



Inferring dynamic genetic networks with low order independencies

Sophie Lèbre

► **To cite this version:**

Sophie Lèbre. Inferring dynamic genetic networks with low order independencies. 2009. <hal-00142109v7>

HAL Id: hal-00142109

<https://hal.archives-ouvertes.fr/hal-00142109v7>

Submitted on 29 May 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Inferring dynamic genetic networks with low order independencies

Sophie Lèbre*
s.lebre@imperial.ac.uk

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

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

Abstract

In this paper, we propose a novel inference method for dynamic genetic networks which makes it possible to deal 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 which we extend here to the case of Dynamic Bayesian Networks. Most of our results are based on the theory of graphical models associated with 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 cope 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 modelling.

Introduction

The development of microarray technology allows to simultaneously measure the expression levels of many genes at a precise time point. Thus it has

become possible to observe gene expression levels across a whole process such as the cell cycle or response to radiation or different 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 case of 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 deal with the dimension of the problem when 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’ genes 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.

Static Modelling Such gene networks were first described using static modelling and mainly non oriented networks. One of the first tools used to describe interactions between genes is the *relevance network* [5] or *correlation network* [36]. Better known as the *covariance graph* [7] in graphical models theory, this undirected 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. However, the correlation between two variables may be caused by linkage with other variables. This creates spurious edges due to indirect dependence relationships.

Consequently, there has been great interest in the *concentration graph* [20], also called the *covariance selection* model, which describes the *conditional* dependence structure between gene expression using 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 (See Figure 1B). The standard theory of estimation in GGMs [20, 46] 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. However, in most microarray gene expression datasets, we have to cope with the opposite situation ($n \ll p$). Thus, the growing interest in “small n , large p ” furthered the development of numerous alternatives (Schäfer and Strimmer [31, 32], Waddell and Kishino [44, 43],

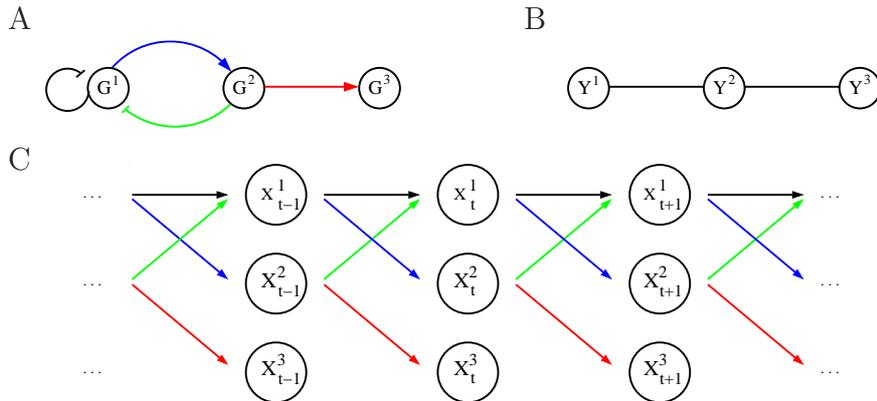


Figure 1: (A) A biological regulation motif. (B) The concentration graph corresponding to the motif A. For all $i \geq 3$, Y^i is a Gaussian variable representing the expression level of gene G^i . Some cycles cannot be represented on the concentration graph. (C) Dynamic network equivalent to the regulation motif A. 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.

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 as in Figure 1A cannot be properly represented.

Contrary to the previous undirected graphs, Bayesian networks (BNs) [13] 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]. The theory of graphical models [46, 9, 20] then 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.

Dynamic Bayesian networks This limitation can be overcome by employing Dynamic Bayesian networks (DBNs) introduced for the analysis of gene expression time series by Friedman et al. [14] and Murphy and Mian [25]. In DBNs, a gene is no longer represented by a single vertex but by as many vertices as time points in the experiment. A dynamic network (Figure 1C) can then be obtained by *unfolding in time* the initial cyclic motif in Figure 1A. The direction according to time guarantees the acyclicity of this dynamic network and consequently allows to define a Bayesian network. The nature of the relationships (positive/negative) does not appear in this DAG but is derived from estimates of the model parameters.

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 measurements of the expression level of the same gene at a given time. Hence, we assume that the process is *homogeneous* across time. This means that the system is considered to be governed by the same rules during the whole experiment. Consequently, the temporal dependencies are homogeneous: any edge is present or absent during the whole process. This is a strong assumption which is not necessarily satisfied. Nevertheless, this condition is necessary to carry out estimation unless we have several measurements of each gene expression at each time point.

Up to now, various DBN representations based on different probabilistic models have been proposed (discrete models [26, 51], multivariate autoregressive process [27], State Space or Hidden Markov Models [29, 49, 30, 3], nonparametric additive regression model [16, 17, 19, 37]). See also Kim et al. [18] for a review of such models. Faced with so much diversity, we introduce in this paper sufficient conditions for a model to admit a DBN representation and we set out a concrete 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. 1C) which describes exactly the full order conditional dependencies given all the remaining *past* variables (See Section 1). This approach extends the principle of the concentration graph showing conditional independencies to the dynamic case.

Dimension reduction Even under the assumption of homogeneity, which enables to use the pairs of successive time point gene expression as repeated measurements, we have to deal with the “curse of dimensionality” when inferring the structure of DAG $\tilde{\mathcal{G}}$. The difficulty lies in coping 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. To name a few, Murphy [24] implemented several Bayesian structure learning procedures for dynamic models in the 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. [3] 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] 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 to model

concentration graphs has gained attention. When $n \ll p$, 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

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

where $r(Y^i, Y^j | Y^k)$ is the partial correlation between Y^i and Y^j given Y^k . Hence, whenever the 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 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. [8] for estimating undirected gene networks from microarray gene expression of the yeast *Saccharomyces cerevisiae*. Castelo and Roverato [6] investigate such undirected q^{th} order partial independence graphs for $q \geq 1$ and present a thorough analysis of their properties. In this paper, we extend this approach by defining q^{th} order conditional dependence DAGs $\mathcal{G}^{(q)}$ for DBN representations. Then, by basing our results on these low order conditional dependence DAGs, we propose a novel inference method for dynamic genetic networks which makes it possible to deal with the “small n , large p ” problem.

The remainder of the paper is organized as follows. In Section 1, we provide sufficient conditions for a DBN modelling of time series describing temporal dependencies. In particular, we 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 2. From conditions on the topology of $\tilde{\mathcal{G}}$ and the faithfulness assumption, we establish inclusion relationships between both DAGs $\tilde{\mathcal{G}}$ and $\mathcal{G}^{(q)}$. In Section 3, we exploit our results on DAGs $\mathcal{G}^{(q)}$

Finally, validation is obtained on both simulated and real data in Section 4. 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

Table 1: Notations

$P = \{1 \leq i \leq p\}$	set of the observed genes,
$P_i = p \setminus \{i\}$	set of the observed genes except gene i ,
$N = \{1 \leq t \leq n\}$	set of observation times,
$X = \{X_t^i; i \in P, t \in N\}$	stochastic process (gene expression 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 the set of full order conditional dependencies,
$\mathcal{G}^{(q)}$	q^{th} order conditional dependence DAG,

by Smith et al. [33].

1 A minimal DBN representation

Let $P = \{1 \leq i \leq p\}$ describe the set of observed genes and $N = \{1 \leq t \leq n\}$ the set 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, the minimal DAG $\tilde{\mathcal{G}}$ is given in the case of an AR(1) model in Subsection 1.2. Most of our results are derived from the theory of graphical models associated with DAGs [20]. Note that, even though we need to consider a homogeneous DBN for the inference of gene interaction networks, the theoretical results introduced in Sections 1 and 2 are valid without assuming homogeneity across time.

1.1 Background

Theory of 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 DAGs [20]. A characterization of a Bayesian Network (BN) representation for a process X is given in Proposition 1.

Definition 1 (Parents, Lauritzen [20]) 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 process X admits a Bayesian Network (BN) representation according to DAG \mathcal{G} whenever its density f factorizes as a product of the conditional density of each variable X_t^i given its parents in \mathcal{G} ,

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

Throughout this paper, a central notion is that of conditional independence of random variables. Two random variables U and V are *conditionally independent* given a third variable W (and we write $U \perp\!\!\!\perp V \mid W$) if they are independent in the joint probability distribution $\mathbb{P}_{U,V,W}$ of the three random variables (U, V, W) . In other words, U and V are conditionally independent given W if for any possible value w of W , variables U and V are independent given the variable $W = w$. This result generalizes to disjoint sets of variables. Such conditional independence relationships can be obtained from a BN representation by using graphical theory associated with DAGs, which is essentially based on the *directed global Markov property* recalled in Proposition 2.

Definition 2 (Moral graph, Lauritzen [20]) 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 the directions of the original edges of \mathcal{G} . For an illustration, Figure 2A displays the moral graph of the DAG in Figure 1C.

Definition 3 (Ancestral set, Lauritzen [20]) 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 2B for an illustration.

Proposition 2 (Directed global Markov property, Lauritzen [20], 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 .

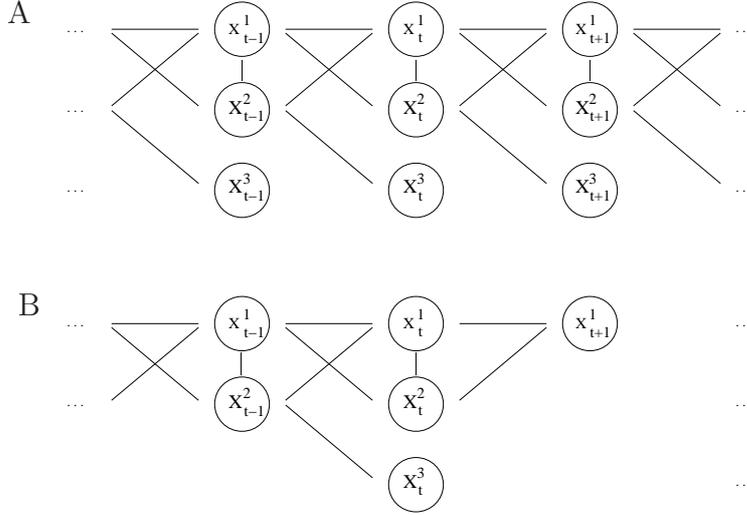


Figure 2: (A) Moral graph of the DAG in Figure 1C. For all $t > 1$, the parents of the variable X_t^1 are ‘married’, that is connected by an undirected edge. (B) Moral graph of the smallest ancestral set containing the variables X_{t+1}^1 , its parents in the DAG in Figure 1C and X_t^3 . As the set (X_t^1, X_t^2) blocks all paths between X_t^3 and X_{t+1}^1 , thus $\{X_t^1, X_t^2\}$ separates X_{t+1}^1 from X_t^3 and we have $X_{t+1}^1 \perp\!\!\!\perp X_t^3 \mid (X_t^1, X_t^2)$.

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 1C). We first assume that the observed process X_t is first-order Markovian (Assumption 1). That is, the expression level of a gene at a 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 gene expression level X_t^i measured at time t is better explained by the previous time expression levels X_{t-1} than by some current expression level X_t^j .

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

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

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

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

Assumptions 1 and 2 allow the existence of a DBN representation of the distribution \mathbb{P} according to DAG $\mathcal{G}_{full} = (X, \{(X_{t-1}^j, X_t^i)\}_{i,j \in P, t > 1})$

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 (See Lemma 1 in Appendix A.1). The direction of the edges according to time guarantees the acyclicity of \mathcal{G}_{full} .

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

Existence and definition Among the DAGs included in \mathcal{G}_{full} , we show that the probability distribution \mathbb{P} factorizes according to a minimal DAG, which we denote by $\tilde{\mathcal{G}}$ (See Lemma 2, Appendix A.1). The set of edges of $\tilde{\mathcal{G}}$ is exactly the set of full order conditional dependencies between successive variables given the past of the process as set up in the Proposition 3 (See Proof in Appendix A.2).

Proposition 3 (Existence of minimal DAG $\tilde{\mathcal{G}}$, the smallest subgraph

of \mathcal{G}_{full} allowing DBN modelling) Let $P_j = P \setminus \{j\}$ and $X_t^{P_j} = \{X_t^k; k \in P_j\}$ refer to the set P_j of $p - 1$ variables observed at time t . Whenever Assumptions 1 and 2 are satisfied, the probability distribution \mathbb{P} admits a DBN 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 DBN representation.

Thus 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 in order to recover potential regulation relationships from gene expression time series. From Proposition 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 . In short, for all i, j in P , for all $t > 1$, we have,

$$(X_{t-1}^j, X_t^i) \notin E(\tilde{\mathcal{G}}) \quad \Leftrightarrow \quad X_t^i \perp\!\!\!\perp X_{t-1}^j | pa(X_t^i, \tilde{\mathcal{G}}).$$

We will make use of this result in Section 2 in order to define low order conditional dependence DAGs for the inference of $\tilde{\mathcal{G}}$.

Minimal DAG $\tilde{\mathcal{G}}$ for an AR(1) process Consider the following first order auto-regressive model (AR(1)) with a diagonal error covariance matrix Σ ,

$$X_1 \sim \mathcal{N}(\mu_1, \Sigma_1) \tag{2}$$

$$\forall t > 1, \quad X_t = AX_{t-1} + B + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma), \tag{3}$$

$$\forall s, t \in N, \quad Cov(\varepsilon_t, \varepsilon_s) = \delta_{ts}\Sigma, \tag{4}$$

$$\forall s > t, \quad Cov(X_t, \varepsilon_s) = 0. \tag{5}$$

where $A=(a_{ij})_{1 \leq i \leq p, 1 \leq j \leq p}$ is a real matrix of size $p \times p$, $B=(b_i)_{1 \leq i \leq p}$ is a real column vector, $\Sigma = Diag(\sigma_{ii}^2)_{1 \leq i \leq p}$ is the diagonal error covariance matrix of size $p \times p$ 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 modelling assumes homogeneity across time (constant matrix A) and linearity of the dependency relationships. From (3) and (5), the model is first order Markovian (Assumption 1). From (4), Assumption 2 is satisfied whenever the error covariance matrix Σ is diagonal. Thus from Proposition 3, the probability distribution of the AR(1) process defined by equations (2-5) factorizes according to the 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

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

As process X is Gaussian, the set of null coefficients of matrix A exactly describes the conditional independencies between successive variables, thus if Σ is diagonal, we have,

$$a_{ij} = 0 \quad \Leftrightarrow \quad \forall t > 1, \quad X_t^i \perp\!\!\!\perp X_{t-1}^j | X_{t-1}^{P_j}.$$

Finally, 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,

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

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 Fig.1C ($p=3$).

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

In this paper, we propose to use the DBN modelling according to DAG $\tilde{\mathcal{G}}$ (introduced in Proposition 3) to model genetic regulatory networks from gene expression time series. Reverse discovery of 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 $X_{t-1}^{P_j}$. However, even under the time homogeneity assumption discussed in the introduction, standard estimation methods do not allow

us to infer the parameters of a regression model for p genes (*i.e.* p^2 possible edges) from np measurements. We still have to face the ‘curse of dimensionality’ since the number of genes p , is much higher than the number of measurements n .

In order to reduce the dimension, we approximate DAG $\tilde{\mathcal{G}}$ by q^{th} order conditional dependence DAGs $\mathcal{G}^{(q)}$ ($q < p$). To such an end, 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 graphs for GGMs in Section). After defining q^{th} order conditional dependence DAGs $\mathcal{G}^{(q)}$ for DBNs, we investigate the manner in which they allow us to approximate the DAG $\tilde{\mathcal{G}}$ describing full order conditional dependencies.

2.1 DAG $\mathcal{G}^{(q)}$ definition

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

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

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

DAGs $\mathcal{G}^{(q)}$ offer a way of producing dependence relationships between the variables, but they are no longer 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 partial order correlation network used by Wille and Bühlmann [47]). Indeed, we consider that including only the q^{th} order dependencies better reflects 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 modelling essentially differs from static correlation network modelling¹.

In general, DAGs $\mathcal{G}^{(q)}$ differ from DAG $\tilde{\mathcal{G}}$. For instance, the approximation of the DAG of Figure 1C by the 1st order conditional dependence

¹In 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. Thus, for instance, when considering the following “V” structure $X_{t-1}^j \rightarrow X_t^i \leftarrow X_{t-1}^k$, no spurious edge can be inferred between the variables X_{t-1}^j and X_{t-1}^k .

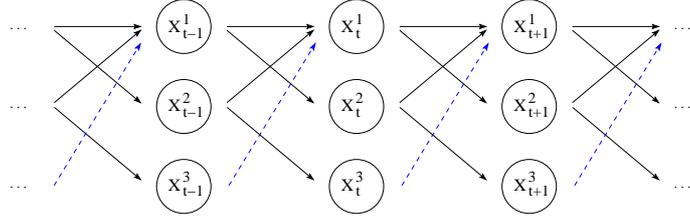


Figure 3: First-order conditional dependence DAG $\mathcal{G}^{(1)}$ (obtained from the DAG in Figure 1C). The spurious dashed arrow may appear in $\mathcal{G}^{(1)}$.

DAG may give rise to the spurious edge $X_t^3 \rightarrow X_{t+1}^1$, for all $t < n$ (See Figure 3). Indeed, X_t^1 (resp. X_t^2) does not separate X_{t+1}^1 from X_t^3 in the smallest moral graph containing the variables $X_{t+1}^1 \cup X_t^3 \cup X_t^1$ (resp. $X_{t+1}^1 \cup X_t^3 \cup X_t^2$) displayed in Figure 2B. Nevertheless, if the vertices of $\tilde{\mathcal{G}}$ have few parents, DAGs $\mathcal{G}^{(q)}$ bring relevant information about the topology of $\tilde{\mathcal{G}}$, even for small values of q . In the following, we give characterizations of low order conditional dependence DAGs $\mathcal{G}^{(q)}$ and analyze the accuracy of the approximations they offer.

2.2 A restricted number of parents

In some known gene regulation mechanisms, it is the case that a few genes regulate many other genes (e.g. the single input modules in the transcriptional regulatory network of *S. Cerevisiae* [21]). However, we do not expect a single gene to be regulated by many 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 $N_{\text{pa}}(X_t^i, \tilde{\mathcal{G}})$ the number of parents of X_t^i in DAG $\tilde{\mathcal{G}}$ and $N_{\text{pa}}^{\text{Max}}(\tilde{\mathcal{G}})$ the maximal number of parents of any variable X_t^i in $\tilde{\mathcal{G}}$,

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

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

Proposition 4 *If $N_{\text{pa}}(X_t^i, \tilde{\mathcal{G}}) \leq q$ then we have,*

$$\left\{ (X_{t-1}^j, X_t^i) \notin E(\tilde{\mathcal{G}}) \right\} \Rightarrow \left\{ (X_{t-1}^j, X_t^i) \notin E(\mathcal{G}^q) \right\}.$$

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

Proposition 5 *Let X be a Gaussian process. If $N_{\text{pa}}^{\text{Max}}(\tilde{\mathcal{G}}) \leq 1$ 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 $X_t^i \cup X_{t-1}^j \cup \text{pa}(X_t^i, \tilde{\mathcal{G}})$, the set of parents $\text{pa}(X_t^i, \tilde{\mathcal{G}})$ separates X_t^i from X_{t-1}^j . From Proposition 2, we have $X_t^i \perp\!\!\!\perp X_{t-1}^j \mid \text{pa}(X_t^i, \tilde{\mathcal{G}})$. The number of parents $\text{pa}(X_t^i, \tilde{\mathcal{G}})$ 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 this case, $\mathcal{G}^{(q)}$ does not contain spurious edges.

The converse inclusion relationship is not true in general². 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.2). 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 $\tilde{\mathcal{G}}$.

2.3 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} has measure zero with respect to Lebesgue measure over $\pi_{\mathcal{G}}^{\mathcal{N}}$ (resp. over $\pi_{\mathcal{G}}^{\mathcal{D}}$).

From Definition 5, whenever the distribution \mathbb{P} is faithful to $\tilde{\mathcal{G}}$, 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. However, Theorem 1, established by Spirtes [35] for the Gaussian distribution and extended to discrete distributions by Meek [23], makes this assumption

²As an illustration, let $X_{t-1}^j \rightarrow X_t^i$ be an edge of $\tilde{\mathcal{G}}$ then in essence (See Prop 3) X_t^i and X_{t-1}^j are conditionally dependent given the remaining variables $X_{t-1}^{P_j}$. There may however exist a subset of q variables X_{t-1}^Q , where Q is a subset of $P \setminus \{j\}$ 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 us to establish some conditional independencies, DAG $\tilde{\mathcal{G}}$ does not necessarily 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 $\tilde{\mathcal{G}}$.

reasonable at least in a measure-theoretic sense. Moreover this assumption remains very reasonable in a modelling framework where the network to be inferred describes actual interaction relationships. The next propositions are derived from the faithfulness of the distribution \mathbb{P} to $\tilde{\mathcal{G}}$ (See proofs in Appendix A.2).

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

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 $(X_{t-1}^j, X_t^i) \in E(\mathcal{G}^{(q)}) \Rightarrow (X_{t-1}^j, X_t^i) \in E(\tilde{\mathcal{G}})$.

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

Whenever \mathbb{P} is faithful to $\tilde{\mathcal{G}}$, DAG $\mathcal{G}^{(q)}$ contains DAG $\tilde{\mathcal{G}}$ (Proposition 6). Even though we expect the number of parents in a gene interaction networks to be bounded above, the exact maximal number of parents $N_{pa}^{Max}(\tilde{\mathcal{G}})$ remains mostly unknown. However, we show 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 (Proposition 7). Thus, if \mathbb{P} is faithful to $\tilde{\mathcal{G}}$, knowledge of the topology of DAG $\mathcal{G}^{(q)}$ only allows us to ascertain some edges of DAG $\tilde{\mathcal{G}}$. From Propositions 6 and 7, we establish that both DAG $\mathcal{G}^{(q)}$ and DAG $\tilde{\mathcal{G}}$ exactly coincide if any node of $\mathcal{G}^{(q)}$ has less than q parents (Corollary 3).

3 *G1DBN*, a procedure for DBN inference

We introduced and characterized the q^{th} order dependence DAGs $\mathcal{G}^{(q)}$, for all $q < p$, for dynamic modelling. We now exploit our results to develop a non-Bayesian inference method for DAG $\tilde{\mathcal{G}}$ defining a DBN representation for process X . 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) in order to develop an inference procedure for $\tilde{\mathcal{G}}$. Indeed, the inference of $\mathcal{G}^{(1)}$ is both faster (complexity) and more accurate (number of tests). Thus we introduce a 2 step-procedure

for DBN inference. In the first step, we infer the 1st 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 4, is implemented in a R package ‘G1DBN’ [1] freely available from the Comprehensive R Archive Network.

3.1 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 the 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 [12]. 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:

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

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 $\text{Max}_{k \neq j}(p_{ij|k})$ of the $p-1$ computed p-values, that is the most favorable result to 1st 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 assigned a score below a chosen threshold α_1 .

3.2 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 Section 1.1. Consequently, the number of parents of each variable in $\hat{\mathcal{G}}^{(1)}$ is much smaller than n . Then model selection can be

Choose either LS, Huber or Tukey estimator and set α_1 and α_2 thresholds.

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),

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

$E(\hat{\mathcal{G}}^{(1)}) = \{(X_{t-1}^j, X_t^i)_{t>1}; i, j \in P, \text{ such that } S_1(i, j) < \alpha_1\}$.

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

If $N_{pa}^{Max}(\hat{\mathcal{G}}^{(1)}) \sim n - 1$, choose a higher threshold α_1 and go to Step1.

For all i such that $N_{pa}(X_t^i, \hat{\mathcal{G}}^{(1)}) \geq 1$, compute the p-value $p_{ij}^{(2)}$ from (9).

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

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

Figure 4: Outline of the 2 step-procedure *G1DBN* for DBN inference.

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,

$$X_t^i = m_i + \sum_{j \in \text{pa}(X_t^i, \hat{\mathcal{G}}^{(1)})} a_{ij}^{(2)} X_{t-1}^j + \eta_t^i, \quad (8)$$

where the rank of the matrix $(X_{t-1}^j)_{t \geq 2, j \in \text{pa}(X_t^i, \hat{\mathcal{G}}^{(1)})}$ is $|\text{pa}(X_t^i, \hat{\mathcal{G}}^{(1)})|$ and the errors $\{\eta_t^i\}_{t \geq 2}$ 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 $p_{ij}^{(2)}$ derived from the significance test,

$$\text{under } (\mathcal{H}_0^{i,j}) : " a_{ij}^{(2)} = 0 ", \quad \frac{\hat{a}_{ij}^{(2)}}{\hat{\sigma}(\hat{a}_{ij}^{(2)})} \sim t(n - 1 - |\text{pa}(X_t^i, \hat{\mathcal{G}}^{(1)})|). \quad (9)$$

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 Precision-Recall curves obtained for simulated data in Figure 5). 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 $\hat{\mathcal{G}}^{(1)}$.

3.3 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 acceptance of the null assumption, *i.e.* the absence of an edge. Standard approaches for multiple testing correction do not apply to choose α_1 threshold. Thus we introduce a heuristic approach to choose α_1 threshold which is detailed in Supplementary Material [2], Section B. Overall, α_1 threshold is chosen so that, after the Step 1, the number of genes having exactly one parent in DAG $\mathcal{G}^{(1)}$ predominates.

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 a lower threshold when a low number of edges is wanted. However, a large number of tests are computed (as many as edges in DAG $\mathcal{G}^{(1)}$). In such multiple testing situations, a set of the predictions are *expected* to be false and it is useful to control this. We control the expected *proportion* of false positives edges, *i.e.* the False Discovery Rate (FDR) with the approach introduced by Benjamini and Hochberg³ [4].

3.4 Complexity of the algorithm

The complexity of this algorithm is $O(p^3)$. However the scores $(S_1(i, j))_{j \in P}$ of the incoming edges of each target gene i can be computed separately by using parallel run. This option is available in the R package *G1DBN* by specifying the target gene i in the function *DBNScoreStep1* dedicated to the 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 DBN inference performed from a real data set by Spellman [34] containing 786 target genes in Section 4.3, the computation of Step 1 required 7 minutes when the set of possible TF genes was restricted to 18 genes (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 4.1). When all the 786 genes can be TFs,

³Let 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: $p_{(1)} \leq \dots \leq p_{(i)} \leq \dots \leq p_{(m)}$. Then reject the null assumption ($H_0^{(i)}$: "Edge i is not DAG \mathcal{G} ") for all $i \leq k$ where k is defined as follows: $k = \max \{i : p_{(i)} \leq \frac{i}{m}q\}$. 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.

the computation was parallel run and required 19 minutes by target gene with *G1DBN* (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 of *G1DBN* is very quick and requires less than 5 seconds for the 786-TF study. Despite the need for more time, inference with *G1DBN* for a data set containing 800 genes is fully computable, especially when parallel running.

4 Validation

4.1 Comparison with two reference methods

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 (Least Absolute Shrinkage and Selection Operator) introduced by Tibshirani [39]. Opgen-Rhein and Strimmer recently proposed a model selection procedure based 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 [10]. Shrinkage inference is performed using the R code for shrinkage estimation⁴ by Opgen-Rhein and Strimmer.

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 performed the lasso with the R package LARS developed by Efron et al. [11].

4.2 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 [15]. Thus we first investigate the accuracy of *G1DBN*, the shrinkage and the lasso inference procedures on simulated data.

Data generation We generated 100 random time series according to a multivariate AR(1) model defined by parameters $(A_{[p \times p]}, B, \Sigma)$ for $p = 50$

⁴available at <http://strimmerlab.org/software.html>.

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 from having more than one parent. Non zero regression coefficients a_{ij} , mean coefficients b_i and error variances σ_i were drawn from uniform distributions ($a_{ij}, b_i \sim \mathcal{U}([-0.95; -0.05] \cup [0.05; 0.95])$, $\sigma_i \sim \mathcal{U}[0.03, 0.08]$). Time series were generated under the corresponding multivariate AR(1) models for $n=20$ to 50.

Evaluation based on PR curves We evaluated the performance of DBN inference procedures using the *Precision-Recall (PR)* curve as plotted in Figure 5. PR curves show the precision, equal to the Positive Predictive Value (PPV) on the ordinate against the recall, equal to the power, on the abscissa. PR curves are drawn by first ordering the edges by decreasing significance, and then by computing the PPV and power for the first selected edge and for each newly included edge successively. We recall the next definitions,

$$\begin{aligned}
 \text{Positive Predictive Value (PPV)} &= \text{True Discovery Rate (TDR)} \\
 &= 1 - \text{False Discovery Rate (FDR)} \\
 &= \frac{TP}{TP + FP} \\
 \text{Recall} = \text{Sensitivity} = \text{Power} &= \frac{TP}{TP + FN}
 \end{aligned}$$

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 (used for simulating the data); 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.

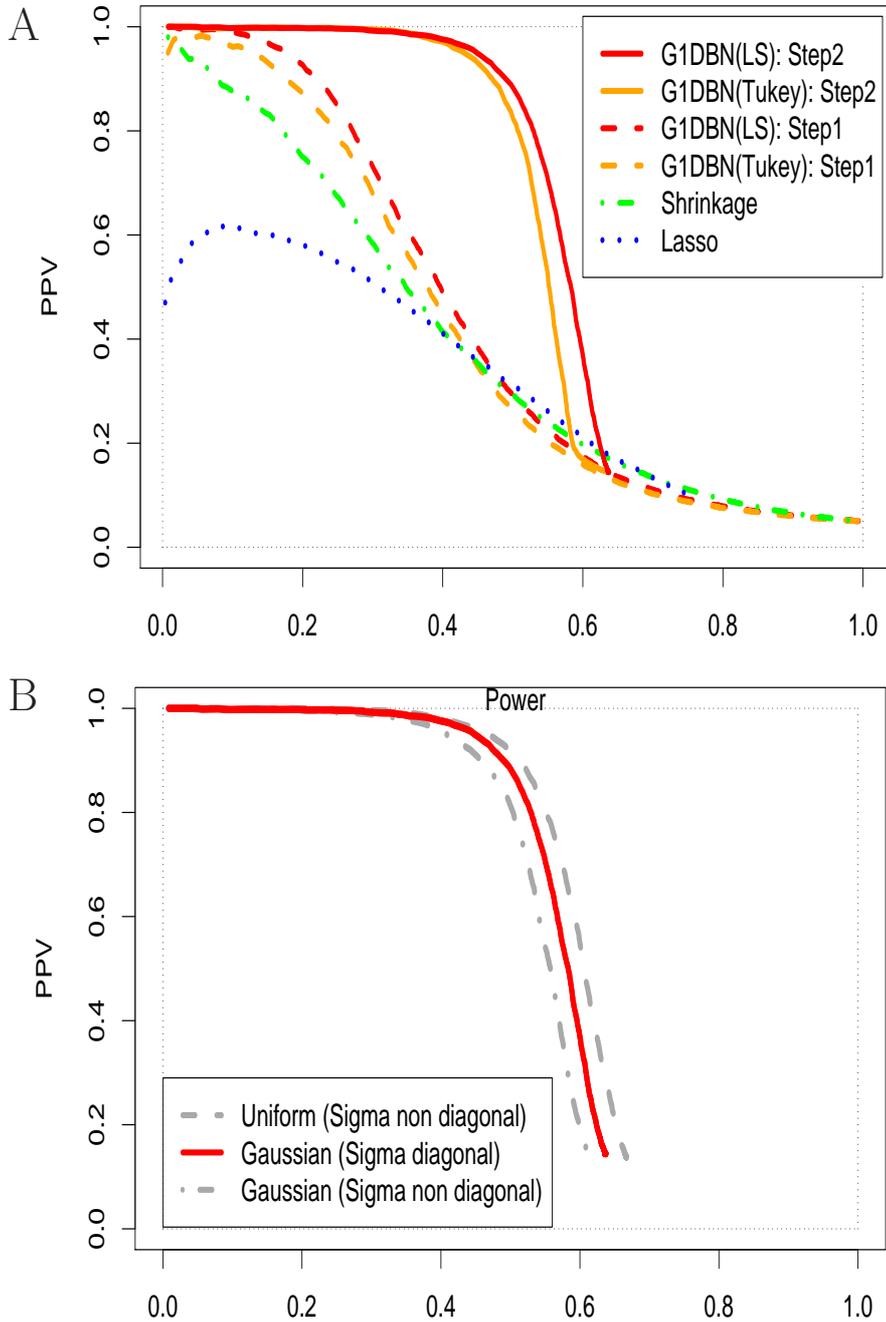


Figure 5: Precision-Recall (PR) curves obtained for network inference from simulated data ($n = 20$). (A) Comparison of the inference procedures: G1DBN (LS or Tukey), shrinkage and lasso. Step 2 of the G1DBN approach drastically improves the results (threshold $\alpha_1 = 0.7$). (B) Impact of noisy data, simulated using a non diagonal matrix Σ with either Gaussian or uniform noise, on the G1DBN procedure (Step 2) computed with LS estimates.

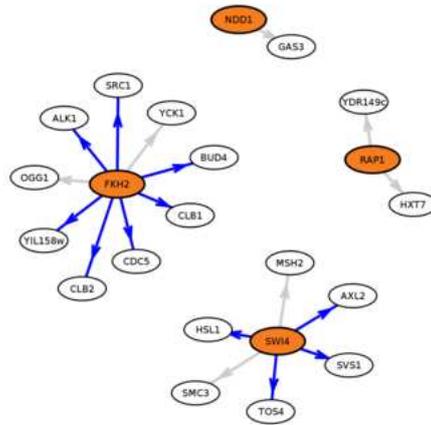
Simulation results We show on Figure 5 the results obtained with $n = 20$, a length one can expect from existing gene expression time series. Figure 5A displays the average Precision Recall (PR) curves obtained with the various inference approaches when the error covariance matrix Σ is diagonal and the noise distribution is Gaussian. The Step 1 of the *G1DBN* procedure computed either with the LS estimator or with the Tukey estimator (dashed lines) gives a very high PPV for the very first selected edges. The Step 2 of the *G1DBN* procedure (solid line) drastically improves the results. It allows to maintain the PPV greater than 95 % while the power goes up to 50%. PR curves computed with the Huber estimates (not shown) led to comparable results. The lasso (dotted line) is clearly outperformed by the other approaches and the shrinkage approach (dashed-dotted line) gives results comparable to the Step 1 of the *G1DBN* procedure only. The results of the three methods are naturally improved for greater values of n but their relative performances are preserved (curves not shown).

We investigated the impact of the violation of the model assumptions. First 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 ($\mathcal{U}[-2; 2]$). As shown on Figure 5B, the accuracy of the *G1DBN* procedure (Step 2) is not strongly affected when these assumptions on the noise distribution are not satisfied. However, it is difficult to get rid of the 1st order Markov Assumption which was chosen in order to reduce the model dimension. 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 2nd 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 (See Supp. Material [2], Section A).

4.3 Analysis of microarray time course data sets

Spellman’s Yeast cell cycle data set We performed dynamic network inference from the *Saccharomyces cerevisiae* cell cycle data collected by Spellman et al. [34]. We used the α Factor-based synchronization data (18 time points) and we focus here on a set of 786 genes which demonstrated consistent periodic changes in transcription level (See Supplementary Material [2], Section D.1 for more details).

A



B

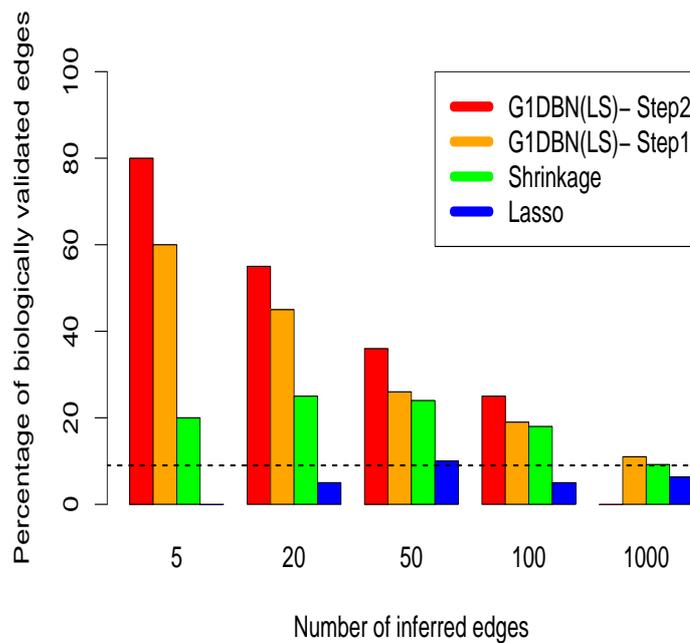


Figure 6: Some results of the 18 TF-survey of *S. cerevisiae* cell cycle. (A) DAG containing the 18 first selected edges with *G1DBN* with LS estimates (PPV=60%). Colored nodes represent the TFs and the dark blue edges are validated by the Yeastract database. (B) Percentage of validated edges out of the first 5 to 1000 edges inferred with the *G1DBN* procedure, after Step 2 or after Step 1 only, the shrinkage or the lasso procedure. The dashed line shows the proportion of validated edges out of the 786×18 possible edges.

We carried out two surveys on this dataset. First, we allow only a subset of 18 genes⁵ identified as putative TFs to be possible parent genes (*i.e.* to have edges pointing out towards other genes in DAG $\tilde{\mathcal{G}}$) and look for their target genes. Then we extend the search for parent genes to the whole dataset of 786 genes in a second survey. We set α_1 threshold for the *G1DBN* procedure according to guidelines detailed in Supplementary Material [2], Section B ($\alpha_1 = 0.1$ for the 18 TF-survey, $\alpha_1 = 0.05$ for the 786 TF-survey).

It is somehow difficult to assert the validity of the results obtained from real data as the whole regulatory machinery is not known yet. However the yeast cell cycle has been studied a lot and many regulation relationships have been recovered. We study the consistency of the first inferred edges with annotations in the *Yeasttract* database [38], a curated repository currently listing found regulatory associations between TFs and target genes in *S. cerevisiae*.

In the 18 TF-survey, the first few selected edges are biologically validated. In the DAG comprising the 18 first selected edges (Figure 6A), 11 edges refer to identified regulatory relationships (thick blue edges). 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 *Yeasttract*.

As introduced in Section 3.3, we chose α_2 threshold in order to keep the False Discovery Rate (FDR) smaller than 1% with the approach by Benjamini and Hochberg [4]. This lead to $\alpha_2 = 0.0059$. The corresponding inferred DAG is shown in Figure 7. 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 Supplementary Material [2], Section D.2, Table 1, where the third 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.

⁵The 18 genes code 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].

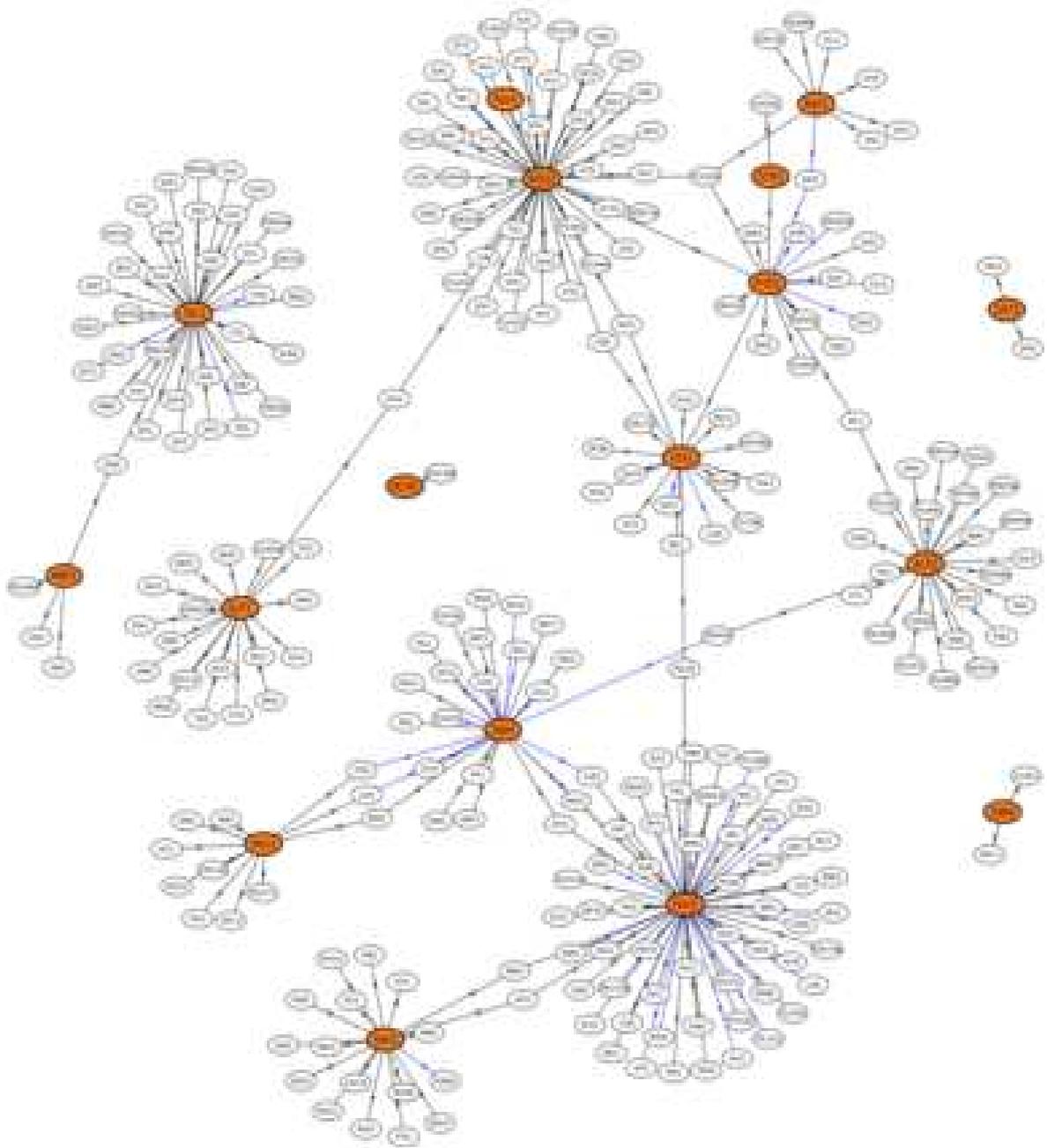


Figure 7: DAG inferred by *G1DBN* with LS estimates, using $\alpha_1 = 0.1$, $\alpha_2 = 0.0059$ (ensuring $\text{FDR} < 0.01$), in the 18 TF-survey of the *S. cerevisiae* cell cycle. The 17 colored nodes represent the 16 TFs selected as parent node out of the 18 TFs under study, plus node FKH1 which is selected as a target of NDD1. The dark blue edges are validated by YeastRACT. This network contains 286 genes and 308 edges. See the complete edges list in Supp. Material [2].

For a comparative overview, the histogram of Figure 6B displays the percentage of validated edges out of the first 5 to 1000 selected edges inferred with each inference procedure. When considering the 1000 first inferred edges, the results are very similar to what could be expected by chance only. Note that, as the Step 2 of G1DBN choose 308 edges only, it is not considered when comparing the 1000 first edges.

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, the proportion of validated edges doesn't exceed 12.5%, obtained with the 2nd step of *G1DBN* procedure among the first selected edges. However, this is still a substantial result as compared with the proportion of validated edges (equal to 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 [4]. The inferred DAG for the 786 TF-survey contains 437 genes and 380 edges. The display of this DAG, as well as the list of its edges and the list of the genes selected as TFs, is available in Supplementary Material [2].

Diurnal cycle on the starch metabolism of *A. Thaliana* We applied the *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 profiles⁶.

Using the heuristic approach detailed in Supplementary Material [2], Section B, 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 [4] (See Section 3.3 for details). We recover the DAG in Figure 8 which has a “hub” connectivity structure. This network contains 206 edges implicating 277 different genes. We may notice that this DAG differs from the one inferred by Opgen-Rhein and Strimmer [27]. However the edges selected by the three inference procedures discussed in this section differ somewhat (See the proportion of edges in common by using the various inference approaches in Supplementary Material [2], Section C) and may, in fact, yield complementary information or insights.

⁶The 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).

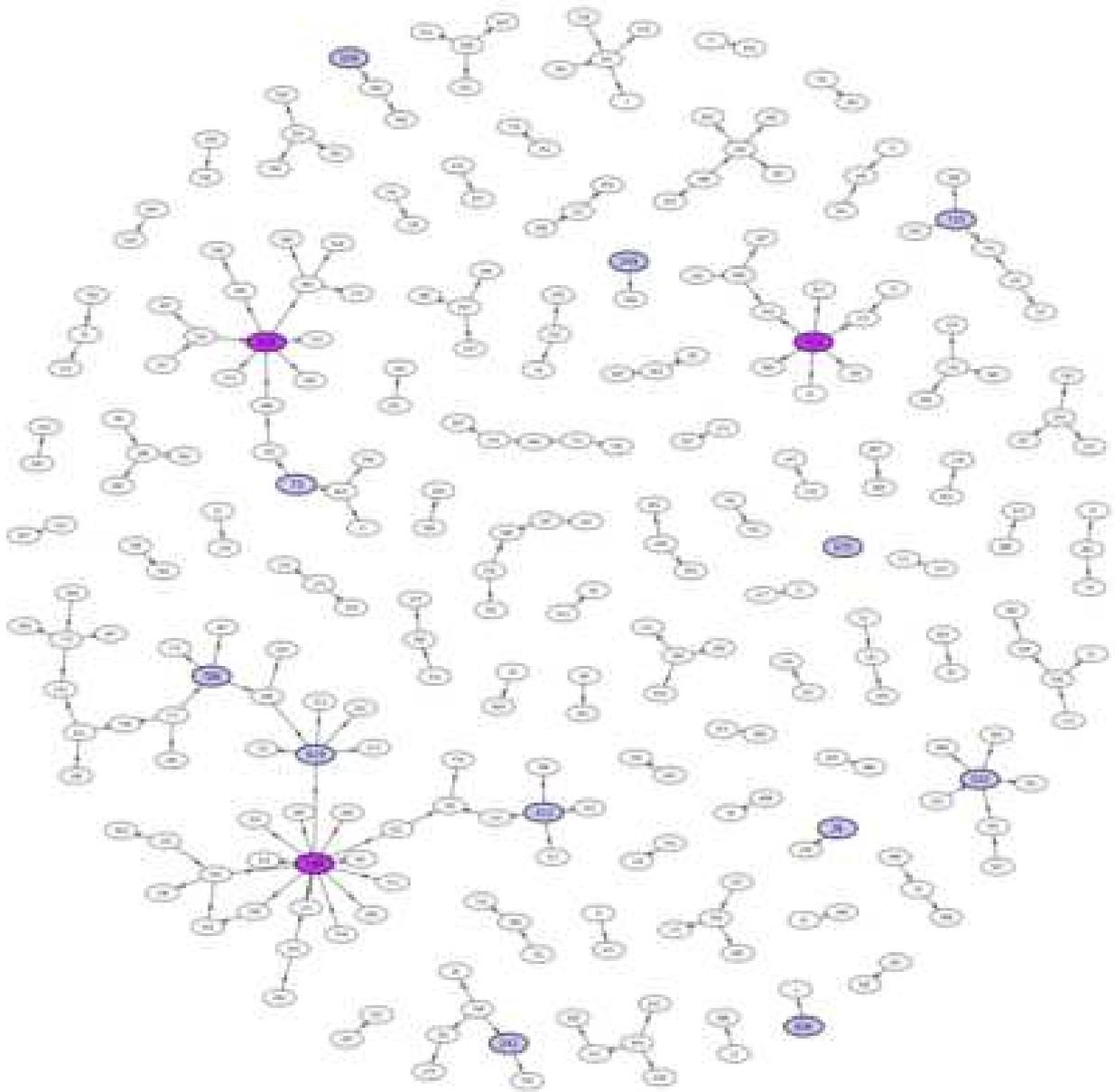


Figure 8: DAG inferred with *G1DBN* from the data by Smith et al. [33] in order to investigate starch metabolism of *A. thaliana* (LS estimates, $\alpha_1 = 0.1$, $\alpha_2 = 0.005$ such that $FDR < 0.01$). The dark colored nodes are the 3 nodes with the most targets, 2 out of them are known for being implicated in starch metabolism. The light colored nodes are parent nodes already identified as TF or DNA binding protein (See Supp. Material[2], Section E, Table 2). This network contains 277 genes and 206 edges. See the edges list in Supp. Material.

Among the ‘parent’ nodes in the inferred DAG displayed in Figure 8, 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, 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) is a transferase (At5g24300) implicated in the starch synthase. Node 702, which is an unknown protein (At5g58220), has also 6 targets. These three nodes are dark-colored in the DAG of Figure 8. Note that there is no prior knowledge regarding the role of each gene (TF or target) in this survey. As a consequence, some edges might be inferred wrong way around⁷. Thus node 799, which is a gene coding for an enzyme (DPE2), is most probably not a TF for its 14 apparent target genes. However node 799 is still the gene whose expression level best explains the expression of the 14 genes. Consequently these genes might be implicated in the same pathway as DPE2. The remaining parent nodes have from 1 to 4 targets. Among them, 9 genes, which are listed in Supplementary Material [2], Section E, Table 2, 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. Potentially implicated in the regulation machinery of starch metabolism, these proteins represent a subset of genes which is relevant for further analyses. See more details on the inferred network displayed in Figure 8 in the Supplementary Material [2].

5 Discussion and conclusion

As more and more gene expression time series has become available, the need for efficient tools to analyze such data has become imperative. In this paper, we first determine sufficient conditions for Dynamic Bayesian Network modelling of gene expression time series. This type of modelling offers a straightforward interpretation: the edges of the DAG $\tilde{\mathcal{G}}$ defining the 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 1st order conditional dependence DAG $\mathcal{G}^{(1)}$.

From these results, we develop *G1DBN*, a novel procedure for DBN

⁷In particular if some assumption of the model is not satisfied. For instance if an essential TF is missing or if the regulation is not transcriptional, *i.e.* does not depend on the amount of mRNA coding for the protein.

inference, which makes it possible to tackle the ‘small n , large p ’ estimation case that occurs 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. With respect to other methods, 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 . However, considering 1st order conditional independence proved to be more efficient for DBN inference in terms of power and PPV on simulated data, and gave promising results on 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 uniformly. As a consequence, the lasso tends to uniformly 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 the 1st order conditional independence consideration, significance tests are performed in a model of dimension 4 (See Section 3.1). This represents a drastic dimension reduction 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 that the edges selected by the three inference procedures differ somewhat (See Supplementary Material [2], Section C). A further relevant study would consist of analyzing in which way these DBN inference procedures could have different strengths and may be complementary. On the other hand, the use of robust estimators like Huber or Tukey bisquare did not allow a noticeable change of the inference approach on 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.

APPENDIX

A Proofs

A.1 Lemmas 1 to 3 and proofs

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 DAG $\mathcal{G}_{full} = (X, \{(X_{t-1}^j, X_t^i)\}_{i,j \in P, t > 1})$ which has edges between any pair of successive variables.*

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

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

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 $f(X_t | X_{t-1})$ can be written 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,

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

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} . ■

Lemma 2 *Assume the joint probability distribution \mathbb{P} of 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 $\mathcal{G}_1 \cap \mathcal{G}_2$.*

From Lemma 2, it is straightforward that, among the DAGs included in \mathcal{G}_{full} , there exists a minimal DAG (denoted by $\tilde{\mathcal{G}}$ in the paper) according to which the probability distribution \mathbb{P} factorizes, thus establishing a BN representation of process X .

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 ,

$$pa_1 = \{j \in P; X_{t-1}^j \in pa(X_t^i, \mathcal{G}_1)\}$$

$$\overline{pa}_1 = P \setminus \{pa_1\}$$

$$pa_2 = \{j \in P; X_{t-1}^j \in pa(X_t^i, \mathcal{G}_2)\}$$

$$\overline{pa}_2 = P \setminus \{pa_2\}$$

- and the densities of the joint or marginal probability distributions of (X_t^i, X_{t-1}) ,

$g : \mathbb{R}^{p+1} \rightarrow \mathbb{R}$ 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 $(X_t^i, X_{t-1}^{pa_1})$

where,

$$X_{t-1}^{pa_1} = pa(X_t^i, \mathcal{G}_1),$$

g^{i,\overline{pa}_2} the density of the joint probability distribution of $(X_t^i, X_{t-1}^{\overline{pa}_2})$

where

$$X_{t-1}^{\overline{pa}_2} = X_{t-1} \setminus \{pa(X_t^i, \mathcal{G}_2)\},$$

etc...

In the following, $y \in \mathbb{R}, x = (x_1, \dots, x_p) \in \mathbb{R}^p$ 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,

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

Equivalent results derived from the factorization according to \mathcal{G}_2 gives,

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

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

$$\begin{aligned} \int g^{i,pa_2}(y, x_{pa_2}) d(x_{pa_2 \cap \overline{pa_1}}) &= \int \frac{g^{i,pa_1}(y, x_{pa_1})}{g^{pa_1}(x_{pa_1})} g^{pa_2}(x_{pa_2}) d(x_{pa_2 \cap \overline{pa_1}}) \\ g^{i,pa_1 \cap pa_2}(y, x_{pa_1 \cap pa_2}) &= \frac{g^{i,pa_1}(y, x_{pa_1})}{g^{pa_1}(x_{pa_1})} g^{pa_1 \cap pa_2}(x_{pa_1 \cap pa_2}) \end{aligned}$$

Finally we have,

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

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 $\mathcal{G}_1 \cap \mathcal{G}_2$. ■

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*

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

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

As an illustration of Lemma 3, assume \mathbb{P} admits a BN representation according to the DAG of Figure 1C. There is no edge between X_t^3 and X_{t+1}^1 in this DAG. Now consider in Figure 2B 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 $X_{t+1}^1 \perp\!\!\!\perp X_t^3 \mid pa(X_{t+1}^1, \mathcal{G})$.

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 $S \subseteq \{X_u^k; k \in P, u < t\}$. ■

A.2 Proof of Propositions 3, 5, 6 and 7

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 (See Appendix A.1) and Prop. 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.

From Lemma 2 (See Appendix A.1), \mathbb{P} admits a BN representation according to the intersection of the DAG $(X, E(\mathcal{G}_{full}) \setminus (X_{t-1}^j, X_t^i))$ for any pair (X_t^i, X_{t-1}^j) such that $X_t^i \perp\!\!\!\perp X_{t-1}^j | X_{t-1}^{P_j}$, that is DAG $\tilde{\mathcal{G}}$.

Also, DAG $\tilde{\mathcal{G}}$ cannot be reduced. Indeed, let (X_{t-1}^l, X_t^k) be an edge of $\tilde{\mathcal{G}}$

and assume that \mathbb{P} admits a BN representation according to $\tilde{\mathcal{G}} \setminus (X_{t-1}^l, X_t^k)$, that is DAG $\tilde{\mathcal{G}}$ with the edge (X_{t-1}^l, X_t^k) removed. From Lemma 3 (Appendix A.1), we have $X_t^k \perp\!\!\!\perp X_{t-1}^l | X_{t-1}^{P_l}$, which contradicts $(X_{t-1}^l, X_t^k) \in V(\tilde{\mathcal{G}})$ (i.e. $X_t^k \not\perp\!\!\!\perp X_{t-1}^l | X_{t-1}^{P_l}$).

■

Proof of Proposition 5.

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

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

$$Cov(X_t^i, X_{t-1}^j | X_{t-1}^{P_j}) \neq 0.$$

Now assume that there exists $k \neq j$, such that $X_t^i \perp\!\!\!\perp X_{t-1}^j | X_{t-1}^k$ ie $(X_{t-1}^j, X_t^i) \notin E(\mathcal{G}^{(1)})$. We are going to prove that this contradicts the nullity of covariance $Cov(X_t^i, X_{t-1}^j | X_{t-1}^{P_j}) \neq 0$.

Let l be an element of $P \setminus \{j, k\}$. The conditional covariance $Cov(ij|k, l) = Cov(X_t^i, X_{t-1}^j | X_{t-1}^k, X_{t-1}^l)$ can be written,

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

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

- since $X_t^i \perp\!\!\!\perp X_{t-1}^j \mid X_{t-1}^k$, then $Cov(X_t^i, X_{t-1}^j \mid X_{t-1}^k) = 0$,
- as $N_{pa}^{\text{Max}}(\tilde{\mathcal{G}}) \leq 1$, X_{t-1}^j is the only parent of X_t^i in $\tilde{\mathcal{G}}$. So the variable X_{t-1}^j and thus also the set (X_{t-1}^j, X_{t-1}^k) blocks all paths between X_{t-1}^l and X_t^i in the moral graph of the smallest ancestral set containing $X_t^i \cup X_{t-1}^{j,k,l}$. Then we have, $X_t^i \perp\!\!\!\perp X_{t-1}^l \mid \{X_{t-1}^j, X_{t-1}^k\}$, that is $Cov(X_t^i, X_{t-1}^l \mid X_{t-1}^k, X_{t-1}^j) = 0$.

Then $Cov(ij|k, l) = 0$. By induction, we obtain $Cov(X_t^i, X_{t-1}^j \mid X_{t-1}^{P_j}) = 0$ leading to a contradiction with $(X_{t-1}^j, X_t^i) \in E(\tilde{\mathcal{G}})$. Therefore $(X_{t-1}^j, X_t^i) \in \mathcal{G}^{(1)}$ and we have $\tilde{\mathcal{G}} \subseteq \mathcal{G}^{(1)}$.

■

Proof of Prop 6 .

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 there 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 $\tilde{\mathcal{G}}$. ■

Proof of Prop 7 .

From faithfulness, $\tilde{\mathcal{G}} \subseteq \mathcal{G}^{(q)}$. Then for all i in P , for all $t > 1$, we have $N_{pa}(X_t^i, \tilde{\mathcal{G}}) \leq N_{pa}(X_t^i, \mathcal{G}^{(q)}) \leq q$.

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 $(X_{t-1}^j, X_t^i) \in E(\mathcal{G}^{(q)}) \Rightarrow (X_{t-1}^j, X_t^i) \in E(\tilde{\mathcal{G}})$.

■

References

- [1] G1DBN: A package performing Dynamic Bayesian Network inference, by S. Lèbre, available from the Comprehensive R Archive Network at <http://cran.r-project.org/web/packages/G1DBN/index.html>.
- [2] Supplementary Material available is at https://lsiit.u-strasbg.fr/fdbt-fr/index.php/g1dbn_supplementary_material.
- [3] 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.

- [4] 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.
- [5] 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. *PNAS*, 97(22):12182–12186, October 2000.
- [6] 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.
- [7] D R. Cox and N. Wermuth. *Multivariate dependencies: Models, analysis and interpretation*. Chapman and Hall, London, 1996.
- [8] 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.
- [9] D. Edwards. *Introduction to Graphical Modelling*. Springer-Verlag, New York, 1995.
- [10] B. Efron. Local false discovery rates. *Technical Report number. Dept. of Statistics, Stanford University.*, 2005.
- [11] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Annals of Statistics*, 32(2):407–499, 2004.
- [12] J. Fox. *An R and S-Plus companion to applied regression*. Sage Publications, Thousand Oaks, CA, USA, 2002.
- [13] 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.
- [14] 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.
- [15] Dirk Husmeier. Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic Bayesian networks. *Bioinformatics*, 19(17):2271–2282, 2003.

- [16] 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.
- [17] 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.
- [18] 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.
- [19] 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.
- [20] S. L. Lauritzen. *Graphical models*. Oxford Statistical Science Series, 1996.
- [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.yeasttract.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.