Inferring dynamic genetic networks with low order independencies

Sophie Lèbre

Inferring dynamic genetic networks with low order independencies

Sophie Lèbre

19th April 2007

Université d'Evry-Val-d'Essone
CNRS UMR 8071
INRA 1152,
Laboratoire Statistique et Génome
523 place des Terrasses
91000 Evry
France.
E-mail: sophielebre@hotmail.com

Abstract

In this paper, we propose a novel inference method for dynamic genetic networks which makes it possible to face with a number of time measurements n much smaller than the number of genes p. The approach is based on the concept of low order conditional dependence graph that we extend here in the particular case of Dynamic Bayesian Networks. Most of our results are based on the theory of graphical models associated with the Directed Acyclic Graphs (DAGs). In this way, we define a DAG $\tilde{\mathcal{G}}$ which describes exactly the *full order conditional dependencies* given the past of the process. Then, to face with the large p and small n estimation case, we propose to approximate DAG $\tilde{\mathcal{G}}$ by considering low order conditional independencies. We introduce partial q^{th} order conditional dependence DAGs and analyze their probabilistic properties. In general, DAGs $\mathcal{G}^{(q)}$ differ from $\tilde{\mathcal{G}}$ but still reflect relevant dependence facts for sparse networks such as genetic networks. By using this approximation, we set out a non-bayesian inference method and demonstrate the effectiveness of this approach on both simulated and real data analysis. The inference procedure is implemented in a R package available on request.
Keywords : conditional independence, Bayesian networks, directed acyclic graphs, dynamic networks inference, time series modeling.

1 Introduction

The development of microarray technology allows to simultaneously measure the expression levels of many genes at a precise time point. Thus it has become possible to observe gene expression levels across a whole process like cell cycle or response to radiation or several treatments. The objective is now to recover gene regulation phenomena from this data. We are looking for simple relationships such as "gene i activates gene j". But we also want to capture more complex scenarios such as auto-regulations, feed-forward loops, multi-component loops... as described by Lee et al. [17] in the transcriptional regulatory network of the yeast Saccharomyces cerevisiae.

To such an aim, we both need to accurately take into account temporal dependencies and to face with the dimension of the problem as the number p of observed genes is much higher than the number n of observation time points. Moreover we know that most of the observed genes are not taking part to the evolution of the system. So we want to determinate which are the few "active" agents, that are the agents being responsible for the evolution of the system and what are the relationships between them. In short, we want to infer a network representing the dependence relationships which govern a multiple elements-system from the observation of this system across short time series.

Such gene networks were firstly described by using static modeling and mainly non oriented networks. One of the first tools used to describe interaction between genes is the *relevance network* [3] or *correlation network* [34]. Better known as *covariance graph* [4] in the graphical models theory, this non directed graph describes the pair-wise correlation between genes. Its topology is derived from the covariance matrix between the gene expression levels; an undirected edge is drawn between two variables whenever they are correlated. Nevertheless, the correlation between two variables may come from the linkage with other variables. This creates spurious edges due to indirect dependence relationships.

Consequently, great interest has been taken in the *concentration graph* [16], also called *covariance selection* model, which describes the *conditional* dependence structure between gene expression in Graphical Gaussian Models (GGMs). Let $Y = (Y_i)_{1 \leq i \leq p}$ be a multivariate Gaussian vector representing the expression levels of p genes. An undirected edge is drawn between two variables Y_i and Y_j whenever they are conditionally dependent given the remaining variables. The standard theory of estimation in GGMs [42], [16] can be exploited only when the number of measurements n is much higher than the number of variables p. This ensures that the sample covariance matrix is positive definite with probability one. Nevertheless, in most of the microarray gene expression data, we have to cope with the opposite situation ($n << p$). Thus, the growing interest for 'small n, large p' furthered the development of numerous alternatives (Schäfer and Strimmer [28] [29] , Waddell and Kishino [40] [39], Toh and Horimoto [37] [38], Wu et al. [46], Wang et al. [41]). Even though concentration graphs allow to point out some dependence relationships between genes, they do not offer an accurate description of the interactions. Firstly, no direction is given to the interactions. Secondly, some motifs containing cycles cannot be properly represented (see Figure 1).

Contrary to the previous undirected graphs, Bayesian networks (BNs) [10] model directed relationships. Based on a probabilistic measure, a BN representation of a model is defined by a Directed Acyclic Graph

Figure 1: A biological regulation motif (left) and the corresponding concentration graph (right). For all $i \geq 3$, Y_i is a Gaussian variable representing the expression level of gene G_i. Some cycles cannot be represented on the concentration graph.

(DAG) and the set of conditional probability distributions of each variable given its parents in the DAG [24]. Then the theory of graphical models [42, 6, 16] allows to derive conditional independencies from this DAG. However the acyclicity constraint in static BNs is a serious restriction given the expected structure of genetic networks.

Here comes the interest of Dynamic Bayesian networks (DBNs) first introduced for the analysis of gene expression time series by Friedman et al. [11] and Murphy and Mian [21]. In DBNs, a gene is not anymore represented by a single vertex but by as much vertices as time points in the experiment. A dynamic network (Figure 2) can then be obtained by *unfolding in time* the initial cyclic motif in Figure 1 (left). The directions according to the time guarantees the acyclicity of this dynamic network and consequently allows to define a Bayesian network. The signs +/- showing the type of regulation in the biological motif do not appear in this DAG but they can be derived from model parameters estimates.

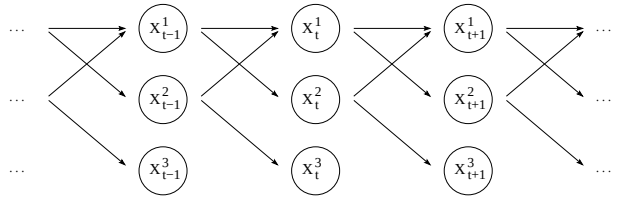

Figure 2: Dynamic network equivalent to the regulation motif in Figure 1 (left). Each vertex X_t^i represents the expression level of gene G_i at time t. This graph is acyclic and allows to define a Bayesian network.

The very high number p of genes simultaneously observed raises a dimension problem. Moreover, a large majority of time series gene expression data contain no or very few repeated measurement(s) of the expression level of the same gene at a given time. Hence, we assume that the process is *homogeneous* across time. This consists of considering that the system is governed by the same rules during the whole experiment. Consequently, the temporal dependencies are homogeneous: any edge is present during the whole process. This is a strong assumption which is not necessarily satisfied. Nevertheless, this condition is necessary to carry out estimation. Indeed, in that case, we observe $n-1$ repeated measurements of the expression level of each gene at two successive time points.

Up to now, various DBN representations based on different probabilistic models have been proposed (discrete models [22, 47], multivariate auto-regressive process [23], State Space or Hidden Markov Models [25, 45, 26, 2], nonparametric additive regression model [12, 13, 15, 35]). See also Kim et al. [14] for a review of such models. Facing with as much diversity, we expose here sufficient condition such that a model admits a DBN representation and we set out a straight interpretation in terms of dependencies between variables by using the theory of graphical models for DAGs. Our DBN representation is based on a DAG $\tilde{\mathcal{G}}$ (e.g. like the DAG of Fig. 2) which describes exactly the full order conditional dependencies given all the remaining *past* variables (see section 2). This approach extends the principle of the concentration graph showing conditional independencies to the dynamic case.

Even under homogeneity assumption, which enables to use the different time points as repeated measurements of the same process, we still have to deal with the 'curse of dimension' to infer the structure of DAG $\tilde{\mathcal{G}}$.

The difficulty lies in facing with the large p and small n estimation case. Several inference methods have been proposed for the estimation of the topology of the various graphs quoted above. Among others, Murphy [20] implemented several bayesian structure learning procedures for dynamic models in the open-source Matlab package BNT (Bayes Net Toolbox); Ong et al. [22] reduce the dimension of the problem by considering prior knowledge; Perrin et al. [25] use an extension of the linear regression; Wu et al. [45] use factor analysis and Beal et al. [2] develop a variational Bayesian method; Zou and Conzen [47] limit potential regulators to the genes with either earlier or simultaneous expression changes and estimate the transcription time lag; Opgen-Rhein and Strimmer [23] recently proposed a model selection procedure based on an analytic shrinkage approach. However, a powerful approach based on the consideration of zero- and first-order conditional independencies recently gained attention to model concentration graphs. When $n << p$, Wille et al. [44, 43] propose to approximate the concentration graph by the graph $\mathcal{G}_{0-1}$ describing zero- and first-order conditional independence. An edge between the variables Y_i and Y_j is drawn in the graph $\mathcal{G}_{0-1}$ if and only if, zero- and first-order correlations between these two variables both differ from zero, that is, if the next conditions are satisfied,

$$Corr(Y_i, Y_j) \neq 0 \quad \text{and} \quad \forall k \in \{1, ..., p\} \backslash \{i, j\},\ Corr(Y_i, Y_j | Y_k) \neq 0, \tag{1}$$

where $Corr(Y_i, Y_j|Y_k)$ is the partial correlation between Y_i and Y_j given Y_k. Hence, whenever the possible correlation between two variables Y_i and Y_j can be entirely explained by the effect of some variable Y_k, no edge is drawn between them.

This procedure allows a drastic dimension reduction: by using first order conditional correlations, estimation can be carried out accurately even with a small number of observations. Even if the graph of zero- and first-order conditional independence differs from the concentration graph in general, it still reflects some measure of conditional independence. Wille et al. show through simulations that the graph $\mathcal{G}_{0-1}$ offers a good approximation of sparse concentration graphs and demonstrate that both graphs even coincide exactly if the concentration graph is a forest ([43], Corollary 1). This approach has also been used by Magwene and Kim [18] and de la Fuente et al. [5] for estimating non-directed gene networks from microarray gene expression of the yeast *Saccharomyces cerevisiae*. Roverato and Castelo [27] investigate such non directed q^{th} order partial independence graphs for $q \geq 1$ and expose a sharp analysis of their properties. In this paper, we extend this approach by defining q^{th} order order conditional dependence DAGs $\mathcal{G}^{(q)}$ for DBN representations. Then, by basing on our results on these low order conditional dependence DAGs, we propose a novel inference method for dynamic genetic networks which makes it possible to face with the 'small n, large p' estimation case.

The remainder of the paper is organized as follows. In section 2, we expose sufficient conditions for a DBN modeling of time series describing temporal dependencies. We notably show the existence of a minimal DAG $\tilde{\mathcal{G}}$ which allows such a DBN representation. To reduce the dimension of the estimation of the topology of $\tilde{\mathcal{G}}$, we propose to approximate $\tilde{\mathcal{G}}$ by q^{th} order conditional dependence DAGs $\mathcal{G}^{(q)}$ and analyze their probabilistic properties in section 3. From conditions on the topology of $\tilde{\mathcal{G}}$ and faithfulness assumption, we establish inclusion relationships between both DAGs $\tilde{\mathcal{G}}$ and $\mathcal{G}^{(q)}$. In section 4, we exploit our results on DAGs $\mathcal{G}^{(q)}$ to develop a non-bayesian estimation procedure. Finally, validation is obtained on both simulated and real data in section 5. We notably expose our results for the analysis of two microarray time course data sets: the Spellman's yeast cell cycle data [32] and the diurnal cycle data on the starch metabolism of Arabidopsis Thaliana collected by Smith et al. [31].

2 A DBN representation

Let $P = \{1 \leq i \leq p\}$ describe the set of observed genes and $N = \{1 \leq t \leq n\}$ the space of observation times. In this paper, we consider a discrete-time stochastic process $X = \{X_t^i; i \in P, t \in N\}$ taking real values and assume the joint probability distribution $\mathbb{P}$ of the process X has density f with respect to Lebesgue measure on $\mathbb{R}^{p \times n}$. We denote by $X_t = \{X_t^i; i \in P\}$ the set of the p random variables observed at time t and $X_{1:t} = \{X_s^i; i \in P, s \leq t\}$ the set of the random variables observed before time t.

In this section, we expose sufficient conditions under which the probability distribution $\mathbb{P}$ admits a BN representation according to a dynamic network (e.g. in Figure 2). The main result is set out in Proposition 3; we show that it exists a BN representation according to a minimal DAG $\tilde{\mathcal{G}}$ whose edges describe exactly the set of direct dependencies between successive variables X_{t-1}^j, X_t^i given the past of the process. For an illustration,

Table 1: Notations

$P = \{1 \leq i \leq p\}$	set of observed genes,
$P_i = P \backslash \{i\}$	
$N = \{1 \leq t \leq n\}$	time space,
$X = \{X_t^i; i \in P, t \in N\}$	stochastic process (gene expression levels time series),
$\mathcal{G} = (X, E(\mathcal{G}))$	a DAG whose vertices are defined by X and edges by $E(\mathcal{G}) \subseteq X \times X$,
$\tilde{\mathcal{G}}$	the "true" DAG describing full order conditional dependencies (Proposition 3),
$\mathcal{G}^{(q)}$	q^{th} order conditional dependence DAG (Definition 4).

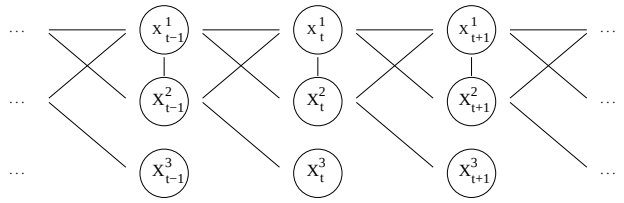

Figure 3: Moral graph of the DAG in Figure 2. For all $t > 1$, the parents of the variable X_t^1 are "married", that is connected by a non directed edge.

minimal DAG $\tilde{\mathcal{G}}_{AR(1)}$ is given in the particular case of an AR(1) model in subsection 2.4. The main interest of a DBN representation is to derive conditional dependence relationships between the variables by using the graphical theory associated with the DAGs. Note that, even though we need to consider a homogeneous DBN for the inference of gene interaction networks, the general framework (sections 2 and 3) is developed without assuming homogeneity.

2.1 Backgrounds

Let $\mathcal{G} = (X, E(\mathcal{G}))$ be a DAG whose vertices are the variables $X = \{X_t^i; i \in P, t \in N\}$ and whose set of edges $E(\mathcal{G})$ is a subset of $X \times X$. We quickly recall here elements of the theory of graphical models associated with the DAGs [16].

Definition 1 *The parents of a vertex X_t^i in $\mathcal{G}$, denoted by $pa(X_t^i, \mathcal{G})$, are the variables having an edge pointing towards the vertex X_t^i in $\mathcal{G}$,*

$$pa(X_t^i, \mathcal{G}) := \{X_s^j \text{ such that } (X_s^j, X_t^i) \in E(\mathcal{G}); j \in P, s \in N\}.$$

Proposition 1 ***(BN representation [24])*** *The probability distribution $\mathbb{P}$ of the process X admits a Bayesian Network representation according to DAG $\mathcal{G}$ whenever its density f factorizes as a product of the conditional density of each variable X_i^t given its parents in $\mathcal{G}$,*

$$f(X) = \prod_{i \in P} \prod_{t \in N} f(X_t^i | pa(X_t^i, \mathcal{G})).$$

Definition 2 ***(Moral graph)*** *The moral graph $\mathcal{G}^m$ of DAG $\mathcal{G}$ is obtained from $\mathcal{G}$ by first 'marrying' the parents (draw an undirected edge between each pair of parents of each variable X_t^i) and then deleting directions of the original edges of $\mathcal{G}$.*

For an illustration, Figure 3 exposes the moral graph of the DAG in Figure 2.

Definition 3 ***(Ancestral set)*** *The subset S is ancestral if and only if, for all $\alpha \in S$, the parents of α satisfy $pa(\alpha, \mathcal{G}) \subseteq S$. Hence, for any subset S of vertices, there is a smallest ancestral set containing S which is denoted by $An(S)$. Then $\mathcal{G}_{An(S)}$ refers to the graph of the smallest ancestral set $An(S)$.*

Throughout this paper, a central notion is that of conditional independence of random variables. Let $\mathbb{P}_{U,V,W}$ be the joint distribution of three random variables (U, V, W). We say that U is *conditionally independent of V given W under $\mathbb{P}_{U,V,W}$* and write $U \perp\!\!\!\perp V \mid W$ whenever the variable U does not depend on V when considering the joint distribution $\mathbb{P}_{U,V,W}$. This result generalizes to sets of disjoint variables. Such conditional independence relationships can be set from a BN representation by using the graphical theory associated with the DAGs. Most of the results are based on the next proposition which is derived from the Directed global Markov property [16].

Proposition 2 ***(Lauritzen [16], Corollary 3.23)*** *Let $\mathbb{P}$ admit a BN representation according to $\mathcal{G}$. Then,*

$$E \perp\!\!\!\perp F \mid S,$$

*whenever all paths from E to F intersect S in $(\mathcal{G}_{An(E\cup F\cup S)})^m$, the moral graph of the smallest ancestral set containing $E \cup F \cup S$. We say that S **separates** E from F.*

2.2 Sufficient conditions for a DBN representation

Assumption 1 *The stochastic process X_t is first-order markovian,*

$$\forall t \geq 3, \quad X_t \perp\!\!\!\perp X_{1:t-2} \mid X_{t-1}.$$

Assumption 2 *For all $t \geq 1$, the random variables $\{X_t^i\}_{i\in P}$ are conditionally independent given the past of the process $X_{1:t-1}$, that is,*

$$\forall t \geq 1, \forall i \neq j, \quad X_t^i \perp\!\!\!\perp X_t^j \mid X_{1:t-1}.$$

We first assume that the observed process X_t is first-order markovian (Assumption 1). That is the expression level of a gene at given time t only depends on the past through the expression level at the previous time $t-1$. Then we assume that the variables observed simultaneously are conditionally independent given the past of the process (Assumption 2). In other words, we consider that time measurements are close enough so that a gene expression level X_t^i measured at time t is better explained by the previous time expression levels X_{t-1} than by some current expression level X_t^j.

From these two assumptions, we establish in the next lemma the existence of a DBN representation of the distribution $\mathbb{P}$ according to DAG $\mathcal{G}_{full}$ which contains all the edges pointing out from a variable observed at some time $t-1$ towards a variable observed at next time t. The direction of the edges according to the time guarantees the acyclicity of $\mathcal{G}_{full}$.

Lemma 1 *Under Assumptions 1 and 2, the probability distribution $\mathbb{P}$ admits a DBN representation, at least according to the DAG $\mathcal{G}_{full} = (X, \{(X_{t-1}^j, X_t^i)\}_{i,j\in P, t>1})$ having edges between any pair of successive variables.*

Proof. From assumption 1, the density f of the joint probability distribution of the process X writes as the product of conditional densities,

$$f(X) = f(X_1) \prod_{t=2}^{n} f(X_t|X_{t-1}), \tag{2}$$

where $f(X_t|X_{t-1})$ refers to the density of the conditional probability distribution of X_t given X_{t-1}.

From Assumption 2, for all $t > 1$, the conditional density $f(X_t|X_{t-1})$ writes as the product of the conditional density of each variable X_t^i given the set of variables X_{t-1} observed at the previous time,

$$f(X_t|X_{t-1}) = \prod_{i\in P} f(X_t^i|X_{t-1}). \tag{3}$$

From equations (2) and (3), the density f writes as the product of the conditional density of each variable X_t^i given its parents in $\mathcal{G}_{full}$. From Proposition 1, the probability distribution $\mathbb{P}$ admits a BN representation according to $\mathcal{G}_{full}$. ■

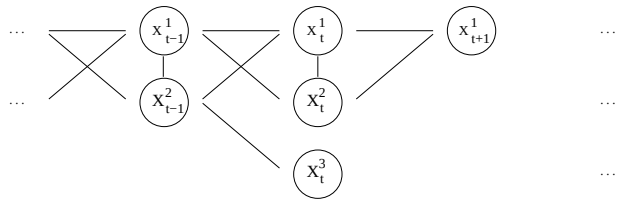

Figure 4: Moral graph of the smallest ancestral set containing the variables X^1_{t+1}, its parents in the DAG in Figure 2 and X^2_t. As the set (X^1_t, X^2_t) blocks all paths between X^3_t and X^1_{t+1}, we have $X^1_{t+1} \perp\!\!\!\perp X^3_t \mid (X^1_t, X^2_t)$.

2.3 Minimal DAG $\tilde{\mathcal{G}}$

Lemma 2 *Assume the joint probability distribution $\mathbb{P}$ of the process X has density f with respect to Lebesgue measure on $\mathbb{R}^{p\times n}$. If $\mathbb{P}$ factorizes according to two different subgraphs of $\mathcal{G}_{full}$, $\mathcal{G}_1$ and $\mathcal{G}_2$, then $\mathbb{P}$ factorizes according to $\mathcal{G}_1 \cap \mathcal{G}_2$.*

Lemma 3 ***(Conditional independence between non adjacent successive variables)*** *Let $\mathcal{G}$ be a subgraph of $\mathcal{G}_{full}$ according to which the probability distribution $\mathbb{P}$ admits a BN representation. For any pair of successive variables (X^j_{t-1}, X^i_t) which are non adjacent in $\mathcal{G}$, we have*

$$X^i_t \perp\!\!\!\perp X^j_{t-1} \mid pa(X^i_t, \mathcal{G}) \quad \text{and} \quad X^i_t \perp\!\!\!\perp X^j_{t-1} \mid pa(X^i_t, \mathcal{G}) \cup S,$$

for all S subset of $\{X^k_u; k \in P, u < t\}$.

The proof of these two lemmas is in Appendix. For an illustration of Lemma 3, assume $\mathbb{P}$ admits a BN representation according to the DAG of Figure 2. There is no edge between X^3_t and X^1_{t+1} in this DAG. Now consider in Figure 4 the moral graph of the smallest ancestral graph containing X^3_t, X^1_{t+1} and the parents (X^1_t, X^2_t) of X^1_{t+1}. The set (X^1_t, X^2_t) blocks all paths between X^3_t and X^1_{t+1}. From Proposition 2, we have $X^1_{t+1} \perp\!\!\!\perp X^3_t \mid pa(X^1_{t+1}, \mathcal{G})$.

It follows directly from Lemma 2 that, among the DAGs included in $\mathcal{G}_{full}$, it exists a minimal DAG, denoted by $\tilde{\mathcal{G}}$, according to which the probability distribution $\mathbb{P}$ factorizes. From Lemma 3, the set of edges of $\tilde{\mathcal{G}}$ is exactly the set of full order conditional dependencies given the past of the process as set up in the next proposition.

Let $P_j = P \backslash \{j\}$. We denote by $X^{P_j}_t = \{X^k_t; k \in P_j\}$ the set of $p-1$ variables observed at time t.

Proposition 3 ***(BN representation according to $\tilde{\mathcal{G}}$, the smallest subgraph of $\mathcal{G}_{full}$)*** *Whenever Assumptions 1 and 2 are satisfied, the probability distribution $\mathbb{P}$ admits a BN representation according to DAG $\tilde{\mathcal{G}}$ whose edges describe exactly the full order conditional dependencies between successive variables X^j_{t-1} and X^i_t given the remaining variables $X^{P_j}_{t-1}$ observed at time $t-1$,*

$$\tilde{\mathcal{G}} = \left(X, \left\{ (X^j_{t-1}, X^i_t);\ X^i_t \not\perp\!\!\!\perp X^j_{t-1} | X^{P_j}_{t-1} \right\}_{i,j\in P, t\in N} \right),$$

Moreover, DAG $\tilde{\mathcal{G}}$ is the smallest subgraph of $\mathcal{G}_{full}$ according to which $\mathbb{P}$ admits a BN representation.

See Proof in Appendix. In DAG $\tilde{\mathcal{G}}$, the set of parents $pa(X^i_t, \tilde{\mathcal{G}})$ of each variable X^i_t is the smallest subset of X_{t-1} such that the conditional densities satisfy $f(X^i_t | pa(X^i_t, \tilde{\mathcal{G}})) = f(X^i_t | X_{t-1})$. The set of parents of each variable can be seen as the only variables on which this variable depends directly. So $\tilde{\mathcal{G}}$ is the DAG we want to infer to recover potential regulation relationships from gene expression time series. From Lemma 3, any pair of successive variables (X^j_{t-1}, X^i_t) which are non adjacent in $\tilde{\mathcal{G}}$ are conditionally independent given the parents of X^i_t,

$$X^i_t \perp\!\!\!\perp X^j_{t-1} \mid pa(X^i_t, \tilde{\mathcal{G}}).$$

We will make use of this result in section 3 in order to define low order conditional independence DAGs for the inference of $\tilde{\mathcal{G}}$.

2.4 DAG $\tilde{\mathcal{G}}_{AR(1)}$ for an AR(1) process

Consider the following first order auto-regressive model,

AR(1) model

$$X_1 \sim \mathcal{N}(\mu_1, \Sigma_1) \tag{4}$$
$$\forall t > 1, \quad X_t = AX_{t-1} + B + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma), \tag{5}$$
$$\forall s, t \in N, \; Cov(\varepsilon_t, \varepsilon_s) = \delta_{ts}\Sigma, \tag{6}$$
$$\forall s > t, \quad Cov(X_t, \varepsilon_s) = 0. \tag{7}$$

where $A = (a_{ij})_{1 \leq i \leq p, 1 \leq j \leq p}$ is a $p \times p$ matrix, $B = (b_i)_{1 \leq i \leq p}$ is a column vector of size p, $\Sigma = (\sigma_{ij})_{1 \leq i \leq p, 1 \leq j \leq p}$ is the error covariance matrix and for all s, t in N, $\delta_{ts} = \mathbb{1}_{\{s=t\}}$. Equation (7) implies that the coefficient matrices are uniquely determined from the covariance function of X_t.

This modeling assumes homogeneity across time (constant matrix A) and linearity of the dependency relationships. From (5) and (7), the model is first order markovian and Assumption 1 is satisfied. From (6), Assumption 2 is satisfied whenever the error covariance matrix Σ is diagonal. Considering non correlated measurement errors between distinct genes is a strong assumption especially since microarray data contain several sources of noise including block effects. Nevertheless, assuming Σ diagonal is still reasonable after a normalization procedure.

From Proposition 3, the probability distribution of this AR(1) process factorizes according to a minimal DAG $\tilde{\mathcal{G}}_{AR(1)}$ whose edges correspond to the non-zero coefficients of matrix A. Indeed, if matrix Σ is diagonal, each element a_{ij} is the regression coefficient of the variable X_t^i on X_{t-1}^j given $X_{t-1}^{P_j}$, that is $a_{ij} = Cov(X_t^i, X_{t-1}^j \mid X_{t-1}^{P_j}) / Var(X_{t-1}^j \mid X_{t-1}^{P_j})$. So the set of null coefficients of the matrix A exactly describes the conditional independencies between successive variables,

$$\text{if } \Sigma \text{ is diagonal, we have} \quad a_{ij} = 0 \;\; \Leftrightarrow \;\; \left\{ \forall t > 1, \;\; X_t^i \perp\!\!\!\perp X_{t-1}^j | X_{t-1}^{P_j} \right\}.$$

So DAG $\tilde{\mathcal{G}}_{AR(1)}$ has an edge between two successive variables X_{t-1}^j and X_t^i, for all $t > 1$, whenever the coefficient a_{ij} of the matrix A differs from zero,

$$\tilde{\mathcal{G}}_{AR(1)} := \left(X, \left\{ (X_{t-1}^j, X_t^i) \text{ such that } a_{ij} \neq 0; \; t > 1, \; i, j \in P \right\} \right). \tag{8}$$

For an illustration, any AR(1) process whose matrix Σ is diagonal and matrix A has the following form,

$$A = \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & 0 & 0 \\ 0 & a_{32} & 0 \end{pmatrix}$$

admits a BN representation according to the dynamic network of Figure 2 ($p = 3$).

3 Approximating $\tilde{\mathcal{G}}$ with DAGs $\mathcal{G}^{(q)}$

From Proposition 3, reverse discovering the DAG $\tilde{\mathcal{G}}$ requires to determine, for each variable X_t^i, the set of variables X_{t-1}^j observed at time $t-1$ which are conditionally dependent on X_t^i given the remaining variables $X_{t-1}^{P_j}$. Even under homogeneity assumption (see section 1), the available data of gene expression time series do not allow such testing. We still have to face the 'curse of dimension' as the number of genes p, is much higher than the number of measurements n. By extending the approach proposed by Wille et al. [44, 43] to DBNs, we propose here an original approach for the inference of dynamic networks of high size by considering low order independencies.

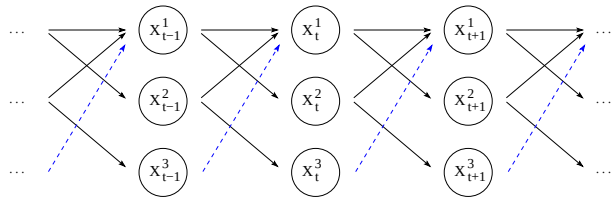

Figure 5: First-order conditional dependence DAG $\mathcal{G}^{(1)}$ (obtained from the DAG in Figure 2). The spurious dashed arrow may appear in $\mathcal{G}^{(1)}$.

3.1 Definition

We approximate DAG $\tilde{\mathcal{G}}$ of full order conditional dependence by the q^{th} order conditional dependence graph $\mathcal{G}^{(q)}$ (with $q < p$). In DAG $\mathcal{G}^{(q)}$, no edge is drawn between two successive variables X^j_{t-1} and X^i_t whenever it exists a subset X^Q_{t-1} of q variables among the $p-1$ variables $X^{P_j}_{t-1}$ such that X^j_{t-1} and X^i_t are conditionally independent given this subset. In short, DAGs $\mathcal{G}^{(q)}$ are defined as follows,

Definition 4 ***qth-order conditional dependence DAG*** $\mathcal{G}^{(q)}$

$$\forall q < p, \quad \mathcal{G}^{(q)} = \left(X, \left\{ (X^j_{t-1}, X^i_t);\ \forall Q \subseteq P_j, |Q| = q,\ X^i_t \not\perp\!\!\!\perp X^j_{t-1} | X^Q_{t-1} \right\}_{i,j \in P, t \in N} \right).$$

Note that DAGs $\mathcal{G}^{(q)}$ offer a way of producing some dependence relationships between the variables but are not anymore associated with a BN representation which would call for more global relationships. However DAG $\tilde{\mathcal{G}}$, which allows a BN representation, corresponds to the $(p-1)^{th}$ order conditional dependence DAG $\mathcal{G}^{(p-1)}$.

In general, DAGs $\mathcal{G}^{(q)}$ differ from DAG $\tilde{\mathcal{G}}$. For instance, the approximation of the DAG of Figure 2 by the 1^{st} order conditional dependence DAG may give birth to the spurious edge $X^3_{t-1} \rightarrow X^1_t$, for all $t > 1$ (see Figure 5). Indeed, neither X^1_{t-1} nor X^2_{t-1} separates X^1_t from X^3_{t-1} in the smallest moral graph containing $X^1_t \cup X^3_{t-1} \cup X^1_{t-1}$ (resp. $X^1_t \cup X^3_{t-1} \cup X^2_{t-1}$). Nevertheless, if the vertices of $\tilde{\mathcal{G}}$ have few parents, DAGs $\mathcal{G}^{(q)}$ can bring relevant information on the topology of $\tilde{\mathcal{G}}$, even for small value of q. In the following, we give characterizations of low order conditional dependence DAGs $\mathcal{G}^{(q)}$ and analyze how good approximations they do offer.

3.2 A restricted number of parents

In the known gene regulation mechanisms, some genes regulate many other genes (e.g. single input modules in the transcriptional regulatory network of S. Cerevisiae [17]). Nevertheless, we do not expect a single gene to be regulated by a lot of genes at the same time. So the number of parents in gene interaction networks is expected to be relatively small. In this section, we analyze the properties of $\mathcal{G}^{(q)}$ when the number of parents in $\tilde{\mathcal{G}}$ is lower than q.

Let us denote by $N_{pa}(X^i_t, \tilde{\mathcal{G}})$ the number of parents of X^i_t in the DAG $\tilde{\mathcal{G}}$ and $N^{Max}_{pa}(\tilde{\mathcal{G}})$ the maximal number of parents of any variable X^i_t in $\tilde{\mathcal{G}}$,

$$N_{pa}(X^i_t, \tilde{\mathcal{G}}) = \left| pa(X^i_t, \tilde{\mathcal{G}}) \right|, \qquad N^{Max}_{pa}(\tilde{\mathcal{G}}) = \underset{i \in P, t \in N}{\text{Max}} \left(N_{pa}(X^i_t, \tilde{\mathcal{G}}) \right).$$

The next results hold when the number of parents in $\tilde{\mathcal{G}}$ is restricted.

Proposition 4 *If* $N_{pa}(X^i_t, \tilde{\mathcal{G}}) \leq q$ *then* $\left\{ (X^j_{t-1}, X^i_t) \notin E(\tilde{\mathcal{G}}) \right\} \ \Rightarrow \ \left\{ (X^j_{t-1}, X^i_t) \notin E(\mathcal{G}^q) \right\}$.

Corollary 1 *For all* $q \geq N^{Max}_{pa}(\tilde{\mathcal{G}})$, *we have* $\tilde{\mathcal{G}} \supseteq \mathcal{G}^{(q)}$.

Proposition 5 *Let* X *be a Gaussian process. If* $N^{Max}_{pa}(\tilde{\mathcal{G}}) \leq 1$ *then* $\tilde{\mathcal{G}} = \mathcal{G}^{(1)}$.

Consider a variable X_t^i having at most q parents in $\tilde{\mathcal{G}}$ ($q < p$). Let X_{t-1}^j be a variable observed at the previous time $t-1$ and having no edge pointing towards X_t^i in $\tilde{\mathcal{G}}$. In the moral graph of the smallest ancestral set containing $X_t^i \cup X_{t-1}^j \cup pa(X_t^i, \tilde{\mathcal{G}})$, the set of parents $pa(X_t^i, \tilde{\mathcal{G}})$ separates X_t^i from X_{t-1}^j. From Proposition 2, we have $X_t^i \perp\!\!\!\perp X_{t-1}^j \mid pa(X_t^i, \tilde{\mathcal{G}})$. The number of parents $pa(X_t^i, \tilde{\mathcal{G}})$ is lower than q, so the edge $X_{t-1}^j \rightarrow X_t^i$ is not in $\mathcal{G}^{(q)}$. This establishes Proposition 4.

Consequently, if the maximal number of parents in $\tilde{\mathcal{G}}$ is lower than q then $\mathcal{G}^{(q)}$ is included in $\tilde{\mathcal{G}}$ (Corollary 1). In that case, $\mathcal{G}^{(q)}$ does not contain spurious edges.

The converse inclusion relationship is not true in general. Let $X_{t-1}^j \rightarrow X_t^i$ be an edge of $\tilde{\mathcal{G}}$, then X_t^i and X_{t-1}^j are conditionally dependent given the remaining variables $X_{t-1}^{P_j}$. It may however exist a subset of q variables X_{t-1}^Q, where Q is a subset of $P \backslash \{j\}$ of size q, such that X_t^i and X_{t-1}^j are conditionally independent with respect to this subset X_{t-1}^Q. Indeed, even though the topology of $\tilde{\mathcal{G}}$ allows to establish some conditional independencies, DAG $\tilde{\mathcal{G}}$ does not necessary allow to derive all of them. Two variables can be conditionally independent given a subset of variables whereas this subset does not separate these two variables in $\tilde{\mathcal{G}}$. Nevertheless, if each variable has at most *one* parent, the converse inclusion $\tilde{\mathcal{G}} \subseteq \mathcal{G}^{(1)}$ is true if the process is Gaussian and $q = 1$ (Proposition 5, see proof in Appendix). At a higher order, we need to assume that all conditional independencies can be derived from $\tilde{\mathcal{G}}$, that is $\mathbb{P}$ is *faithful* to $\tilde{\mathcal{G}}$.

3.3 Faithfulness

Definition 5 ***(faithfulness, Spirtes [33])*** *A distribution $\mathbb{P}$ is* ***faithful*** *to a DAG $\mathcal{G}$ if all and only the independence relationships true in $\mathbb{P}$ are entailed by $\mathcal{G}$ (as set up in Proposition 2).*

Theorem 1 ***(Measure zero for unfaithful Gaussian (Spirtes [33]) and discrete (Meek [19]) distributions)*** *Let $\pi_{\mathcal{G}}^{\mathcal{N}}$ (resp. $\pi_{\mathcal{G}}^{\mathcal{D}}$) be the set of linearly independent parameters needed to parametrize a multivariate normal distribution (resp. discrete distribution) $\mathbb{P}$ which admits a factorization according to a DAG $\mathcal{G}$. The set of distributions which are unfaithful to $\mathcal{G}$ is measure zero with respect to Lebesgue measure over $\pi_{\mathcal{G}}^{\mathcal{N}}$ (resp. over $\pi_{\mathcal{G}}^{\mathcal{D}}$).*

If distribution $\mathbb{P}$ is faithful to $\tilde{\mathcal{G}}$, then any subset $X_{t-1}^Q \subseteq X_{t-1}$, with respect to which X_t^i and X_{t-1}^j are conditionally independent, separates X_t^i and X_{t-1}^j in the moral graph of the smallest ancestral set containing $X_t^i \cup X_{t-1}^j \cup X_{t-1}^Q$. Under this assumption, we can derive interesting properties on $\tilde{\mathcal{G}}$ from the topology of low order dependence DAGs $\mathcal{G}^{(q)}$. As there is no way to assess a probability distribution to be faithful to a DAG, this assumption has often been criticized. Nevertheless, Theorem 1, established by Spirtes [33] for Gaussian distribution and extended to discrete distribution by Meek [19], makes this assumption reasonable at least in a measure-theoretic sense. Given that we consider a single distribution inherent to the studied process, the distribution $\mathbb{P}$ is not necessary faithful to $\tilde{\mathcal{G}}$. Nevertheless, this assumption appears very reasonable and calls for careful interest. The next propositions are derived from faithfulness to $\tilde{\mathcal{G}}$.

Proposition 6 *Assume $\mathbb{P}$ is faithful to $\tilde{\mathcal{G}}$. For all $q < p$, we have $\tilde{\mathcal{G}} \subseteq \mathcal{G}^{(q)}$.*

Proof. Let $(X_{t-1}^j, X_t^i) \in E(\tilde{\mathcal{G}})$. Assume that $(X_{t-1}^j, X_t^i) \notin E(\mathcal{G}^{(q)})$ then it exists a subset of q variables X_{t-1}^Q with respect to which X_{t-1}^j and X_t^i are conditionally independent. From faithfulness, the subset X_{t-1}^Q separates X_{t-1}^j and X_t^i in the moral graph of the smallest ancestral set containing $X_t^i \cup X_{t-1}^j \cup X_{t-1}^Q$. This contradicts the presence of the edge (X_{t-1}^j, X_t^i) in $\tilde{\mathcal{G}}$. ■

Corollary 2 *Assume $\mathbb{P}$ is faithful to $\tilde{\mathcal{G}}$. For all $q \geq N_{pa}^{Max}(\tilde{\mathcal{G}})$, we have $\tilde{\mathcal{G}} = \mathcal{G}^{(q)}$.*

Proposition 7 *Assume $\mathbb{P}$ is faithful to $\tilde{\mathcal{G}}$. If $N_{pa}(X_t^i, \mathcal{G}^{(q)}) \leq q$ then $(X_{t-1}^j, X_t^i) \in E(\mathcal{G}^{(q)}) \Rightarrow (X_{t-1}^j, X_t^i) \in E(\tilde{\mathcal{G}})$.*

Proof. From faithfulness, $\tilde{\mathcal{G}} \subseteq \mathcal{G}^{(q)}$. Then for all X_t^i, $N_{pa}(X_t^i, \tilde{\mathcal{G}}) \leq N_{pa}(X_t^i, \mathcal{G}^{(q)}) \leq q$. From Proposition 4, $(X_{t-1}^j, X_t^i) \notin E(\tilde{\mathcal{G}}) \Rightarrow (X_{t-1}^j, X_t^i) \notin E(\mathcal{G}^{(q)})$,that is $(X_{t-1}^j, X_t^i) \in E(\mathcal{G}^{(q)}) \Rightarrow (X_{t-1}^j, X_t^i) \in E(\tilde{\mathcal{G}})$.

■

Choose either LS, Huber or Tukey estimator and set the thresholds α_1 and α_2.
  Step 1: inferring $\mathcal{G}^{(1)}$.
For all $i \in P$,
For all $j \in P$, for all $k \neq j$, compute the p-value $p_{ij|k}$ from (9),
$S_1(i,j) = Max_{k \neq j}(p_{ij|k})$.
$E(\hat{\mathcal{G}}^{(1)}) = \{(X^j_{t-1}, X^i_t)_{t>1};\ i,j \in P,$ such that $S_1(i,j) < \alpha_1\}$.
  Step 2: inferring $\tilde{\mathcal{G}}$ from $\hat{\mathcal{G}}^{(1)}$.
If $N^{Max}_{pa}(\hat{\mathcal{G}}^{(1)}) \sim n-1$, choose a higher threshold α_1 and go to Step 1.
For all i such that $N_{pa}(X^i_t, \hat{\mathcal{G}}^{(1)}) \geq 1$, compute the p-value $p^{(2)}_{ij}$ from (10).

$$S_2(i,j) = \begin{cases} p^{(2)}_{ij} & \text{for all } i,j \in P \text{ such that } (X^j_{t-1}, X^i_t)_{t>1} \in \hat{\mathcal{G}}^{(1)}, \\ 1 & \text{otherwise.} \end{cases}$$

$E(\tilde{\mathcal{G}}) = \{(X^j_{t-1}, X^i_t)_{t>1}; i \in P, (i,j) \in P$ such that $S_2(i,j) < \alpha_2\}$.

Figure 6: Algorithm

Corollary 3 *Assume $\mathbb{P}$ is faithful to $\tilde{\mathcal{G}}$. For all $q \geq N^{Max}_{pa}(\mathcal{G}^{(q)})$, we have $\tilde{\mathcal{G}} = \mathcal{G}^{(q)}$.*

Even though we expect the number of parents in a gene interaction networks to be upper bounded, the exact maximal number of parents $N^{Max}_{pa}(\tilde{\mathcal{G}})$ remains unknown. Nevertheless, if $\mathbb{P}$ is faithful to $\tilde{\mathcal{G}}$, some edges of $\tilde{\mathcal{G}}$ can still be derived from the topology of q^{th} order conditional dependence DAGs $\mathcal{G}^{(q)}$ without knowing the maximal number of parents in $\tilde{\mathcal{G}}$. Indeed, from Proposition 7, the edges of DAG $\mathcal{G}^{(q)}$ pointing towards a variable having less than q parents in $\mathcal{G}^{(q)}$ are edges of $\tilde{\mathcal{G}}$ too.

4 Inferring $\tilde{\mathcal{G}}$

We introduced and characterized the q^{th} order dependence DAGs $\mathcal{G}^{(q)}$, for all $q < p$, in dynamic modeling. We now exploit our results to develop a non-bayesian inference method for DAG $\tilde{\mathcal{G}}$. Let q_{max} be the maximal number of parents in $\tilde{\mathcal{G}}$. From Corollary 3, inferring $\tilde{\mathcal{G}}$ amounts to inferring $\mathcal{G}^{(q_{max})}$. However, the inference of $\mathcal{G}^{(q_{max})}$ requires to check, for each pair (i,j), if there exists a subset $Q \subseteq P_j$ of dimension q_{max} such that $X^i_t \perp\!\!\!\perp X^j_{t-1} | X^Q_{t-1}$ for all $t > 1$. So, for each pair (i,j), there are $\binom{q_{max}}{p-1}$ potential sets that can lead to conditional independence. To test each conditional independence given any possible subset of q_{max} variables is questionable both in terms of complexity and multiple testings.

To circumvent these issues, we propose to exploit the fact that the true model $\tilde{\mathcal{G}}$ is a subgraph of $\mathcal{G}^{(1)}$ (Proposition 6) to develop an inference procedure. Indeed, the inference of $\mathcal{G}^{(1)}$ is both the faster (complexity) and the most accurate (number of tests). So we set out a two step procedure: first to infer $\mathcal{G}^{(1)}$, second to infer $\tilde{\mathcal{G}}$ from the estimated DAG $\hat{\mathcal{G}}^{(1)}$. Nevertheless, DAG $\hat{\mathcal{G}}^{(1)}$ already offers a very good approximation of $\tilde{\mathcal{G}}$ when it is sparse (see Figure 7, left). We develop here the 2 step-procedure which is summarized in Figure 6.

4.1 Step 1: inferring $\mathcal{G}^{(1)}$

We evaluate the *likelihood* of an edge (X^j_{t-1}, X^i_t) by measuring the conditional dependence between the variables X^j_{t-1} and X^i_t given any variable X^k_{t-1}. Let $a_{ij|k}$ be the partial regression coefficient defined as follows,

$$X^i_t = m_{ijk} + a_{ij|k} X^j_{t-1} + a_{ik|j} X^k_{t-1} + \eta^{i,j,k}_t,$$

where the rank of the matrix $(X^j_{t-1}, X^k_{t-1})_{t \geq 2}$ equals 2 and the errors $\{\eta^{i,j,k}_t\}_{t \geq 2}$ are centered, have same variance and are not correlated.

We chose to measure the conditional dependence between the variables X^j_{t-1} and X^i_t given any variable X^k_{t-1} by testing null assumption $\mathcal{H}^{i,j,k}_0$: "$a_{ij|k} = 0$". To such an aim, we use one out of three M-estimators for this coefficient: either the familiar Least Square (LS) estimator, the *Huber* estimator, or the *Tukey bisquare* (or *biweight*) estimator. The two latter are robust estimators [9]. Then for each $k \neq j$, we compute the

estimates $\hat{a}_{ij|k}$ according to one of these three estimators and derive the p-value $p_{ij,k}$ from the standard significance test:

$$under\ (\mathcal{H}_0^{i,j,k}):\ \text{“}a_{ij|k}=0\text{”}, \qquad \frac{\hat{a}_{ij|k}}{\hat{\sigma}(\hat{a}_{ij|k})} \sim t(n-4), \tag{9}$$

where $t(n-4)$ refers to a student probability distribution with $n-4$ degrees of freedom and $\hat{\sigma}(\hat{a}_{ij|k})$ is the variance estimates for $\hat{a}_{ij|k}$.

Thus, we assign a score $S_1(i,j)$ to each potential edge (X_{t-1}^j, X_t^i) equal to the maximum $Max_{k\neq j}(p_{ij|k})$ of the $p-1$ computed p-values, that is the most favorable result to 1^{st} order conditional independence. This procedure does not derive p-values for the edges but allows to order the possible edges of DAG $\mathcal{G}^{(1)}$ according to how likely they are. The smallest scores point out the most significant edges for $\mathcal{G}^{(1)}$. The inferred DAG $\hat{\mathcal{G}}^{(1)}$ contains the edges having a score below a chosen threshold α_1. We compare the three estimators used for the inference of $\tilde{\mathcal{G}}$ in a simulation study in the next section (Figure 7, right).

4.2 Step 2: inferring $\tilde{\mathcal{G}}$

We use the inferred DAG $\hat{\mathcal{G}}^{(1)}$ as a reduction of the search space. Indeed, from faithfulness, $\tilde{\mathcal{G}} \subseteq \mathcal{G}^{(1)}$ (Proposition 6). Moreover, when DAG $\tilde{\mathcal{G}}$ is sparse, there are far fewer edges in $\mathcal{G}^{(1)}$ than in the complete DAG $\mathcal{G}_{full}$ defined in subsection 2.2. Consequently, the number of parents of each variable in $\hat{\mathcal{G}}^{(1)}$ is much lower than n. Then model selection can be carried out by using standard estimation and tests among the edges of $\hat{\mathcal{G}}^{(1)}$. For each pair (i,j) such that the set of edges $(X_{t-1}^j, X_t^i)_{t>1}$ is in $\hat{\mathcal{G}}^{(1)}$, we denote $a_{ij}^{(2)}$ the following regression coefficient,

$$X_t^i = m_i + \sum_{j\in pa(X_t^i,\hat{\mathcal{G}}^{(1)})} a_{ij}^{(2)}\ X_{t-1}^j + \eta_t^i,$$

where the rank of the matrix $(X_{t-1}^j)_{t\geq 2, j\in pa(X_t^i,\hat{\mathcal{G}}^{(1)})}$ is $|pa(X_t^i,\hat{\mathcal{G}}^{(1)})|$ and the errors $\{\eta_t^i\}_{t\geq 2}$ are centered, have same variance and are not correlated. We assign to each edge of $\hat{\mathcal{G}}^{(1)}$ the score $S_2(i,j)$ equal to the p-value $p_{ij}^{(2)}$ derived from the significance test,

$$under\ (\mathcal{H}_0^{i,j}):\ \text{“}a_{ij}^{(2)}=0\text{”}, \qquad \frac{\hat{a}_{ij}^{(2)}}{\hat{\sigma}(\hat{a}_{ij}^{(2)})} \sim t(n-1-|pa(X_t^i,\hat{\mathcal{G}}^{(1)})|). \tag{10}$$

The score $S_2(i,j)=1$ is assigned to the edges that are not in $\hat{\mathcal{G}}^{(1)}$. The smallest scores point out the most significant edges. The inferred DAG for $\tilde{\mathcal{G}}$ contains the edges whose score is below a chosen threshold α_2. This inference procedure and analysis tools are implemented in a R package which is available on request.

When $\tilde{\mathcal{G}}$ is sparse, Step 1 of the procedure gives a good estimation of $\tilde{\mathcal{G}}$ already (see ROC curves of Figure 7, left). Even better results can be obtained with the 2 step-procedure which requires to tune two parameters α_1 and α_2. Parameter α_1 is the selection threshold of the edges of $\hat{\mathcal{G}}^{(1)}$ in step 1 (that is the dimension reduction threshold), whereas parameter α_2 is the selection threshold for the edges of $\tilde{\mathcal{G}}$ among the edges of DAG $\hat{\mathcal{G}}^{(1)}$. We study the impact of these parameters on the accuracy of the procedure through simulations in the next section.

5 Validation

5.1 Simulation study

We investigate the accuracy of the various approaches we proposed to recover DAG $\tilde{\mathcal{G}}$ for a multivariate AR(1) model. We randomly generate 100 sets of parameters ($A_{[p\times p]}$, B, Σ) for $p=50$. The gene regulation networks are known to be sparse. In accordance with this biological knowledge, each matrix A contains 2 % of non zero coefficients (sampled from uniform distribution). While keeping the number of parents low, this does not prevent it to be higher than one. Non zero coefficients were generated as follows, $a_{ij} \sim$

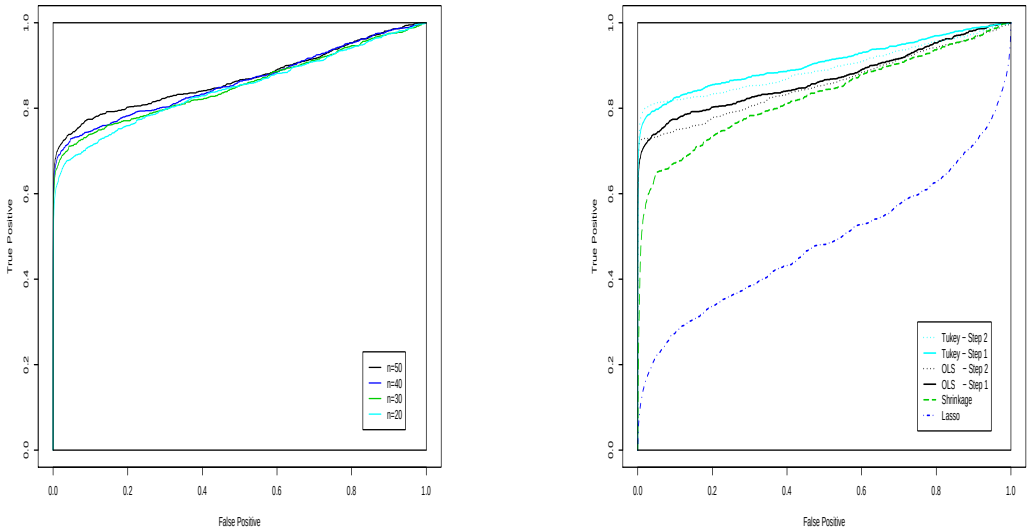

Figure 7: ROC curves for the inference of $\tilde{\mathcal{G}}$. Left: ROC curves obtained by $\mathcal{G}^{(1)}$ approximation (Step 1) with the Least Square estimator when $n = 20$ to 50. Right: improvement obtained with both robust estimation and Step 2 of the procedure ($\alpha_1 = 0.9$, $n = 50$); comparison with Shrinkage and Lasso regression.

$\mathcal{U}([-1.5; -0.5] \cup [0.5; 1.5])$, and we drew the mean $b_i \sim \mathcal{U}(0, 1)$ and error variance $\sigma_i \sim \mathcal{U}[0.03, 0.08]$ from uniform distributions. Time series were simulated under the corresponding multivariate AR(1) models for $n = 20$ to 50.

The left panel of Figure 7 displays the average ROC curves for the inference of DAG $\tilde{\mathcal{G}}$ obtained by $\mathcal{G}^{(1)}$ approximation (Step 1) with the LS estimator for $n = 20$ to 50. We ordered the edges (i, j) according to increasing maximal p-values $Max_{k\neq j}(p_{ij|k})$ for the significance tests of the partial regression coefficient estimates (see section 4.1 for details). For a very low false positive (FP) rate, the true positive (TP) rate rises 70 % for the longer time series. Even when $n = 20$, which is about the maximal length of the available time series gene expression data, the TP rate reaches almost 60 % whereas the FP rate remains almost null.

These results can still be improved. As an illustration, the right panel of Figure 7 displays average ROC curves obtained after the first or the second step of the procedure with either LS or Tukey bisquare estimates when $n = 50$. The ROC curves obtained with Huber estimates are very close to the Tukey bisquare curves and do not appear on this graph for sake of clarity. The solid black curve is the ROC curve obtained after step 1 computed with the LS estimator. Still with Step 1 only, the Tukey estimator allows to obtain better results (see the solid light line). In both cases, Step 2 (dotted lines) still higher the ROC curves for the first selected edges. So both Step 2 and robust estimation allows to higher the ROC curves, at least while keeping the FP rate very small. A similar improvement is obtained for others values of n.

We now recall the definitions:

$$\text{TP}= \frac{\text{Nb of true positive edges}}{\text{Nb of edges in the model}}, \qquad \text{PPV}=\frac{\text{Nb of true positive edges}}{\text{Nb of selected edges}},$$

and examine more precisely the interest of Step 2. We notably analyze the impact of both thresholds α_1 and α_2. The first step (inference of $\mathcal{G}^{(1)}$) allows to obtain a good TP/FP ratio already. Nevertheless the Positive Predictive Value (PPV) deteriorates very quickly when the threshold α_1 increases. As an illustration, see the dotted lines in the left panel of Figure 8 which shows the value of both the true positive rate and the PPV after Step 1 according to the threshold α_1. After Step 1 only, the PPV is high for small values of α_1, but then only a rather small percentage of edges is detected: for $\alpha_1 = 0.05$, $PPV = 90\%$ and $TP = 60\%$. On the contrary, when α_1 is high, even though the PPV is very small, the TP rate can reach very high values (up to 83 % for $\alpha_1 = 0.9$). Here comes the interest of the second step of the procedure. Indeed, even for high value of α_1, the number of edges in $\hat{\mathcal{G}}^{(1)}$ - and consequently the number of parents - is much lower than the initial number of potential edges ($\mathcal{G}_{full}$). The dimension is then reduced in proportion and a second selection

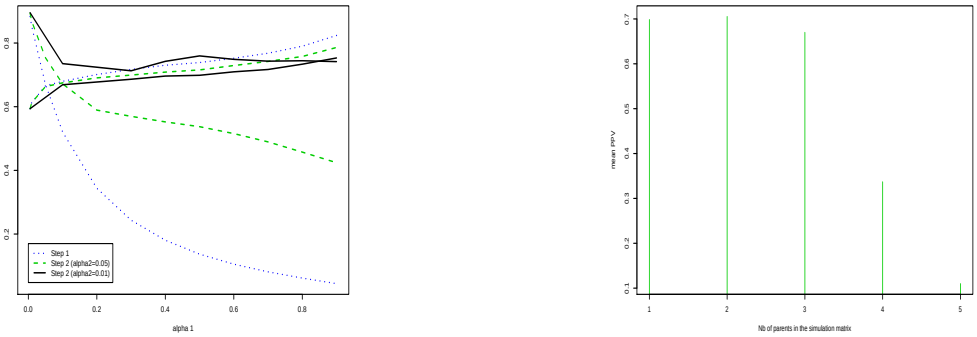

Figure 8: Left: TP rate (increasing curves) and PPV (decreasing curves) according to the threshold α_1 (n=50). Right: PPV according to the number of parents in the simulation DAG in the particular case $\alpha_1 = 0.9$, $\alpha_2 = 0.05$ (n=50).

step can be carried out by using standard significance tests (see section 4.2). As appearing in black solid lines on Figure 8 (left), Step 2 allows to increase the PPV while the TP rate stays high. For $\alpha_1 = 0.85$ and $\alpha_2 = 0.01$, the true positive rate and the PPV both reach 75%. Both thresholds have to be tuned according to the objective.

Note that the procedure performs well even when there are several parents in the true DAG $\tilde{\mathcal{G}}$. The right panel of Figure 8 shows the positive predictive value (PPV) for the inferred DAG $\tilde{\mathcal{G}}$ according to the number of parents in the simulation model for $\alpha_1 = 0.9$ and $\alpha_2 = 0.05$. Up to 3 parents, the PPVs are comparable and reach 70% on average.

Finally, we compare our approach with two reference methods for model selection in multivariate AR(1) process: the shrinkage approach by Opgen-Rhein and Strimmer and the Lasso regression.

Opgen-Rhein and Strimmer [23] recently proposed a model selection procedure based on an analytic shrinkage approach. The procedure first consists in computing the partial correlation coefficients from the shrinkage estimates of the partial regression coefficients, and second in selecting the edges with a *local* false discovery rate approach [7]. We carried out model selection in the simulated data with the R package they implemented. The ROC curve obtained by this shrinkage approach appears in dashed line (- - -) in the right panel of figure 7.

The L1 regression (Lasso) [36] combines shrinkage and model selection. This approach offers the advantage that it automatically sets many regression coefficients to zero. We carried out Lasso regression with the LARS package [8]. We chose the penalty by cross-validation. As proposed by Opgen-Rhein and Strimmer, we computed partial correlation coefficients from the Lasso estimates and drew ROC curves by ordering the edges according to the absolute value of the corresponding partial correlation coefficient. The ROC curve for the Lasso approach appears in dashed-dotted line (-·-·) in the right panel of figure 7.

Our procedure outperforms these two approaches. When using the 2 step-procedure with robust estimation, we reach 80% TP whereas the FP rate is almost null. The accuracy of our procedure comes from the increase of precision thanks to the dimension reduction. Indeed, this selection approach is based on 1^{st} order conditional independence consideration. This allows to carry out significance testing in a model of dimension 4 (see section 4.1). This represents a drastic dimension reduction and makes the testing much more powerful. Indeed, even if there are more edges in $\mathcal{G}^{(1)}$ than in the true DAG $\tilde{\mathcal{G}}$ (Proposition 6), Step 1 of the procedure is very sensitive already (see Figure 7).

Even though the Shrinkage approach improves a lot the precision of the estimation of each partial correlation coefficient in comparison with standard methods, to consider 1^{st} order conditional independence seems to be more powerful for the edge detection.

As for the Lasso, one major drawback lies in the fact that the edge selection is done vertex by vertex whereas the DAG $\tilde{\mathcal{G}}$ is globally but not uniformally sparse. As a consequence, the Lasso tends to uniformally reduce the number of parents of each vertex instead of only keeping small the total number of edges.

Table 2: First selected edges ($\alpha_1 = 0.05$) and validation (1/0=True/False regulation relationship).

TF	Target	$p_{ij}^{(2)}$	Validation
FKH2	KIP2	6.53e-06	1
SWI4	SVS1	6.87e-06	1
SWI4	AXL2	2.81e-05	0
SWI4	RKM1	3.85e-05	0
FKH2	CDC5	5.13e-05	1
FKH2	OGG1	5.21e-05	0
SWI4	SMC3	7.96e-05	0
FKH2	CLB2	9.10e-05	1
FKH2	SRC1	9.73e-05	1
SWI4	MSH2	1.16e-04	0

Table 3: Results of the inference method applied to the 792 genes of Spellman's yeast cell cycle data ($\alpha_1 = 0.05$). Tuning α_2 allows to choose between TP rate and PPV.

α_2	TP edges	PPV
10^{-3}	25	40 %
10^{-2}	47	30 %
10^{-1}	60	18 %

5.2 Analysis of microarray time course data sets

5.2.1 Spellman's Yeast cell cycle data set

We apply the proposed method to the Saccharomyces cerevisiae cell cycle data collected by Spellman et al. [32]. In the α Factor-based synchronization data (18 time points), we focus on the data set containing the 792 genes that demonstrated consistent periodic changes in transcription level. We only allow the 9 identified Transcription Factors (ACE2, FKH1, FKH2, MBP1, MCM1, NDD1, SWI4, SWI5, SWI6) that have been identified to have roles in regulating transcription of yeast genes [30] to be the possible regulators and try to infer their targets. The score is computed from LS estimates as explained in the previous section. We set the threshold α_1 to 0.05 for the inference of $\mathcal{G}^{(1)}$. The inferred DAG $\mathcal{G}^{(1)}$ contains 324 edges. From this DAG, we compute the p-value $p_{ij}^{(2)}$ to infer the true DAG $\tilde{\mathcal{G}}$. The first selected edges appear in Table 2. The 4^{th} column indicates whether or not a selected edge is a known regulation relationship. Validation is obtained from both the Yeastract database [1] and the targets of the cell cycle activators identified by Simon et al. [30].

For $\alpha_2 = 0.001$, this procedure allows to detect 25 known regulation relationships (TP edges) with a PPV of 40%. The results for different values of α_2 appear in Table 3. When increasing α_2, more edges are detected while the specificity stays acceptable comparative with other studies. Indeed, in one of the last in date DBN inference approach applied to the yeast cell cycle, Zou and Conzen [47] reduced their analysis to a subset of only 116 regulated genes and compare their approach with Murphy's Bayesian Network Toolbox [21]. By specifying the 9 identified TFs, Zou and Conzen correctly identify 46 edges with a PPV of 40% with their own procedure whereas they only obtain 18 correct edges with a PPV of 11% with Murphy's BNT.

5.2.2 Diurnal cycle on the starch metabolism of Arabidopsis Thaliana

We applied our inference procedure to expression time series data generated by Smith et al. [31] to investigate the impact of the diurnal cycle on the starch metabolism of *Arabidopsis Thaliana*. We restricted our study to the 800 genes selected by Opgen-Rhein and Strimmer [23] as having periodic expression profiles. The data are available in the GeneNet R package at `http://strimmerlab.org/software/genenet/html/arth800.html`.

We choose a high 1^{st} Step threshold $\alpha_1 = 0.1$ in order to maximize the chance that $\mathcal{G}^{(1)}$ contains the true DAG $\tilde{\mathcal{G}}$. For a 2^{nd} Step threshold $\alpha_2 = 0.001$, we obtain the DAG $\tilde{\mathcal{G}}$ which appears in Figure 9. This DAG

Table 4: List of the 5 proteins selected as parents which have been identified as Transcription Factor or DNA binding.

Node	Gene Name	Description
73	AT2G43010-TAIR-G	PIF4 (PHYTOCHROME INTERACTING FACTOR 4); DNA binding / transcription factor; Isolated as a semidominant mutation defective in red -light responses. Encodes a nuclear localized bHLH protein that interacts with active PhyB protein. Negatively regulates phyB mediated red light responses.
242	AT1G05900-TAIR-G	DNA binding / endonuclease; endonuclease-related, similar to endonuclease III (Homo sapiens) GI:1753174; contains Pfam profile PF00633: Helix-hairpin-helix motif
606	AT5G10400-TAIR-G	DNA binding; histone H3, identical to several histone H3 proteins, including Zea mays SP—P05203, Medicago sativa GI:166384, Encephalartos altensteinii SP—P08903, Pisum sativum SP—P02300; contains Pfam profile PF00125 Core histone H2A/H2B/H3/H4
725	AT5G65360-TAIR-G	DNA binding; histone H3, identical to histone H3 from Zea mays SP—P05203, Medicago sativa GI:166384, Encephalartos altensteinii SP—P08903, Pisum sativum SP—P02300; contains Pfam profile PF00125 Core histone H2A/H2B/H3/H4
788	At4g14410-MinT-G	putative bHLH transcription factor (bHLH104)

contains 168 edges implicating 236 different genes. 100 nodes are parent, 159 are child, 23 are both parent and child. The network differs from the network inferred by Opgen-Rhein and Strimmer [23] but we still recover a network with a "hub" connectivity structure.

Among the 'parent' nodes in the network $\tilde{\mathcal{G}}$, the two proteins having the most 'children' (node 799 and 628) are known to be implicated in the starch metabolism. Indeed, node 799, which has 11 'children' in $\tilde{\mathcal{G}}$, refers to DPE2 (DISPROPORTIONATING ENZYME 2), an essential component of the pathway from starch to sucrose and cellular metabolism in leaves at night. Then node 628 (6 children in $\tilde{\mathcal{G}}$) is a transferase (At5g24300) implicated in the starch synthase. Node 702 which is an unknown protein (At5g58220) has also 6 children in $\tilde{\mathcal{G}}$. These three nodes are dark-colored in the DAG of figure 9. The remaining 'parent' nodes have from 1 to 4 'children'. Among them, two are already identified as TFs and three as DNA binding proteins (see Table 4). These five nodes are light-colored in the DAG of figure 9. Finally a list of 28 unknown proteins have been selected as parents in the inferred DAG $\tilde{\mathcal{G}}$.

Complete results appear in the supplementary information available at `http://stat.genopole. cnrs.fr/~slebre/arth8` This notably displays the complete list of the unknown proteins selected as parents in the inferred DAG (section 2), the list of the parent nodes according to their number of target nodes (section 3) and the list of the edges ordered by decreasing significance (section 4) and by increasing past node number (section 5). The description of the 800 genes can be obtained from the GeneNet R package or at `http://stat.genopole.cnrs.fr/~slebre/arth8`

6 Conclusion

In this paper, we first introduce a DBN modeling of gene expression time series which offers straight interpretation in terms of conditional dependence between gene expression levels. Then we define and characterize low order conditional dependence DAGs for dynamic networks. They offer a very good approximation of sparse DAGs.

From these results, we develop a novel inference method for dynamic genetic networks which makes it possible to face with the 'small n, large p' estimation case. Our procedure proved to be powerful on both simulated and real data analysis. This approach based on the consideration of low order conditional

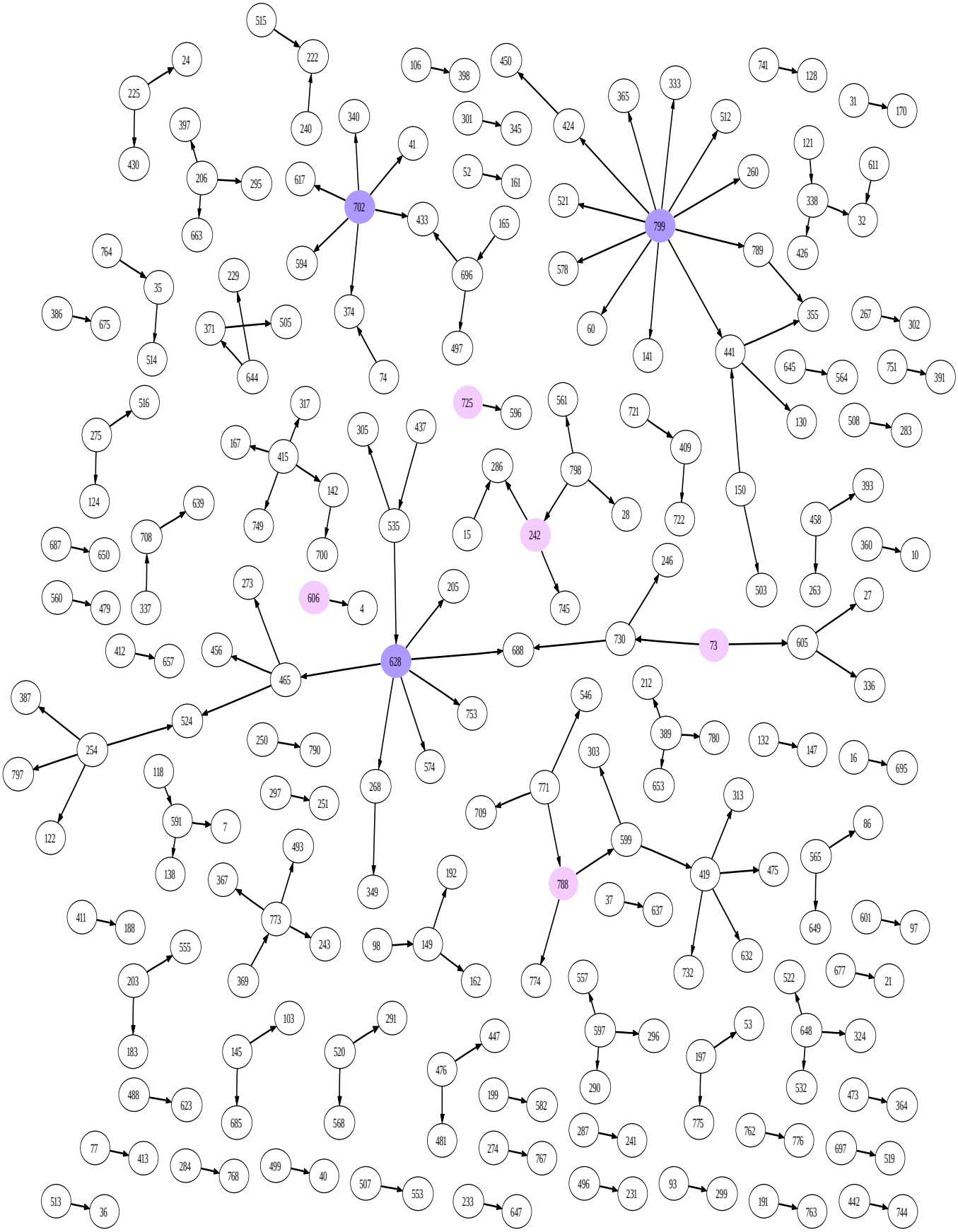

Figure 9: Inferred DAG $\tilde{\mathcal{G}}$ for $\alpha_1 = 0.1, \alpha_2 = 0.001$ (168 edges).

dependence notably outperforms model selection based on shrinkage or lasso estimates.

We point out that robust estimators appeared very efficient for the detection of the edges. An interesting direction for further research lies in investigating which measures of the dependence in gene expression data are the more pertinent.

7 Acknowledgments

I would like to thank Catherine Matias and Bernard Prum for many stimulating and constructive discussions on this work.

APPENDIX

Proof of Lemma 2. Consider a discrete-time stochastic process $X = \{X_t^i;\ i \in P, t \in N\}$ whose joint probability $\mathbb{P}$ distribution has the density f with respect to Lebesgue measure on $\mathbb{R}^{p\times n}$.

Let $\mathcal{G}_1$ and $\mathcal{G}_2$ be two different subgraphs of $\mathcal{G}_{full}$ according to which the joint probability distribution $\mathbb{P}$ factorizes. Let i in P, t in N, we consider the random variable X_t^i.

We denote as follows,

- the following subsets of P,
  $pa_1 = \{j \in P; X_{t-1}^j \in pa(X_t^i, \mathcal{G}_1)\}$
  $\overline{pa}_1 = P\backslash\{pa_1\}$
  $pa_2 = \{j \in P; X_{t-1}^j \in pa(X_t^i, \mathcal{G}_2)\}$
  $\overline{pa}_2 = P\backslash\{pa_2\}$
- and the densities of the joint or marginal probability distributions of (X_t^i, X_{t-1}),
  $g : \mathbb{R}^{p+1} \to \mathbb{R}$ the density of the joint probability distribution of (X_t^i, X_{t-1}),
  g^i the density of the probability distribution of X_t^i,
  g^P the density of the joint probability distribution of (X_{t-1}),
  g^{i,pa_1} the density of the joint probability distribution of $(X_t^i, X_{t-1}^{pa_1}) = (X_t^i, pa(X_t^i, \mathcal{G}_1))$,
  $g^{i,\overline{pa}_2}$ the density of the joint probability distribution of $(X_t^i, X_{t-1}^{\overline{pa}_2}) = (X_t^i, X_{t-1}\backslash\{pa(X_t^i, \mathcal{G}_2)\})$,
  etc...

In the following, $y \in \mathbb{R}$, $x = (x_1, ..., x_p) \in \mathbb{R}^p$ and we denote by $x_{pa_1} = \{x_j; j \in pa_1\} \in \mathbb{R}^{|pa_1|}$ (Thus $x = (x_{pa_1}, x_{\overline{pa}_1}) = (x_{pa_2}, x_{\overline{pa}_2}) \in \mathbb{R}^p$). As the probability distribution $\mathbb{P}$ factorizes according to $\mathcal{G}_1$, we derive from the DAG theory the conditional independence,

$$X_t^i \perp\!\!\!\perp X_{t-1}^{\overline{pa}_1} | X_{t-1}^{pa_1},$$

that is,

$$\forall y \in \mathbb{R}, \forall x \in \mathbb{R}^p, \quad \frac{g(y,x)}{g^P(x)} = \frac{g^{i,pa_1}(y, x_{pa_1})}{g^{pa_1}(x_{pa_1})}.$$

Equivalent results can be derived from the factorization according to $\mathcal{G}_2$ giving,

$$\forall y \in \mathbb{R}, x \in \mathbb{R}^p, N\ g^{i,pa_2}(y, x_{pa_2}) = \frac{g^{i,pa_1}(y, x_{pa_1})}{g^{pa_1}(x_{pa_1})} g^{pa_2}(x_{pa_2}).$$

By taking the integral with respect to $x_{pa_2 \cap \overline{pa}_1}$, we write for all $y \in \mathbb{R}$, for all $x_{pa_1 \cup pa_2} \in \mathbb{R}^{|pa_1 \cup pa_2|}$,

$$\int g^{i,pa_2}(y, x_{pa_2}) d(x_{pa_2 \cap \overline{pa}_1}) = \int \frac{g^{i,pa_1}(y, x_{pa_1})}{g^{pa_1}(x_{pa_1})} g^{pa_2}(x_{pa_2}) d(x_{pa_2 \cap \overline{pa}_1})$$

$$g^{i,pa_1 \cap pa_2}(y, x_{pa_1 \cap pa_2}) = \frac{g^{i,pa_1}(y, x_{pa_1})}{g^{pa_1}(x_{pa_1})} g^{pa_1 \cap pa_2}(x_{pa_1 \cap pa_2})$$

Finally we have,

$$\forall y \in \mathbb{R}, \forall x \in \mathbb{R}^p, \quad \frac{g(y,x)}{g^P(x)} = \frac{g^{i,pa_1 \cap pa_2}(y, x_{pa_1 \cap pa_2})}{g^{pa_1 \cap pa_2}(x_{pa_1 \cap pa_2})},$$

that is the conditional density of the probability distribution of X_t^i given X_{t-1} is the conditional density of the probability distribution of X_t^i given $X_{t-1}^{pa_1 \cap pa_2}$. Then $\mathbb{P}$ factorizes according to $\mathcal{G}_1 \cap \mathcal{G}_2$. ■

Proof of Lemma 3. Assume $\mathbb{P}$ admits a BN representation according to $\mathcal{G}$, a subgraph of $\mathcal{G}_{full}$. Let X_{t-1}^j and X_t^i be two *non adjacent* vertices of $\mathcal{G}$ (there is no edge between them in $\mathcal{G}$) and consider the moral graph $(\mathcal{G}_{An\left(X_t^i \cup X_{t-1}^j \cup pa(X_t^i, \mathcal{G})\right)})^m$ of the smallest ancestral set containing the variables X_t^i, X_{t-1}^j and the parents $pa(X_t^i, \mathcal{G})$ of X_t^i in $\mathcal{G}$. As DAG $\mathcal{G}$ is a subgraph of $\mathcal{G}_{full}$, the set of parents $pa(X_t^i, \mathcal{G})$ blocks all paths between X_{t-1}^j and X_t^i in the moral graph $(\mathcal{G}_{An\left(X_t^i \cup X_{t-1}^j \cup pa(X_t^i, \mathcal{G})\right)})^m$. From Proposition 2, this establishes the conditional independence $X_t^i \perp\!\!\!\perp X_{t-1}^j \mid pa(X_t^i, \mathcal{G})$.

This result holds for the conditioning according to any subset $S \subseteq \{X_u^k; k \in P, u < t\}$. ■

Proof of Proposition 3.

First, we show that $\mathbb{P}$ admits a BN representation according to $\tilde{\mathcal{G}}$. Let $i, j \in P$ such that $X_t^i \perp\!\!\!\perp X_{t-1}^j | X_{t-1}^{P_j}$, then we have,

$$f(X_t^i | X_{t-1}) = f(X_t^i | X_{t-1}^{P_j}).$$

Under Assumptions 1 and 2, from Lemma 1 and Proposition 1, $\mathbb{P}$ admits a BN representation according to the DAG $(X, E(\mathcal{G}_{full}) \setminus (X_{t-1}^j, X_t^i))$ which has the edges of $\mathcal{G}_{full}$ except for the edge (X_{t-1}^j, X_t^i). This holds for any pair of successive variables that are conditionally independent. Consequently, from Lemma 2, $\mathbb{P}$ admits a BN representation according to the intersection of the DAG $(X, E(\mathcal{G}_{full}) \setminus (X_{t-1}^j, X_t^i))$ for any pair (X_t^i, X_{t-1}^j) such that $X_t^i \perp\!\!\!\perp X_{t-1}^j | X_{t-1}^{P_j}$, that is DAG $\tilde{\mathcal{G}}$.

Second, DAG $\tilde{\mathcal{G}}$ cannot be reduced. Indeed, let (X_{t-1}^l, X_t^k) be an edge of $\tilde{\mathcal{G}}$ and assume that $\mathbb{P}$ admits a BN representation according to $\tilde{\mathcal{G}} \backslash (X_{t-1}^l, X_t^k)$, that is $\tilde{\mathcal{G}}$ reduced from the edge (X_{t-1}^l, X_t^k). From Lemma 3, we have $X_t^k \perp\!\!\!\perp X_{t-1}^l | X_{t-1}^{P_l}$, which contradicts $(X_{t-1}^l, X_t^k) \in V(\tilde{\mathcal{G}})$ (i.e. $X_t^k \not\perp\!\!\!\perp X_{t-1}^l | X_{t-1}^{P_l}$).

■

Proof of Proposition 5.

First, from Corollary 1, $\tilde{\mathcal{G}} \supseteq \mathcal{G}^{(1)}$.

Second, let X be a Gaussian process and $(X_{t-1}^j, X_t^i) \in E(\tilde{\mathcal{G}})$, then according to Proposition 3, $X_t^i \not\perp\!\!\!\perp X_{t-1}^j \mid X_{t-1}^{P_j}$. Since X is Gaussian, this implies $Cov(X_t^i, X_{t-1}^j | X_{t-1}^{P_j}) \neq 0$.

Now assume that it exists $k \neq j$, such that $X_t^i \perp\!\!\!\perp X_{t-1}^j \mid X_{t-1}^k$ ie $(X_{t-1}^j, X_t^i) \notin E(\mathcal{G}^{(1)})$. We are going to prove that this contradicts $Cov(X_t^i, X_{t-1}^j | X_{t-1}^{P_j}) \neq 0$. Let l be an element of $P \backslash \{j, k\}$. The conditional covariance $Cov(ij|k,l) = Cov(X_t^i, X_{t-1}^j \mid X_{t-1}^k, X_{t-1}^l)$ writes,

$$\begin{aligned} Cov(ij|k,l) &= Cov(X_t^i, X_{t-1}^j \mid X_{t-1}^k) - \frac{Cov(X_t^i, X_{t-1}^l \mid X_{t-1}^k) Cov(X_{t-1}^j, X_{t-1}^l | X_{t-1}^k)}{Var(X_{t-1}^l | X_{t-1}^k)}, \\ &= Cov(X_t^i, X_{t-1}^j \mid X_{t-1}^k) \times \left[1 - \frac{(Cov(X_{t-1}^j, X_{t-1}^l | X_{t-1}^k))^2}{Var(X_{t-1}^j | X_{t-1}^k) Var(X_{t-1}^l | X_{t-1}^k)}\right] \end{aligned}$$

$$- \frac{Cov(X^j_{t-1}, X^l_{t-1}|X^k_{t-1})Cov(X^i_t, X^l_{t-1}\ |X^k_{t-1}, X^j_{t-1})}{Var(X^l_{t-1}|X^k_{t-1})}.$$

However both terms in the latter expression of $Cov(ij|k,l)$ are null:

- since $X^i_t \perp\!\!\!\perp X^j_{t-1} \mid X^k_{t-1}$, then $Cov(X^i_t, X^j_{t-1}\ |X^k_{t-1}) = 0$,
- as $N^{Max}_{pa}(\tilde{\mathcal{G}}) \leq 1$, X^j_{t-1} is the only parent of X^i_t in $\tilde{\mathcal{G}}$. So the variable X^j_{t-1} and thus also the set (X^j_{t-1}, X^k_{t-1}) blocks all paths between X^l_{t-1} and X^i_t in the moral graph of the smallest ancestral set containing $X^i_t \cup X^{j,k,l}_{t-1}$. Then we have, $X^i_t \perp\!\!\!\perp X^l_{t-1} \mid \left\{X^j_{t-1}, X^k_{t-1}\right\}$, that is $Cov(X^i_t, X^l_{t-1}\ |X^k_{t-1}, X^j_{t-1}) = 0$.

Then $Cov(ij|k,l) = 0$. By induction, we obtain $Cov(X^i_t, X^j_{t-1}|X^{P_j}_{t-1}) = 0$ leading to a contradiction with $(X^j_{t-1}, X^i_t) \in E(\tilde{\mathcal{G}})$. Therefore $(X^j_{t-1}, X^i_t) \in \mathcal{G}^{(1)}$ and $\tilde{\mathcal{G}} \subseteq \mathcal{G}^{(1)}$.

■

References

[1] Yeastract (yeast search for transcriptional regulators and consensus tracking) [http://www.yeastract.com].

[2] M.J. Beal, F.L. Falciani, Z. Ghahramani, C. Rangel, and D. Wild. A bayesian approach to reconstructing genetic regulatory networks with hidden factors. *Bioinformatics*, 21:349–356, 2005.

[3] A. J. Butte, P. Tamayo, D. Slonim, T. R. Golub, and I. S. Kohane. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc Natl Acad Sci U S A*, 97(22):12182–12186, October 2000.

[4] D R. Cox and N. Wermuth. *Multivariate dependencies: Models, analysis and interpretation.* Chapman and Hall, London, 1996.

[5] A. De la Fuente, N. Bing, I. Hoeschele, and P. Mendes. Discovery of meaningful associations in genomic data using partial correlation coefficients. *Bioinformatics*, 20:3565–3574, 2004.

[6] D. Edwards. *Introduction to Graphical Modelling.* Springer-Verlag, New York, 1995.

[7] B. Efron. Local false discovery rates. *Technical Report number. Dept. of Statistics, Stanford University.*, 2005.

[8] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32(2):407–499, 2004.

[9] J. Fox. *An R and S-Plus companion to applied regression.* Sage Publications, Thousand Oaks, CA, USA, 2002.

[10] N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using bayesian networks to analyse expression data. *Journal of computational biology*, 7(3-4):601–620, 2000.

[11] N. Friedman, K. Murphy, and S. Russell. Learning the structure of dynamic probabilistic networks. In *Proceedings of the 14th conference on the Uncertainty in Artificial Intelligence*, pages 139–147, SM, CA, USA, Morgan Kaufmann, 1998.

[12] S. Imoto, T. Goto, and S. Miyano. Estimation of genetic networks and functional structures between genes by using bayesian networks and nonparametric regression. In *Pacific Symposium on Biocomputing 7*, pages 175–186, 2002.

[13] S. Imoto, S. Kim, T. Goto, S. Aburatani, K. Tashiro, S. Kuhara, and S. Miyano. Bayesian network and nonparametric heteroscedastic regression for nonlinear modeling of genetic network. *Journal of Bioinformatics Computational Biology*, 2:231–252, 2003.

[14] S. Kim, S. Imoto, and S. Miyano. Inferring gene networks from time series microarray data using dynamic bayesian networks. *Briefings in Bioinformatics*, 4(3):228, 2003.

[15] S. Kim, S. Imoto, and S. Miyano. Dynamic bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data. *Biosystems*, 75(1-3):57–65, 2004.

[16] S. L. Lauritzen. *Graphical models*. Oxford Statistical Science Series, 1996.

[17] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J. B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, and R. A. Young. Transcriptional regulatory networks in saccharomyces cerevisiae. *Science*, 298(5594):799–804, 2002.

[18] P. M. Magwene and J. Kim. Estimating genomic coexpression networks using first-order conditional independence. *Genome Biology*, 5(12), 2004.

[19] C. Meek. Strong completeness and faithfulness in bayesian networks. In *Proc. of the 11th Annual Conference on Uncertainty in Artificial Intelligence*, SF, CA, USA, Morgan Kaufmann Publishers, 1995.

[20] K. Murphy. The bayes net toolbox for matlab. *Computing Science and Statistics*, 33, 2001.

[21] K. Murphy and S. Mian. Modelling gene expression data using dynamic bayesian networks. *Technical report, Computer Science Division, University of California, Berkeley, CA.*, 1999.

[22] I. M. Ong, J. D. Glasner, and D. Page. Modelling regulatory pathways in e. coli from time series expression profiles. *Bioinformatics*, 18(Suppl 1):S241–S248, 2002.

[23] R. Opgen-Rhein and K. Strimmer. Learning causal networks from systems biology time course data: an effective model selection procedure for the vector autoregressive process. *BMC Bioinformatics*, 8, 2007.

[24] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. SF, CA, USA, Morgan Kaufmann Publishers., 1988.

[25] B.-E. Perrin, L. Ralaivola, A. Mazurie, S. Bottani, J. Mallet, and F. d'Alché Buc. Gene networks inference using dynamic bayesian networks. *Bioinformatics*, 19(Suppl 2):S138–S148, 2003.

[26] C. Rangel, J. Angus, Z. Ghahramani, M. Lioumi, E. Sotheran, A. Gaiba, D. L. Wild, and F. Falciani. Modeling t-cell activation using gene expression profiling and state-space models. *Bioinformatics*, 20(9):1361–1372, 2004.

[27] A. Roverato and R. Castelo. Graphical model search procedure in the large p and small n paradigm with applications to microarray data. *Journal of Machine Learning Research*, 7:2621–2650, 2006.

[28] J. Schäfer and K. Strimmer. An empirical bayes approach to inferring large-scale gene association networks. *Bioinformatics*, 21:754–764, 2005.

[29] J. Schäfer and K. Strimmer. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology*, 21, 2005.

[30] I. Simon, J. Barnett, N. Hannett, C. Harbison, N. Rinaldi, J. Volkert, T.and Wyrick, J. Zeitlinger, Gifford D. K., Jaakkola T. S., and R. A. Young. Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell*, 106:697–708, 2001.

[31] S. M. Smith, D. C. Fulton, T. Chia, D. Thorneycroft, A. Chapple, H. Dunstan, C. Hylton, S. C. Zeeman, and A. M. Smith. Diurnal changes in the transcriptome encoding enzymes of starch metabolism provide evidence for both transcriptional and posttranscriptional regulation of starch metabolism in arabidopsis leaves. *Plant Physiol.*, 136(1):2687–2699, 2004.

[32] P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher. Comprehensive identification of cell cycle-regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization. *Mol Biol Cell*, 9(12):3273–3297, 1998.

[33] P. Spirtes, C. Glymour, and R. Scheines. *Causation, prediction and search*. Springer Verlag, New York (NY), 1993.

[34] R. Steuer, J. Kurths, O. Fiehn, and W. Weckwerth. Observing and interpreting correlations in metabolomic networks. *Bioinformatics*, 19(8):1019–1026, 2003.

[35] N. Sugimoto and H. Iba. Inference of gene regulatory networks by means of dynamic differential bayesian networks and nonparametric regression. *Genome Informatics*, 15(2):121–130, 2004.

[36] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society B*, 58:267–288, 1996.

[37] H. Toh and K. Horimoto. Inference of a genetic network by a combined approach of cluster analysis and graphical gaussian modeling. *Bioinformatics*, 18:287–297, 2002.

[38] H. Toh and K. Horimoto. System for automatically inferring a genetic network from expression profiles. *J. Biol. Physics*, 28:449–464, 2002.

[39] P. J. Waddell and H. Kishino. Cluster inference methods and graphical models evaluated on nci60 microarray gene expression data. *Genome Informatics*, 11:129–140, 2000.

[40] P. J. Waddell and H. Kishino. Correspondence analysis of genes and tissue types and finding genetics links from microarray data. *Genome Informatics*, 11:83–95, 2000.

[41] J. Wang, O. Myklebost, and E. Hovig. Mgraph: graphical models for microarray data analysis. *Bioinformatics*, 19(17):2210–2211, 2003.

[42] J. Whittaker. *Graphical models in applied multivariate statistics*. Wiley, NY, 1990.

[43] A. Wille and P. Bühlmann. Low-order conditional independence graphs for inferring genetic networks. *Statist. Appl. Genet. Mol. Biol*, 4(32), 2006.

[44] A. Wille, P. Zimmermann, E. Vranova, A. Fürholz, O. Laule, and S. Bleuler. Sparse graphical gaussian modeling for genetic regulatory network inference. *Genome Biol*, 5(11), 2004.

[45] F. X. Wu, W. J. Zhang, and A. J. Kusalik. Modeling gene expression from microarray expression data with state-space equations. In *Pacific Symposium on Biocomputing*, pages 581–592, 2004.

[46] X. Wu, Y. Ye, and K. R. Subramanian. Interactive analysis of gene interactions using graphical gaussian model. *ACM SIGKDD Workshop on Data Mining in Bioinformatics*, 3:63–69, 2003.

[47] M. Zou and S. D. Conzen. A new dynamic bayesian network (dbn) approach for identifying gene regulatory networks from time course microarray data. *Bioinformatics*, 21(1):71–79, 2005.