Inferring dynamic genetic networks with low order independencies

Sophie Lèbre

Inferring dynamic genetic networks with low order independencies

Sophie Lèbre *
s.lebre@imperial.ac.uk
March 13, 2008

Université d'Evry-Val-d'Essone, CNRS UMR 8071, INRA 1152,
Laboratoire Statistique et Génome
523 place des Terrasses, 91000 Evry, France.

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 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 the R package 'G1DBN' which is available from the CRAN archive.

Keywords: conditional independence, Dynamic Bayesian Network, Directed Acyclic Graph, 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

*New address: Centre for Bioinformatics, Division of Molecular Biosciences, Imperial College London, South Kensington Campus, SW7 2AZ London, UK.

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. [21] 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 genes whose expression has been monitored using microarrays are not taking part in the temporal evolution of the system. So we want to determine the few "active" gene that are involved in the regulatory machinery, as well as the relationships between them. In short, we want to infer a network representing the dependence relationships which govern a system composed of several agents from the observation of their "activity" 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* [36]. Better known as *covariance graph* [5] 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* [18], 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 [46], [18] 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 [31] [32] , Waddell and Kishino [44] [43], Toh and Horimoto [40] [41], Wu et al. [50], Wang et al. [45]). 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) [11] model directed relationships. Based on a probabilistic measure, a BN representation of a model is defined by a Directed Acyclic Graph (DAG) and the set of conditional probability distributions of each variable given its parents in the DAG [28]. Then the theory of graphical models [46, 7, 18] 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.

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.

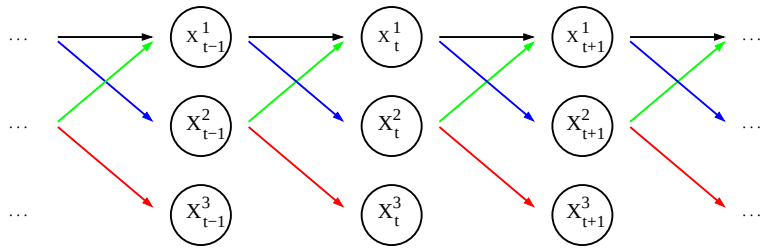

Figure 2: Dynamic network equivalent to the regulation motif in Figure 1 (left). Each vertex X_t^i represents the expression level of gene G_i at time t. This graph is acyclic and allows to define a Bayesian network.

Here comes the interest of Dynamic Bayesian networks (DBNs) first introduced for the analysis of gene expression time series by Friedman et al. [12] and Murphy and Mian [25]. 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.

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 [26, 51], multivariate auto-regressive process [27], State Space or Hidden Markov Models [29, 49, 30, 1], nonparametric additive regression model [14, 15, 17, 37]). See also Kim et al. [16] for a review of such models. Facing with as much diversity, we introduce in this paper sufficient conditions 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 DAG defining the various DBNs quoted above. Among others, Murphy [24] implemented several Bayesian structure learning procedures for dynamic models in the open-source Matlab package BNT (Bayes Net Toolbox); Ong et al. [26] reduce the dimension of the problem by considering prior knowledge; Perrin et al. [29] use an extension of the linear regression; Wu et al. [49] use factor analysis and Beal et al. [1] develop a variational Bayesian method; Zou and Conzen [51] limit potential regulators to the genes with either earlier or simultaneous expression changes and estimate the transcription time lag; Opgen-Rhein and Strimmer [27] 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. [48, 47] 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,

$$r(Y^i, Y^j) \neq 0 \quad \text{and} \quad \forall k \in \{1, ..., p\} \backslash \{i, j\},\ r(Y^i, Y^j | Y^k) \neq 0, \tag{1}$$

where $r(Y^i, Y^j|Y^k)$ 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 ([47], Corollary 1). This approach has also been used by Magwene and Kim [22] and de la Fuente et al. [6] for estimating non-directed gene networks from microarray gene expression of the yeast *Saccharomyces cerevisiae*. Castelo and Roverato [4] investigate such non directed q^{th} order partial independence graphs for $q \geq 1$ and introduce 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 particularly 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 implemented in the R package 'G1DBN' [19]. Finally, validation is obtained on both simulated and real data in Section 5. We use our inference procedure for the analysis of two microarray time course data sets: the Spellman's yeast cell cycle data [34] and the diurnal cycle data on the starch metabolism of *Arabidopsis Thaliana* collected by Smith et al. [33].

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.

The main result of this section is set out in Proposition 3; we show that process X admits a DBN 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, minimal DAG $\tilde{\mathcal{G}}_{AR(1)}$ is given in the case of an AR(1) model in Subsection 2.2.2. Most of our results are derived from the theory of graphical models associated with the DAGs [18]. Note that, even though we need to consider a homogeneous DBN for the inference of gene interaction networks, the theoretical results introduced in Sections 2.2 and 3 are valid without assuming homogeneity across time.

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,
$\mathcal{G}^{(q)}$	q^{th} order conditional dependence DAG,

2.1 Backgrounds

2.1.1 Theory of the graphical models associated with DAGs

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 [18].

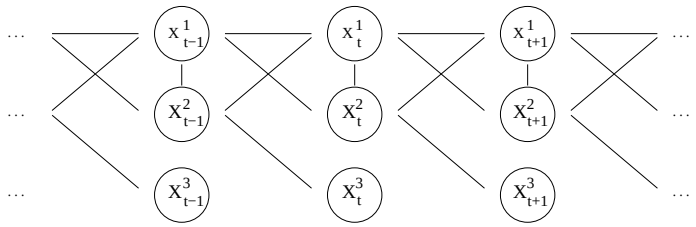

Figure 3: Moral graph of the DAG in Figure 2. For all $t > 1$, the parents of the variable X_t^1 are "married", that is connected by a non directed edge.

Definition 1 ***(Parents, Lauritzen [18])*** *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, Pearl [28])*** *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, Lauritzen [18])*** *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 displays the moral graph of the DAG in Figure 2.*

Definition 3 ***(Ancestral set, Lauritzen [18])*** *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 $An(S)$. See Figure 4 for an illustration.*

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 [18].

Proposition 2 ***(Lauritzen [18], 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.*

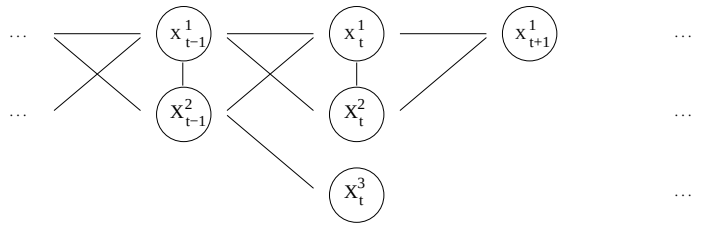

Figure 4: Moral graph of the smallest ancestral set containing the variables X^1_{t+1}, its parents in the DAG in Figure 2 and X^2_t. As the set (X^1_t, X^2_t) blocks all paths between X^3_t and X^1_{t+1}, we have $X^1_{t+1} \perp\!\!\!\perp X^3_t \mid (X^1_t, X^2_t)$.

2.1.2 Sufficient conditions for DBNs representation

We recall here sufficient conditions under which the probability distribution $\mathbb{P}$ of process X admits a BN representation according to a dynamic network (e.g. in Figure 2). 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 gene expression levels observed 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^i_t measured at time t is better explained by the previous time expression levels X_{t-1} than by some current expression level X^j_t.

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 $\{X^i_t\}_{i \in P}$ are conditionally independent given the past of the process $X_{1:t-1}$, that is,*

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

Lemma 1 *Under Assumptions 1 and 2, the probability distribution $\mathbb{P}$ admits a DBN representation according to a DAG whose edges only join nodes representing variables observed at two successive time points, at least according to the DAG $\mathcal{G}_{full} = (X, \{(X^j_{t-1}, X^i_t)\}_{i,j \in P, t>1})$ which has edges between any pair of successive variables.*

Assumptions 1 and 2 allow 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 the next time t (Lemma 1, see proof in Appendix A). The direction of the edges according to the time guarantees the acyclicity of $\mathcal{G}_{full}$.

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

2.2.1 Existence and definition

We demonstrate here the existence of a minimal DAG $\tilde{\mathcal{G}}$ according to which process X admits a DBN representation. This DAG describes exactly the full order conditional dependencies between successive variables given the past of the process (Proposition 3).

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_{t-1}^j, X_t^i) which are non adjacent in $\mathcal{G}$, we have*

$$X_t^i \perp\!\!\!\perp X_{t-1}^j \mid pa(X_t^i, \mathcal{G}) \quad \textit{and} \quad X_t^i \perp\!\!\!\perp X_{t-1}^j \mid pa(X_t^i, \mathcal{G}) \cup S,$$

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

The proofs of lemmas 2 and 3 are shown in Appendix A. As 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_t^3 and X_{t+1}^1 in this DAG. Now consider in Figure 4 the moral graph of the smallest ancestral graph containing X_t^3, X_{t+1}^1 and the parents (X_t^1, X_t^2) of X_{t+1}^1. The set (X_t^1, X_t^2) blocks all paths between X_t^3 and X_{t+1}^1. From Proposition 2, we have $X_{t+1}^1 \perp\!\!\!\perp X_t^3 \mid pa(X_{t+1}^1, \mathcal{G})$.

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 $X_t^{P_j} = \{X_t^k; k \in P_j\}$ the set of $p-1$ variables observed at time t.

Proposition 3 ***(Existence of minimal DAG $\tilde{\mathcal{G}}$, the smallest subgraph of $\mathcal{G}_{full}$ allowing DBN modeling)*** *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_{t-1}^j and X_t^i given the remaining variables $X_{t-1}^{P_j}$ observed at time $t-1$,*

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

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 A. In DAG $\tilde{\mathcal{G}}$, the set of parents $pa(X_t^i, \tilde{\mathcal{G}})$ of a variable X_t^i is the smallest subset of X_{t-1} such that the conditional densities satisfy $f(X_t^i|pa(X_t^i, \tilde{\mathcal{G}})) = f(X_t^i|X_{t-1})$. The set of parents of a 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_{t-1}^j, X_t^i) which are non adjacent in $\tilde{\mathcal{G}}$ are conditionally independent given the parents of X_t^i,

$$X_t^i \perp\!\!\!\perp X_{t-1}^j \mid pa(X_t^i, \tilde{\mathcal{G}}).$$

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

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

Consider the following first order auto-regressive model,

AR(1) model

$$X_1 \sim \mathcal{N}(\mu_1, \Sigma_1) \tag{2}$$
$$\forall t > 1, \quad X_t = AX_{t-1} + B + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma), \tag{3}$$
$$\forall s, t \in N, \; Cov(\varepsilon_t, \varepsilon_s) = \delta_{ts}\Sigma, \tag{4}$$
$$\forall s > t, \quad Cov(X_t, \varepsilon_s) = 0. \tag{5}$$

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 (5) 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 (3) and (5), the model is first order Markovian and Assumption 1 is satisfied. From (4), 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

$$a_{ij} = Cov(X_t^i, X_{t-1}^j \mid X_{t-1}^{P_j}) / Var(X_{t-1}^j \mid X_{t-1}^{P_j}).$$

As process X is Gaussian, the set of null coefficients of 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,

$$\tilde{\mathcal{G}}_{AR(1)} := \left(X, \left\{(X_{t-1}^j, X_t^i) \text{ such that } a_{ij} \neq 0;\ t > 1,\ i, j \in P\right\}\right). \tag{6}$$

As 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 Introducing q^{th} order dependence DAGs $\mathcal{G}^{(q)}$ for DBNs

In this paper, we propose to use the DBN modeling according to DAG $\tilde{\mathcal{G}}$ (introduced in Proposition 3) to model genetic regulatory networks from gene expression time series. Reverse discovering DAG $\tilde{\mathcal{G}}$ requires to determine, for each variable X_t^i, the set of variables X_{t-1}^j observed at time $t-1$ on which variable X_t^i is conditionally dependent given the remaining variables $X_{t-1}^{P_j}$.

Even under the homogeneity assumption discussed in the introduction, the available gene expression time series data do not allow such testing. Indeed, we still have to face the 'curse of dimension' as the number of genes p, is much higher than the number of measurements n. Thus we extend to DBNs the approach based on the consideration of low order independencies introduced by Wille et al.[48, 47] for GGM approximation (see more details on low order independence graph for GGMs in the introduction section). After defining q^{th} order conditional dependence DAGs $\mathcal{G}^{(q)}$ ($q < p$) for DBNs, we investigate in which manner they allow us to approximate the DAG $\tilde{\mathcal{G}}$ describing full order conditional dependencies.

3.1 DAG $\mathcal{G}^{(q)}$ Definition

Let q be smaller than p. In the q^{th} order dependence DAG $\mathcal{G}^{(q)}$, no edge is drawn between two successive variables X_{t-1}^j and X_t^i whenever it exists a subset X_{t-1}^Q of q variables among the $p-1$ variables $X_{t-1}^{P_j}$ such that X_{t-1}^j and X_t^i are conditionally independent given this subset. In short, DAGs $\mathcal{G}^{(q)}$ are defined as follows,

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

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

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. Note that the definition of q-th order partial dependence DAG $\mathcal{G}^{(q)}$ is based on exact q-th order independencies (not on all partial independencies lower than q as in the

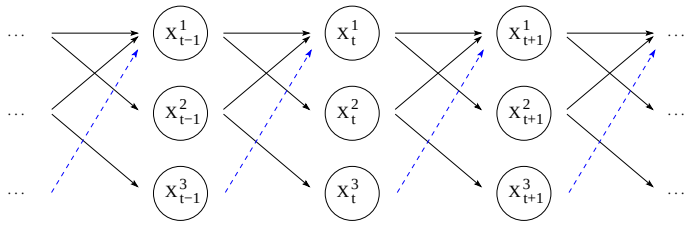

Figure 5: First-order conditional dependence DAG $\mathcal{G}^{(1)}$ (obtained from the DAG in Figure 2). The spurious dashed arrow may appear in $\mathcal{G}^{(1)}$.

partial order correlation network used by Wille and Bühlmann [47]). Indeed, we consider that including the q-th order dependencies only reflects somehow better the true DAG $\tilde{\mathcal{G}}$. In particular, for p variables, DAG $\mathcal{G}^{(p-1)}$ is DAG $\tilde{\mathcal{G}}$. This definition is possible for DBNs because dynamic modeling essentially differs from static correlation networks modeling. In particular, contrary to the case of correlation network, the "V" structures (or structures with multiple parents) do not generate spurious edges in the case of DBN since the definition of the DAG $\tilde{\mathcal{G}}$ defining full order dependencies does not allow edges between variables observed at the same time. Then, for instance, when considering the following "V" structure $X^j_{t-1} \rightarrow X^i_t \leftarrow X^k_{t-1}$, no spurious edge can be inferred between the variables X^j_{t-1} and X^k_{t-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, X^1_{t-1} (resp. X^2_{t-1}) does not separate X^1_t from X^3_{t-1} in the smallest moral graph containing the variables $X^1_t \cup X^3_{t-1} \cup X^1_{t-1}$ (resp. $X^1_t \cup X^3_{t-1} \cup X^2_{t-1}$) displayed in Figure 4. Nevertheless, if the vertices of $\tilde{\mathcal{G}}$ have few parents, DAGs $\mathcal{G}^{(q)}$ 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 accurate 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. the single input modules in the transcriptional regulatory network of *S. Cerevisiae* [21]). 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 $q \geq N_{pa}^{Max}(\tilde{\mathcal{G}})$, we have $\tilde{\mathcal{G}} \supseteq \mathcal{G}^{(q)}$.*

Proposition 5 *Let X be a Gaussian process. If $N_{pa}^{Max}(\tilde{\mathcal{G}}) \leq 1$ 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 smaller 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 A). 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 [35])*** *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 [35]) and discrete (Meek [23]) distributions)*** *Let $\pi_{\mathcal{G}}^{\mathcal{N}}$ (resp. $\pi_{\mathcal{G}}^{\mathcal{D}}$) be the set of linearly independent parameters needed to parameterize 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 [35] for Gaussian distribution and extended to discrete distribution by Meek [23], 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 remains 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 $(X_{t-1}^j, X_t^i) \in E(\mathcal{G}^{(q)}) \Rightarrow (X_{t-1}^j, X_t^i) \in E(\tilde{\mathcal{G}})$.
■

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

From Proposition 6, whenever $\mathbb{P}$ is faithful to $\tilde{\mathcal{G}}$, DAG $\mathcal{G}^{(q)}$ contains DAG $\tilde{\mathcal{G}}$. Then deriving Corollary 1 from Proposition 6 and Corollary 2, we show that both DAG $\mathcal{G}^{(q)}$ and DAG $\tilde{\mathcal{G}}$ exactly coincide if any node of $\tilde{\mathcal{G}}$ has less than q parents. Even though we expect the number of parents in a gene interaction networks to be upper bounded, the exact maximal number of parents $N_{pa}^{Max}(\tilde{\mathcal{G}})$ remains mostly unknown. However, we establish in Proposition 7, that 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. Thus, if $\mathbb{P}$ is faithful to $\tilde{\mathcal{G}}$, the knowledge of the topology of DAG $\mathcal{G}^{(q)}$ only allows us to ascertain some edges of DAG $\tilde{\mathcal{G}}$.

4 *G1DBN*, a procedure for DBNs inference

We introduced and characterized the q^{th} order dependence DAGs $\mathcal{G}^{(q)}$, for all $q < p$, for 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 $X_t^i \perp\!\!\!\perp X_{t-1}^j | X_{t-1}^Q$ 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 DAG $\tilde{\mathcal{G}}$ is a subgraph of $\mathcal{G}^{(1)}$ (Proposition 6) to develop an inference procedure for $\tilde{\mathcal{G}}$. Indeed, the inference

of $\mathcal{G}^{(1)}$ is both the faster (complexity) and the most accurate (number of tests). Thus we introduce a 2 step-procedure for DBN inference. In the first step, we infer the 1^{st} order dependence DAG $\mathcal{G}^{(1)}$, then we infer DAG $\tilde{\mathcal{G}}$ from the estimated DAG $\hat{\mathcal{G}}^{(1)}$. This 2 step-procedure, summarized in Figure 6, is implemented in a R package 'G1DBN' [19] freely available from the Comprehensive R Archive Network.

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

We evaluate the *likelihood* of an edge (X_{t-1}^j, X_t^i) by measuring the conditional dependence between the variables X_{t-1}^j and X_t^i given any variable X_{t-1}^k. Assuming linear dependencies, we consider the partial regression coefficient $a_{ij|k}$ defined as follows,

$$X_t^i = m_{ijk} + a_{ij|k} X_{t-1}^j + a_{ik|j} X_{t-1}^k + \eta_t^{i,j,k},$$

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

We measure the conditional dependence between the variables X_{t-1}^j and X_t^i given any variable X_{t-1}^k by testing null assumption $\mathcal{H}_0^{i,j,k}$: "$a_{ij|k} = 0$". 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 [10]. 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:

$$under\ (\mathcal{H}_0^{i,j,k}):\ \text{"}a_{ij|k} = 0\text{"}, \qquad \frac{\hat{a}_{ij|k}}{\hat{\sigma}(\hat{a}_{ij|k})} \sim t(n-4), \tag{7}$$

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.

4.2 Step 2: inferring $\tilde{\mathcal{G}}$ from $\mathcal{G}^{(1)}$

We use the inferred DAG $\hat{\mathcal{G}}^{(1)}$ as a reduction of the search space. Indeed, from faithfulness, we know that $\tilde{\mathcal{G}} \subseteq \mathcal{G}^{(1)}$ (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 2.1.2. Consequently, the number of parents of each variable in $\hat{\mathcal{G}}^{(1)}$ is much smaller than n. Then model selection can be carried out 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 by $a_{ij}^{(2)}$ the regression coefficient,

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 (7),

$S_1(i,j) = Max_{k \neq j}(p_{ij|k})$.

$E(\hat{\mathcal{G}}^{(1)}) = \{(X^j_{t-1}, X^i_t)_{t>1};\ i,j \in P,$ 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 (9).

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

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

Figure 6: Outline of the 2 step-procedure *G1DBN* for DBN inference.

$$X^i_t = m_i + \sum_{j \in pa(X^i_t, \hat{\mathcal{G}}^{(1)})} a^{(2)}_{ij}\ X^j_{t-1} + \eta^i_t, \tag{8}$$

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

$$\text{under } (\mathcal{H}^{i,j}_0):\ \text{“}a^{(2)}_{ij} = 0\text{”}, \qquad \frac{\hat{a}^{(2)}_{ij}}{\hat{\sigma}(\hat{a}^{(2)}_{ij})} \sim t(n-1-|pa(X^i_t, \hat{\mathcal{G}}^{(1)})|). \tag{9}$$

The score $S_2(i,j) = 1$ is assigned to the edges that are not in $\hat{\mathcal{G}}^{(1)}$. The smallest scores indicate the most significant edges. The inferred DAG for $\tilde{\mathcal{G}}$ contains those edges whose score is below a chosen threshold α_2.

When $\tilde{\mathcal{G}}$ is sparse, Step 1 of *G1DBN* inference procedure gives already a good estimation of $\tilde{\mathcal{G}}$ (see Percision-Recall curves obtained for simulated data in Figure 7). 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 $\hat{\mathcal{G}}^{(1)}$.

4.3 Choice of the thresholds

The choice of thresholds is often something non trivial, especially when using multiple testing. However, Step 1 of the procedure is conservative by construction. Indeed, the definition of score S_1 (equal to the maximum of $p-1$ p-values computed for testing 1st-order conditional independence) clearly supports the acceptation of the null assumption, i.e. the absence of an

edge. Standard approaches for multiple testing correction do not apply to choose threshold α_1. Thus we introduce a heuristic approach to choose α_1 threshold in Subsection 5.3.2.

The choice of α_2 threshold is less problematic. Indeed, the second Step of the inference procedure is a standard multivariate regression. Then the usual thresholds 1%, 5% or 10 % can be chosen or even lower threshold when a low number of edges is wanted. However, this implies the computation of multiple testing (as many tests as are edges in DAG $\mathcal{G}^{(1)}$). Then a set of predictions is *expected* to be false predictions and it is interesting to control it. It is possible for instance to control the chance of any false positives as Bonferroni or random field methods do. For a more comprehensive approach, we chose to control the expected *proportion* of false positives edges, i.e. the False Discovery Rate (FDR) with the approach introduced by Benjamini and Hochberg [2] summarized as follows. Let m be the number of remaining edges after Step 1, then Step 2 requires to compute m tests. Choose a maximal FDR level q and order the set of m observed p-values: $p_{(1)} \leq \cdots \leq p_{(i)} \leq \cdots \leq p_{(m)}$. Then reject the null assumption ($\mathrm{H}_0^{(i)}$: "Edge i is not DAG $\tilde{\mathcal{G}}$") for all $i \leq k$ where k is defined as follows,

$$k = max \left\{ i : p_{(i)} \leq \frac{i}{m} q \right\}.$$

If no such i exists, reject no hypothesis. Benjamini and Hochberg (1995) showed that this procedure ensures the FDR is lower than $q\frac{m_0}{m} \leq q$ where m_0 is the number of true null hypotheses.

4.4 Complexity of the algorithm

The complexity of this algorithm is $O(p^3)$. Note however that the score for the edges of each target gene can be computed separately. Thus parallel run is straightforward and is made possible with the R package 'G1DBN' by specifying a target gene in the function *DBNScoreStep1* for Step 1 computation.

All the computations were performed on Redhat WS 4 AMD opteron 270 (2GHz). The computation time mostly depends on the number of TF genes, i. e. the genes allowed to be parents in the DAG to be inferred. For an illustration based on DBNs inference performed from a real data set by Spellman [34] containing 786 target genes in Section 5.3.1, the computation of Step 1 required 7 minutes when allowing only 18 genes to be TFs (resp. 4 minutes with the lasso [39] and 7 seconds with the shrinkage procedure [27], which are two alternative approaches for DBN inference introduced in Section 5.1). When all the 786 genes can be TFs, the computation was parallel run and required 19 minutes by target gene (resp. 8 minutes by target gene with the lasso and 5 minutes for the whole set of 786 target genes with the shrinkage procedure). Step 2 is very quick and did not require more than 5 seconds in both studies. Even though the *G1DBN* procedure is slower than the others, inference with *G1DBN* for a data set containing 800 genes is fully computable, especially when parallel running.

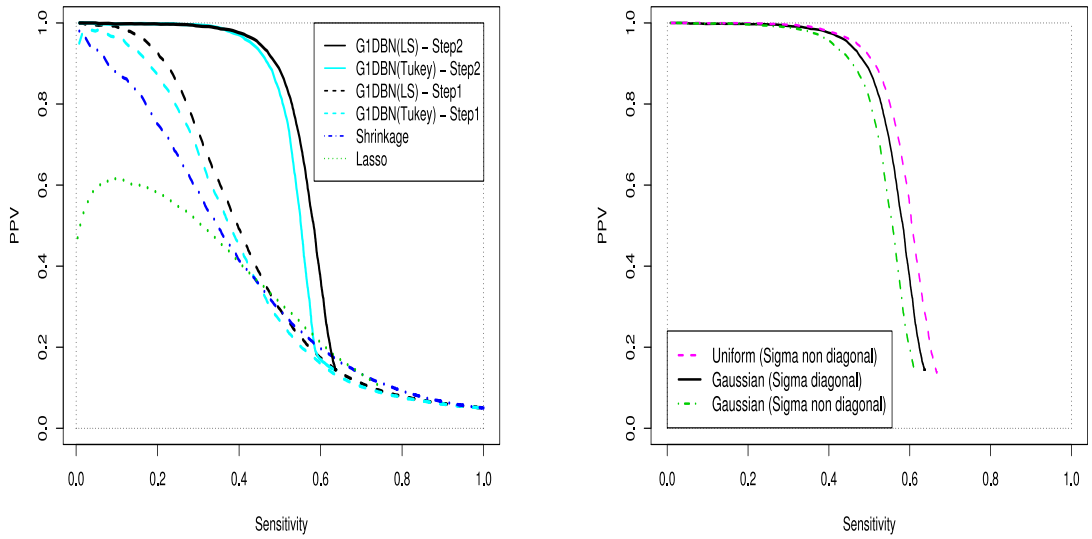

Figure 7: Percision-Recall (PR) curves obtained for network inference from simulated data ($n = 20$). Left: Comparison of the inference procedures: G1DBN (LS or Tukey), shrinkage and lasso. Step 2 of the G1DBN approach drastically improves the results (threshold $\alpha_1 = 0.7$). Right: Impact of noisy data, simulated using a non diagonal matrix Σ with either Gaussian or uniform noise, on the G1DBN procedure (Step 2) computed with LS estimates.

5 Validation

5.1 Evaluation and comparison

The criterion used to evaluate the performance of DBN inference procedures is the Precision-Recall (PR) curve as plotted in Figure 7. In PR curves, the precision, equal to the Positive Predictive Value (PPV), is displayed on the ordinate and the recall, equal to the sensitivity, on the abscissa. We recall here the next definitions,

$$PPV = \frac{TP}{TP + FP} \qquad Sensitivity = \frac{TP}{TP + FN}$$

where TP refers to the number of true positive edges, i.e. the number of edges which are selected by the inference procedure and actually belongs to the true DAG (either the DAG used for simulating the data or the validation DAG established from biological knowledge); FP refers to the number of false positive edges, i.e. the edges which are selected by the procedure but are not in the true DAG and FN refers to the number of false negative edges, i.e. the number of edges which are not selected by the procedure but are in the true DAG. PR curves are drawn by first ordering the edges by decreasing significance, and then by computing the PPV and sensitivity for the first selected edge and for each newly included edge successively.

We compare the *G1DBN* inference procedure with two reference methods for model selection for multivariate AR(1) process: the shrinkage approach by Opgen-Rhein and Strimmer [27] and the lasso (for Least Absolute Shrinkage and Selection Operator) introduced by Tibshirani [39]. Opgen-Rhein and Strimmer recently proposed a model selection procedure based

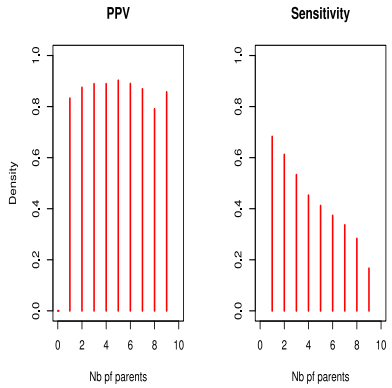

Figure 8: PPV and sensitivity according to the number of parents in the simulation matrix A obtained with G1DBN inference procedure with LS estimates (thresholds $\alpha_1 = 0.7$, $\alpha_2 = 0.01$).

on an analytic approach using James-Stein-Type shrinkage. The procedure consists of first computing the partial correlation coefficients, $r(X_t^i, X_{t-1}^j | X_{t-1}^{P_j})$, from the shrinkage estimates of the partial regression coefficients, and second, selecting the edges with a *local* false discovery rate approach [8]. Shrinkage inference is performed using the R functions implemented by Opgen-Rhein and Strimmer, available on request from the authors.

The lasso (also called L1 shrinkage) combines shrinkage and model selection. The lasso estimates are obtained by minimizing the residual sum of squares subject to the sum of the absolute values of the coefficients being less than a constant. This approach offers the advantage that it automatically sets many regression coefficients to zero. We carried out the lasso with the LARS package developped by Efron et al. [9]. Edges are ordered by decreasing absolute value of the partial regression coefficient lasso estimates.

5.2 Simulation study

As the discovery of genetic regulatory interaction is a field in progress, validation of predictions made on real gene expression data is only partial, which may render the estimation of true and false positive detection rate not fully reliable [13]. Thus we first investigate the accuracy of *G1DBN*, shrinkage and lasso and inference procedure using simulated data. We randomly generated 100 time series according to a multivariate AR(1) model defined by parameters ($A_{[p\times p]}$, B, Σ) for $p = 50$ genes. Since gene regulation networks are sparse, each matrix A contains 5 % of non zero coefficients. While keeping the number of parents low, this does not prevent a vertex to have more than one parent. Non zero regression coefficients a_{ij}, mean coefficients b_i and error variances σ_i were drawn from uniform distributions $(a_{ij}, b_i \sim \mathcal{U}([-0.95; -0.05] \cup [0.05; 0.95])$, $\sigma_i \sim \mathcal{U}[0.03, 0.08])$. Then time series were generated under the corresponding multivariate AR(1) models for $n = 20$ to 50.

In this paper, we show the results obtained with $n = 20$, which is a length one can expect from existing gene expression time series. The left panel of Figure 7 displays the average Precision Recall (PR) curves obtained with the various inference approaches (see PR curve definition in Subsection 5.1). We plotted the PR curves obtained with the *G1DBN* procedure after both Step 1 (dashed lines) and Step 2 (solid lines); with either LS (black

lines) or Tukey bisquare (light blue lines) estimates. The PR curves obtained with Huber estimates are very similar to the Tukey bisquare curves and do not appear for sake of clarity. Step 1 computed with the LS estimator gives a very high PPV for the very first selected edges. Initially, the PPV is greater than 95% while sensitivity goes up to 20%, but then the PPV decreases almost linearly against the sensitivity. However, Step 2 of the *G1DBN* procedure drastically improves the results. It allows to maintain the PPV greater than 95 % while sensitivity goes up to 50%. PR curves computed with the Tukey estimator led to comparable results. By basing on these simulated datasets, the *G1DBN* inference procedure clearly outperforms the lasso (dotted line) and the shrinkage approach (dashed-dotted line) gives results comparable to the first step of the *G1DBN* procedure only. The results of the three methods are naturally improved for greater values of n but their relative perfomances are preserved (curves not shown).

We also investigated the impact of the violation of the assumptions on the noise distribution: Gaussian distribution and diagonal covariance matrix. Thus we performed DBN inference on simulated data where the error covariance matrix Σ is not diagonal (3% of the coefficients outside the diagonal differ from 0) and the noise distribution is either Gaussian or uniform ($\mathcal{U} \sim ([-2; 2])$). As shown on the right panel of Figure 7, the accuracy of the *G1DBN* procedure (Step 2) is not strongly affected in both cases. However, it is difficult to get rid of the 1^{st} order Markov Assumption. Indeed, in order to face the dimension, the approach is dedicated to the inference of constant time delayed dependence relationships only. When simulating an AR(2) model, the 2-order time dependencies existing in the model are missed. However, the 1-order time dependencies existing in the model are still recovered. Then, when considering a 2^{nd} order Markov process, an approximation can still be performed by successively inferring 1- and 2-order time dependencies.

Note that the procedure also performs well when the number of parents in the true DAG $\tilde{\mathcal{G}}$ is greater than one. Figure 8 displays the positive predictive value (PPV) and the sensitivity according to the number of parents in the simulation model. The number of parents of each gene i is the number of non zero coefficient of i^{th} row of matrix A. The PPV is very stable and greater than 80% whatever the number of parents is. The sensitivity, even though decreasing when the number of parents increases, is still greater than 50% up to 3 parents. We chose a low α_2 threshold ($\alpha_2 = 0.01$) in order to be confident in the selected edges. When choosing α_2 thresholds greater the results are overall similar but with lower PPV and higher sensitivity.

5.3 Analysis of microarray time course data sets

5.3.1 Spellman's Yeast cell cycle data set

We apply the *G1DBN* inference procedure to the *Saccharomyces cerevisiae* cell cycle data collected by Spellman et al. [34]. 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. Out of this dataset, 6 gene expression profiles are duplicated (YCL014W, YCL061C, YCL063W, YJL019W, YML034W). We computed the mean expression profile of the duplicated genes and reduced the dataset to 786 genes. Then we included 4 genes (coding for FKH2, MBP1, MCM, SWI6) that are among the 9 transcription factors (TFs) pointed out by Lee et al. [21] as being responsible for the cell cycle regulation

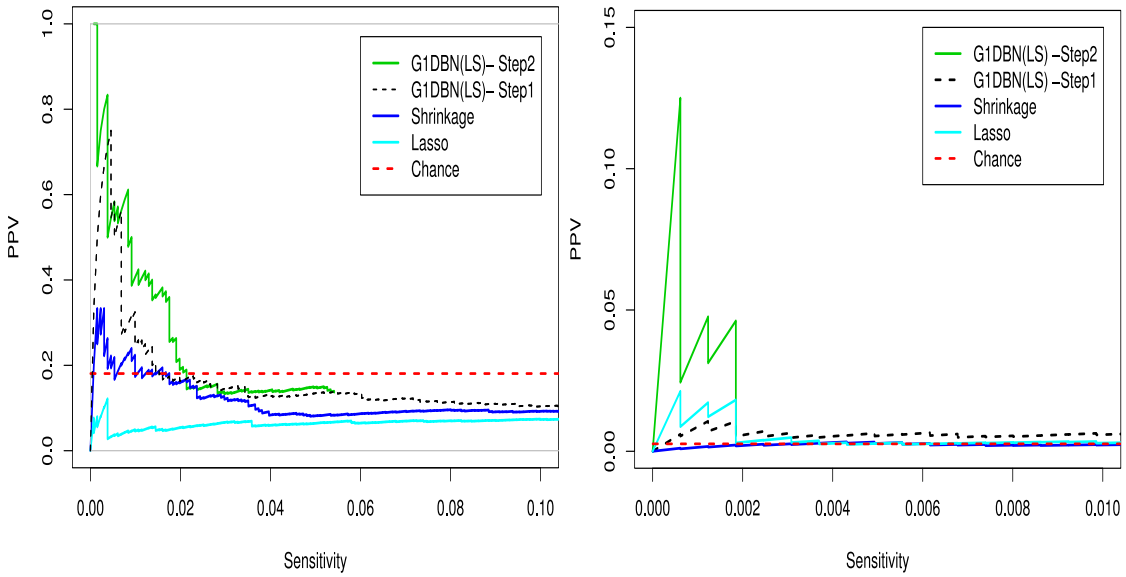

Figure 9: Zoom on Precision-Recall (PR) curves obtained from various inference procedures in two analyses carried out on Spellman's *Saccharomyces cerevisiae* cell cycle dataset (786 genes). Left: the set of putative TFs is reduced to 18 genes identified as potential TFs in [42]; Right: all the 786 genes are eligible for being TFs. The horizontal dashed line, referred to as "Chance", represents the PR curve we can expect to obtain by selecting the edges at random (9% of validated edges in the 18 TFs case, 0.0026 % in the 786 TFs case).

and are missing in this dataset.

The validation of the edges is obtained from the *Yeastract* database [38], a curated repository currently listing more than 30990 regulatory associations between TFs and target genes in Saccharomyces cerevisiae, based on more than 1000 bibliographic references. We took out of the dataset 4 genes (YCL013W, YCL022C, YCLX09W, YCRX05W) that were unknown to this database. The final dataset which contains the expression of the 786 genes is available online at [20].

We carry out two surveys on this dataset. First, we allow only a subset of 18 genes to be putative TFs (i. e. to have edges pointing out towards other genes in DAG $\tilde{\mathcal{G}}$) and look for their target genes. These 18 genes (coding for proteins ACE2, FKH1, FKH2, GAT3, MBP1, MCM1, MIG2, NDD1, PHD1, RAP1, RME1, STB1, SUT1, SWI4, SWI5, SWI6, TEC1 and YOX1) consist of the overlap between the 786 genes under study and the 50 genes identified as putative TFs in a recent study by Tsai et al. [42]. Then we extend the search for TFs to the whole dataset of 786 genes in a second survey.

In both cases, we compute the score S1, i.e. the maximal p-value for first order conditional independence, for each edge (Step 1). We set a threshold α_1 according to guidelines detailed in Subsection 5.3.2 ($\alpha_1 = 0.1$ for the 18 TF-survey, $\alpha_1 = 0.05$ for the 786 TF-survey). Edges whose score is below α_1 stand for the edges of DAG $\mathcal{G}^{(1)}$, describing the first order conditional dependencies. Second, for each gene i, we consider the regression model defined by equation (8) where the predictors are the parents of gene i in DAG $\mathcal{G}^{(1)}$. We compute the score S2, i.e. the p-value, $p_{ij}^{(2)}$, for each regression coefficient (Step 2), thus obtaining the score of the edges of the full order conditional dependencies DAG $\tilde{\mathcal{G}}$.

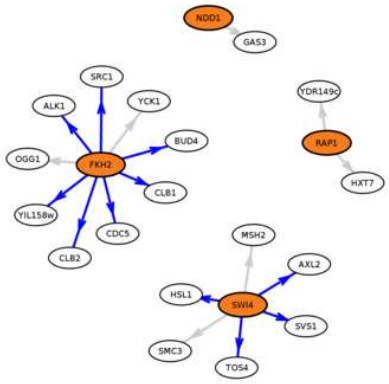

Figure 10: DAG containing the 18 first selected edges with *G1DBN* with LS estimates in the 18 TF-survey of *S. cerevisiae* cell cycle (PPV=60%). Colored nodes represent the TFs and the blue edges are validated by the Yeastract database.

The results of the *G1DBN* inference procedure, after both Step 1 and Step 2, are compared with the shrinkage approach by Opgen-Rhein and Strimmer and the lasso, introduced in Subsection 5.1. Figure 9 displays the PR curves obtained with each procedure in the two different surveys (Left: 18 TFs, Right: 786 TFs). The horizontal dashed line, referred to as "Chance", represents the PR curve which we can expect to obtain by selecting the edges at random (respectively 9% of validated edges in the 18 TFs case, 0.26 % in the 786 TFs case). In both cases, the 2-Step *G1DBN* procedure outperforms the other approaches in terms of PPV and sensitivity (according to Yeastract validation). Note however that, the shrinkage approach outperforms the lasso when considering 18 TFs only but the lasso give better results when all the 786 genes can be elected as TFs.

Surprisingly, the PR curves obtained with *G1DBN* using robust estimates failed to improve the results (compared to the LS estimates). Not plotted for sake of clarity, PR curves obtained with Huber estimates are similar to these obtained with LS estimates and Tukey estimates performed slightly worse.

In the 18 TF-survey, the first few selected edges are biologically validated (PPV=1). When considering the 18 first selected edges, the PPV is still 60 %. The corresponding inferred DAG appears in Figure 10, where the blue edges are validated by Yeastract. The first detected TFs are the genes coding for proteins FKH2, NDD1, RAP1 and SWI4. In particular, the proteins FKH2 (known as a TF with a major role in the expression of G2/M phase genes) and SWI4 (TF regulating late G1-specific transcription of targets) are pointed out as being essential TFs; they have the most target genes and the high majority (73%) of these regulatory relationships is listed in Yeastract.

As introduced in Subsection 4.3, we chose α_2 threshold in order to keep the False Discovery Rate (FDR) smaller than 1% with the approach by Benjamini and Hochberg [2]. This lead to $\alpha_2 = 0.0059$. The corresponding inferred DAG is shown Figure 11. The two proteins FKH2 and SWI4 are still part of the TFs having the most targets, together with NDD1, which is an essential component of the activation of the expression of a set of late-S-phase-specific genes and TEC1, a transcription factor required for full Ty1 expression and Ty1-mediated gene activation (Ty transposable-element own for causing cell-type-dependent activation of adjacent-gene expression). The set of selected TFs is listed in Table 2, where the third

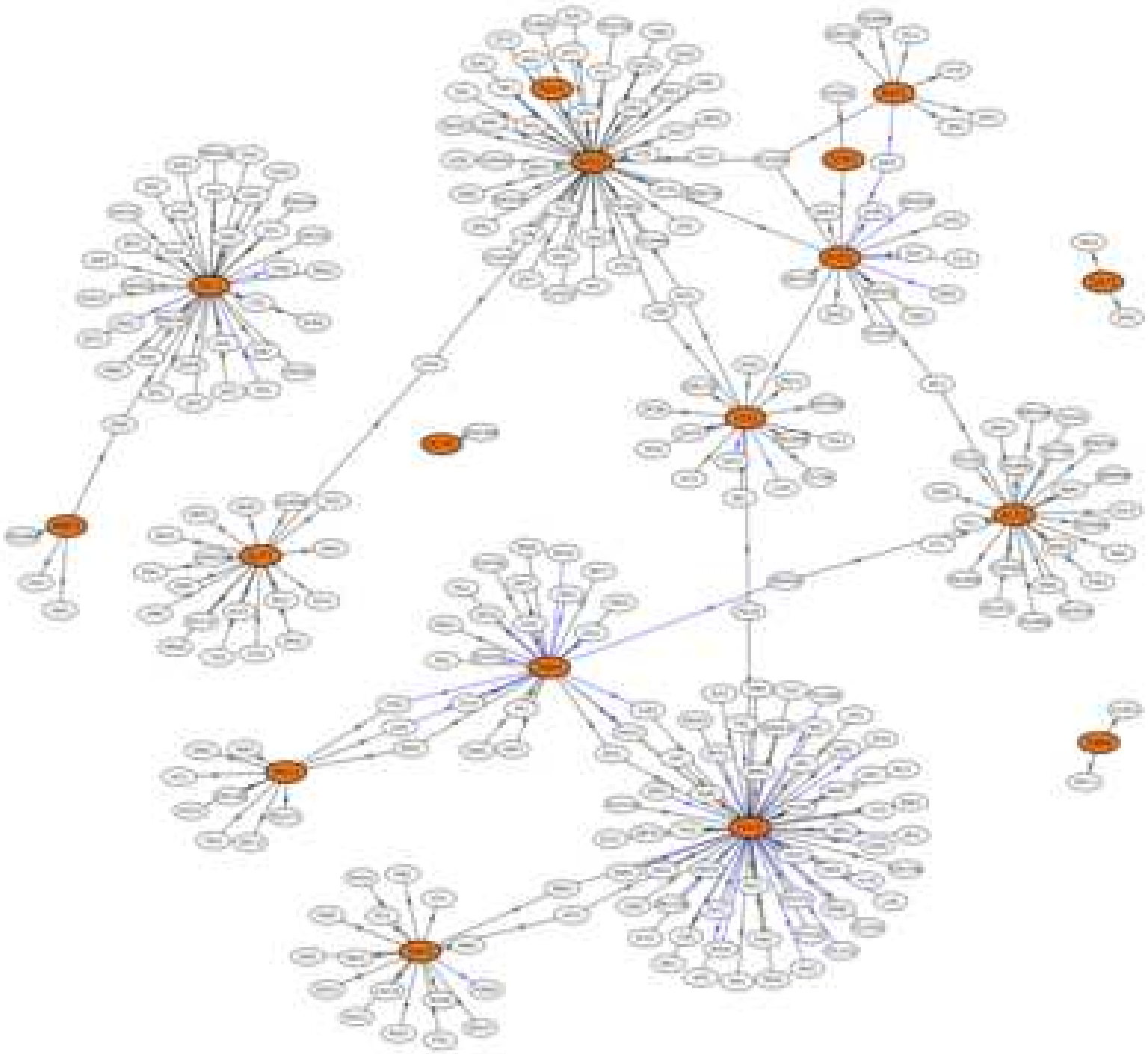

Figure 11: Inferred DAG $\tilde{\mathcal{G}}$ with *G1DBN* with LS estimates, for $\alpha_1 = 0.1, \alpha_2 = 0.0059$ (ensuring FDR< 0.01) in the 18 TF-survey of the *S. cerevisiae* cell cycle. The colored nodes

Table 2: List of the 16 genes identified as TFs with *G1DBN* with LS estimates, $\alpha_1 = 0.1$ and $\alpha_2 = 0.0059$ (ensuring FDR< 0.01) in the 18 TF-survey of the *S. cerevisiae* cell cycle. The second and third columns respectively describe the number of inferred target genes and the number of validated targets via Yeastract. These 16 TFs are the parents (colored nodes) in the DAG of Figure 11.

TF	Nb of Targets	Validated
FKH2	64	17
NDD1	51	0
TEC1	38	5
SWI4	26	12
ACE2	25	0
SUT1	20	0
SWI5	19	1
PHD1	18	4
YOX1	17	2
MIG2	11	0
RAP1	9	1
MBP1	4	0
GAT3	2	0
SWI6	2	0
MCM1	1	0
STB1	1	0

column indicates the number of validated edges out of the selected ones. Except for NDD1, for which no target gene is listed in yeastract, one forth of the targets genes of the top four TFs are validated.

In the second survey including all the 786 genes as putative TFs, the dimension is far higher and the results are consequently more restricted. Indeed, in the PR curves of the right panel of Figure 9, the PPV doesn't exceed 12.5% (obtained with the 2nd step of *G1DBN* procedure). However, this is still a subtantial result as compared with the proportion of validated edges (0.26%). In order to keep the FDR smaller than 0.01, we chose $\alpha_2 = 0.0067$ by following the Benjamini and Hochberg approach [2]. The inferred DAG for the 786 TF-survey contains 437 genes and 380 edges. This DAG, as well as the list of its edges and the list of the genes selected as TFs, is available in online supplementary information [20].

5.3.2 A heuristic approach for the choice of α_1 threshold

As the score S1 is not a usual statistical value, standard approaches to choose α_1 threshold cannot be used (see Subsection 4.3). Thus we introduce a heuristic approach to choose α_1 threshold. We carried out inference with *G1DBN* for different values of α_1, ranging from 0.01 to the maximal possible value allowing the number of parents in DAG $\mathcal{G}^{(1)}$ to be small enough as compared to the number of time points ($\alpha_1 = 0.9$ for the 18 TF-survey, $\alpha_1 = 0.5$ for the 786 TF-survey). For real data, the best results are obtained for small values of α_1,

thus substantially reducing the number of edges, but it is not entirely clear how small α_1 should be. Thresholds giving relevant results differ according to the data. However a gene is expected to have a limited number of TFs. Then a good manner to choose α_1 is to base on the distribution of the number of parents in the DAG $\mathcal{G}^{(1)}$ (obtained by keeping only the edges with a score S1 lower than α_1). Indeed, in the two studies performed here on the yeast cell cycle, the results of Step 2 are clearly related to the distribution of the number of parents in the DAG $\mathcal{G}^{(1)}$ (inferred after Step 1). The PR curves obtained for distributions generating significantly different PR curves are shown in Figures 12 and 13, respectively for the 18 TF and 786 TF-survey. Let us call q-parent genes the genes having q parents, i.e. the genes having q incoming edges in DAG $\mathcal{G}^{(1)}$. The distinctive distributions have either a dominating number of 0 parent-genes, 1 parent-genes or 2 parent-genes. The right panel of the Figures 12 and 13 displays the distribution of the number of parents in the inferred DAG $\mathcal{G}^{(1)}$ for the various values of α_1.

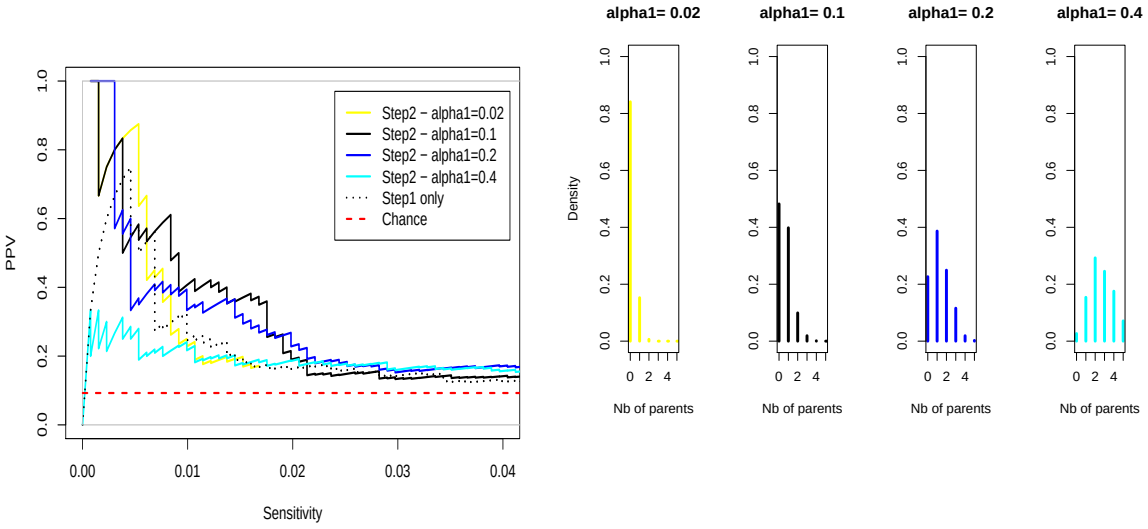

Figure 12: Impact of α_1 threshold in the 18 TF-survey on the yeast cell cycle. Left: PR curves obtained after the Step 2 of *G1DBN* for various values of α_1. Right: The distribution of the number of parents in DAG $\mathcal{G}^{(1)}$ according to the various α_1 values.

Note that the value of α_1 corresponding to a particular distribution differs in the two studies, but the PR curves are related to the distribution in the same manner. In both cases when the number of 0-parent genes clearly dominates, the PPV is great for the very first edges only. When the number of 1 parent-genes dominates, the PR curve is already slightly lower and clearly falls down when the number of 2 parent-genes dominates. However, in both the 18 TF and the 786 TF-survey, the PR curve is overall greater when the number of 0 parent-genes dominates but the number of 1 parent-genes remains far greater than the other, i.e. almost as great for the 18 TF-survey or half as great when allowing the whole set of 786 genes to be TFs. The corresponding PR curves and distributions are black plotted in Figures 12 and 13.

This does not account for a theoretical proof but still represents an empirical result observed twice in studies with different number of putative TFs. Up to now, this heuristic

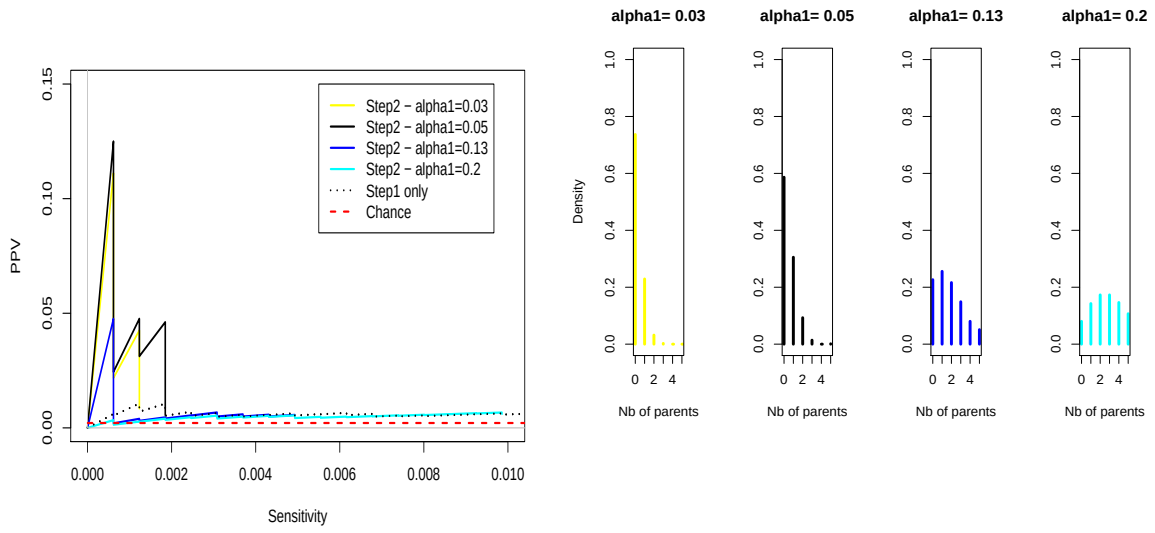

Figure 13: Impact of α_1 threshold in the 786 TF-survey on the yeast cell cycle. Left: PR curves obtained after the Step 2 of *G1DBN* for various values of α_1. Right: The distribution of the number of parents in DAG $\mathcal{G}^{(1)}$ according to the various α_1 values.

approach remains the best guideline to choose α_1 threshold according to the expectation: when an overall greatest PPV is wanted, α_1 threshold is chosen so that, in the DAG $\mathcal{G}^{(1)}$ inferred after Step 1, the number of 0 parent-genes dominates but the number of 1 parent-genes is still far greater than the rest. When only a very small number of edges is wanted, α_1 threshold can be chosen so that the number of 0 parent-genes clearly dominates. Note however that the step 1 performs already well for a small number of TF.

Table 3: List of the 9 proteins selected as parents in the DAG of Figure 14 which have been identified as Transcription Factor or DNA binding by *A. thaliana*.

Node	Gene Name	Description
26	AT2G45820-TAIR-G	DNA binding
73	AT2G43010-TAIR-G	PIF4; DNA binding / transcription factor
242	AT1G05900-TAIR-G	DNA binding / endonuclease
249	AT1G01250-TAIR-G	DNA binding / transcription factor
509	AT4G31720-TAIR-G	TAFII15; transcription factor
570	At3g57600-MinT-G	AP2 transcription factor
606	AT5G10400-TAIR-G	DNA binding
725	AT5G65360-TAIR-G	DNA binding
788	At4g14410-MinT-G	putative bHLH transcription factor

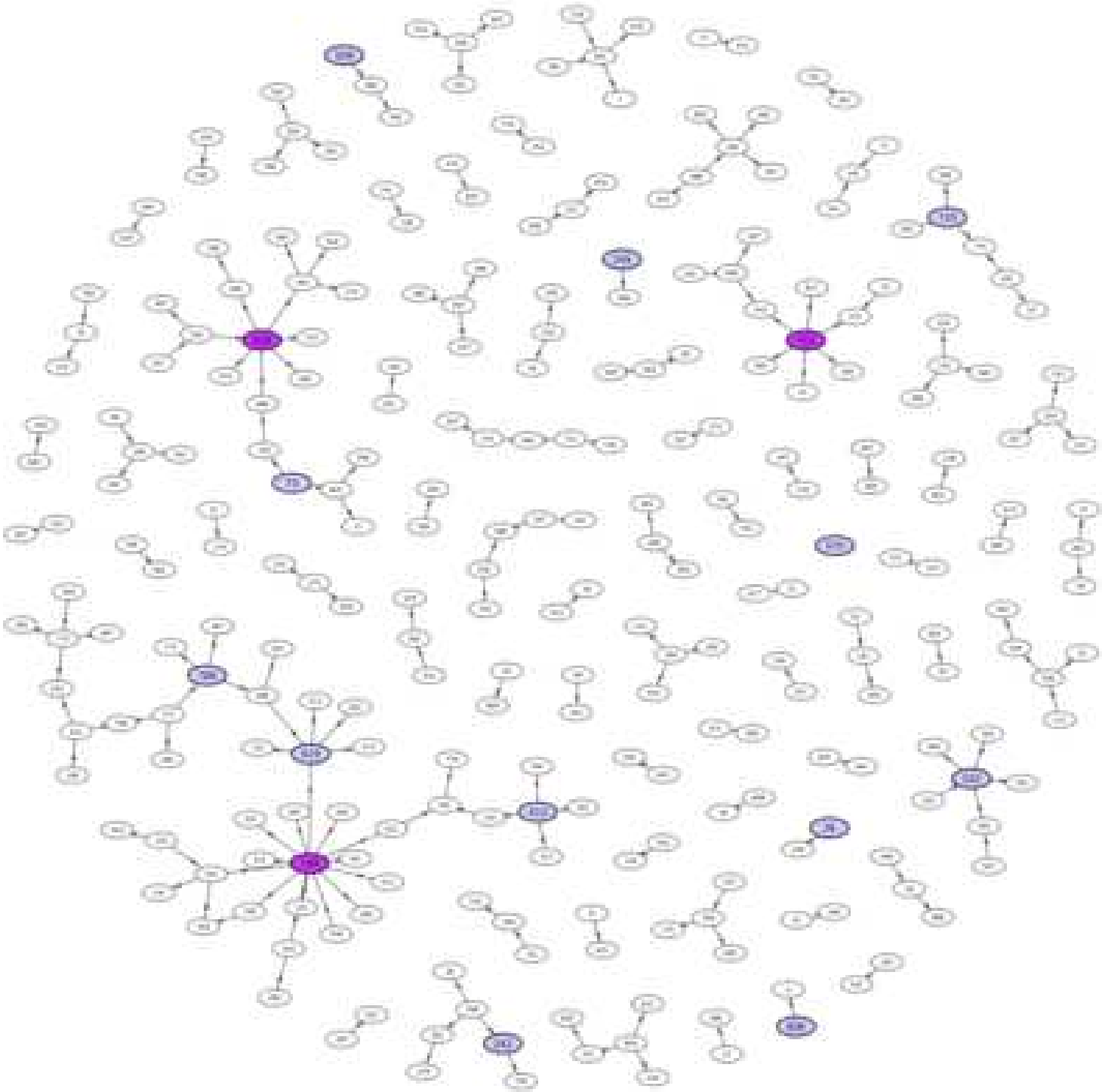

Figure 14: Inferred DAG $\tilde{\mathcal{G}}$ with *G1DBN* with LS estimates from the data by Smith et al. [33] to investigate starch metabolism of *A. thaliana* ($\alpha_1 = 0.1, \alpha_2 = 0.005$ such that FDR< 0.01) The dark colored nodes are the three nodes with the most targets, two out of them are known for being implicated in starch metabolism. The light colored nodes are parent nodes already identified as TF or DNA binding protein (see Table 3). This network contains 277 genes and 206 edges.

5.3.3 Diurnal cycle on the starch metabolism of *Arabidopsis Thaliana*

We applied our *G1DBN* inference procedure to the expression time series data generated by Smith et al. [33] 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 [27] as having periodic expression profiles. The data are available in the GeneNet R package at `http://strimmerlab.org/software/genenet/html/ar th800.html` or in our R package *G1DBN* (arth800line).

Using the heuristic approach introduced in the previous subsection, we choose threshold $\alpha_1 = 0.02$ allowing the distribution of the number of parents in the DAG $\mathcal{G}^{(1)}$ having the number of 0-parent genes to dominate and the number of 1-parent genes to be half as large. We set $\alpha_2 = 0.005$ in order to maintain the False Discovery Rate smaller than 0.01 by using the approach by Benjamini and Hochberg [2] (See Subsection 4.3 for details). We obtain the DAG $\tilde{\mathcal{G}}$ plotted in Figure 14 containing 206 edges implicating 277 different genes.

This DAG differs from the one inferred by Opgen-Rhein and Strimmer [27]. However we may notice that the edges selected by the three inference procedures discussed in this section differ somewhat. See the proportion of common edges selected by the various approaches in Appendix B. The three approaches may, in fact, yield complementary information or insights.

However we still recover a network with a "hub" connectivity structure. Among the 'parent' nodes in the DAG $\tilde{\mathcal{G}}$, two nodes (799 and 628) out of the three having the most target refers to proteins that are known to be implicated in starch metabolism. Indeed, node 799, which has 14 'target' nodes in $\tilde{\mathcal{G}}$, refers to DPE2 (DISPROPORTIONATING ENZYME 2), which is an essential component of the pathway from starch to sucrose and cellular metabolism in plant leaves at night. Node 628 (6 targets in $\tilde{\mathcal{G}}$) is a transferase (At5g24300) implicated in the starch synthase. Node 702, which is an unknown protein (At5g58220), has also 6 targets in $\tilde{\mathcal{G}}$. These three nodes are dark-colored in the DAG of figure 14. The remaining 'parent' nodes have from 1 to 4 targets. Among them, 9 genes, listed in Table 3, have already been identified as TFs or as DNA binding proteins. These 9 nodes are light-colored in the displayed DAG. Finally a list of 37 unknown proteins have been selected as parents in the inferred DAG $\tilde{\mathcal{G}}$. Potentially implicated in the regulation machinery of starch metabolism, these proteins represent a subset of genes which is relevant for further analyses.

See all the details on the inferred network displayed in Figure 14 in the online supplementary information available at [20]. In particular, the description of the 800 genes, the list of the 37 unknown proteins selected as parents in the inferred DAG, the list of the parent nodes according to their number of target nodes and the list of the edges ordered by decreasing significance are listed.

6 Discussion and conclusion

As more and more gene expression time series are available, the need for efficient tools to analyze such data has become imperious. In this paper, we first determine sufficient conditions for a Dynamic Bayesian Network modeling for gene expression time series which offers straightforward interpretation: the edges of the DAG $\tilde{\mathcal{G}}$ defining this DBN exactly describe the set of conditional dependencies between successive gene expression levels. Having defined

and characterized low order conditional dependence DAGs for DBNs, we point out relevant characteristics for the approximation of sparse DAGs. In particular, under faithfulness assumption, DAG $\tilde{\mathcal{G}}$ is included in the 1^{st} order conditional dependence DAG $\mathcal{G}^{(1)}$.

From these results, we develop *G1DBN*, a novel procedure for DBNs inference, which makes it possible to face the 'small n, large p' estimation case occuring with genetic time series data. Based on the consideration of low order conditional dependencies, the *G1DBN* procedure proved to be powerful on both simulated and real data analysis. In particular, this approach outperforms previous model selection approaches using shrinkage or lasso estimates, respectively. The shrinkage approach considerably improves the precision of the overall estimation of the partial correlation coefficients when the number of observations n is small compared to the number of genes p (with respect to standard methods). However, considering 1^{st} order conditional independence proved to be more efficient in terms of PPV and sensitivity on both simulated and real data analysis. As for the lasso, one might notice that a drawback lies in the fact that the edge selection is done vertex by vertex whereas the DAG $\tilde{\mathcal{G}}$ is globally sparse but not uniformally. As a consequence, the lasso tends to uniformally reduce the number of parents of each vertex instead of only keeping the total number of edges contained.

The power of the *G1DBN* procedure comes from the accuracy improvement of the testing made possible by the dimension reduction. Indeed, as the first step selection is based on 1^{st} order conditional independence consideration, significance tests are performed in a model of dimension 4 (see Section 4.1). This represents a drastic dimension reduction as compared to full order independence testing and makes the testing much more accurate. Thus, even if there are more edges in the DAG $\mathcal{G}^{(1)}$ than in the true DAG $\tilde{\mathcal{G}}$ (Proposition 6), Step 1 of the procedure is already very predictive.

Throughout the analyses performed for this paper, we point out two major directions for further research. On the one hand, we noticed than the edges selected by the three inference procedures differ somewhat (see Appendix B). A further relevant study would consist of analyzing in which way these DBN inference procedures could have different strenghts and may be complementary. On the other hand, the use of robust estimators like Huber or Tukey bisquare did not allow a noticeable improvement of the inference approach on the analysis of real data. Another interesting survey lies in the investigation of which measures of dependence, like non linear or other robust estimates, are the more pertinent to analyze gene expression data.

Acknowledgments

I would like to thank Catherine Matias and Bernard Prum for many stimulating and constructive discussions on this work. I also thank Michael Stumpf for his relevant advices and Thomas Thorne for the illustration of the networks. Finally, I thank the referees for their comments and suggestions which contributed to great improvement to this paper.

APPENDIX

A Some more proofs

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

$$f(X) = f(X_1) \prod_{t=2}^{n} f(X_t|X_{t-1}), \tag{10}$$

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,

$$f(X_t|X_{t-1}) = \prod_{i \in P} f(X_t^i|X_{t-1}). \tag{11}$$

From equations (10) and (11), 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 $\mathcal{G}_{full}$. ■

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 $(\mathcal{G}_{An(X_t^i \cup X_{t-1}^j \cup pa(X_t^i, \mathcal{G}))})^m$ 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 $(\mathcal{G}_{An(X_t^i \cup X_{t-1}^j \cup pa(X_t^i, \mathcal{G}))})^m$. 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 $(X, E(\mathcal{G}_{full}) \setminus (X^j_{t-1}, X^i_t))$ for any pair (X^i_t, X^j_{t-1}) such that $X^i_t \perp\!\!\!\perp X^j_{t-1} | X^{P_j}_{t-1}$, that is DAG $\tilde{\mathcal{G}}$.

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

■

Proof of Proposition 5.

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

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

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

$$
\begin{aligned}
Cov(ij|k,l) &= Cov(X^i_t, X^j_{t-1} \,|X^k_{t-1}) - \frac{Cov(X^i_t, X^l_{t-1} \,|X^k_{t-1}) Cov(X^j_{t-1}, X^l_{t-1} | X^k_{t-1})}{Var(X^l_{t-1} | X^k_{t-1})}, \\
&= Cov(X^i_t, X^j_{t-1} \,|X^k_{t-1}) \times \left[1 - \frac{(Cov(X^j_{t-1}, X^l_{t-1} | X^k_{t-1}))^2}{Var(X^j_{t-1} | X^k_{t-1}) Var(X^l_{t-1} | X^k_{t-1})} \right] \\
&\quad - \frac{Cov(X^j_{t-1}, X^l_{t-1} | X^k_{t-1}) Cov(X^i_t, X^l_{t-1} \,| X^k_{t-1}, X^j_{t-1})}{Var(X^l_{t-1} | X^k_{t-1})}.
\end{aligned}
$$

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 $\tilde{\mathcal{G}} \subseteq \mathcal{G}^{(1)}$.

■

B Comparison of the first selected edges according to the chosen inference procedure

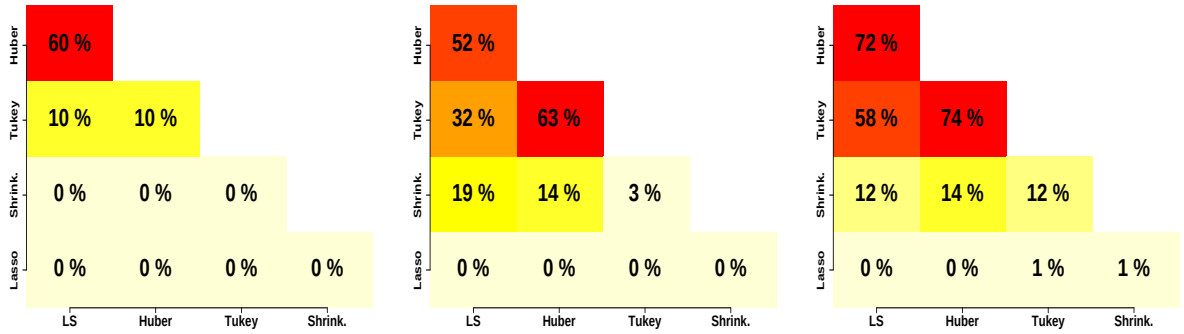

Figure 15: Proportion of shared edges among the 10, 100 and 200 first selected edges (from left to right) using different inference procedures from the yeast cell cycle data (18 TFs authorized only). Huber (resp. Tukey) refers to G1DBN with Huber (resp. Tukey) estimates.

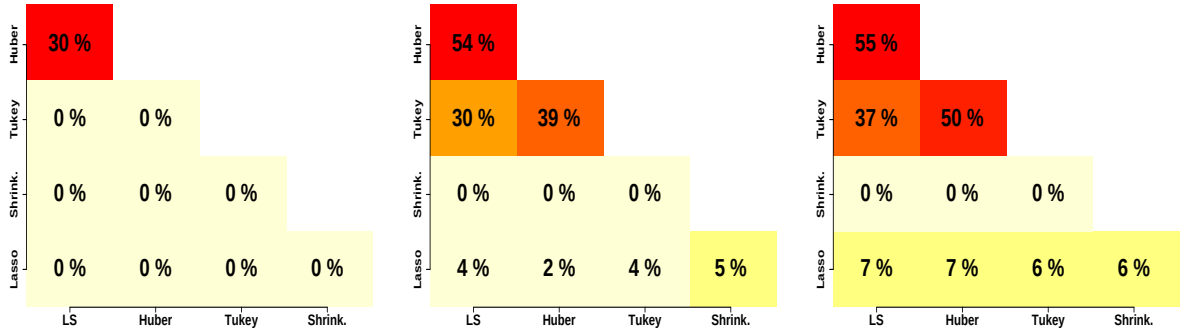

Figure 16: Proportion of shared edges among the 10, 100 and 200 first selected edges (from left to right) using different inference procedure from the yeast cell cycle data (all 786 TFs authorized).

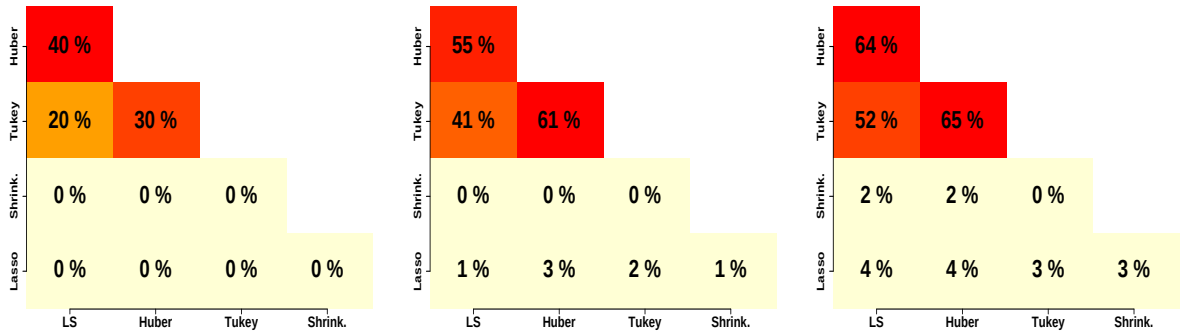

Figure 17: Proportion of shared edges among the 10, 100 and 200 first selected edges (from left to right) using different inference procedures from *A. thaliana* data (800 genes).

References

[1] 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.

[2] Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Serie B*, 57:289–300, 1995.

[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] R. Castelo and A. Roverato. 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.

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

[6] 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.

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

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

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

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

[11] 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.

[12] 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.

[13] Dirk Husmeier. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic bayesian networks. *Bioinformatics*, 19(17):2271–2282, 2003.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

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

[19] S. Lebre. G1DBN: A package performing dynamic bayesian network inference available from the Comprehensive R Archive Network at. *http://cran.r-project.org/web/packages/G1DBN/index.html*, 2008.

[20] S. Lebre. Supplementary information available from. *http://www3.imperial.ac.uk/theoreticalgenomics/people/sophielebre*, 2008.

[21] 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.

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

[23] 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.

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

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

[26] 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.

[27] 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(Suppl. 2):S3, 2007.

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

[29] 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.

[30] 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.

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

[32] 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*, 4(32), 2005.

[33] 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.

[34] 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.

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

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

[37] 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.

[38] M. C. Teixeira and P. Monteiro. The YEASTRACT database: a tool for the analysis of transcription regulatory associations in saccharomyces cerevisiae [`http://www.yeastract.com`]. *Nucleic Acids Research*, 34:D446–D451, 2006.

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

[40] 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.

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

[42] H.-K. Tsai, H. Horng-Shing Lu, and W.-H. Li. Statistical methods for identifying yeast cell cycle transcription factors. *PNAS*, 102(Sep):13532 – 13537, 2005.

[43] 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.

[44] 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.

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

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

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

[48] 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.

[49] 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.

[50] 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.

[51] 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.