International Journal of Approximate Reasoning 77 (2016) 87–104

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Bayesian selection of graphical regulatory models

Silvia Liverani [a,*], Jim Q. Smith [b]

[a] Department of Mathematics, Brunel University London, UK
[b] Department of Statistics, University of Warwick, Coventry, UK

ARTICLE INFO

Article history:
Received 15 October 2015
Received in revised form 5 May 2016
Accepted 31 May 2016
Available online 15 June 2016

Keywords:
Clustering
Bayesian inference
Causality
Time-course microarray experiments

ABSTRACT

We define a new class of coloured graphical models, called regulatory graphs. These graphs have their own distinctive formal semantics and can directly represent typical qualitative hypotheses about regulatory processes like those described by various biological mechanisms. They admit an embellishment into classes of probabilistic statistical models and so standard Bayesian methods of model selection can be used to choose promising candidate explanations of regulation. Regulation is modelled by the existence of a deterministic relationship between the longitudinal series of observations labelled by the receiving vertex and the donating one. This class contains longitudinal cluster models as a degenerate graph. Edge colours directly distinguish important features of the mechanism like inhibition and excitation and graphs are often cyclic. With appropriate distributional assumptions, because the regulatory relationships map onto each other through a group structure, it is possible to define a conditional conjugate analysis. This means that even when the model space is huge it is nevertheless feasible, using a Bayesian MAP search, to a discover regulatory network with a high Bayes Factor score. We also show that, like the class of Bayesian Networks, regulatory graphs also admit a formal but distinctive causal algebra. The topology of the graph then represents collections of hypotheses about the predicted effect of controlling the process by tearing out message passers or forcing them to transmit certain signals. We illustrate our methods on a microarray experiment measuring the expression of thousands of genes as a longitudinal series where the scientific interest lies in the circadian regulation of these plants.

1. Introduction

There has been a recent explosion in the development of new techniques in exploratory data analysis for huge data sets fuelled, for example, by intense activities in the study of genome sequences and DNA microarray analyses. Examples of such exploratory tools are the fast Bayesian methods [7,21,8], designed to cluster tens of thousands of longitudinal series of measurements. The outputs from these methods, especially the displays of the cluster means (see e.g. Fig. 1) of a MAP estimated model, have proved to be particularly helpful to biologists when they explore hypotheses about which sets of genes might be passing a message to another.

However these Bayesian cluster methods begin with the assumption that clusters express independently of one another. This is clearly not usually a credible hypothesis for regulation. For under a hypothesis that one cluster regulates another it

* Corresponding author.
E-mail address: silvia.liverani@brunel.ac.uk (S. Liverani).

http://dx.doi.org/10.1016/j.ijar.2016.05.007

is typically believed that the shape of the profile of the receiving cluster will be related to the profile of the sending cluster in a relatively predictable way.

Various authors have therefore modelled the likely dependences between the time courses of gene expression profiles using a variety of standard graphical models: often variants of the Bayesian Network (BN) [9]. Thus [12], acknowledging the consequent computational and methodological limitations of their techniques, used dynamic BNs to model subsets of series like those we model here. Albeit in a non-exploratory context [19] analysed variations of a baseline model representing known dependence structures. Other graphical studies of regulatory models include [4].

Although these graphical models are clearly very useful, particularly at later stages of the analysis of a given process, they are hardly ever integrated into an early exploratory data analysis [2,13,20]. Furthermore we argue below that the semantics of the BN is not well matched to modelling dependences induced by regulation. Rather than representing conditional independence relationships, ideally we would like our graph to be explicitly regulatory: i.e. that a directed edge from a parent node v_1 is connected to a child v_2 if and only if the model embodies the hypothesis that the object v_1 regulates v_2. Clearly this is not in general true of a BN. In particular there is no way its acyclic graph could embody any cyclical regulation, the most important type of relationship in our running example.

In this paper we develop a new graphical model for depicting regulatory relationships within a Gaussian family, building on a class of Bayesian cluster models that have been widely and successfully implemented [3,7]. The model class embeds the class of Bayesian longitudinal cluster models as a special case. We are aware of no similar work in the literature, in scope nor design. We demonstrate how the method provides:

1. a feasible exploratory technique for regulatory networks of longitudinal high-dimensional data;
2. a formal semantics that embodies most of the useful properties of the BN;
3. a framework for a conditional conjugate analysis;
4. a formal graphical representation of a MAP model which can represent the types of dependences the scientist might hypothesise; this graphical representation will be in a form which closely corresponds to the less formal graphs she might currently use to depict that model's regulatory mechanism;
5. an associated causal algebra enabling the scientist, under certain assumptions, to generate predictions of the effect of certain types of control on the system.

This work is motivated by our collaboration with biologists so our running example concerns the study of the statistical models of the regulation mechanism of the circadian clock in plants. However, the applicability of this new class of models extends beyond genetics and bioinformatics.

In Section 1.1 we introduce the data used in the paper and in Section 2 we briefly review the form of Bayesian cluster analysis which we plan to generalise. Then in Section 3 we use a group action which is hypothesised to embody the relationship of the profile shape of a receiving cluster and the profile of its regulating cluster. This group defines equivalence classes of clusters called supraclusters. These supraclusters will define the disconnected components of the graph of our process. In Section 4 we discuss graphical representations for supraclusters. We complete our methodological development in Section 5 where we provide a causal interpretation for our graphs of supraclusters. We then demonstrate our method using two examples in Section 6: a simulated scenario and a real system. In this way we illustrate not only that our methods are feasible and evocative but also how the output graphs provide a helpful and familiar framework through which results can be fed back to the scientist.

1.1. Data

We demonstrate that our method successfully identifies dependence between clusters of genes in two examples: a simulated running example and a more detailed analysis of a genuine microarray experiment on the plant *Arabidopsis thaliana*. These experiments were designed to detect genes whose expression levels, and hence functionality, might be connected with circadian rhythms. The aim was to investigate the possible identity and role of those genes involved in the regulation of the circadian clock of a plant. For our running example we simulated 5 clusters of 30 genes each, of the type discussed in [5]. Here the gene expression was measured at $T = 13$ time points over two days. Constant white light was shone on the plants for 26 hours before the first microarray was taken, with samples every four hours. Thus, there are two cycles of data for each of the *Arabidopsis* microarray chip. The data y_{z_i} was simulated from

$$y_i = \lambda_{z_i} B(\theta_{z_1}) R_{z_i} \boldsymbol{\beta} + \varepsilon_i$$

where the subscript $z_i = j$ if individual i belongs to cluster j with $j = 1, \ldots, 5$ and $\varepsilon_i \sim \text{Normal}(0, 4)$. The matrix $B(\theta_{z_i})$ is a Fourier basis function (see Equation (2) for more details) and R_{z_i} is a rotation matrix, that is, a block diagonal matrix where each block is a matrix $M(k, \theta_{z_i})$ as

$$M(k, \theta_{z_i}) = \begin{pmatrix} \cos(k\theta) & -\sin(k\theta) \\ \sin(k\theta) & \cos(k\theta) \end{pmatrix}$$

with k corresponding to the index of the Fourier coefficients. The generating values of λ_{z_i}, θ_{z_i} and $\boldsymbol{\beta}$ are given in Section 6. For the second example we analysed this dataset in full (22,810 genes whose expressions are each observed over 13 time point) as well as other datasets, as we discuss in Section 6.

2. Cluster models for longitudinal data

The study of the family of regulatory graphs proposed in this paper began with our work on Bayesian cluster models. In the applications of the proposed methodology there are N – usually tens of thousands – longitudinal sets of measurements of the activity of units taken at particular discrete points over time. To understand the underlying processes it is therefore important to learn which of these processes are copies of each other, as evidenced by the fact that their profiles appear to be replicates. In this way we can reduce the tens of thousands of time series into a much smaller number of K clusters.

Typically in our applications K depends on how we set the hyperparameters of our models. However for our real dataset there are usually about one hundred longitudinal clusters: a still large but more manageable number. Each of these longitudinal clusters is then treated as a candidate regulator. The underlying dependence structure can be represented graphically and it can be further elaborated to describe hypotheses about the effects of the types of controls eluded to above.

We first need to set up some notation. Let $\mathcal{C} \triangleq \{C_1, C_2, \dots C_K\}$ denote a partition of indices $\{1, 2, \dots, N\}$, $N \geq K$, where N_k denotes the cardinality of cluster C_k. Let $\boldsymbol{Y}_i$, $i = 1, 2, \dots, N$, denote a set of N T-vectors of real valued observations of the continuous time development of a unit at T given time points – henceforth called a *profile*. Let $\boldsymbol{D}_C$, the $T \times N_j$ matrix of values be defined by $\boldsymbol{D}_C = \{\boldsymbol{Y}_i : i \in C\}$ for some subset C of $\{1, 2, \dots, N\}$. Let the distribution of $\{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \dots, \boldsymbol{Y}_N\}$ be parametrised by $\boldsymbol{\theta} \in \Theta$, and fixed hyperparameters $\boldsymbol{\varphi} \in \Psi$. Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots \boldsymbol{\theta}_K)$ where $\boldsymbol{\theta}_k$, for $k = 1, 2, \dots, K$ is the parameter vector specific to the k-th cluster with N_k elements. Note that then $\sum_{k=1}^{K} N_k = N$. The first Bayesian model we review is one which admits no dependence or causal structure between any of the objects of interest – the cluster profiles.

Definition 1. A *Bayes cluster model* $\mathcal{C}$ of a set of profiles $\{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \dots, \boldsymbol{Y}_N\}$ assumes that:

1. profiles in different clusters and their associated parameters are independent given the hyperparameters i.e.
$$\amalg_{k=1}^{K} \left(\boldsymbol{D}_{C_k}, \boldsymbol{\theta}_k\right) | \boldsymbol{\varphi}$$
2. within each cluster the profiles $\boldsymbol{Y}_i | \boldsymbol{\theta}_k, \boldsymbol{\varphi}$ are identically distributed, $i = 1, 2, \dots, N_k$, $k = 1, 2, \dots, K$ and conditionally on each cluster's parameter vector $\boldsymbol{\theta}_k$ and the hyperparameters $\boldsymbol{\varphi}$ these observations are taken independently: i.e.
$$\amalg_{i=1}^{N_k} \boldsymbol{Y}_i | \boldsymbol{\theta}_k, \boldsymbol{\varphi}$$

There now are several such Bayesian cluster models which have been successfully and usefully implemented [7,16,5,17]. Many take the form that for cluster C_k the time series of observations is modelled by a linear regression

$$\boldsymbol{Y}_i^{(k)} = \boldsymbol{X}^{(k)} \boldsymbol{\beta}^{(k)} + \varepsilon_i^{(k)} \tag{1}$$

for $k = 1, \dots, K$ and $i = 1, \dots, N_k$ where $\boldsymbol{\beta}^{(k)} \in \mathbb{R}^p$ is the vector of parameters with $p \leq T$, where $\varepsilon_i^{(k)}$ are independent identically distributed measurement errors with variance σ_k^2 and the design matrix, or basis function, $\boldsymbol{X}^{(k)}$, a matrix of size $N_k T \times p$, is *customised* to the hypothesised underlying process. In our running example where interest lies in regulators acting over a twenty four hour cycle, an appropriate basis function is Fourier. So then, for example when p is even,

$$y_{it}^{(k)} = \beta_1^{(k)} + \sum_{i=1}^{p/2} \beta_{2i}^{(k)} \cos\left(2\pi t(2i)/\mathcal{T}\right) + \sum_{i=1}^{p/2} \beta_{2i+1}^{(k)} \sin\left(2\pi t(2i)/\mathcal{T}\right) + \varepsilon_{it}^{(k)}. \tag{2}$$

Here the parameters $\boldsymbol{\theta}_k$ associated with cluster C_k are the corresponding regression parameters $\boldsymbol{\beta}_k$ and an error variance parameter linked to the within cluster diversity.

Within these classes it is often possible, using an appropriate conjugate analysis, to construct a score function based on the corresponding Bayes Factor – a MAP score which can be derived in closed form conditional on the hyperparameters for each partition of the profiles. Thus if errors and prior parameter distributions are assumed Gaussian, and the noise variance a priori inverse Gamma distributed then conditional on a fixed noise-to-signal ratio matrix, the marginal likelihood of each possible cluster partition is a product of multivariate t densities. For the details of how these methods work see e.g. [3,7, 17] and [10]. It has now been established that these and related methodologies can successfully and feasibly identify well supported models within these typically huge search spaces.

The methods discussed in this section are widely used in the literature on clustering of microarray experiments. Fig. 1 depicts the expression profiles within a few typical clusters together with graphs of the clusters' mean Fourier coefficients in the analysis given in [5], one of the many analyses performed using this methodology.

Our experience using these methods have demonstrated that with the careful setting of hyperparameters the outputs of such models are remarkably robust, both formally and practically, when it comes to the estimation of parameters that might be linked to regulation [5,17,10,11].

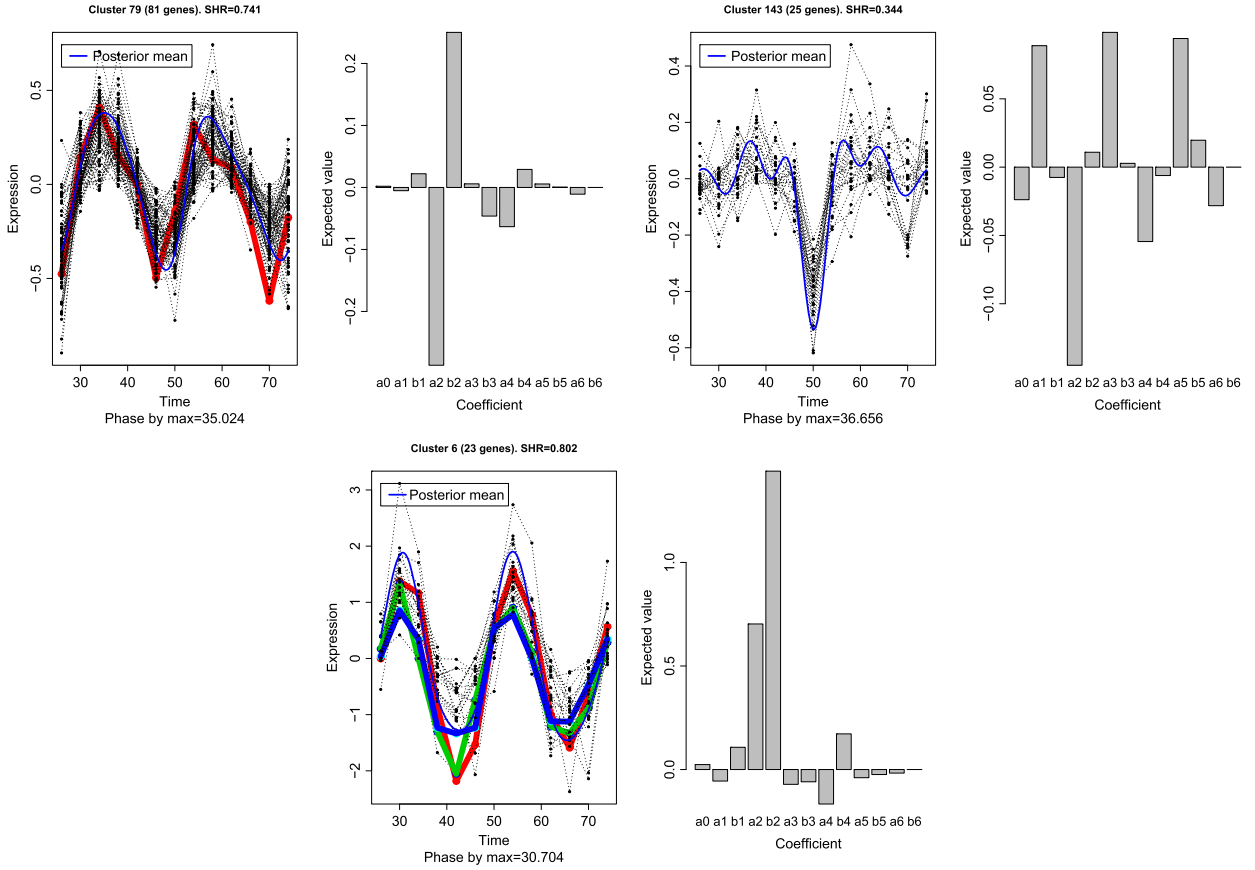


Fig. 1. Four of the clusters obtained clustering the data in [5], our real data application. The blue line represents the posterior mean, the thicker coloured lines are well known genes and the barplot on the left are the posterior estimates of the Fourier coefficients. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Supraclustering and regulatory models

This type of cluster analysis can be extended to model directly the dependences between clusters if a regulatory relationship between them can be specified as follows. A transmission function is defined which maps the parameters of one unit's profile on to the parameters of the profile it regulates. These functions are then coded as a group of transmission matrices, customised to mirror the types of dependences induced by a particular hypothesised regulating mechanism.

In this section we fix the value of the cluster hyperparameter $\boldsymbol{\varphi} \in \Psi$ and for clarity suppress this parameter in the indexing. Then for the transition hyperparameters $\boldsymbol{\phi} \in \Phi$, there are two groups used to map the shape of one profile onto another ($\mathcal{L}^*$) and one cluster's parameters onto another ($\mathcal{L}$).

For some integer m let $\mathcal{L}^* \triangleq \left\{L^*(\boldsymbol{\phi}) : \boldsymbol{\phi} \in \Phi \subseteq \mathbb{R}^m\right\}$ be a family of $T \times T$ transition matrices used to define a regulatory relationship from a profile $\boldsymbol{Y}_{i_1}$ of a given unit i_1 to a second unit with the profile $\boldsymbol{Y}_{i_2}$, where the transition hyperparameter $\boldsymbol{\phi}$ indexes each matrix in this family. Associated with $\mathcal{L}^*$ is a set $\mathcal{L} \triangleq \left\{L(\boldsymbol{\phi}) : \boldsymbol{\phi} \in \Phi \subseteq \mathbb{R}^m\right\}$ of $(p+1) \times (p+1)$ matrices mapping the parameters $\boldsymbol{\theta}_{j_1}$, $i \in C_{j_1}$ to the parameters of the regulated unit $\boldsymbol{\theta}_{j_2}$, $i_2 \in C_{j_2}$, such that

$$\boldsymbol{\theta}_{j_2} \triangleq L(\boldsymbol{\phi}_{j_1 \to j_2})\boldsymbol{\theta}_{j_1}$$

The family of transition functions L parametrised by $\phi_{j_1 \to j_2}$ must first be elicited from the domain expert since this must be determined from those features in the data she believes will be indicative of a regulatory relationship. In our running example such relationships were communicated to us: "The time profile of expression of a regulated gene usually appears to be a short translation of the profile of the regulating cluster of genes, usually rescaled and sometimes, if inhibited rather than excited, reversed". This determined a 3 parameter transition function with a parameter capturing the phase change, another expressing a positive rescaling and a third representing inhibition, which corresponded to a simple sign change. In this paper, both for simplicity and because the subclass seemed to fit well, we confined ourselves to search over this parameter class, although other of our more complex models could also model damping effects. Of course in other settings and with different interpretations of basis functions the group of transformations could be very different.

For the purposes of this paper we will assume, as in our running example, that both $\mathcal{L}^*$ and $\mathcal{L}$ form a group under matrix multiplication. Write $i_2 \backsim i_1$ where $i_2 \in C_{j_2}$, $i_1 \in C_{j_1}$ and $C_{j_1}, C_{j_2} \in \mathcal{C}$ iff $\exists \boldsymbol{\phi}_{j_1 \to j_2} \in \Phi$ such that

Table 1
Summary table of the notation.

Profile name (i_1) and its cluster membership (j_1)	$i_1 \in C_{j_1}$
Profile for observation i_1	$\boldsymbol{Y}_{i_1}$
Parameter vector for cluster j_1	$\boldsymbol{\theta}_{j_1}$
Transmission matrix on the profile space from cluster j_1 to cluster j_2	$L^*(\boldsymbol{\phi}_{j_1\to j_2})$
Transmission matrix on the parameter space from cluster j_1 to cluster j_2	$L(\boldsymbol{\phi}_{j_1\to j_2})$

$$\boldsymbol{Z}_{i_2}|\boldsymbol{\theta}_{j_2}, \boldsymbol{\phi}_{j_1\to j_2} \equiv L^*(\boldsymbol{\phi}_{j_1\to j_2})\boldsymbol{Y}_{i_1}|\boldsymbol{\theta}_{j_2}, \boldsymbol{\phi}_{j_1\to j_2}$$

has the same distribution as $\boldsymbol{Y}_{i_1}|\boldsymbol{\theta}_{j_1}$, where $\boldsymbol{\theta}_{j_2}$ is defined above. See Table 1 for a summary of the notation above.

Note that for known values of transmission parameters, these linearly transformed profiles remain in the same conjugate families used in the conjugate analyses discussed in the last section. It is this property which enables us to exploit the technology originally designed for fast model selection over clusters with exchangeable elements generalising this to models reflecting more structured and nuanced mutual dependences.

Both the shapes and the maps $\mathcal{L}$ will be chosen to reflect the types of relationships that the scientist is interested in and expects to see, if they exist in the dataset at hand. For example in our running example the shapes of clusters the scientist is particularly interested in will have profiles which exhibit a 24 hour cycle. The types of dependences of interest will be ones where associated genes might regulate each other. This will be reflected, for example, by small phase changes – modelling a short delay due to the time it takes to pass on a message – together with a change in amplitude perhaps combined with a low frequency filter. Although in principle these relationships can be expressed at the unit level it is usually more convenient to think of relationships between clusters.

Clearly when $\mathcal{L}$ is a group $\backsim$ defines an equivalence relation. It is therefore possible to coarsen the partition $\mathcal{C}$ so that certain sets of its clusters are placed within the same equivalence class under the group action above. So let $\mathcal{B} = \{B_1, B_2, \ldots, B_S\}$ denote the partition of the K clusters into these $S \leq K$ equivalence classes. Call $\mathcal{B}$ the set of *supraclusters* and let K_s denote to cardinality of the supracluster B_s. Now to simplify the notation we will pick an arbitrary cluster $C_0^{(s)}$ – called the *seed* cluster – from each supracluster B_s, $s = 1, 2, \ldots, S$. In practice in our examples for explanatory purposes it is usually helpful to choose a $C_0^{(s)}$ containing an established regulatory gene, or failing that one with a high amplitude in its associated profile. We will suppress the s index below if no confusion arises.

We can now let $L(\boldsymbol{\phi}_j) \triangleq L(\boldsymbol{\phi}_{j_0\to j})$ denote the transformation of the profile of the seed cluster C_0 into the profile of a different cluster C_j in the same supracluster. Note that since the transform L defines a group, under this notation

$$L(\boldsymbol{\phi}_{j_1\to j_2}) = L(\boldsymbol{\phi}_{j_0\to j_1})^{-1}L(\boldsymbol{\phi}_{j_0\to j_2}) = L(\boldsymbol{\phi}_{j_1})^{-1}L(\boldsymbol{\phi}_{j_2}).$$

Let $\boldsymbol{\theta} \triangleq (\boldsymbol{\theta}_s : s = 1, 2, \ldots, S)$ where $\boldsymbol{\theta}_s$ are the parameters of the seed cluster in supracluster B_s of partition $\mathcal{B}$. Let

$$\boldsymbol{\phi} \triangleq \left(\boldsymbol{\phi}_s : s = 1, 2, \ldots, S\right)$$

where $\boldsymbol{\phi}_s \in \mathbb{R}^{K_s - 1}$ is a concatenation of the vectors of all the m transmission hyperparameters associated with every cluster in the supracluster B_s which is not a seed cluster.

In a cluster model two units in the same cluster of $\mathcal{C}$ are exchangeable. In contrast the parameters $\boldsymbol{\theta} \in \Theta$ of two units in the same supracluster B_s are deterministically and linearly related via a matrix $L(\boldsymbol{\phi}_j)$ where $\boldsymbol{\phi}_j$ is typically a short vector of hyperparameters. The parameters $\boldsymbol{\theta}_j \in \Theta$ themselves define the distribution of the profile vector $\boldsymbol{Y}_i$ such that the linearly transformed profiles

$$\boldsymbol{Z}_i \triangleq L^*(\boldsymbol{\phi}_j)\boldsymbol{Y}_i$$

all share the same shape parameter vector $\boldsymbol{\theta}_s$ of the relevant seed cluster $C_0^{(s)}$. Note that under this construction, since L^* forms a group of different choices of seed, clusters have different but equivalent associated parametrisations within their supracluster.

We now have the formal definition of the supracluster below.

Definition 2. A *Bayes supracluster partition* $\mathcal{B}$ of a set of profiles $\{\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_N\}$ is a family of joint distributions on these observations parametrised by $\boldsymbol{\theta} \in \Theta$ with global hyperparameters $\boldsymbol{\varphi} \in \Psi$ and transition hyperparameters $\boldsymbol{\phi} \in \Phi$. It has the property that under the transformations of parameters and profiles defined above,

$$\coprod_{s=1}^{S} \left(\boldsymbol{Z}_{B_s}, \boldsymbol{\theta}_s\right)|\boldsymbol{\phi}, \boldsymbol{\varphi}, \mathcal{C}, \mathcal{B}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots \boldsymbol{\theta}_S)$ and

$$\coprod_{i=1}^{K_s} \boldsymbol{Z}_i|\boldsymbol{\theta}_s, \boldsymbol{\phi}_s, \boldsymbol{\varphi}, \mathcal{C}, \mathcal{B}$$

where $\boldsymbol{Z}_i|\boldsymbol{\theta}_s, \boldsymbol{\phi}_s, \boldsymbol{\varphi}, \mathcal{C}, \mathcal{B}$ are identically distributed for $i = 1, 2, \ldots, K_s$ and $s = 1, 2, \ldots, S$.

Remark 3. The Bayes supracluster model is fully and uniquely specified by

$$\mathcal{C}, \mathcal{B}, \boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\varphi}$$

where $\mathcal{B}$ is a coarsening of the partition $\mathcal{C}$. Here $\boldsymbol{\theta}$ denotes a concatenation of the parameters defining the distribution of the profile of the seed cluster and $\boldsymbol{\phi}$ denotes a concatenation of all the transition parameters of each non-seed cluster in each supracluster again concatenated over all supraclusters, as defined above, and $\boldsymbol{\varphi}$ are the cluster hyperparameters containing for example hypothesised fixed noise-to-signal ratios of observational noise and parameter variability within cluster.

Note from our definition that only the parameters of units in different supraclusters are independent. Under typical regulatory hypotheses about the conservation of shape under regulation, each supracluster B_s would then represent an independent regulatory mechanism. The clusters within each supracluster represent collections of different units co-expressing different roles within that mechanism, relative to the role of the reference cluster. This is described by the values of the hyperparameters of the different components of $\boldsymbol{\phi}_s$. This will be illustrated below.

Note that the original Bayes cluster model simply sets $\mathcal{B} = \mathcal{C}$. The supracluster coarsening is then the trivial one. Therefore the class of Bayes cluster models is contained in a much larger class of models which embodies certain types of dependence that is often hypothesised to occur in a regulatory environment. In this sense the class of supracluster models *generalises* the class of cluster models. In the search method we use below we find the MAP model assuming the trivial supracluster. We then progressively refine that model searching for a model with regulatory dependences having a higher Bayes Factor score.

4. Transmission matrices and regulatory graphs

Regulatory hypotheses can give rise to relationships of two types. The first is when two clusters with different shapes are hypothesised to perform the same function. For example two clusters in the same supracluster might have identical shapes over time but somewhat different amplitudes. Then it may well be appropriate to believe that these have potentially the same regulatory functions even though their component genes have different trajectories. When this happens we say that they lie in the same *hypercluster* $H \in \mathcal{H}$ where the partition $\mathcal{H}$ is a coarsening of $\mathcal{C}$ but a refinement of $\mathcal{B}$. If $\mathcal{H} = \mathcal{B}$ so that all clusters in a supracluster perform the same function then the supraclustering partition $\mathcal{B}$ is sufficient to describe the system.

However, more usually, because of their shape, certain hyperclusters within a supracluster have a different role: for example, from the relationship of the shapes of two profiles one might be naturally assumed to excite the other. This can be conjectured when the profile of the receiving cluster has a short forward shift in phase accompanied by a possible reduction in amplitude and a low pass filtering effect. Usually it is these types of asymmetric relationship which are the most interesting ones for the scientist to discover.

In this paper we assume that the scientist specifies a priori how plausible she believes a regulatory relationship from cluster $C_{j_1} \to C_{j_2}$ might be, as reflected through a real positive function $t(\boldsymbol{\phi}_{j_1 \longrightarrow j_2})$ – called the *transmission separation* – of the transmission parameters $\boldsymbol{\phi}_{j_1 \longrightarrow j_2}$. Smaller values of $t(\boldsymbol{\phi}_{j_1 \longrightarrow j_2})$ correspond to stronger evidence for direct message passing from $C_{j_1} \to C_{j_2}$. So for example in our running example of circadian regulation, $t(\boldsymbol{\phi}_{j_1 \longrightarrow j_2})$ might measure the difference in the phase of the profile of C_{j_2} from the phase of C_{j_1}. For example a common hypothesis is that the shorter the time period to lapse in this transition the more likely it is that a message has passed directly from $C_{j_1} \to C_{j_2}$. We shall let $\delta(\boldsymbol{\phi}_{j_1, j_2}) = \min\left\{t(\boldsymbol{\phi}_{j_1 \longrightarrow j_2}), t(\boldsymbol{\phi}_{j_2 \longrightarrow j_1})\right\}$ denote the *transmission length* between clusters C_{j_1} and C_{j_2}.

Definition 4. For each supracluster $B \in \mathcal{B}$, let $\Delta(B)$ – called the *transmission matrix* – denote the matrix of transmission separations from $C_{j_1} \to C_{j_2}$, with $C_{j_1}, C_{j_2} \in B$, whose (j_1, j_2) entry is $\delta(\boldsymbol{\phi}_{j_1, j_2})$. Define $\delta(\boldsymbol{\phi}_{j,j}) = 0$.

Definition 5. Say $\Delta(B)$ is *zeroed* if for any three clusters $C_{j_1}, C_{j_2}, C_{j_3}$

$$\delta(\boldsymbol{\phi}_{j_1, j_2}) = 0 \quad \text{and} \quad \delta(\boldsymbol{\phi}_{j_2, j_3}) = 0 \quad \Rightarrow \quad \delta(\boldsymbol{\phi}_{j_1, j_3}) = 0.$$

Note that if $\Delta(B)$ is zeroed then the clusters whose separations are zero form an equivalence class. When $\Delta(B)$ is zeroed the interpretation of δ given above means in particular that the hyperclusters $H \in \mathcal{H}$ are of the form $H = \left\{C_{j_1}, C_{j_2} \in \mathcal{C} : \delta(\boldsymbol{\phi}_{j_1, j_2}) = 0\right\}$. For the development below it is useful to identify one cluster C_{j_0} within H_j called the *designated cluster* of H_j.

Definition 6. Say $\Delta(B)$ is *graphical* if it is zeroed and has the property that if any pairs of clusters $C_{j_1}, C_{j_2} \in H_j$ and $C_{k_1}, C_{k_2} \in H_k$ then for any j, k

$$\delta(\boldsymbol{\phi}_{j_1, k_1}) = \delta(\boldsymbol{\phi}_{j_2, k_2})$$

where $H_j, H_k \in B$ and $B \in \mathcal{B}$.

The condition that $\Delta(B)$ is graphical is a substantive one. However, it holds for all examples in this paper. If $\Delta(B)$ is graphical we can define a new *hyperseparation matrix* $\Delta_H(B)$ where the rows are labelled by the hyperclusters and the distances between hyperclusters are inherited from the transmission separations between any two clusters, one in each of the hyperclusters. Note that for a graphical $\Delta(B)$, $\Delta(B)$ is a function of $\Delta_H(B)$. So no information is lost through this transformation.

Clusters in the same hyperclusters can be interpreted as co-expressing in an appropriately coarse sense, in the context of our running example. Suppose two clusters have a large amplitude. Then they will lie in the same hypercluster if they have the same phase and shape profile, differing only in their amplitudes. This reflects a belief that two highly expressed genes are indistinguishable from each other in terms of their likely regulatory function if their amplitudes are different. If this is not so of course we can modify our inferences by adopting a more appropriate transmission matrix.

When $\Delta(B)$ is graphical, the transmission matrix allows us to define graphs which provide particularly useful summaries of putative regulatory relationships between hyperclusters.

Definition 7. For each $B \in \mathcal{B}$ a *regulatory graph* over B, $\mathcal{G}_B$ is any directed connected graph with no directed two cycles whose vertex set $V(\mathcal{G}_B) \triangleq \mathcal{H}_B$ and which has the property that whenever the directed edge $e(j_1, j_2) \in E(\mathcal{G}_B)$ then $\delta(\boldsymbol{\phi}_{j_1 \longrightarrow j_2}) \geq \delta(\boldsymbol{\phi}_{j_2 \longrightarrow j_1})$.

Note in particular that the direction of an edge in a regulatory graph is determined by the direction of the shorter transmission separation. The edges we include will be used to represent different causal hypotheses: see Section 5. A useful property of regulatory graphs is that we can annotate its vertices and edges to highlight important features of the regulatory process being described. Thus the hypercluster vertices can be annotated by the posterior estimate of its designated cluster. By the interpretations above all units within the hypercluster have a regulatory function potentially equivalent to a unit with this profile. In a similar way edges can be annotated by the directional distance δ between the clusters.

We next collate together the graphs $\{\mathcal{G}_B : B \in \mathcal{B}\}$ by making these the disconnected components of a new graph $\mathcal{G}_{\mathcal{B}}$.

Definition 8. Given a supracluster partition $\mathcal{B}$, let a *regulatory graph* (RG) $\mathcal{G}_{\mathcal{B}} = (V(\mathcal{G}_{\mathcal{B}}), E(\mathcal{G}_{\mathcal{B}}))$ be defined by setting

$$V(\mathcal{G}_{\mathcal{B}}) = \cup_{B \in \mathcal{B}} V(\mathcal{G}_B)$$
$$E(\mathcal{G}_{\mathcal{B}}) = \cup_{B \in \mathcal{B}} E(\mathcal{G}_B).$$

4.1. Annotating an extended regulation graph

By definition, given $\boldsymbol{\varphi} \in \Psi$, the distribution of all observations is then fully specified by the clusters and the superclusters $(\mathcal{C}, \mathcal{B})$ together with the associated parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ defined above. Recall that a BN has the property that its graph can be annotated by conditional probability tables. This means that – conditional on these parameters/conditional probabilities – the joint distribution can be unambiguously and fully described by the values assigned to these conditional probability tables. This is a useful property because the graph can be seen as expressing some important features of the distribution which, if necessary, can always be more fully expanded into a complete model. We next show that any RG which faithfully describes the beliefs of the scientist admits an analogous embellishment. The only difference is that the parameters annotating the vertices and edges of an RG are rather different to the conditional probability tables of a BN and customise to the different families of hypotheses associated with regulation rather than conditional independence. This embellishment is particularly elegant when it is possible for the regulatory group to admit an orthogonal decomposition of its transmission matrices. Then the group action can be separated into two distinct types of actions measured by two different components of $\boldsymbol{\phi}$.

Write $\boldsymbol{\phi} = (\boldsymbol{\phi}^V, \boldsymbol{\phi}^E)$. The first components $\boldsymbol{\phi}^V$ parametrise the first group action and determine the responsiveness of the receiving cluster relative to the donating cluster. The second components $\boldsymbol{\phi}^E$ determine the strength of transmission from the donating cluster. So, for example, $\boldsymbol{\phi}^E_{j_1 \to j_2}$ quantifies the nature of the putative regulatory role of one or more units in C_{j_1} on one or more units in C_{j_2} whilst $\boldsymbol{\phi}^0_j$ is determined by the responsiveness of cluster C_j relative to its seed.

Definition 9. Say L is *R-separable* if

$$L(\boldsymbol{\phi}_{j_1 \to j_2}) = L(\boldsymbol{\phi}^V_{j_1 \to j_2}, \boldsymbol{\phi}^E_{j_1 \to j_2}) = L_1(\boldsymbol{\phi}^V_{j_1 \to j_2}) L_2(\boldsymbol{\phi}^E_{j_1 \to j_2})$$

where L_1 is a function only of $\boldsymbol{\phi}^V_{j_1 \to j_2}$, L_2 is a function only of $\boldsymbol{\phi}^E_{j_1 \to j_2}$ and L_1 and L_2 commute.

A simple example of an R-separable transmission matrix is given in the running example below, where ϕ^V_j determines the amplitude of cluster C_j relative to its seed whilst $\phi^E_{j_1 \to j_2}$ is the phase transition from C_{j_1} to C_{j_2}.

Definition 10. Say the transmission matrix $\{\Delta(B) : B \in \mathcal{B}\}$ admits an *orthogonal decomposition* if it is possible to choose a parametrisation $\boldsymbol{\phi} = (\boldsymbol{\phi}^V, \boldsymbol{\phi}^E)$ of the transmission vector where

$$\delta(C_{j_1}, C_{j_2}) = 0 \quad \Leftrightarrow \quad \boldsymbol{\phi}^E(C_{j_1}) = \boldsymbol{\phi}^E(C_{j_2}).$$

When $\{\Delta(B) : B \in \mathcal{B}\}$ admits such an orthogonal decomposition $\boldsymbol{\phi}^V$ parametrises – together with parameters $\boldsymbol{\theta}$ used to describe the profile of a typical cluster in the hypercluster – the variety of shapes allowed within that hypercluster vertex. On the other hand $\boldsymbol{\phi}^E$ parametrises the relational parts of the model and so they can be *directly* linked to the edges of the RG. Because all our examples admit an orthogonal decomposition we will only consider such RG's in this paper.

Explicitly, the topology of an RG admitting an orthogonal decomposition can be unambiguously expanded into a full probability model by annotating its graph in the following way. First note that the supraclusters in $\mathcal{B}$ can be identified as the sets of clusters contained in each connected component of an RG $\mathcal{G}$. Now fix the hyperparameters $\boldsymbol{\varphi} \in \Psi$. Then by definition the joint marginal distribution of units of any hypercluster $H_0 \in \mathcal{H}_k$ containing the seed of the supracluster $B_k \in \mathcal{B}$ that it belongs to, is fully and unambiguously specified by:

1. the profile parameters $\boldsymbol{\theta}_k$ of the designated cluster (which is also a seed of the supracluster); and
2. the parameter vectors $\boldsymbol{\phi}_0^V$ of the non-seed clusters in H_0 which determine the distribution of each cluster in the seed hypercluster.

On the other hand the margins of the clusters in any $H_h \in \mathcal{H}_k$ not containing the seed of the supracluster $B_k \in \mathcal{B}$ that it belongs to, is fully and unambiguously specified by:

1. the transmission parameter vectors $\boldsymbol{\phi}_h^E$ which together with $\boldsymbol{\theta}_k$ fully determine the profile parameter vector of H_h; and
2. the parameter vectors $\boldsymbol{\phi}_h^V$ of the non-designated clusters in H_h which determine the distribution of the remaining profile parameter vectors.

The sets of parameters above can therefore be used to annotate the vertices of $\mathcal{G}$ – i.e. to fully describe the signature of the components in the regulatory mechanism which contain a message that might be transmitted.

We now use the edge parameters to represent the joint distribution of the whole regulation network. By definition, since – conditional on their parameters – all observations lying in different supraclusters of the RG are independent of each other we need only show how the joint distribution of the clusters in each supraclusters can be fully and unambiguously defined from annotating edge parameters. Then the joint distribution of the whole data set can be obtained by multiplying the corresponding densities together. Suppose that for each $B_k \in \mathcal{B}$ we annotate an edge from hypercluster H_{k_1} to $H_{k_2} \in B_k$ by $\boldsymbol{\phi}^E_{k_1 \to k_2}$ where $\boldsymbol{\phi}_{k_1 \to k_2} = \left(\boldsymbol{\phi}^V_{k_1 \to k_2}, \boldsymbol{\phi}^E_{k_1 \to k_2}\right)$ are the corresponding transmission parameters from the designated cluster in H_j to the designated cluster in H_{k_2}. Since L is a group, and all hyperclusters within a supracluster are deterministically related and connected in $\mathcal{G}$

$$\boldsymbol{\phi}(\mathcal{G}_{B_k}) \triangleq \left\{\boldsymbol{\phi}_{k_1 \to k_2} : (k_1, k_2) \in E(\mathcal{G}), H_{k_1}, H_{k_2} \in B_k\right\}$$

are sufficient to specify the dependence relationship between *all* pairs of designated clusters $H_{k_1}, H_{k_2} \in B_k$. Furthermore the fact that these are invertible deterministic relationships means that the joint distribution of all clusters in any given pair of hyperclusters is fully and unambiguously defined by the corresponding two vertex parameters of the hyperparameters and $\boldsymbol{\phi}(\mathcal{G}_{B_k})$. Again because all these pairs of relationships are deterministic the full joint distribution of the unit profiles in all clusters are fully specified by the vertex parameters above and $\boldsymbol{\phi}(\mathcal{G}_{B_k})$.

In our worked examples we illustrate how functions of such an annotation can highlight further useful information about the process whose broad qualitative framework is given in the RG. In particular these graphs can be presented to biologists as they are more familiar with these graphs and the express critical information they contain. These allow us to explicitly embellish an RG to differentiate explicitly important features of dependence in this domain, namely, clusters potentially coregulating with a seed cluster (red), being excited by a seed (black) or being inhibited by a seed (green). This motivates the following definition.

Definition 11. Given a supracluster partition $\mathcal{B}$, an *extended regulatory graph* (ERG) $\mathcal{G}_B$ for a supracluster B is a coloured mixed graph whose vertex set $V(\mathcal{G}_B) \triangleq \mathcal{C}$. There is a directed edge $e(j_0, k_0) \in E(\mathcal{G}_B)$ from $C_{j_0} \to C_{k_0}$ whenever $t(\boldsymbol{\phi}_{j_0 \longrightarrow k_0}) \geq t(\boldsymbol{\phi}_{k_0 \longrightarrow j_0})$ and $\delta(\boldsymbol{\phi}_{j_0,k_0}) \neq 0$ where C_{j_0} and C_{k_0} are the respective designated clusters of H_j and H_k. An edge between two seed clusters is coloured black if it is associated with a positive scale change, green if it has a negative change. All other (non-designated) clusters are connected by a directed red edge from its respective seed cluster.

In Section 6 we give some illustrations of ERG's where the vertices are further labelled by their profile signatures. Note that all ERG can be transformed into an RG. Furthermore note the all annotated RG's which are R-separable with its parameter vector $\boldsymbol{\phi}^V$ containing a scale parameter can be represented as an ERG. In the next section for simplicity and more generality all causal analysis is defined in terms of the RG and its annotation. However from the comment above we could

obviously also represent the graphical causal models within our running examples by defining the hypotheses in terms of an ERG. We illustrate this later in Section 6 where we also present graphs explicitly expressing phase changes.

4.2. Regulatory graphs and Bayesian networks

Conditioning on the hyperparameter vector $\boldsymbol{\psi}$ throughout in an RG allows us to make a strong link between the conditional independences implicit in a BN and those in an RG.

Theorem 12. *An acyclic regulatory graph $\mathcal{G}$ is a valid BN on component vectors $\{\boldsymbol{\theta}(C,\mathcal{B}), C \in \mathcal{C}\}$.*

Proof. If two designated clusters C_{j_1} to C_{j_2} are in unconnected components of $\mathcal{G}$ then they are in different supraclusters and so their associated profile parameters are independent of each other by definition. So without loss assume C_{j_1} and C_{j_2} are in the same supracluster. Then for $\mathcal{G}$ to be valid we must be able to assert that

$$\boldsymbol{\theta}(C_{j_1},\mathcal{B}) \amalg \left\{\boldsymbol{\theta}(C,\mathcal{B}) : C \in R_{j_2}(\mathcal{B})\right\} \mid \left\{\boldsymbol{\theta}(C,\mathcal{B}) : C \in Q_{j_2}(\mathcal{B})\right\}$$

where $Q_{j_2}(\mathcal{B})$, $R_{j_2}(\mathcal{B})$ are respectively the set of all parent and non-parents of C_{j_2} in $\mathcal{B}$. But since by definition $Q_{j_2}(\mathcal{B})$ is non-empty this is immediate since again by definition of a supracluster, given $(\boldsymbol{\phi}(\mathcal{C},\mathcal{B}),\boldsymbol{\varphi})$, $\boldsymbol{\theta}(C_{j_1},\mathcal{B})$ is a *function* of each of its individual parents. □

Thus if a regulatory graph is acyclic it is in particular a valid, if degenerate, BN on the profile parameters of the designated clusters in its hyperclusters. Of course, because by definition it also contains many functional relationships, the corresponding BN just expresses some of the distributional relationships implicit in the RG. Furthermore RG's are often cyclic so then the theorem above will not apply. However in this case note that if we take a spanning tree whose pattern is its skeleton and add directions to the edges of that tree so that the resultant BN is acyclic, then again that tree is a valid BN. So even in the cyclic case there is a corresponding logical link between the topology of the RG and a valid BN whose skeleton is a subgraph of the skeleton of the RG.

5. Tearing, causality and regulation graphs

The consequence of the results in the last section is that the class of statistically equivalent graphs is very large and it is not possible to differentiate them just on the basis of the data observed in one experiment. However the causal extension of BNs enables the modeller to distinguish BNs that are statistically equivalent by allowing the topology of the graph to express additional hypotheses about what might happen to the system were it subjected to certain types of control. We now show how an analogous collection of hypotheses allows us to discriminate the different topologies of otherwise statistically equivalent RG's.

There are two complementary types of control to which a typical regulatory process might be subjected and will form the basis of our semantics. These are analogous but not the same as those developed for BN's and their extensions to CBN's. The first is a manipulation called *tearing* which is specific to regulatory relationships and closely relates to their robustness to the destruction of a particular potential message passing hypercluster. The second control called *doing* is more closely analogous to an intervention as defined in a causal BNs. We describe both these types of control below. Henceforth we follow [15] and call the regulatory system as *idle* when no controls are applied.

5.1. Tearing

Definition 13. A set of hyperclusters $\mathcal{H}$ is said to be *torn* from a regulatory network represented by an RG $\mathcal{G}$ if the set of hyperclusters is forcibly removed so that it no longer can have a message passing function within the represented network.

Controlling a system by tearing clearly has a very close relationship with choosing to omit a cluster in the description of an idle system. Because of the deterministic nature of the embedded hypotheses, it is easily checked that, unlike for many other graphical models, the probability model represented by an RG omitting a subset of the vertices (here hyperclusters) often retains the integrity of the probability model. This is so provided that the associated RG remains connected: the annotating parameters of its edges and vertices simply being inherited from the larger system. Tearing performs an analogous change, but instead of simply ignoring a hypercluster, that cluster is actively knocked out.

Definition 14. Let $\mathcal{G}^{\mathcal{H}} \triangleq (V(\mathcal{G}^{\mathcal{H}}), E(\mathcal{G}^{\mathcal{H}}))$ denote the subgraph of a regulatory graph $\mathcal{G} \triangleq (V(\mathcal{G}), E(\mathcal{G}))$ where $V(\mathcal{G}^{\mathcal{H}}) = V(\mathcal{G})\backslash\mathcal{H}$ and $e \in E(\mathcal{G}^{\mathcal{H}})$ iff $e \in E(\mathcal{G})$ and e does not have an $H \in \mathcal{H}$ as an end point. Let $\left(\mathcal{C}^{\mathcal{H}}, \mathcal{B}^{\mathcal{H}}, \boldsymbol{\theta}^{\mathcal{H}}, \boldsymbol{\phi}^{\mathcal{H}}, \boldsymbol{\varphi}\right)$ denote the clusters in the hyperclusters $V(\mathcal{G}^{\mathcal{H}})$, their supraclusters and their associated vertex and edge parameters. We say that $\mathcal{G}^{\mathcal{H}}$ *inherits* the distribution of $\mathcal{G}$ when $\mathcal{C}^{\mathcal{H}}$ consists of all clusters in the hyperclusters $V(\mathcal{G})\backslash\mathcal{H}$, $\mathcal{B}^{\mathcal{H}}$ are the sets of clusters inherited by inclusion from $\mathcal{B}$, the profile parameters of the remaining clusters $\boldsymbol{\theta}(\mathcal{C}^{\mathcal{H}}, \mathcal{B}^{\mathcal{H}})$ and their associated marginal

distributions are equal to those labelling the same cluster within the larger vector $\boldsymbol{\theta}(\mathcal{C},\mathcal{B})$ and the values of parameters annotating edges and vertices in $E(\mathcal{G}^{\mathcal{H}})$ are copied from $\boldsymbol{\phi}(\mathcal{C},\mathcal{B})$.

Definition 15. Say an RG $\mathcal{G} \triangleq (V(\mathcal{G}), E(\mathcal{G}))$ is *collapsible* if tearing any set of hyperclusters $\mathcal{H}$ from the regulatory network gives rise to a new valid RG whose topology and distribution is defined by $\mathcal{G}^{\mathcal{H}} \triangleq (V(\mathcal{G}^{\mathcal{H}}), E(\mathcal{G}^{\mathcal{H}}))$ above which inherits its distribution from $\mathcal{G}$.

Note from our previous comments that whenever $\mathcal{G}^{H}$ has the same number of disconnected components as $\mathcal{G}$ a collapsible RG model will assign the same distribution as if we simply ignored these clusters in the supraclustering. However if the tearing splits a component of the graph (and hence its associated supracluster) into two or more disconnected subgraphs then the profile parameters of clusters in different fragments become (functionally and statistically) independent of each other, whilst still inheriting their marginal distributions from the original model.

On the other hand the effect of tearing is assumed to be entirely local. In particular the relationships and distributions of all hyperclusters not immediately linked to elements in $\mathcal{H}$ are unaffected by such an action. So tearing cuts communications between hyperclusters by knocking out their links but the remaining dependence structure and parameters remain unaffected.

The assumption of collapsibility of an RG allows us to discriminate the meaning of many RG's with different topologies which in the idle system are statistically equivalent. Furthermore the type of control expressed through tearing often has a clear physical meaning, and sometimes controls that are actually possible to enact in some future experiment. Recall that the skeleton of a mixed graph is the undirected graph obtained by replacing any directed edge by an undirected one.

Theorem 16. *Two collapsible RG's $\mathcal{G}_1$ and $\mathcal{G}_2$ are probabilistically equivalent under all controls only if they have the same skeleton.*

Proof. Let $\mathcal{H}(H,\mathcal{G})$ denote the set of hyperclusters connected to H by in-going or out-going edges, i.e. the neighbours of H in the skeleton of $\mathcal{G}$. Then unless $\mathcal{G}_1$ and $\mathcal{G}_2$ have the same skeleton there must exist an H for which $\mathcal{H}(H,\mathcal{G}_1) \neq \mathcal{H}(H,\mathcal{G}_2)$. Without loss assume that $\mathcal{H}(H,\mathcal{G}_1) \subsetneq \mathcal{H}(H,\mathcal{G}_2)$. Then tearing $\mathcal{H}(H,\mathcal{G}_2)$ in $\mathcal{G}_2$ will make H independent of all other hyperclusters, whilst in $\mathcal{G}_1$ this is not so. Therefore, the tearing control of the two graphs predict different collections of effects. □

5.2. Doing for regulatory graphs

To capture the directionality of a graph we need to imagine being able to perform another type of control inspired by the development of Causal Bayesian Networks (CBN) from BN's, where instead of imagining tearing away the clusters in a hypercluster, the scientist forces them all to transmit an artificial signal. In a rather different context [15] refers to this sort of control as doing. A plausible effect of imposing such a control in an RG is to give the message to the receiving children who then pass the message on as if it were naturally occurring. Notice that such a hypothesis is about a *direction* effect since parents $\mathcal{G}$ of a controlled hypercluster will only receive the artificial message if a directed path exists to it from the manipulated vertex via its children, and so allows us to differentiate unique sets of hypothesised about the functionality of the regulatory system through the chosen regulatory graph $\mathcal{G}$.

Each hypercluster has a set of identically located uncertain profiles of the same shape, or equivalently the vector of parameters $\{\boldsymbol{\theta}_C : C \in H\}$ corresponds to a vector amenable to *atomic* manipulation. The dependence relationships from one hypercluster to another are represented by a directed edge along which a message is passed. Therefore a potential definition of an atomic manipulation in the context of regulatory hypotheses might be one which inhibited all but the regulation of the manipulated hypercluster and then set the value of the profile to a particular value.

Without loss assume that we have transformed to a parametrisation of the problem where the manipulated hypercluster contains the seed cluster of the corresponding supracluster. This provokes the following definition. Let H be a hypercluster and H_R be one of its children or parents. Let (H, H_R) denote the set of children and parents of H which are not H_R.

Definition 17. Say that the hypercluster H_R in the RG $\mathcal{G}$ is controlled by *doing* H if we tear $\mathcal{H}(H,H_R)$ and then intervene to set the profiles $\{\boldsymbol{\theta}_C : C \in H\} = \left\{\widehat{\boldsymbol{\theta}}_C : C \in H\right\}$ in a way that is consistent with the definition of that hypercluster.

Definition 18. Call a RG $\mathcal{G}$ *causal* (CRG) if it is collapsible and has the additional property that for all H and each of its neighbours H_R, $\mathcal{G}^{\mathcal{H}(H,H_R)}$. When H_R is a parent of H then doing H to any $\left\{\widehat{\boldsymbol{\theta}}_C : C \in H\right\}$ that retains the definition of H as a hypercluster, produces the joint distribution of parameters as in $\mathcal{G}^{\mathcal{H}(H,H_R)\cup\{H\}}$ where $\mathcal{H}(H,H_R)\cup\{H\}$ is torn from the idle system. On the other hand when H_R is a child of H then the distributions of $\{\boldsymbol{\theta}_C : C \in H_R\}$ are different from those of $\mathcal{G}^{\mathcal{H}(H,H_R)\cup\{H\}}$ for at least one of the possible values of $\left\{\widehat{\boldsymbol{\theta}}_C : C \in H\right\}$ consistent with H being a hypercluster.

So essentially if $\mathcal{G}$ is weakly causal then doing will not affect the distribution of any of its parents if we isolate the relationship by tearing away the other vertices. On the other hand it is possible to affect the distribution of its children by changing the designated hypercluster to take a particular profile. Now we have the following result.

Theorem 19. *The topologies of two distinct CRG's are associated with different hypotheses about the underlying idle system and the associated hypotheses about the effect of the controls of tearing and doing.*

Each CRG is hypothesised to behave differently under the controls of tearing and doing.

Proof. Since each CRG is collapsible if two CRG's perform the same under control they must in particular perform the same under tearing which implies by the previous theorem that they must have the same skeleton. So different CRG's $\mathcal{G}_1$ and $\mathcal{G}_2$ must be only distinguished by the directions of their edges. But if they are topologically different then they must have at least two hyperclusters H and H_R where H_R is a child of H in $\mathcal{G}_1$ and a parent of H in $\mathcal{G}_2$. So by the definition above we can distinguish the two models by doing H over all possible values of the profiles of a designated cluster to $\{\widehat{\boldsymbol{\theta}}_C : C \in H\}$ – and all other clusters within the hypercluster consistently with this. In $\mathcal{G}_1$, for some value of $\{\widehat{\boldsymbol{\theta}}_C : C \in H\}$ this manipulation is believed to change the distribution of the profile of the designated cluster in H_R whilst in $\mathcal{G}_2$ it will not. □

Therefore by performing a set of subsequent experiments it is possible to determine whether the hypotheses within a given CRG are indeed true and to discriminate between different ones. Of course the hypothesis that a regulatory system is a CRG is even stronger than collapsibility. On the other hand the ideas it embodies are close to those a scientist might hold and certainly have a scientific meaning. It should be noted that microarray cluster experiments are often conducted so as to identify which genes can be turned off to prevent certain activities in an organism (for example growth).

5.3. From observational studies to causal hypotheses

The final question is what guidance should be given as to the MAP estimate for the CRG associated with different statistically equivalent RG's. Commonly, especially in biological application, there is a scientific hypothesis that the CRG might be a prior expected to have a higher probability if it is sufficiently sparse, given that certain clusters have a role and that certain functions must be fulfilled. This corresponds to the hypothesis that

1. there are only a few units that are actually controlling the system and that by knocking these out the regulation will be destroyed; and
2. that if one hypercluster regulates another then the transmission times will be short.

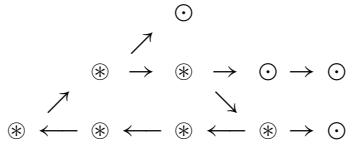
However in our chosen scientific context it is important to impose certain caveats about known features of the problem. To make this possible we partition hyperclusters into three categories: those *intrinsic* ($H \in \mathcal{I}$) – i.e. those containing a unit known to have a regulatory function, so having at least one parent and one child in $\mathcal{G}(\mathcal{B})$; the *juxtaposed* – hyperclusters ($H \in \mathcal{J}$) hypothesised not to contain a unit passing on other information, i.e. have no child in $\mathcal{G}(\mathcal{B})$; and the rest ($H \in \mathcal{K}$). Note that by setting $\mathcal{B} = \mathcal{K}$ no such conditioning caveats are imposed. Under such hypotheses – all other things being equal – the following class of CRGs are particularly attractive.

Definition 20. For each $B \in \mathcal{B}$ a *minimal regulatory graph (MRG)* $\mathcal{G}^R(\mathcal{B}, \mathcal{I}, \mathcal{J})$ is a regulatory graph with the property that there is an edge into all hyperclusters and into and out of all intrinsic hyperclusters $H \in B$ and where

$$\delta(\mathcal{G}_B) = \sum_{(H_{j_1}, H_{j_2}) \in E} \delta(\boldsymbol{\phi}_{j_1, j_2})$$

is minimised over all $\mathcal{G}(\mathcal{B}, \mathcal{I}, \mathcal{J})$, $\mathcal{G}(B, I)$ where E denotes the edge set of $\mathcal{G}(\mathcal{B}, \mathcal{I}, \mathcal{J})$.

We note that all cluster containing genes with a known regulatory function should be labelled as intrinsic. However with no such prior knowledge, and for exploratory purposes, we suggest that the statistician begins by assuming no clusters are intrinsic, as we do in our worked example. In this case it is easily checked that the MRG defined below becomes a minimum length spanning graph. In other circumstances it is appropriate to label all clusters as intrinsic. Notice that unlike BN's MRG's can be cyclic. Indeed they are always cyclic if all their vertices are intrinsic. However they are also constrained to be sparse, for example they can contain no acyclic complete subgraphs on 3 vertices.



An MRG with ⊛ nodes intrinsic

6. The circadian model selection graph in action

We now describe our model search algorithm and demonstrate how our method can be applied to a simulated and a real dataset and it can provide us with an unprecedented overview into complex data.

6.1. The implementation of a search for regulatory graphs

For a given cluster $\mathcal{C}$ let the hyperparameters $\boldsymbol{\phi}$ for all possible coarsenings $\mathcal{B}$ of $\mathcal{C}$ be appended to the original hyperparameters $\boldsymbol{\varphi}$ to define a new (huge) class of models. Then at least in principle, just as for Bayes cluster models, we can search the space algebraically and find the MAP model over the traversed candidate models. From a technical point of view this search is much harder than the corresponding search over the cluster space because the associated model space is orders of magnitude larger. The conclusions drawn from such a search must therefore be more tentative than in the analogous partition search, the evaluated models being an extremely sparse subset of the full class. On the other hand, at least for the types of regulatory models we have examined so far, it appears that the broad conclusions associated with the regulatory mechanisms from even naive extensions of standard cluster search algorithms can be remarkably robust. Our proposed initial fast procedure is as follows:

1. Initialise by using a standard search method – such as AHC – over a Bayesian cluster model maximising over a coarse grid of hyperparameters, as suggested by [7] to discover putative clusters of shapes $\mathcal{C}$.
2. Use statistics associated with the profiles annotating the clusters of $\mathcal{C}$ to provide a coarse putative matrix of estimates of the transition hyperparameters $\boldsymbol{\phi}_{j_1 \to j_2}$ between two clusters were one supracluster to exist. Assume that these estimates are appropriate for *any* coarsening $\mathcal{B}$.
3. Treating these estimates as known search over the space $(\mathcal{C}, \mathcal{B}, \boldsymbol{\phi})$ using a standard search method such as AHC (Agglomerative Hierarchical Clustering, see [10]) but on the superclusters rather than on the clusters. Note that the score of any candidate supracluster can be evaluated in closed form using the score function given above.
4. Retain the highest scoring model of this class which is the MAP model within those models traversed in the search.

It is of course possible to subsequently embellish the search in a number of ways where necessary. See [10] for details of the implementation of such methods in an analogous context.

6.2. An illustration using data simulated from a circadian model

To first illustrate our methods when we know the truth we simulated the longitudinal series of the 30 genes in each of 5 clusters, as in the [5] experiment, as shown in Fig. 2. These data y_{z_i} were simulated from

$$y_i = \lambda_{z_i} B(\theta_{z_1}) R_{z_i} \boldsymbol{\beta} + \varepsilon_i$$

where the subscript $z_i = j$ if individual i belongs to cluster j with $j = 1, \ldots, 5$ and $\varepsilon_i \sim \text{Normal}(0, 4)$. The matrix $B(\theta_{z_i})$ is a Fourier basis function as by Equation (2) and R_{z_i} is a rotation matrix, that is, a block diagonal matrix where each block is a matrix $M(k, \theta_{z_i})$ as

$$M(k, \theta_{z_i}) = \begin{pmatrix} \cos(k\theta) & -\sin(k\theta) \\ \sin(k\theta) & \cos(k\theta) \end{pmatrix}$$

with k corresponding to the index of the Fourier coefficients. The generating values for the rest of the parameters are given by

$$\lambda = (1, 1.1, 1, 1, -1, 0.5)$$

$$\theta = (6\pi/25, 9\pi/25, 14\pi/25, 18\pi/25, 23\pi/25)$$

$$\boldsymbol{\beta} = (1, 1, 0, 0.5, 0.5, 0.3, 3, 7.6, 0, 0.2, 2.3, 0.2, 0).$$

We specify a joint prior distribution for $(\boldsymbol{\beta}^{(k)}, \sigma_k^2)$. We adopt the convenient conjugate prior specification where

$$f(\boldsymbol{\beta}^{(k)} | \sigma_k^2) \equiv \text{N}(\boldsymbol{m}, \sigma_k^2 \boldsymbol{V}), \qquad f(\sigma_k^2) \equiv \text{IGamma}(\frac{a}{2}, \frac{b}{2})$$

$\boldsymbol{m}$ is $p \times 1$, $\boldsymbol{V} = v\boldsymbol{I}$ is $p \times p$ positive definite and symmetric, I is a $p \times p$ identity matrix and all other parameters are scalar. The parameters a, b and v are fixed, while the rest are random.

In fact, our method goes a step further and identifies relationships between these clusters too. For this dataset we are interested in three types of dependence between clusters:

1. Phase transition: clusters have similar shapes but there is a delay in the gene expression of one of the clusters. We take a Fourier transform of the data in the two clusters and we estimate the shift with Bayes factors over a grid of possible values of the rotation angle.
2. Inhibition: clusters have similar shapes but upside-down and slightly shifted. This corresponds to multiplying by -1 and also have a phase transition.
3. Amplitude change: clusters have similar shapes but different magnitude. This corresponds to a cluster being multiplied by a factor λ such that $\lambda > 0$.

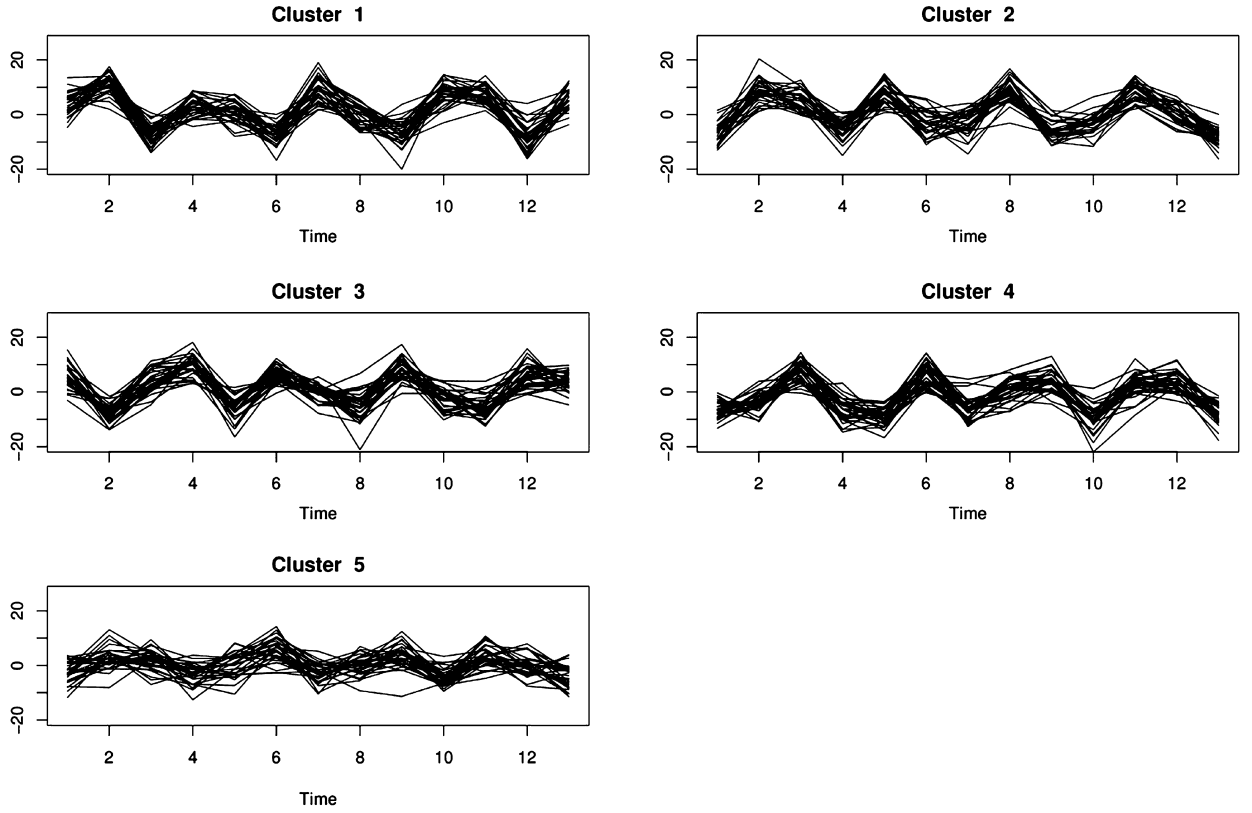


Fig. 2. The five simulated clusters identified by our method as described in Section 6.2.

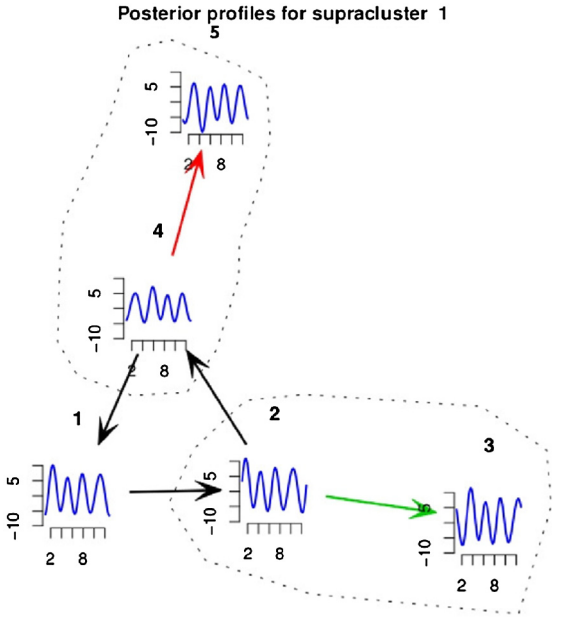


Fig. 3. The posterior means of the 5 simulated clusters and the relationships between them. The black arrow represents a phase transition, a green arrow represents inhibition and a red arrow represents amplitude change. Two hyperclusters are identified by the dashed lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The ERG of the MRG between the 5 clusters of the MAP estimate of this model is given in Fig. 3. Note that this depicts the two most important features of the data relative to this type of regulation: the expression patterns of different clusters as vertices and the most likely message passing links (their colours distinguishing excitation, inhibition and equivalent

Table 2
Standard deviations of the mean of the posterior distribution of all elements of $\boldsymbol{\beta}$ for 50 runs with values of v between 1 and 1,000.

	Standard deviation
β_1	0.0182
β_2	0.0235
β_3	0.0355
β_4	0.0047
β_5	0.0415
β_6	0.0813
β_7	0.0536
β_8	0.2563
β_9	0.1323
β_{10}	0.0463
β_{11}	0.0966
β_{12}	0.0078
β_{13}	0.0202

expression). The method successfully reconstructs the generating process which places all the clusters in one supracluster, something that is really difficult to do by eye even for this small number of units. Plots like this are easy to draw automatically in R.

We also studied the sensitivity of the partition and the posterior distribution of the parameters to different choices of the hyperparameters. We found our results to be very robust. We show here the results of 50 runs with values of v between 1 and 1,000. The resulting clustering was robust and the standard deviations of the mean of the posterior distribution of all elements of $\boldsymbol{\beta}$ were small, as shown in Table 2.

6.3. An application to an experiment on circadian gene regulation

We apply our method on the original experiment by [5], using the 175 high amplitude clusters discovered in the Bayes cluster analysis reported in [10]. In [5] the gene expression of 22,810 genes of the plant *Arabidopsis thaliana* was measured by Affymetrix microarrays at $T = 13$ time points over two days. Constant white light was shone on the plants for 26 hours before the first microarray was taken and the light remained on for the rest of experiment. Samples were taken every four hours. Thus, there are two cycles of data (13 time points) for each of the 22,810 genes available on the *Arabidopsis* microarraychip. The aim was to identify the genes which may be connected with the circadian clock of the plant. [10] analysed this dataset by proposing a search algorithm which is guided by the scientific interest of each cluster. We apply our method to the 175 high amplitude clusters discovered in the Bayes cluster analysis reported in [10].

To demonstrate our method without the use of existing contextual information, in this study, we chose the default options and labelled no cluster as intrinsic. Our AHC method enabled us to discover a MAP model with many supraclusters, the majority of which were singletons and therefore less likely to be communicating with each other. However several of the supraclusters were not trivial. The most pertinent nature of the most likely dependence relationships are evocatively expressed through these embellishments of the associated RG.

However this is not just a useful depiction of an apposite well-supported statistical model. If we are prepared to allow that the process is driven by a CRG and that the MAP model that we have discovered is indeed generating the idle process, then identifying the disconnected components of the system allows us to immediately make assertions about the impact of various controls we might apply to this regulatory process – just as we can were we to believe the model was a causal extension of a BN. In the context of microarrays, the objective of clustering is to identify patterns among the data and decide which genes to focus on in further, more gene-specific, experiments. It is therefore necessary for the scientist to make such causal conjectures about the effect of controls available to her on the expressions reflecting the underlying regulatory process she studies. These conjectures can be universal or nuanced by evoking ideas of parsimony.

First there are deductions that will apply to *any* CRG whose RG is in the same equivalence class as the discovered MAP MRG. These deductions are analogous to those provided by colliders in a BN search [15,18]. For example if we were to plan an experiment which was to tear out all the genes in a trivial cluster – i.e. one that was not connected by an edge to any other cluster – then under the causal extension of the model this would not change the relationships between the genes in the other clusters.

The second type of conjectures are those that need to evoke ideas of simplicity as reflected in sparser RCGs. In Fig. 4 we illustrate this on supracluster B which is the MRG of the MAP estimate and compare this with two alternative models (Figs. 5 and 6) equivalent in the uncontrolled system. Thus for example, let us consider the three representations of supracluster B as B1, B2 and B3. If the scientist sets up an experiment to test B1, she could tear hypercluster $H(3)$ and control hypercluster $H(4)$, as in Fig. 7(a). In this case she could test the hypothesis that hypercluster $H(4)$ is independent of the remaining ones. If the scientist had a different set of beliefs, as in B2 or B3, tearing $H(3)$ and controlling $H(4)$ would allow her to confirm whether there really is a causal relationship between $H(4)$ and $H(10)$.

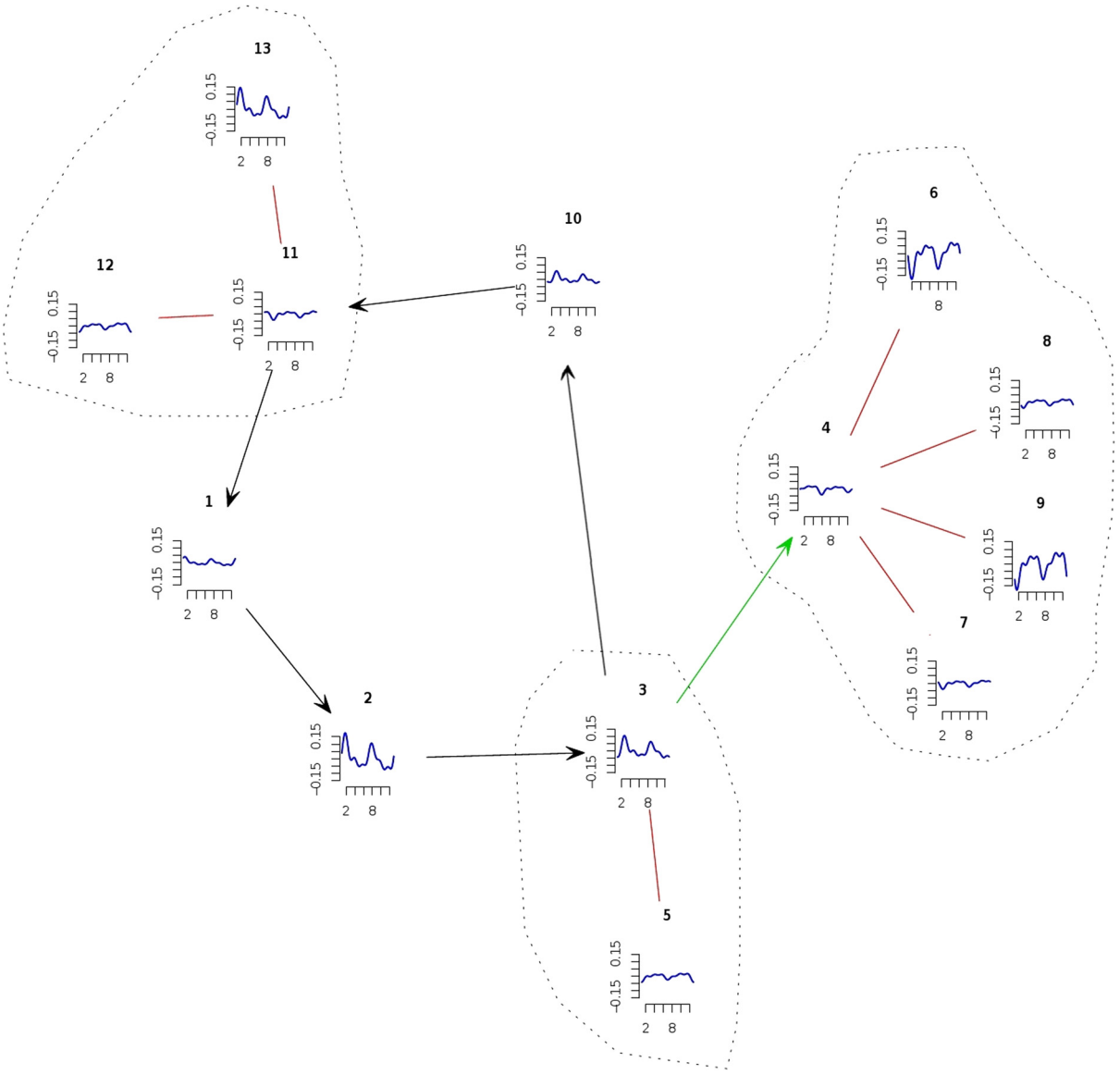


Fig. 4. The posterior means of the 13 clusters which belong to supracluster B and the relationships between them. The black arrow represents a phase transition, a green arrow represents inhibition and a red arrow represents amplitude change. Three hyperclusters are identified by the dashed lines. Hypercluster $H(3)$ includes clusters 3 and 5; hypercluster $H(4)$ includes clusters 4, 6, 7, 8, 9; hypercluster $H(11)$ includes clusters 11, 12 and 13. The arrows identify the B1 representation of supracluster B. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The MRG makes additional assertions about the process which cannot be deduced from statistical considerations at all. The idea is that manipulating the far right hypercluster to take a given value would have no effect on the underlying mechanism. If this were not thought plausible an equivalent graph with the mechanism depicted in Figs. 5 and 6 could be conjectured.

Of course such causal analyses are highly conjectural. In particular any deductions like the ones above depend heavily on the scientists' beliefs about how regulation is evidenced through the cluster profiles. In this example the hypothesis that the evidenced relationship can be expressed simply through amplitude and phase changes is a heroic one and could usefully be nuanced by allowing a low pass filter to relate the donating cluster profile to the receiving one. This would allow some of the superclusters to merge. However these sorts of adjustments are possible: we simply use a different richer class of transmission matrices to describe the relationships, and conditional conjugacy is still preserved albeit with some extra hyperparameters to estimate. In the context of the models we are currently analysing, the output arising from these rather naive classes appears to provide considerable new insights about the nature of the regulatory relationships between these genes.

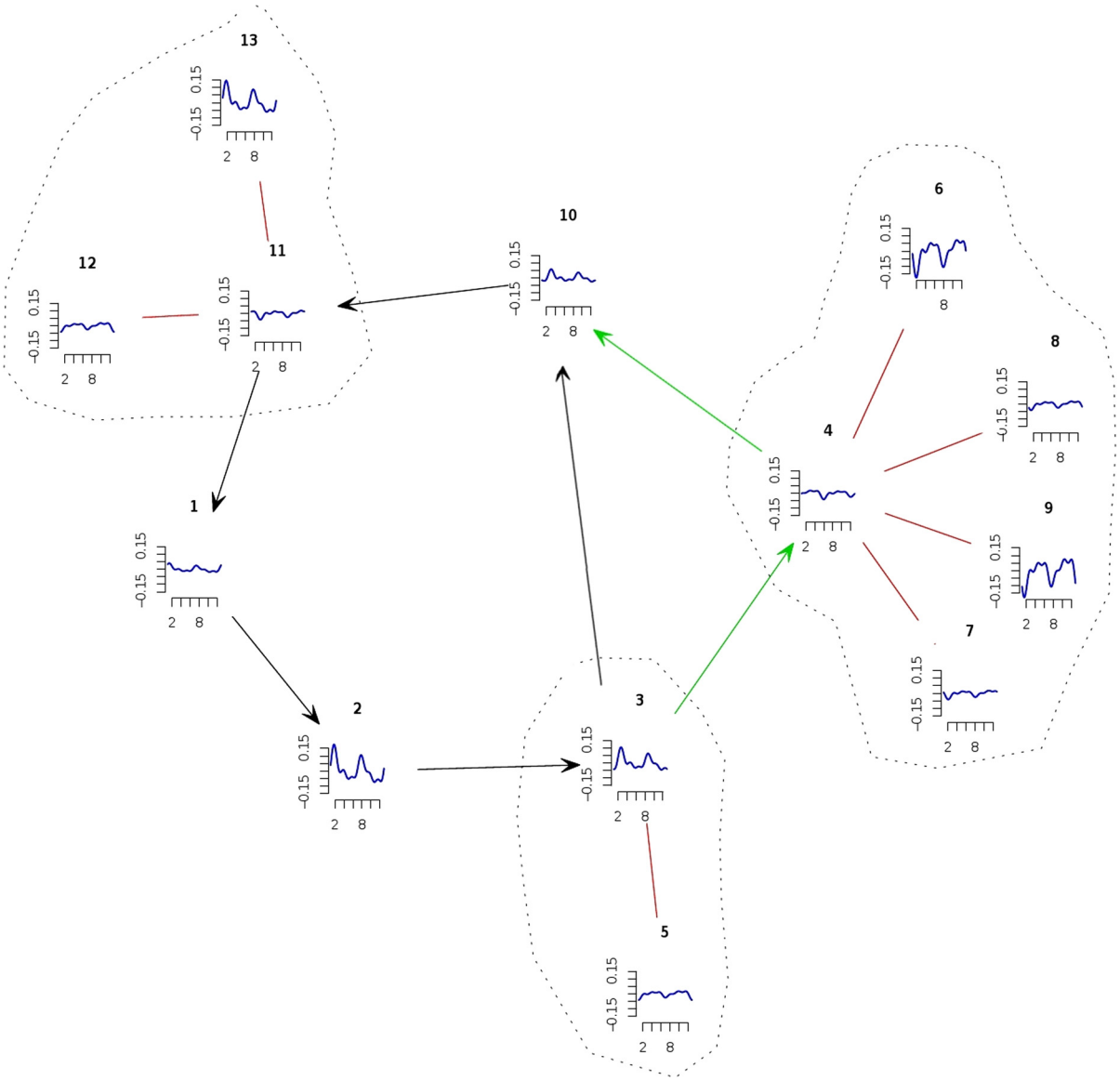


Fig. 5. The posterior means of the 13 clusters which belong to supracluster B and the relationships between them. The black arrow represents a phase transition, a green arrow represents inhibition and a red arrow represents amplitude change. Three hyperclusters are identified by the dashed lines. Hypercluster $H(3)$ includes clusters 3 and 5; hypercluster $H(4)$ includes clusters 4, 6, 7, 8, 9; hypercluster $H(11)$ includes clusters 11, 12 and 13. The arrows identify the B2 representation of supracluster B. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

7. Discussion

We conclude the paper with three brief remarks. First, because we intend the methods below to be exploratory, they needed to be simple. At the moment we are using an agglomerative hierarchical clustering algorithm together with a grid search over hyperparameters, similar to the one that [7] used in their cluster search method. This can obviously be improved [1]. Furthermore once a small class of candidate models has been found it is then sensible to invest time in refining the parameter distributions of the candidate models. We note that because our models are all formal Bayesian models and exhibit certain conditional conjugacy properties numerical methods are feasible to refine our analyses using slower but more accurate numerical methods. In particular we are currently developing methods to stochastically estimate transmission hyperparameters using ideas from [6,14].

However, even with these developments, of course the analysis we describe in this paper would still be exploratory in nature. In particular like many analyses of this type it ignores the existence of many sources of uncertainty – here associated with the inferred graph and the partition of clusters it finally adopts – which from a formal point of view need to be folded

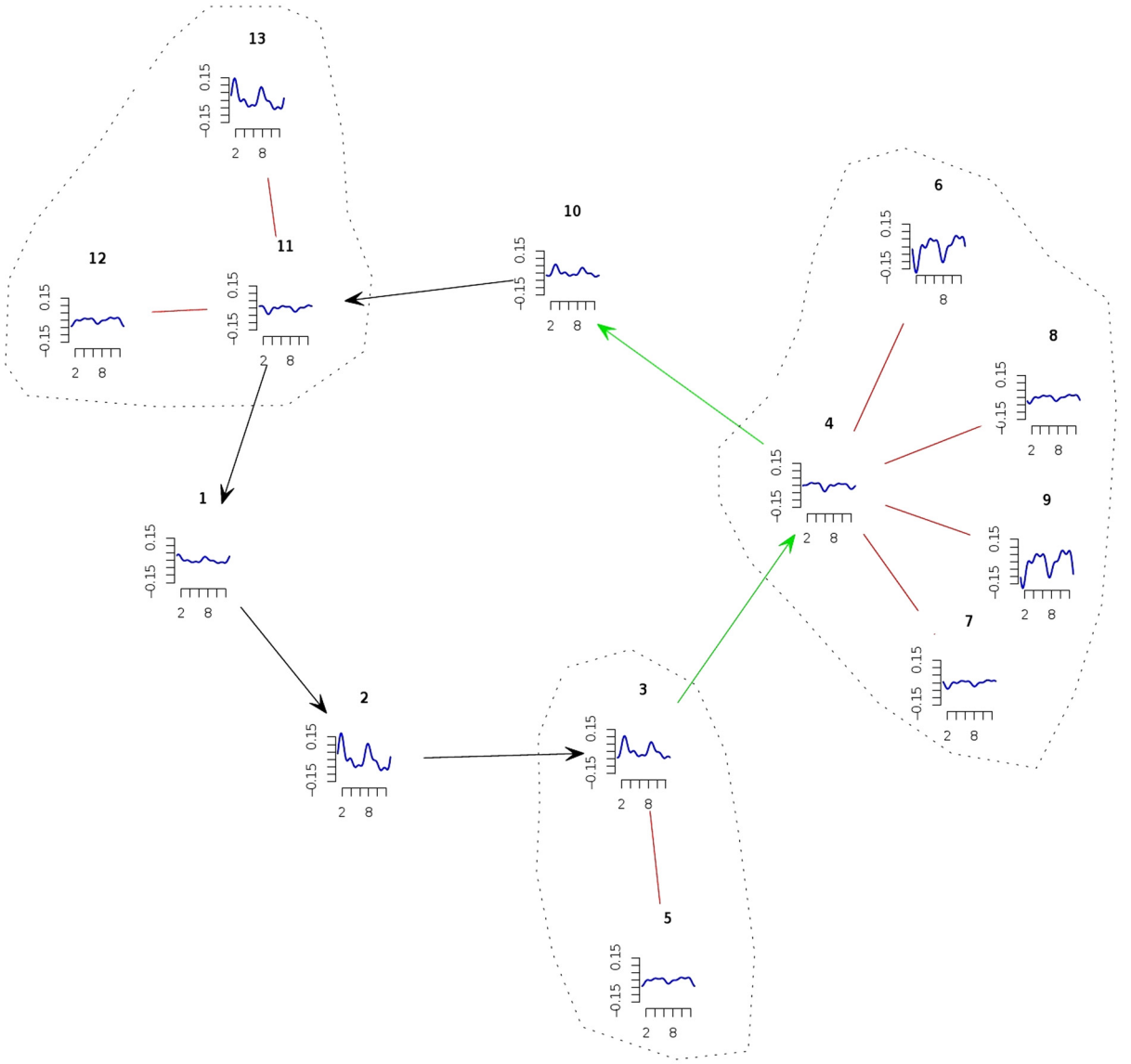


Fig. 6. The posterior means of the 13 clusters which belong to supracluster B and the relationships between them. The black arrow represents a phase transition, a green arrow represents inhibition and a red arrow represents amplitude change. Three hyperclusters are identified by the dashed lines. Hypercluster $H(3)$ includes clusters 3 and 5; hypercluster $H(4)$ includes clusters 4, 6, 7, 8, 9; hypercluster $H(11)$ includes clusters 11, 12 and 13. The arrows identify the B3 representation of supracluster B. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

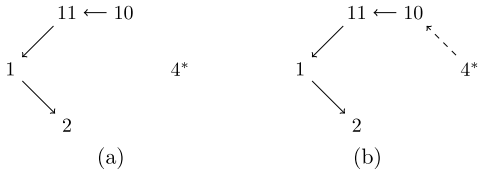


Fig. 7. Examples of manipulations of supracluster B.

into the analysis. The omission of these sources should always be born in mind and the depicted relationships therefore seen simply as indicative of possible regulatory or causal relationships for future closer scrutiny and investigation and not as firmly and formally based inferences. Our analyses are nevertheless proving useful as providing an evocative depiction to domain experts of some of the features appearing in the data.

Second, here we have focused on just one type of regulatory process: circadian regulation. There are many other types of regulation. Many of these appear to have an associated group action albeit one different from the scale rotations illustrated above. When this is so obviously our methods can be adjusted to define supraclusters and transmission matrices exactly analogous to those defined above but customised to these different hypothesised mechanisms expressed in this new group. We plan to explore such examples in future papers.

Finally with reference to the circadian experiments above, we are currently comparing experiments from different laboratories performing replicate experiments in order to reconcile their results. Previous analyses have tended to demonstrate that cluster containment across genes is rather disappointingly inconsistent. However in our preliminary analyses it appears that the qualitative structure of their associated regulations graphs are much more comparable. This suggests that the mechanism itself is more resilient than the actual identities of genes contained in sets of message passers and their co-expressing genes. We are working with biologists and using these graphical methods to compare the regulatory processes behind different species of plant. We will report these developments in a later paper.

The analysis of the properties of this new class of graphical models is still in its infancy. However we hope that we have demonstrated in this paper how graphical models which embody in their topology the probabilistic structure implied by a specific scientific domain and natural hypotheses therein can be built. Furthermore it is feasible to do this for the data-rich problems currently faced by scientists and these models provide the basis of useful and formally justifiable summary graphs of the processes under study.

Acknowledgements

Silvia Liverani acknowledges support from the Leverhulme Trust (ECF-2011-576) and EPSRC (EP/D063485/1). The authors would like to thank Andrew Millar for his comments on an early version of this paper.

References

[1] H.A. Chipman, E.I. George, R.E. McCulloch, Bayesian treed models, Mach. Learn. 48 (1–3) (2002) 299–320.
[2] H. de Jong, Modeling and simulation of genetic regulatory systems: a literature review, J. Comput. Biol. 9 (1) (2002) 67–103.
[3] D.G.T. Denison, C.C. Holmes, B.K. Mallick, A.F.M. Smith, Bayesian Methods for Nonlinear Classification and Regression, Wiley Series in Probability and Statistics, John Wiley and Sons, 2002.
[4] A. Dobra, C. Hans, B. Jones, J. Nevins, G. Yao, M. West, Sparse graphical models for exploring gene expression data, J. Multivar. Anal. 90 (1) (2004) 196–212.
[5] K.D. Edwards, P.E. Anderson, A. Hall, N.S. Salathia, J.C.W. Locke, J.R. Lynn, M. Straume, J.Q. Smith, A.J. Millar, FLOWERING LOCUS C mediates natural variation in the high-temperature response of the *Arabidopsis* circadian clock, Plant Cell 18 (2006) 639–650.
[6] P. Green, K. Mardia, Bayesian alignment using hierarchical models, with applications in protein bioinformatics, Biometrika 93 (2) (2006) 235.
[7] N.A. Heard, C.C. Holmes, D.A. Stephens, A quantitative study of gene regulation involved in the immune response of Anopheline Mosquitoes: an application of Bayesian hierarchical clustering of curves, J. Am. Stat. Assoc. 101 (473) (2006) 18–29.
[8] J.W. Lau, P.J. Green, Bayesian model-based clustering procedures, J. Comput. Graph. Stat. 16 (3) (2007) 526.
[9] S. Lauritzen, Graphical Models, Oxford University Press, 1996.
[10] S. Liverani, P.E. Anderson, K.D. Edwards, A.J. Millar, J.Q. Smith, Efficient utility-based clustering over high dimensional partition spaces, J. Bayesian Anal. 4 (3) (2009) 539–572.
[11] A. Monnier, S. Liverani, R. Bouvet, B. Jesson, J. Mosser, F. Corellou, J.Q. Smith, F.-Y. Bouget, Light-regulated transcriptional networks in Ostreococcus provides insight into the biology and physiology of the marine eukaryotic picophytoplankton, BMC Genomics 11 (192) (2010).
[12] K. Murphy, S. Mian, Modelling gene expression data using dynamic Bayesian networks, Technical report, University of California, Berkeley, 1999.
[13] M.E.J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2) (2003) 167–256.
[14] C.J. Oates, J.Q. Smith, S. Mukherjee, J. Cussens, Exact estimation of multiple directed acyclic graphs, Stat. Comput. (2015) 1–15.
[15] J. Pearl, Causality: Models, Reasoning and Inference, second ed., Cambridge University Press, Cambridge, 2009.
[16] S. Ray, B. Mallick, Functional clustering by Bayesian wavelet methods, J. R. Stat. Soc., Ser. B 68 (2) (2006) 305–332.
[17] J.Q. Smith, P.E. Anderson, S. Liverani, Separation measures and the geometry of Bayes factor selection for classification, J. R. Stat. Soc., Ser. B 70 (5) (2008) 957–980.
[18] P. Spirtes, C. Glymour, R. Scheines, Causation, Prediction, and Search, vol. 81, The MIT Press, 2000.
[19] D. Telesca, P. Müller, G. Parmigiani, R.S. Freedman, et al., Modeling dependent gene expression, Ann. Appl. Stat. 6 (2) (2012) 542–560.
[20] E.P. van Someren, L.F.A. Wessels, E. Backer, M.J.T. Reinders, Genetic network modeling, Pharmacogenomics 3 (4) (2002) 507–525.
[21] C. Zhou, J.C. Wakefield, L.L. Breeden, Bayesian analysis of cell-cycle gene expression data, in: K.-A. Do, P. Müller, M. Vannucci (Eds.), Bayesian Inference for Gene Expression and Proteomics, Cambridge University Press, 2006, pp. 177–200.