta$ for all k, l). Prior elicitation in this reduced class of models therefore requires the specification of hyperparameters λ and η. The impact of the choice of hyperparameters on the MAP estimators is clarified in the following:

Property 1. *(a) When $\eta = 0$, $\hat{G}^{(1:K)}$ consists of DAGs equal to those computed using independent estimation. (b) For $\eta > 0$ we have $(k, l) \notin \hat{A} \Rightarrow \hat{G}^{(k)} \quad \hat{G}^{(l)}$. (c) For fixed η there exists $\lambda^* \in [0, \infty)$ such that whenever $\lambda > \lambda^*$ we have $(k, l) \in \hat{A} \Rightarrow \hat{G}^{(k)} = \hat{G}^{(l)}$. (d) There exists $\eta^* \in [0, \infty)$ such that $\hat{A}$ is the complete network whenever $\eta > \eta^*$.*

The above result deals with the extremes of the parameter space; intuitively we would expect non-trivial values of (λ, η) to interpolate "smoothly" between these extremes. The following shows that this intuition is not strictly true. Specifically, as λ is monotonically increased, it is possible for a particular edge to enter and exit the MAP estimator $\hat{G}^{(k)}$ multiple times and furthermore, non-monotonicity is also exhibited by the network estimator $\hat{A}$:

Property 2. *(a) Fix a network $A \in \mathcal{A}$ and consider varying the hyperparameter λ. If A is non-empty, then there exist values of the sufficient statistics $s^{(k)}(i, \pi)$ such that $[j \in \hat{G}_i^{(k)}]$ is not monotonic in λ. (b) Fix the hyperparameter λ, and consider unknown A with hyperparameter η. Then there exist values of the sufficient statistics $s^{(k)}(i, \pi)$ such that $[(j, i) \in \hat{A}]$ is not monotonic in η.*

Thus the joint MAP, like other penalised likelihood approaches (including the GLASSO for GGMs; Friedman *et al.*, 2008) does not obey a monotonicity property. Property 2 makes it surprising that exact algorithms exists in this nontrivial setting. In practice and in results below we have found that, like the GLASSO, monotonicity holds approximately.

2.6 Elicitation of tuning parameters; practice

The elicitation of hyperparameters such as λ, η should principally be driven by the scientific context, the nature of the data and the rôle that inferences are to play in future work. For example, if the estimated networks are the basis for features within a classification algorithm, then elicitation of hyperparameters should target the classification error. However in some settings, including our illustrative example, the non-availability of relevant ancillary data (e.g. the class labels in classification) precludes such an objective elicitation. Below we therefore illustrate diagnostics that could form the basis for subjective elicitation in quite general settings, based on retrospective inspection of the posterior.

The analysis of resting state fMRI data is an emerging area of research (Cole *et al.*, 2010) and currently neither the source nor the extent of subject-specific variation are well-understood. If the extent of variability at resting state was known, this could be directly leveraged to facilitate the objective elicitation of hyperparameters. However this is not currently the case and subjective elicitation is required. The biological knowledge that forms the basis for elicitation is qualitative in nature, as we explain below: Firstly, connectivity should not change within a subject over the brief time period under which the fMRI experiments were conducted. Secondly, recent studies (e.g. Ringach, 2009) indicate that the notion of "resting state" is poorly defined and can correspond to several contrasting neurological activity profiles; we would therefore not expect to obtain identical DAGs under a replication experiment that is unable to control for the precise nature of the resting state.

A subjective analysis can be obtained using diagnostics based on retrospective examination of the posterior, that we describe below. Specifically, for our fMRI dataset, we performed exact estimation of the joint MAP based on four technical replicate datasets obtained from the first two subjects under identical laboratory conditions. To inform elicitation for the regularity parameter λ, we fixed the network A such that $(k, l) \in A$ if and only if datasets k

and *l* were both technical replicates derived from the same subject (Fig. 4a). This corresponds to placing an exchangeability assumption on the technical replicates, but prohibiting the sharing of information between subjects. We then computed the total structural Hamming distance (SHD; Tsamardinos *et al.*, 2006) between all pairs of DAGs that are technical replicates (Fig. 4b). This diagnostic could be used as the basis for subjective elicitation of λ in general situations where replicate data are available. Below for illustration we focus on one such value, λ = 4, that attributes approximately 50% of variability between technical replicates to extrinsic noise resulting from the experimental design. Examination of the Bayes factor as a function of λ provides a second diagnostic to assist with elicitation that may be useful to highlight over-regularisation. In this case the value λ = 4 scores considerably better compared to the alternative that assigns the same DAG to all replicate datasets (log-Bayes factor ≈ 900, Fig. 4c). Additional diagnostics for the subjective elicitation of η are discussed in the subsequent sections.

2.7 Learning multiple DAGs with exchangeability

Based on the elicitation λ = 4, for illustration, we employed exact estimation for the joint MAP $\hat{G}^{(1:K)}|A$ under the exchangeability assumption that A is the complete network (Eqn. 7). In order to limit scope, here we simply consider one dataset per subject (i.e. no technical replicates were included), but data aggregation is naturally accommodated in the methodology we present (see discussion). Results in Figure 2b demonstrate that the estimated DAG structures are substantially more similar that our original estimate obtained using independent inference (Fig. 2a), with a 23% decrease in total SHD between DAGs. This estimate can be expected to more closely represent the true subject-specific neural connectivity patterns, based on the empirical conclusions of Oates *et al.* (2014). We note however that validation of this inferred connectivity remains extremely challenging (e.g. Stein *et al.*, 2007).

2.8 Learning multiple DAGs without exchangeability

The scientific motivation for multi-subject analysis is typically to elucidate differential connectivity between subjects, either in a purely unsupervised context for exploratory investigation, or in a supervised context to determine whether certain features of connectivity are associated with auxiliary covariates of interest such as disease status. In these cases a statistical model that assumes exchangeability between subjects may be inappropriate and "regularise away" the differential connectivity that is of interest. We therefore proceed to jointly estimate both subject-specific DAGs $G^{(k)}$ and the network A that describes relationships between the subjects themselves (Eqn. 9).

Elicitation of the hyperparameter η, that controls density of the network A, could again be performed by retrospective inspection of the posterior. For our resting state fMRI dataset we would proceed by requiring (i) a moderate amount of similarity between subjects, motivated by expectation that connectivity should not differ substantially between subjects, and (ii) a moderate amount of heterogeneity between subjects, since we aim to highlight any potential differences between the neural connectivity of different subjects. Results in Figure 3 demonstrate that for η = 60 the six subjects are regularised into three distinct components {1, 4}, {2, 3}, {5, 6}, whilst for the higher value η = 70 the subjects are regularised into two

distinct components {1, 2, 3, 4}, {5, 6}. (When $\eta = 80$ the network A is complete and subject-specific DAGs coincide with Fig. 2b.) Examination of the Bayes factor as a function of η demonstrates that the values $\eta = 60, 70$ provide considerably better estimates compared to the DAGs obtained under an exchangeability assumption (log-Bayes factor $\approx 200, 180$ respectively). This suggests that group and sub-group structure may be present amoung the subjects at the level of neural connectivity and provides evidence against exchangeability of the subjects.

Finally we illustrate an alternative and novel approach to learning similarities between subjects, called k-means clustering of DAGs, that does not assume exchangeability of the subjects. In brief, additional latent DAGs $G^{(K+1)}, \ldots, G^{(K+L)}$ are introduced that represent cluster centres or "prototypes", summarising the typical DAG structure within their cluster. The (unknown) network A on the extended vertex set $1, \ldots, K + L$ is constrained to have edges that connect each of the vertices $1, \ldots, K$ to precisely one of the vertices $K+1, \ldots, K+L$, so that estimation of A corresponds exactly to Bayesian model-based clustering. Our methodology thereby facilitates joint estimation of both subject-specific DAGs and their optimal cluster assignment (Oates *et al.*, 2014). (Note that, like in any mixture model, the optimal cluster assignment A is defined only up to permutation of the cluster labels $K + 1, \ldots, K + L$.) Here we applied k-means clustering of DAGs to the six subjects using $L = 2$ clusters (Fig. 5a) and $L = 3$ clusters (Fig. 5b). The optimal cluster assignment with $L = 3$ recovers the three distinct components {1, 4}, {2, 3}, {5, 6} that were obtained above via joint estimation of A, whilst the optimal cluster assignment with $L = 2$ was {1, 2, 3}, {4, 5, 6} which differs from the assignment obtained above in the position of the fourth subject only. This analysis provides an alternative route to investigate similarity between the subjects and offers an alternative route to subjective elicitation of the η hyperparameter. We note that the prototypes that summarise cluster-specific graphical structure may be useful as summary statistics for the purposes of dimensionality reduction.

3 Discussion

In neuroscience experiments it is increasingly common for data to be collected from multiple subjects whose neural connectivities are likely to be related but non-identical. To uncover the causal mechanisms that underpin neural signalling it is necessary to work within a formal statistical theory for inferred causation, the most well-studied of which is rooted in DAGs (Pearl, 2009). Yet until recently exact estimation for multiple related DAGs was computationally infeasible. In this letter we have illustrated, using a small fMRI dataset, how recent algorithmic advances enable sophisticated causal inference using multi-subject experimental data. In particular we have seen how novel statistical models, that do not assume exchangeability between subjects, achieve both a better description of the data (in terms of Bayes factors) and enable the more robust inference of subject-specific connectivity.

The model class that we discuss is large and allows for multiple opportunities to integrate prior knowlege, pertaining to both (i) the connectivity between specific neural regions, and (ii) additional covariates that might associate with subject-specific connectivity, such as age, gender or disease status. The integration of auxiliary covariate data and the more general

experimental validation of our techniques require an extensive and thorough investigation involving many more subjects than we analyse here; this is now the focus of our ongoing research.

We focused on a particularly simple formulation with two tuning parameters and illustrated through application how both tuning parameters could be elicited retrospectively through examination of MAP estimates. This methodology extends naturally to highly structured datasets, for example where each subject is asked to provide multiple fMRI time courses. In these cases a combination of the techniques discussed above would permit all data on a particular subject to be aggregated into a single "prototype" and then estimation to proceed on the basis of these prototypes.

At present an analysis involving K 10 subjects and DAGs of size P 10 requires a few minutes' serial computation on a 2.10GHz Intel Core i7 CPU Windows host with 8GB memory. Our ongoing research focuses on reducing this computational burden so that exact estimation becomes feasible for much larger datasets that include hundreds of neural regions. Recent advances in estimation of single DAGs involving thousands of nodes suggests that much progress can be made in this direction (Bartlett and Cussens, 2013; Sheehan *et al.*, 2014).

Causal inference for neural connectivity is central to the study of brain functionality (Smith *et al.*, 2011; Friston, 2011) and we envisage that the techniques presented here will play an important rôle in the future analysis of multi-subject experimental data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Appendix

Proof of Property 1. (a) This follows since when $\eta = 0$ the DAGs $G^{(1)}$, …, $G^{(k)}$ are *a priori* independent. Since the likelihood also factorises over k it follows that the DAGs $G^{(1)}$, …, $G^{(k)}$ are independent in the posterior.

(b) The objective that we wish to maximise can be written as

$$\boldsymbol{f}^T \boldsymbol{x} = \eta [(k,l) \in A] - \lambda \sum_{i=1}^{P} \sum_{j=1}^{P} [(j \in G_i^{(k)}) \oplus (j \in G_i^{(l)})] \cap [(k,l) \in A] + C \quad (13)$$

where $\eta > 0$ and C does not depend on $[(k, l) \in A]$. When $\hat{G}^{(k)} = \hat{G}^{(l)}$ it follows that the middle term is zero and hence to maximise this objective we must take $\hat{A}$ such that $[(k, l) \in \hat{A}] = 1$.

(c,d) To prove both statements we can take

$$\lambda^* = \eta^* = \sum_{k=1}^{K} \sum_{i=1}^{P} \left[\max_{\pi \in \{1:P\} \setminus \{i\}} s^{(k)}(i, \pi) \right] - \left[\min_{\pi \in \{1:P\} \setminus \{i\}} s^{(k)}(i, \pi) \right]. \quad (14)$$

For (c) note that if $\lambda > \lambda^*$ and $(k, l) \in A$, then the choice $G^{(k)} = G^{(l)}$ strictly maximises the objective function, since a selection $G^{(k)} \neq G^{(l)}$ incurs a penalty of at least λ^* that cannot be compensated for by an increase in the likelihood term

$\sum_{k=1}^{K} \sum_{i=1}^{P} \sum_{\pi \subseteq \{1:P\} \setminus \{i\}} s^{(k)}(i, \pi) [G_i^{(k)} = \pi]$. Similarly for (d), we have that $(k, l) \notin A$ incurs a penalty of at least η^* that cannot be compensated for by an increase in the likelihood term.

Proof of Property 2. (a) Consider the following simple system with two variables and two individuals. Individual 1 has parent set scores $s^{(1)}(1, \{\}) = 0$, $s^{(1)}(1, \{2\}) = -3$ for variable 1 and $s^{(1)}(2, \{\}) = 0$, $s^{(1)}(2, \{1\}) = 1$ for variable 2. Individual 2 has parent set scores $s^{(2)}(1, \{\}) = 0$, $s^{(2)}(1, \{2\}) = 4$ for variable 1 and $s^{(2)}(2, \{\}) = 0$, $s^{(2)}(2, \{1\}) = 1$ for variable 2. Then it is directly verified that for $0 \leq \lambda < 1$, $\hat{G}^{(1)} = 1 \rightarrow 2$ and $\hat{G}^{(2)} = 2 \rightarrow 1$, for $1 < \lambda < 2$, $\hat{G}^{(1)}$ has no edges and $\hat{G}^{(2)} = 2 \rightarrow 1$ and for $\lambda > 2$, $\hat{G}^{(1)} = 1 \rightarrow 2$ and $\hat{G}^{(2)} = 1 \rightarrow 2$. In particular, the edge $(1, 2)$ is present in $\hat{G}^{(1)}$ for $\lambda \in [0, 1) \cup (2, \infty)$ but absent for $\lambda \in (1, 2)$.

To embed the above example in a larger system with P variables and K individuals we proceed as follows: Without loss of generality, assume $A(1, 2) = 1$. For all variables in $G^{(1)}$ and $G^{(2)}$, assign scores $-\infty$ to any parent set π that contains variables from both $\{1, 2\}$ and $\{3, \ldots, P\}$. For variables $\{3, \ldots, P\}$ in $G^{(1)}$ and $G^{(2)}$, and all variables in individuals $\{3, \ldots, K\}$, take all scores to be zero (i.e. non-informative). Then the above proof demonstrates that the edge $(1, 2)$ is present in $\hat{G}^{(1)}$ for $\lambda \in [0, 1) \cup (2, \infty)$ but absent for $\lambda \in (1, 2)$.

(b) Consider the following simple system with two variables and four individuals. Individual 1 has parent set scores $s^{(1)}(1, \{\}) = 0$, $s^{(1)}(1, \{2\}) = 0$ for variable 1 and $s^{(1)}(2, \{\}) = 0$, $s^{(1)}(2, \{1\}) = 1$ for variable 2. Individual 2 has parent set scores $s^{(2)}(1, \{\}) = 0$, $s^{(2)}(1, \{2\}) = 0$ for variable 1 and $s^{(2)}(2, \{\}) = 0$, $s^{(2)}(2, \{1\}) = 2$ for variable 2. Individual 3 has parent set scores $s^{(3)}(1, \{\}) = 0$, $s^{(3)}(1, \{2\}) = 2$ for variable 1 and $s^{(3)}(2, \{\}) = 0$, $s^{(3)}(2, \{1\}) = 0$ for variable 2. Individual 4 has parent set scores $s^{(4)}(1, \{\}) = 0$, $s^{(4)}(1, \{2\}) = 3$ for variable 1 and $s^{(4)}(2, \{\}) = 0$, $s^{(4)}(2, \{1\}) = 0$ for variable 2. Take $\lambda > \lambda^*$, as defined in Property 1, so that $\hat{G}^{(k)} = \hat{G}^{(l)}$ whenever $k \sim_{\hat{A}} l$. Then it is directly verified that for $0 \leq \eta < 1$, $\hat{G}^{(1)} = 1 \rightarrow 2$, $\hat{G}^{(2)} = 1 \rightarrow 2$, $\hat{G}^{(3)} = 2 \rightarrow 1$, $\hat{G}^{(4)} = 2 \rightarrow 1$, and $\hat{A} = \{(1, 2), (3, 4)\}$; for $1 < \eta < 3/2$, $\hat{G}^{(1)} = 2 \rightarrow 1$, $\hat{G}^{(2)} = 1 \rightarrow 2$, $\hat{G}^{(3)} = 2 \rightarrow 1$, $\hat{G}^{(4)} = 2 \rightarrow 1$, and $\hat{A} = \{(1, 3), (3, 4), (4, 1)\}$; for $\eta > 3/2$, $\hat{G}^{(1)} = 2 \rightarrow 1$, $\hat{G}^{(2)} = 2 \rightarrow 1$, $\hat{G}^{(3)} = 2 \rightarrow 1$, $\hat{G}^{(4)} = 2 \rightarrow 1$, and $\hat{A}$ is the complete network. In particular, the edge $(1, 2)$ is present in $\hat{A}$ for $\lambda \in [0, 1) \cup (3/2, \infty)$ but absent for $\lambda \in (1, 3/2)$.

References

Badillo S, Vincent T, Ciuciu P. Group-level impacts of within- and between-subject hemodynamic variability in fMRI. Neuroimage. 2013; 82:433–448. [PubMed: 23735261]

Bartlett, M.; Cussens, J. Advances in Bayesian Network Learning using Integer Programming; Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence; 2013. p. 182-191.

Beckmann CF, Smith SM. Probabilistic independent component analysis for functional magnetic resonance imaging. IEEE T Med. Imaging. 2002; 23:137–152.

Chandrasekaran V, Parrilo PA, Willsky AS. Latent Variable Graphical Model Selection via Convex Optimisation. Ann. Stat. 2012; 40(4):1935–1967.

Chen J, Chen Z. Extended bayesian information criteria for model selection with large model spaces. Biometrika. 2008; 95(3):759–771.

Chickering DM. Optimal structure identification with greedy search. J. Mach. Learn. Res. 2003; 3:507–554.

Claassen, T.; Heskes, T. A Bayesian Approach to Constraint Based Causal Inference; Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence; 2012. p. 207-216.

Cole DM, Smith SM, Beckmann CF. Advances and pitfalls in the analysis and interpretation of resting-state fMRI data. Front. Syst. Neurosci. 2010; 4:8. [PubMed: 20407579]

Consonni G, La Rocca L. Moment Priors for Bayesian Model Choice with Applications to Directed Acyclic Graphs. Bayesian Statistics. 2010; 9(9):119–144.

Costa L, Smith JQ, Nichols T. On the selection of Multiregression Dynamic Models of fMRI networked time series. CRiSM Working Paper, University of Warwick. 2013; 13:6.

Costa L, Smith J, Nichols T, Cussens J. Searching Multiregression Dynamical Models of Resting-State fMRI Networks Using Integer Programming. CRiSM Working Paper, University of Warwick. 2013; 13:20.

Danaher P, Wang P, Witten DM. The joint graphical lasso for inverse covariance estimation across multiple classes. J. R. Statist. Soc. B. 2014; 76(2):373–397.

Dawid AP, Lauritzen SL. Hyper Markov laws in the statistical analysis of decomposable graphical models. Ann. Stat. 1993; 21(3):1272–1317.

Dawid AP. Beware of the DAG! J. Mach. Learn. Res. W & CP. 2010; 6:59–86.

Diekman C, Dasgupta K, Nair V, Unnikrishnan KP. Discovering Functional Neuronal Connectivity from Serial Patterns in Spike Train Data. Neural Comput. 2014; 26:1263–1297. [PubMed: 24708373]

Ellis B, Wong WH. Learning Causal Bayesian Network Structures From Experimental Data. J. Am. Stat. Assoc. 2008; 103(482):778–789.

Foygel R, Drton M. Bayesian model choice and information criteria in sparse generalized linear models. Technical report, University of Chicago. 2013 arXiv:1403.3374.

Friedman N, Koller D. Being Bayesian about Network Structure: A Bayesian Approach to Structure Discovery in Bayesian Networks. Mach. Learn. 2003; 50(1–2):95–126.

Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008; 9(3):432–441. [PubMed: 18079126]

Friston KJ. Functional and Effective Connectivity: A review. Brain Connectivity. 2011; 1(1):13–36. [PubMed: 22432952]

Hill S, Lu Y, Molina J, Heiser LM, Spellman PT, Speed TP, Gray JW, Mills GB, Mukherjee S. Bayesian Inference of Signaling Network Topology in a Cancer Cell Line. Bioinformatics. 2012; 28(21):2804–2810. [PubMed: 22923301]

Li J, Wang ZJ, Palmer SJ, McKeown MJ. Dynamic Bayesian network modeling of fMRI: A comparison of group-analysis methods. Neuroimage. 2008; 41(2):398–407. [PubMed: 18406629]

Loh P-L, Wainwright MJ. Structure Estimation for Discrete Graphical Models: Generalized Covariance Matrices and Their Inverses. Ann. Stat. 2013; 41(6):3022–3049.

Marquand AF, Brammer M, Williams SC, Doyle OM. Bayesian multi-task learning for decoding multi-subject neuroimaging data. Neuroimage. 2014; 92:298–311. [PubMed: 24531053]

Mechelli A, Penny WD, Price CJ, Gitelman DR, Friston KJ. Effective connectivity and intersubject variability: Using a multisubject network to test differences and commonalities. Neuroimage. 2002; 17(3):1459–1469. [PubMed: 12414285]

Meinshausen N, Bühlmann P. High-dimensional graphs and variable selection with the lasso. Ann. Stat. 2006; 34(3):1436–1462.

Mumford JA, Nichols T. Simple group fMRI modeling and inference. Neuroimage. 2009; 47(4):1469–1475. [PubMed: 19463958]

Nemhauser, GL.; Wolsey, LA. Integer and Combinatorial Optimization. New York: Wiley; 1988.

Oates CJ, Smith JQ, Mukherjee S, Cussens J. Exact Estimation of Multiple Directed Acyclic Graphs. CRiSM Working Paper, University of Warwick. 2014; 14:07.

Oyen, D.; Lane, T. Bayesian Discovery of Multiple Bayesian Networks via Transfer Learning; Proceedings of the IEEE 13th International Conference on Data Mining; 2013. p. 577-586.

Pearl, J. Causality: models, reasoning and inference. 2nd ed.. Cambridge University Press; 2009.

Poldrack, RA.; Mumford, JA.; Nichols, TE. Handbook of fMRI Data Analysis. Cambridge University Press; 2011.

Queen CM, Smith JQ. Multiregression dynamic models. J. R. Statist. Soc. B. 1993; 55(4):849–870.

Queen CM, Albers CJ. Intervention and causality: Forecasting traffic flows using a dynamic Bayesian network. J. Am. Stat. Assoc. 2009; 104(486):669–681.

Ramsey J, Hanson S, Hanson C, Halchenko Y, Poldrack R, Glymour C. Six problems for causal inference from fMRI. Neuroimage. 2009; 49:1545–1558. [PubMed: 19747552]

Ringach DL. Spontaneous and driven cortical activity: implications for computation. Curr. Opin. Neurobiol. 2009; 19:439–444. [PubMed: 19647992]

Sanyal N, Ferreira MA. Bayesian hierarchical multi-subject multiscale analysis of functional MRI data. Neuroimage. 2012; 63(3):1519–1531. [PubMed: 22951257]

Scott JG, Berger JO. Bayes and Empirical-Bayes Multiplicity Adjustment in the Variable-Selection Problem. Ann. Stat. 2010; 38(5):2587–2619.

Sheehan NA, Bartlett M, Cussens J. Maximum Likelihood Reconstruction of Very Large Pedigrees. Theoretical Population Biology. 2014 Submitted.

Smith SM, Miller KL, Salimi-Khorshidi G, Webster M, Beckmann CF, Nichols TE, Woolrich MW. Network modelling methods for fMRI. Neuroimage. 2011; 54(2):875–891. [PubMed: 20817103]

Smith SM, Beckmann CF, Andersson J, Auerbach EJ, Bijsterbosch J, Douaud G, Glasser MF. Resting-state fMRI in the Human Connectome Project. Neuroimage. 2013; 80:144–168. [PubMed: 23702415]

Stein JL, Wiedholz LM, Bassett DS, Weinberger DR, Zink CF, Mattay VS, Meyer-Lindenberg A. A validated network of effective amygdala connectivity. Neuroimage. 2007; 36(3):736–745. [PubMed: 17475514]

Sugihara G, Kaminaga T, Sugishita M. Interindividual uniformity and variety of the "Writing center": A functional MRI study. Neuroimage. 2006; 32(4):1837–1849. [PubMed: 16872841]

Tsamardinos I, Brown LE, Aliferis CF. The max-min hill-climbing Bayesian network structure learning algorithm. Mach. Learn. 2006; 65(1):31–78.

Valdes-Sosa PA, Roebroeck A, Daunizeau J, Friston K. Effective connectivity: influence, causality and biophysical modeling. Neuroimage. 2011; 58(2):339–361. [PubMed: 21477655]

Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K. The WU-Minn Human Connectome Project: An Overview. Neuroimage. 2013; 80:62–79. [PubMed: 23684880]

Waldorp L, Christoffels I, van de Ven V. Effective connectivity of fMRI data using ancestral graph theory: Dealing with missing regions. Neuroimage. 2011; 54(4):2695–2705. [PubMed: 21047558]

Werhli AV, Husmeier D. Gene regulatory network reconstruction by Bayesian integration of prior knowledge and/or different experimental conditions. Journal of Bioinformatics and Computational Biology. 2008; 6(3):543–572. [PubMed: 18574862]

West, M.; Harrison, PJ. Bayesian Forecasting and Dynamic Models. 2nd ed. New York: Springer-Verlag; 1997.
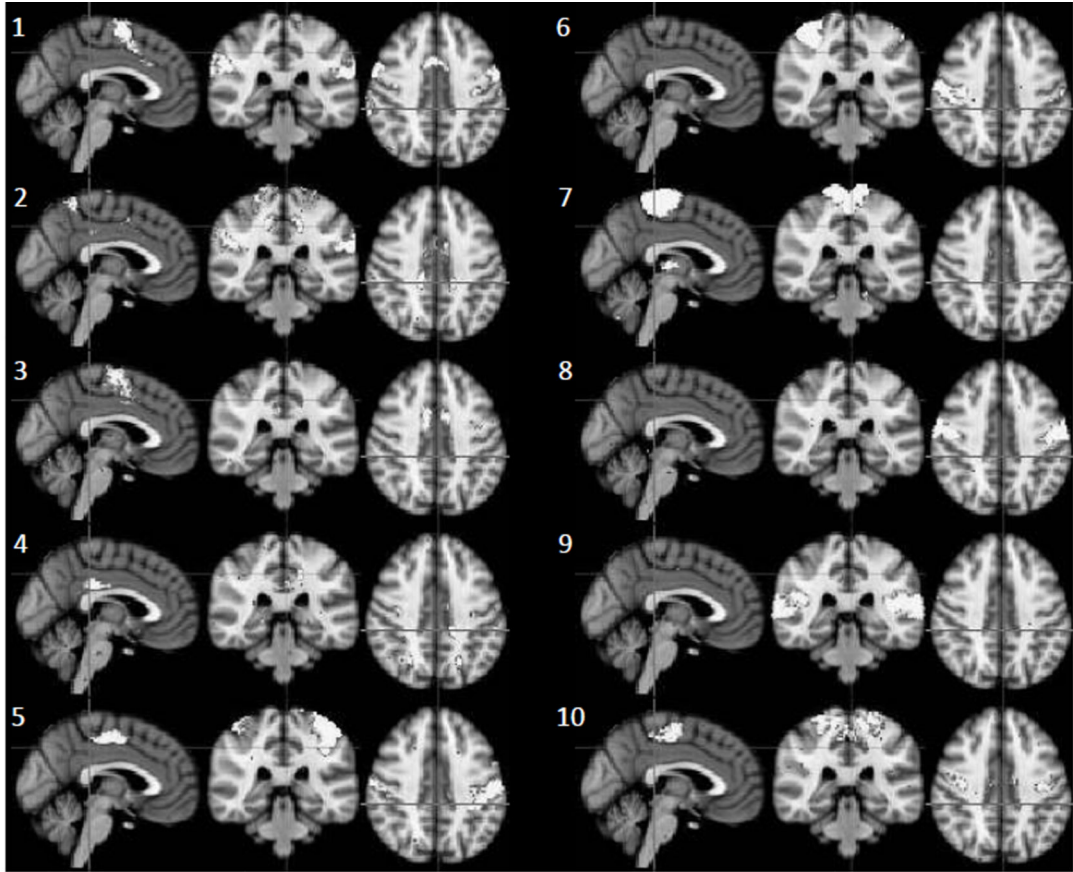
(a) Estimated ICA components

Node Number	Symmetry	Summary
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

(b) Functionality of neural regions
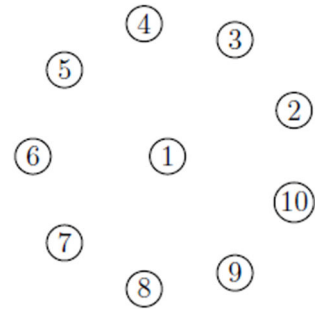
(c) Vertex layout

Figure 1.

Illustrative resting-state fMRI dataset. We consider 10 spatial modes obtained using ICA, show in (a) and described in (b), each having a corresponding time series for each subject. All graphs that we present will adopt the vertex layout shown in (c). [A colour version of this figure is available in the online supplemental material.]
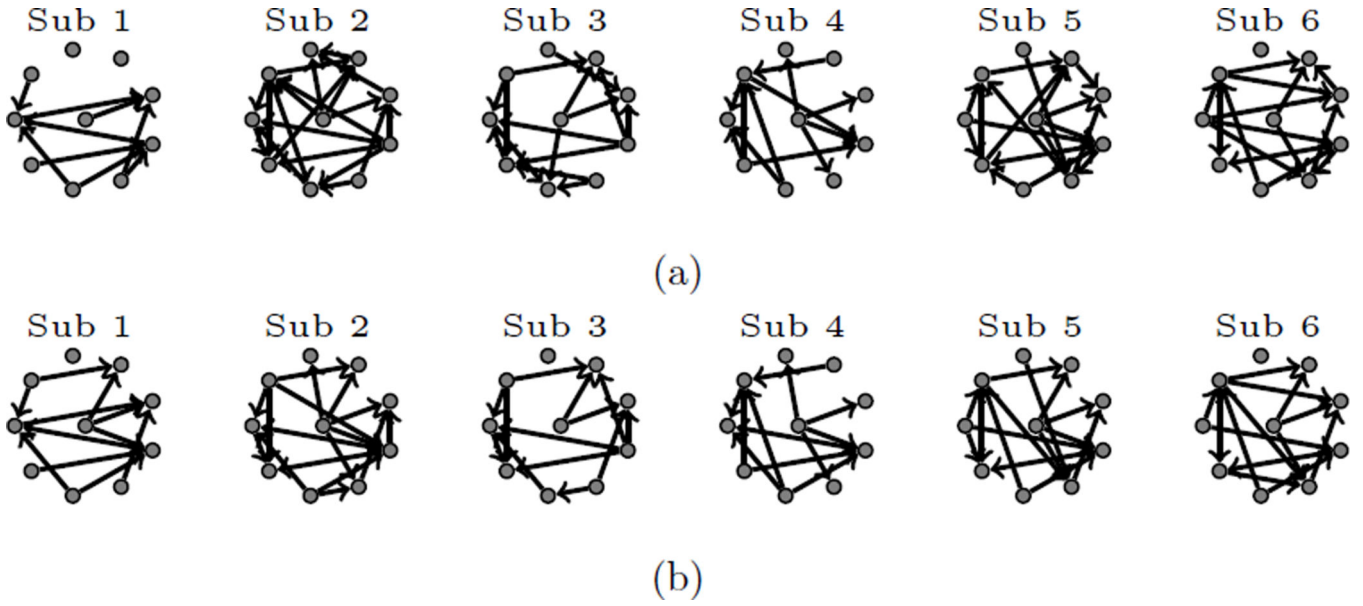
(a)

(b)

Figure 2.
Illustrative fMRI dataset; here time course data were obtained on six subjects. The subject-specific connectivity between ten distinct regions of the brain was estimated using multiregression dynamical models applied (a) to each subject separately, (b) to all subjects jointly with regularity hyperparameter $\lambda = 4$. The graphs in (b) are 23% more similar compared to the graphs in (a), as explained in the main text. [Figure 1c provides a key.]
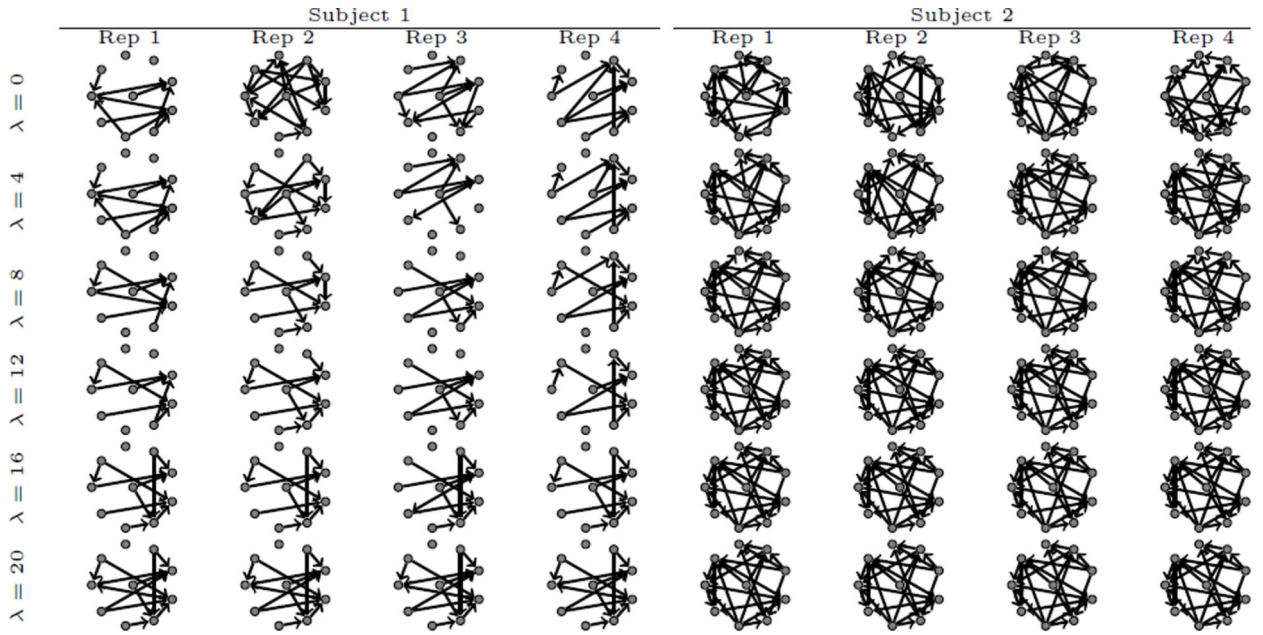
(a)
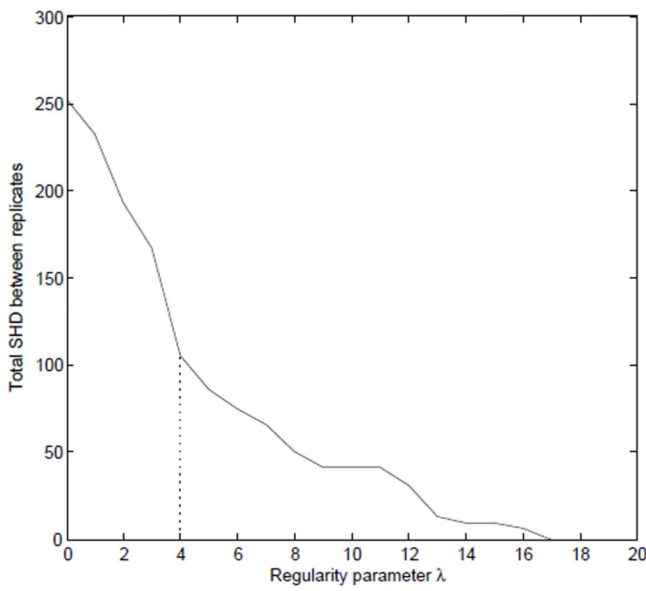
(b)
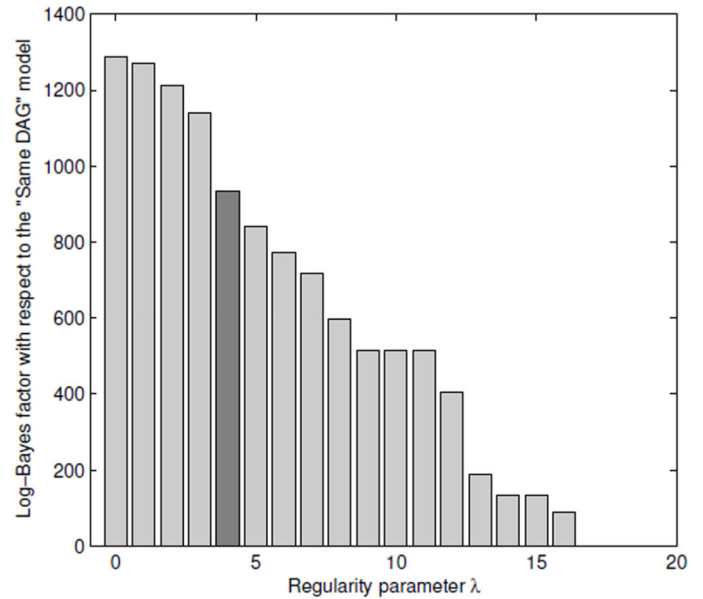
(c)

Figure 3.
Illustrative fMRI dataset; diagnostics for eliciting the regularity parameter λ based on technical replicate data and retrospective inspection of the posterior. (a) Here two subjects each provided four technical replicate datasets. The DAGs shown are joint MAP estimates for varying λ, such that replicates were assumed to be exchangeable but subjects were treated independently. As λ is increased the DAGs corresponding to technical replicates become increasingly similar until they are identical at λ 17. (b) Here we plot the total SHD between DAGs corresponding to technical replicates against the regularity parameter λ. The

dashed line indicates the value $\lambda = 4$ that reduces the between-subject variability by approximately 50%. (c) Comparing the Bayes factor corresponding to model $\lambda = 4$ against independent estimation ($\lambda = 0$) and estimation that forces all DAGs to be identical (λ ‾ 17).
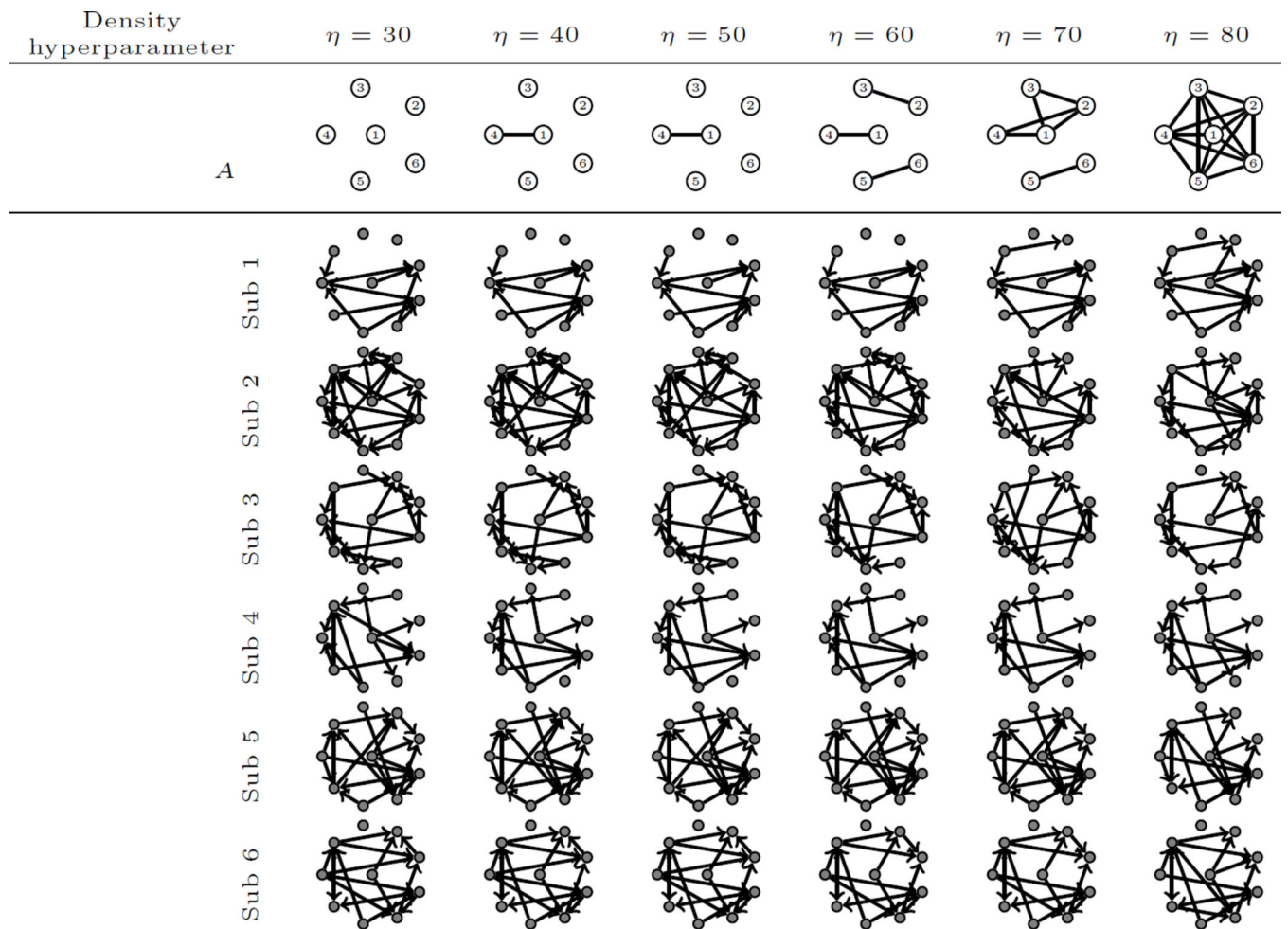
Figure 4.
Illustrative fMRI dataset; learning multiple DAGs *without* an exchangeability assumption.
[Here we simultaneously estimate both subject-specific DAGs and the network *A* that relates
subjects. The regularity hyperparameter $\lambda = 4$ was fixed whilst the density hyperparameter η
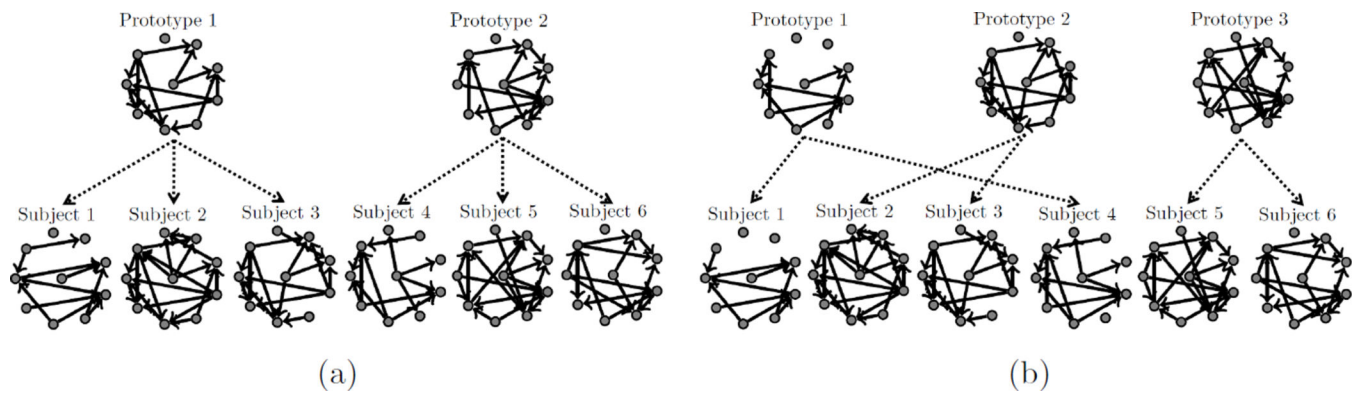was varied.]

Figure 5.
Illustrative fMRI dataset; *k*-means clustering of DAGs with (a) $L = 2$ clusters and (b) $L = 3$ clusters. [We simultaneously estimate subject-specific DAGs, their cluster assignments (dashed edges) and the prototypes that summarise graphical structure within each cluster. The regularity hyperparameter was fixed at $\lambda = 4$.]