# Computational approaches for reconstruction of time-varying biological networks from Omics data

Vinay Jethava, Chiranjib Bhattacharyya and Devdatt Dubhashi

**Abstract** This chapter presents a survey of recent methods for reconstruction of time-varying biological networks such as gene interaction networks based on time series node observations (e.g. gene expressions) from a modeling perspective. Time series gene expression data has been extensively used for analysis of gene interaction networks, and studying the influence of regulatory relationships on different phenotypes. Traditional correlation and regression based methods have focussed on identifying a single interaction network based on time series data. However, interaction networks vary over time and in response to environmental and genetic stress during the course of the experiment. Identifying such time-varying networks promises new insight into transient interactions and their role in the biological process. A key challenge in inferring such networks is the problem of high-dimensional data i.e. the number of unknowns $p$ is much larger than the number of observations $n$. We discuss the computational aspects of this problem; and examine recent methods that have addressed this problem. These methods have modeled the relationship between the latent regulatory network and the observed time series data using the framework of probabilistic graphical models. A key advantage of this approach is natural interpretability of network reconstruction results; and easy incorporation of domain knowledge into the model. We also discuss methods that have addressed the problem of inferring such time-varying regulatory networks by integrating multiple sources or experiments including time series data from multiple perturbed networks. Finally, we mention software tools that implement some of the methods discussed in this chapter. With next generation sequencing promising yet further growth in publicly available -omics data, the potential of such methods is significant.

Vinay Jethava
Chalmers University of Technology, Göteborg, Sweden, e-mail: jethava@chalmers.se

Chiranjib Bhattacharyya
Indian Institute of Science, Bangalore, India, e-mail: chiru@csa.iisc.ernet.in

Devdatt Dubhashi,
Chalmers University of Technology, Göteborg, Sweden, e-mail: dubhashi@chalmers.se

## 1 Introduction

Most cellular components exert their functions through interactions with other cellular components, which can be located either in the same cell or across cells, and even across organs. In humans, the potential complexity of the resulting network – the human *interactome* – is daunting: with about 25,000 protein–coding genes, 1000 metabolites and an undefined number of distinct proteins and functional RNA molecules, the number of cellular components that serve as the nodes of the interactome easily exceeds 100,000.

It is increasingly recognized that an understanding of a gene's network context is essential in determining the phenotypic impact of defects that affect it. To understand the behaviour of any one gene in the context of human disease, individual genes must be understood in the context of molecular networks that define the disease states. Following on from this principle, a key hypothesis is that a disease phenotype is rarely a consequence of an abnormality in a single effector gene product, but reflects various pathobiological processes that interact in a complex network. A corollary of this widely held hypothesis is that the interdependencies among a cell's molecular components lead to deep functional, molecular and causal relationships among apparently distinct phenotypes [9, 34, 54]. Analysis based on microarray expression experiments has been used extensively for exploring these interdependencies.

The past decade has seen exponential growth in publicly available datasets for analysis of gene regulatory networks, predominantly in the form of time series gene expression data. Reflecting this trend, the focus has shifted from traditional perturbation experiments (knockout/knockdown) in which a single gene or a pair of genes are inactivated and the downstream effects are studied [68, 67]; to a more holistic approach aimed at studying the influence of several regulators simultaneously based on time series gene expression data [2, 8, 14].

Initial methods for analysis of time series gene expression data focussed on correlation-based methods to identify the regulatory relationships (see [4] for an early survey focussing on correlation-based methods). Gitter et. al. [27] present a recent survey discussing methods using lagged correlation and regression analysis for inferring gene regulatory networks. They also discuss methods for combining time series gene expression data with static data for reconstruction of the regulatory network.

One of the key challenges in using time series gene expression data for inference of gene regulatory networks is the relatively small number of observations compared to the number of unknown variables [28]. For example, gene expression time series data is much smaller (usually less than 8 time points, [22]) compared to the number of possible interactions between genes at different time points. Traditional methods for time-series analysis [4] fail to address this. The above problem arises in a number of domains, and is often referred to as the curse of high-dimensional data [19, 16]. This refers to the scenario when the number of unknown variables, typically represented by $p$ is larger than the number of observations, typically rep-

resented by $n$ i.e. large $p$, small $n$. This problem is ill-posed, and cannot be solved without additional assumptions.

Several regularization methods have been investigated for addressing this problem. One popular choice is to introduce $\ell_1$ penalty on the model parameters (interaction strengths) which makes the resulting problem well-posed. The $\ell_1$ penalty is known to yield sparse networks i.e. most of the interactions are absent in the resulting network. This is especially well-suited for reconstruction of gene regulatory networks, since it is known that regulatory networks are sparse i.e. only a handful of genes act as regulators for a single gene [17]. Recent advances [13, 20] in theoretical understanding of $\ell_1$-based regularization have led to the development of multiple optimization methods and related applications collectively referred to using the encompassing term *compressed sensing*. See e.g. [24] for a textbook introduction to $\ell_1$-based regularization and related techniques, and [12] for a more advanced treatment.

The problem of high-dimensional data is exacerbated in the case of reconstruction of time-varying networks since reconstruction of the network at a time point depends on the single observation i.e. expression data at that time point. In principle, the interaction networks at different time points could be very dissimilar. However, a natural consequence of the underyling biological process is that networks at nearby time points are largely similar; except in the case of sharp changes in response to external stimula [43]. Modeling the time-varying evolution of the network is a key aspect of time-varying network reconstruction. Current methods have modeled the dynamic evolution of the network under different assumptions e.g. smooth variation i.e. the network is changing slowly over time [58], or sharp changes in network structure in response to external stimuli [1]. Other methods have used additional domain knowledge like presence or absence of certain motifs [29] and functional roles of the genes [36].

One solution to the high-dimensional data problem in the context of time-varying network reconstruction is to perform multiple measurements at each time point under similar experimental conditions. However, this is largely infeasible. A related approach is to combine data from multiple related sources or experiments. Such data and the underlying networks often exhibit commonalities such as the presence of a large common subnetwork. On the other hand, there might be significant differences in some part of the networks either to due experimental conditions or genetic perturbations. Integrating expression data for network reconstruction poses a twofold challenge, namely, modeling the common subnetwork and the variation across the networks corresponding to different data sources, and capturing the impact of network variation on the node observations. In some cases, there is additional information available such as the genetic perturbations in the networks. Recent methods [30, 36] have focussed on modeling the network variation under the assumption that there is a large common subnetwork.

Gitter et. al. [27] discuss computational methods for reconstruction of regulatory networks providing a broad overview of the different approaches that have been used. In this chapter, we survey recent methods for reconstruction of time-varying networks from short time-series data from a modeling perspective. These meth-

ods use the framework of probabilistic graphical models to model the dependence between node observations and the underlying interaction network. Such methods have to make additional assumptions on network structure as well as network dynamics (how the interactions are varying with time) in order to make the inference tractable. We explore the connection between the regularization techniques and the underlying biological processes that justify these assumptions.

### Organization

The remainder of this chapter is organized as follows: In Section 2, we discuss the framework of probabilistic graphical models in the context of network reconstruction. Section 3 provides a brief discussion of the $\ell_1$ penalty and its usage in reconstruction of gene regulatory networks. In Section 4, we discuss recent methods for reconstruction of time-varying interaction networks. Section 5 presents methods for integrating time series information from multiple sources into the graphical model framework. Section 6 presents the relevant software tools for reconstruction of dynamic interaction networks. Finally, Section 7 presents our conclusions as well as a discussion of open problems in this area.

### Notation

We use $\det(\cdot)$ and $\mathrm{Tr}(\cdot)$ to denote the matrix determinant and trace (sum of diagonal elements) of a matrix respectively. We use $\|X\|_{\ell_0}$ to denote $\ell_0$-norm of a matrix $X$ which is equal to the number of non-zero entries of $X$

$$\|X\|_{\ell_0} := \#\{X_{ij} : X_{ij} \neq 0\} \tag{1}$$

Similarly, we use $\|X\|_{\ell_1}$ to denote the $\ell_1$-norm of a matrix $X$ which is given by the sum of absolute values of entries in $X$

$$\|X\|_{\ell_1} := \sum_{i,j} |X_{ij}| \tag{2}$$

## 2 Background

This section describes the problem of network reconstruction from a modeling perspective focussing on time series gene expression data. Suppose gene expressions are measured for $p$ genes denoted by $V := \{1,\ldots,p\}$ at $n$ different time points $T := \{1,\ldots,n\}$. We denote the gene expressions at time $t$ as a random variable $X^{(t)} := [X_1^{(t)},\ldots,X_p^{(t)}]^\top$ where $X_i^{(t)} \in \mathbb{R}$ denotes gene expression for gene $i$ at time $t$.

A network-based approach models the multiple interactions among the different components in a biological system using a graph. A graph $G = (V,E)$ consists of set

of nodes (or *vertices*) $V = \{1, \ldots, p\}$ representing different components in the biological systems; and set of edges $E \subseteq V \times V$ representing the dependence between the different components. In the case of a time-varying network, this is equivalent to having a different underlying graph $G^{(t)} = (V, E^{(t)})$ at each time $t \in \{1, \ldots, n\}$ wherein the set of edges varies with time.

For gene regulatory relationships, the nodes of the graph correspond to the set of genes; and the edges in the graph represent the regulatory relationships. The node observations correspond to the gene expressions $X^{(t)}$ at each time $t \in \{1, \ldots, n\}$. It is well-known that the gene regulatory network $E$ has an impact on the observed gene expression profile $X^{(t)}$. More precisely, the correlation between the gene expression values $X_i^{(t)}$ and $X_j^{(t)}$ measured at different times $t \in \{1, \ldots, n\}$ is a good indicator for the interaction $(i, j)$ being present in the regulatory network. Several methods have been developed for identifying gene interaction networks which are based on the correlation of gene expressions in gene microarray experiments. See [27] for a recent survey on different methods for analysis of dynamic regulatory networks and [4] for an earlier survey.

However, random effects such as noise, measurement error, etc. may lead to false correlation between the expression profiles at two genes $i$ and $j$. A probabilistic approach to model this uncertainty is to treat the gene expression profile $X^{(t)}$ at time $t$ as a random variable drawn from a parametric distribution with unknown parameter. For static network reconstruction, the expression profiles $X^{(1)}, \ldots, X^{(n)}$ are assumed to be drawn independently and identically distributed (i.i.d.) from the same distribution. A major subtask in network reconstruction in this framework is to estimate the parameters of the underlying distribution which best fit the measured expression profiles.

More importantly, there is a strong duality between a distribution over several random variables $X = [X_1, \ldots, X_p]$ and a graph which describes the dependence between individual random variables $X_i$ and $X_j$. Formally, this has been studied using the framework of *Probabilistic Graphical Models* (PGM). See e.g. [39] for a recent textbook providing a comprehensive introduction on the subject.

Several well-known models such as Hidden Markov Models (HMM), Bayesian Networks (BN), Dynamic Bayesian Networks (DBN), etc. are instances of graphical models, and have been successfully used for static network reconstruction [56, 49, 31]. However, HMMs require the number of observations ($n$) to be larger than the number of variables ($p$); and therefore, cannot be used in the case of short time-series data. DBNs also suffer from the curse of dimensionality (large $p$, small $n$), and a number of regularization methods have been investigated in order to address this [48, 37, 32, 57, 72]. More fundamentally, DBNs can only be used to identify relationships in a directed acyclic graph. In effect, while dependence between expression profiles $X_i^{(t)}$ and $X_j^{(t+1)}$ between any two genes $i$ and $j$ at different time points $t$ and $(t+1)$ can be easily captured; the dependence between expression profiles $X_i^{(t)}$ and $X_j^{(t)}$ at the same time instant cannot be fully modeled as this may lead to cycles in the resulting directed graph. This problem can be addressed using undirected probabilistic graphical models, which we discuss below.

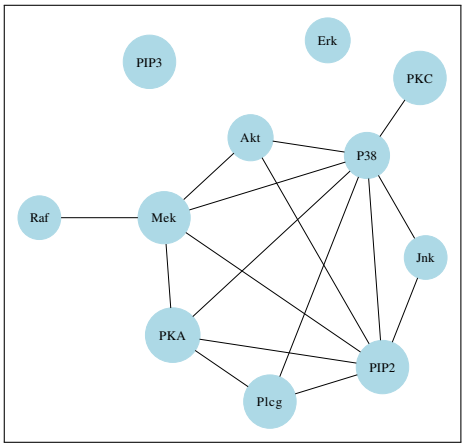



**Fig. 1** Association network estimated from flow cytometry dataset with $p = 11$ proteins measured on $n = 7466$ cells. A missing edge between nodes e.g. *Raf* and *PKA* means the expression levels of the two nodes is conditionally independent given the remaining expression levels. (Adapted with permission from [23])

The dependence between the expressions levels $X = [X_1, \ldots, X_p]^\top$ and the interaction strengths $W = \{W_{ij} : (i, j) \in E\}$ in network $G = (V, E)$ has been modeled using the conditional probability distribution [25, 66, 55, 65, 44, 59, 58]

$$P(X = x|W = w) = \frac{1}{Z(w)} \exp\left(-\frac{1}{2} \sum_{i,j \in V} w_{ij} x_i x_j\right) \tag{3}$$

where $Z(w)$ is a normalization constant. It has the property that whenever $w_{ij} = 0$, the node expressions for nodes $i$ and $j$ are *conditionally independent* given other expression levels. This has been used to construct gene association network where missing edges encode conditional independence.

Formally, the gene association network is constructed by considering the graph $G = (V, E)$ where $E \subseteq V \times V$ denotes the set of edges; with edge $(i, j) \notin E$ whenever genes $i$ and $j$ are conditionally independent ($w_{ij} = 0$). Figure 1 shows an association network between $p = 11$ proteins constructed using $n = 7466$ measurements [53, 23]. In Figure 1, the absence of an edge between two nodes e.g. *Raf* and *PKA* indicates that these are conditionally independent given the remaining nodes.

The problem of model selection is to infer $W$ based on i.i.d. samples $X^{(i)}$ drawn from the conditional distribution in (3). Static network reconstruction methods using time series data model the observations at different times $X^{(1)}, \ldots, X^{(n)}$ as being i.i.d. samples from an unknown static network $W$ i.e.

$$P(X^{(t)} = x^{(t)}|W = w) = \frac{1}{Z(w)} \exp\left(-\frac{1}{2} \sum_{i,j \in V} w_{ij} x_j^{(t)} x_j^{(t)}\right) \tag{4}$$

This is not biologically consistent since it is known that the underlying network is not static during the course of the experiment. Rather, the network is varying across time and in response to external and internal stimuli [43].

Therefore, one should instead consider $W^{(t)} = \{W_{ij}^{(t)} : (i,j) \in E^{(t)}\}$ as the interactions strengths in the network $G^{(t)} = (V, E^{(t)})$ at time $t$. The dependence between gene expressions $X^{(t)}$ and the instantaneous interaction strengths $W^{(t)}$ is given by

$$P(X^{(t)} = x^{(t)} | W^{(t)} = w^{(t)}) = \frac{1}{Z(w^{(t)})} \exp\left(-\frac{1}{2} \sum_{i,j \in V} w_{ij}^{(t)} x_j^{(t)} x_j^{(t)}\right) \tag{5}$$

The problem of time-varying interaction network reconstruction is to identify the interaction strengths $W^{(t)}$ at different times $t \in \{1,\ldots,n\}$ based on the node observations $X^{(1)},\ldots,X^{(n)}$. However, this problem is ill-posed since the number of unknowns ($W^{(t)}$) is much larger than the number of observations ($X^{(t)}$).

In the following subsection, we discuss covariance selection - a classical method for reconstructing a static network $W$ based on real-valued node observations $X^{(t)}$ modeled as i.i.d. samples drawn from conditional distribution in (4). However, this method can fail if the number of observations $n$ is smaller than the numbe of unknowns $p$. Further, the method yields a large number of false positives i.e. multiple interactions $W$ even though it is known that underlying biological network is sparse. Section 3 discusses regularization method using $\ell_1$ penalty which addresses the above-mentioned problems. Section 4 presents methods which extend these methods to inference of time-varying networks under mild assumptions on the temporal evolution of the underlying network.

## *2.1 Static network reconstruction using covariance selection*

Whenever the node observations (gene expressions) $X_i \in \mathbb{R}$ and interaction strengths $W_{ij} \in \mathbb{R}$ are treated as real values, the conditional distribution in (3) is equivalent to $X$ being drawn from a multivariate Gaussian distribution with mean 0 and covariance $\Sigma := W^{-1}$ i.e.

$$X \sim \mathcal{N}(0, \Sigma) \tag{6}$$

Equivalently, the conditional probability of $X^{(t)}$ conditioned on $W$ is given by

$$P(X^{(t)}|W) = \frac{1}{\sqrt{(2\pi)^p \det(W^{-1})}} \exp\left(-\frac{1}{2} \sum_{i,j \in V} X_i^{(t)} X_j^{(t)} W_{ij}\right) \tag{7}$$

This model is commonly referred to as the Gaussian Graphical Model [41]. We note that the normalization constant $Z(w)$ has a closed form expression given by $\sqrt{(2\pi)^p \det(W^{-1})}$ for GGMs.

Construction of gene association networks requires inference of the concentration matrix $W$ based on observations $X^{(1)},\ldots,X^{(n)}$. This problem is known as co-

variance selection and was first studied by Dempster [18]. It involves computing the Maximum Likelihood Estimate (MLE) of $W$ given observations $X^{(1)},\ldots,X^{(n)}$. The log-likelihood of the observations is given by

$$\mathscr{L}(W) = \log P(X^{(1)},\ldots,X^{(n)}|W) = \sum_{t=1}^{n} \log P(X^{(t)}|W) \tag{8}$$

$$= \frac{n}{2}\log\det W - \frac{n}{2}\mathrm{Tr}(SW) - \frac{np}{2}\log(2\pi) \tag{9}$$

where $S \in \mathbb{R}^{p\times p}$ is the empirical covariance matrix for observations $X^{(1)},\ldots,X^{(n)}$ given by

$$S_{ij} = \frac{1}{n}\sum_{t=1}^{n} X_i^{(t)} X_j^{(t)} \tag{10}$$

The Maximum Likelihood Estimate $W^*$ is given by

$$W^* = \arg\max_{W\succ 0} \log\det W - \mathrm{Tr}(SW) \tag{11}$$

$$= \arg\min_{W\succ 0} \mathrm{Tr}(SW) - \log\det W \tag{12}$$

where $W \succ 0$ stands for positive-definiteness of $W$, and $\mathrm{Tr}(SW)$ denotes the trace of the matrix. The positive definiteness constraint ensures that $W^{-1}(=\hat{\Sigma})$ is an invertible covariance matrix. Minimization of the negative log-likelihood in (12) is a convex optimization problem [10].

An exhaustive search for finding the non-zero elements of $W^*$ is computationally prohibitive for moderate and large networks (more than $30-40$ genes). A number of methods based on greedy search have been studied for solving this problem [60, 41]. However, these are not suited whenever the number of observations $n$ is smaller than number of genes $p$ in the network [11, 45]. Another aspect of concern is that whenever $n$ is much smaller than $p$, the solution $W^*$ obtained by greedy methods has a large number of non-zero elements. On the other hand, it is known that the underlying interaction network is sparse [69, 61], i.e. each gene is regulated by a small number of genes.

## *2.2 Discretization of gene expression levels*

Microarray measurements are noisy estimates of the gene expression level. In certain applications, the qualitative level of gene expression is a better indicator of up or down regulation than the microarray measurement (which is a rough estimate of the gene expression). Therefore, the gene expression levels are sometimes quantized to a discrete set $\mathscr{X}$. For example, in the case of cDNA microarray, a choice of $\mathscr{X} = \{-1,1\}$ corresponds to the gene being downregulated $(-1)$ or upregulated $(+1)$. For some applications, the relative strengths of the interactions are of interest rather than the actual values. In such a setting, the weights are quantized to some dis-

crete set $\mathscr{W}$. For example, a choice of $\mathscr{W} = \{-1,0,1\}$ would correspond to the gene interaction being activator $(-1)$, conditionally independent (0) or repressed $(-1)$. A choice of $\mathscr{X} = \{-1,1\}$ and $\mathscr{W} = \{-1,0,1\}$ yields the Ising model which is well-studied in statistical physics [47], wherein $W = 0$ state is modeled as the edge being absent in the network.

The normalization constant $Z(w)$ in (3) is given by

$$Z(w) = \sum_{x_i \in \mathscr{X}} \sum_{x_2 \in \mathscr{X}} \dots \sum_{x_p \in \mathscr{X}} \exp\left(-\frac{1}{2} \sum_{i,j \in V} w_{ij} x_i x_j\right) \tag{13}$$

whenever the gene expressions are discretized to the set $\mathscr{X}$. The network reconstruction procedure is more complicated since the log-likelihood function does not have a closed form expression as in (8) for Gaussian Graphical Models.

## 3 Sparse network reconstruction based on $\ell_1$-regularization

As discussed in Section 2.1, traditional methods for covariance selection are computationally prohibitive and inconsistent in the high-dimensional setting i.e. whenever number of genes $p$ is much larger than the number of microarray measurements $n$. Recent work [45, 7, 23, 40, 52] has addressed this by introducing an additional regularization term based on $\ell_1$-norm into the optimization problem. The technique, commonly known as LASSO (least absolute shrinkage and selection operator), was first studied by Tibshirani [62] in the context of linear regression. The lasso penalty can be understood as a relaxation of the $\ell_0$ norm which induces model sparsity (fewer interactions in the network). This is discussed below in the context of gene interaction network.

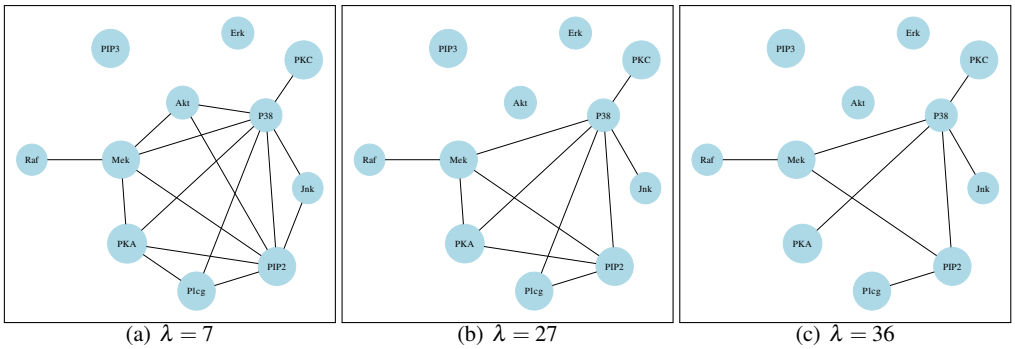
As mentioned earlier, the number of interactions in a gene interaction network is few compared to the total number of possible edges. In other words, a gene is regulated by few other genes. This can be ensured by introducing an additional constraint in (12) as follows

$$\begin{array}{ll} \arg\min_{W \succ 0} & \mathrm{Tr}(SW) - \log\det W \\ & \|W\|_{\ell_0} \leq t \end{array} \tag{14}$$

The solution to (14) has atmost $t$ interactions, where $t$ is a parameter chosen based on domain knowledge. However, the optimization in (14) is an instance of mixed integer programming [10], and is computationally intractable for moderate and large networks ($p \geq 30$). The lasso penalty replaces the $\ell_0$-norm with $\ell_1$-norm to obtain the relaxed convex optimization problem given by

$$\begin{array}{ll} \arg\min_{W \succ 0} & \mathrm{Tr}(SW) - \log\det W \\ & \|W\|_{\ell_1} \leq t \end{array} \tag{15}$$

One can obtain the equivalent Lagrangian formulation [10] as




(a) $\lambda = 7$ (b) $\lambda = 27$ (c) $\lambda = 36$

**Fig. 2** Association network estimated from flow cytometry dataset with $p = 11$ proteins measured on $n = 7466$ cells with different penalty parameter $\lambda$ in (16) (Adapted with permission from [23]). Increasing the parameter $\lambda$ yields a network with fewer interactions.

$$\arg\min_{W \succ 0} \operatorname{Tr}(SW) - \log\det W + \lambda \|W\|_{\ell_1} \tag{16}$$

where $\lambda$ is a user-specified parameter.

The parameter $\lambda$ regulates the penalty incurred by choosing a non-sparse interaction network having many gene-gene interactions. A higher value of $\lambda$ ensures higher sparsity (fewer non-zero interactions), while a choice of $\lambda = 0$ corresponds to covariance selection without any penalty. Figure 2 shows an association network between $p = 11$ proteins constructed using $n = 7466$ measurements with different penalty parameter $\lambda$ in (16) [23]. As the penalty parameter increases, the number on non-zero interactions decreases. A choice of $\lambda = 0$ (no regularization) yields the fully connected network.

Several methods [45, 70, 7, 23] have focussed on efficient computation of the solution in (16). Meinshausen and Bühlmann [45] first explored the connection to $\ell_1$ regularization. They devised a neighbourhood selection procedure which used lasso regression to obtain a set of neighbours (non-zero interaction strength) for each gene based on local conditional likelihoods. In other words, for each gene, their approach chooses a small set of "most-likely neighbour" genes that have non-zero interactions, while remaining interactions are set to zero. These local neighbourhoods are used to construct the association network using either an 'AND' or an 'OR' final step procedure. This procedure corresponds to a modified penalty term in (16) [3].

Banerjee et. al. [7] and Friedman et. al. [23] improved the previous approach by casting it into the penalized maximum likelihood (PML) framework in (16). They used block coordinate descent to solve the resulting optimization. This means the optimization in (16) is solved by iteratively updating one of the rows (or columns) of $W$ till certain convergence criteria is reached. In practice, each iteration (row or column update) requires solving a lasso problem (linear regression with $\ell_1$ regularization term). This procedure is closely related to local neighbourhood search described below. Wainwright et. al. [64, 50] studied network estimation in Ising models and

discrete graphical models based on a similar local neighbourhood search using lasso penalty. Section 3.1 provides a brief description of the local neighbourhood search procedure based on logistic regression.

Wainwright et. al [64, 50] show from a statistical perspective that under mild conditions on sparsity of $W$, local neighbourhood seach recovers the interaction network correctly given a small number of measurements $n \sim d^3 \log p$ where $d$ is the maximum number of interactions (maximum degree) of any gene with other genes in the network. This is often referred to as the *sparsistency* (sparse consistency) property (of the estimator of $W$). A similar condition holds for Penalized Maximum Likelihood framework. In effect, the conditions can be understood as follows: if the underlying model was a Gaussian Graphical Model with few interactions (sparse network); then the methods will recover the interactions $W$ precisely given enough observations $n$.

However, the above conditions do not apply in the context of biological networks since GGM is at best an approximation to the dependence between interactions and node observations. Instead, the reconstructed network is compared to past biological findings; and previously unknown interactions predicted by the model have to be experimentally verified.

## *3.1 Local neighbourhood search*

We describe the local neighbourhood search procedure focussing on the special case of Ising models. This corresponds to situations where gene expression levels are discretized to the set $\mathscr{X} = \{-1, 1\}$ i.e. genes are either downregulated or upregulated.

The basic idea of the method is to iteratively estimate the concentration matrix $W$ by updating a single row (or column) at a time. Let $W_{\backslash i} := [W_{ij}]^\top \in \mathbb{R}^{p-1}$ be a vector of length $(p-1)$ constructed by considering the $i^{th}$ row (or column) of $W$ except the diagonal element. Notice that the set of non-zero elements of $W_{\backslash i}$ represent the interactions of gene $i$ with other genes in the network i.e. the local neighbourhood of gene $i$ given by

$$\mathscr{N}_i = \{j \in V : (W_{\backslash i})_j \neq 0\} \tag{17}$$

Therefore, estimating $W_{\backslash i}$ yields insight into which genes interact with the $i^{th}$ gene. Sparsity in $W_{\backslash i}$ is ensured by introducing an additional $\ell_1$ regularization term. The resulting optimization is a convex optimization of the form

$$W^*_{\backslash i} = \arg \min_{W_{\backslash i} \in \mathbb{R}^{p-1}} \ell(W_{\backslash i}) + \lambda \|W_{\backslash i}\|_{\ell_1} \tag{18}$$

where $\ell(W_{\backslash i})$ is the negative rescaled log likelihood $\ell(W_{\backslash i})$ is given by

$$\ell(W_{\backslash i}) := -\frac{1}{n} \sum_{t=1}^{n} \log P(X_i^{(t)} | X_{\backslash i}^{(t)}, W_{\backslash i}) \tag{19}$$

and $P(X_i^{(t)}|X_{\setminus i}^{(t)}, W_{\setminus i})$ is the conditional probability distribution of gene expression $X_i^{(t)}$ conditioned on the gene expressions of the other genes $X_{\setminus i}^{(t)} := \{X_j^{(t)} : j \in V \setminus i\}$ and interaction strengths $W_{\setminus i}$ is given by

$$P(X_i^{(t)}|X_{\setminus i}^{(t)}, W_{\setminus i}) = \frac{1}{1+\exp\left(x_j^{(t)} \sum_{j \in V \setminus i} w_{ij} x_j^{(t)}\right)} \tag{20}$$

The optimization in (18) can be solved efficiently using convex solvers [38, 42, 21].

The method iteratively estimates $W_{\setminus i}$, and consequently the local neighbourhood, for different genes $i \in V$ till some convergence criteria is met, usually in terms of the size of the increments in log likelihood.

The interaction network can be constructed by either considering an "AND" configuration wherein an edge $(i, j)$ is present in the network if $i$ is in the local neighbourhood of $j$ and vice versa,

$$E = \{(i, j) \in V \times V : i \in \mathcal{N}_j \text{ and } j \in \mathcal{N}_i\} \tag{21}$$

An alternative construction is using "OR" configuration wherein an edge $(i, j)$ is present in the network if $i$ is in the local neighbourhood of $j$ or vice versa,

$$E = \{(i, j) \in V \times V : i \in \mathcal{N}_j \text{ or } j \in \mathcal{N}_i\} \tag{22}$$

## 4 Reconstruction of time-varying regulatory networks

Sections 2.1 and 3 describe the covariance selection problem in the static setting i.e. the gene expressions $X^{(1)}, \ldots, X^{(n)}$ are modeled as independent samples from an multivariate normal distribution with fixed but unknown concentration matrix $W$. The zero elements of $W$ correspond to genes whose gene expressions are conditionally independent conditioned on other genes, while non-zero elements indicate strength of interaction, either activating or repressing depending on the sign of $W_{ij}$ *for all observation time instants*.

Nevertheless, it is known that rewiring occurs in gene interaction networks in response to environmental and genetic stress [43]. For example, genes implicated in yeast metabolism undergo significant rewiring in response to changes in nutrient availability [15]. A biologically plausible modeling of the interaction network, therefore, should incorporate interaction dynamics. This poses new challenges from a modeling perspective, which we discuss below.

The dynamics (temporal variation) of the interaction strengths should depend on the time elapsed between observation instants. For example, if we take observations at very short intervals, the interactions between nearby time instants should not differ a lot. In other words, there is sparsity in the interaction dynamics as the interaction network does not drastically change from one observation time point for

the most part. Nonetheless, there might be instances at which major changes can occur in the network, often in response to environmental or genetic stress.

On the other hand, it is known that gene interactions are a vital part of functional roles performed by a gene. Indeed, one could posit that it is the changing functional requirements imposed by internal development or external stress that which might drive gene rewiring. Thus, a gene interaction network may get rewired in order to satisfy a new functional role. There is a rich literature which relates the functional roles of the genes to the interaction network.

A number of recent approaches [71, 58, 1, 36] have addressed this problem. These methods can be broadly categorized into two classes: optimization-based and model-based methods. Optimization-based methods [1, 71, 58] introduce an additional term into the optimization in (16) which ensures that there are not many changes in the network between consecutive observation times. These methods do not incorporate additional information such as knowledge about the functional roles of the genes, network motifs, etc. into the network reconstruction. Model-based methods [29, 26, 36] incorporate additional knowledge such as information about presence or absence of specific network motifs [29], functional roles of the genes [36], etc. into network dynamics.

Inference of time-varying interactions characterizes the activity of individual genes and predicts their interactions with other genes, including unknown predictions which could serve as test candidates for experimental testing. At a broader level, it yields new insight by implicating groups of genes, often characterized by different functional roles performed by them, interacting among each other and with other groups at critical stages of a biological process [58, 36]. For example, the analysis in [58] reveals high activity between genes related to metamorphosis, wing margin morphogenesis, wing vein morphogenesis and apposition of wing surfaces during early embryonic stage in *Drosophila Melanogaster* (fruit fly). Such interaction is typically visible during the transition from pupa stage to adult stage when wing development occurs [5]. This behaviour could potentially indicate diverse functionality of these genes. Similarly, the analysis in [36] implicates genes related to complex/cofactor binding as active during transition from glucose starvation to nitrogen starvation in *S. Cerevisiae* (baker's yeast).

## *4.1 Optimization-based methods for modeling interaction dynamics*

Zhou et. al. [71] first studied estimation of time-varying Gaussian Graphical Models using $\ell_1$ regularization. Their approach extends the static model in Section 2.1 by assuming the observation $X^{(t)}$ at each time $t$ to be drawn from a Gaussian distribution $\mathcal{N}(0,\Sigma^{(t)})$ independent of other observations. The concentration matrix at time $t$ is given by $W^{(t)} := {\Sigma^{(t)}}^{-1}$. Consequently, the interaction network at time $t$ is given by $G^{(t)} = (V,E^{(t)})$ where $E^{(t)} := \{(i,j) \in V \times V : W_{ij}^{(t)} \neq 0\}$ is the set of edges with non-zero interaction strengths. The conditional probability distribution of $X^{(t)}$

conditioned on interaction strengths $W^{(t)}$ is given by

$$P(X^{(t)}|W^{(t)}) = \frac{1}{Z(W^{(t)})} \exp\left(-\frac{1}{2} \sum_{(i,j)\in E^{(t)}} X_i^{(t)} X_j^{(t)} W_{ij}^{(t)}\right) \tag{23}$$

where $Z(W^{(t)})$ is a normalization constant.

Since the observation at each time is independent of other times, covariance selection procedures are not directly applicable since the empirical covariance $S$ cannot be computed by treating different observations as i.i.d. samples. The method addresses this by constructing a weighted covariance matrix $\hat{S}^{(t)}$ at each time instant $t$ where decreasing weights are assigned to observations with increasing time gap. At a given time $t$, the weight corresponding to observation at time $i$ is defined using a symmetric non-negative kernel $K$ as

$$w^{(t)}(i) = \frac{K(|t-i|/h)}{\sum_{i=1}^{n} K(|t-i|/h)} \tag{24}$$

Then, the weighted empirical covariance $\hat{S}^{(t)}$ at time $t$ is given by

$$\hat{S}^{(t)} = \frac{1}{C} \sum_{i=1}^{n} w_i^{(t)} X^{(i)} {X^{(i)}}^{\top} \tag{25}$$

where $C := \sum_{i=1}^{n} w^{(t)}(i)$ is the scaling term. For example, if the function $K(x)$ used to assign weights to observations is given by

$$K(x) = \begin{cases} 2^{-|x|} & \text{if } x \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

then measurements $X^{(t-1)}$ and $X^{(t+1)}$ made at time instances $(t-1)$ and $(t+1)$ respectively are assigned weight $1/2$ in computation of $\hat{S}^{(t)}$. The empirical covariance $\hat{S}^{(t)}$ at time $t$ is computed as

$$\hat{S}^{(t)} = \frac{1}{2}\left(\frac{1}{2} X^{(t-1)} {X^{(t-1)}}^{\top} + X^{(t)} {X^{(t)}}^{\top} + \frac{1}{2} X^{(t+1)} {X^{(t+1)}}^{\top}\right) \tag{27}$$

This allows estimation of the concentration matrix ${W^{(t)}}^{*}$ at time $t$ using weighted empirical covariance matrix $\hat{S}^{(t)}$ by solving the following optimization problem

$${W^{(t)}}^{*} = \arg\max_{W \succ 0} \operatorname{Tr}(\hat{S}^{(t)} W) - \log\det W + \lambda \|W\|_{\ell_1} \tag{28}$$

The optimization in (27) can be solved independently at each time $t \in \{1, \ldots, n\}$ using the methods described in Section 3.

Zhou et. al. [71] show that the above procedure correctly recovers $W^{(t)}$ under mild smoothness conditions on $\Sigma^{(t)}$ in addition to sparsity assumption of $W^{(t)}$ (few non-zero interactions). More precisely, if $\Sigma^{(t)} = [\sigma_{ij}^{(t)}]$, where $\sigma_{ij}^{(t)} \in C^{\infty}$ is a smooth

function denoting the instantaneous covariance between genes $i$ and $j$ at time $t$; then $\sigma_{ij}^{(t)}$ has bounded first and second order derivatives at all times i.e.

$$\max_{i,j} \sup_t |\frac{\partial}{\partial t}\sigma_{ij}^{(t)}| \leq C_1 \tag{29}$$

$$\max_{i,j} \sup_t |\frac{\partial^2}{\partial t^2}\sigma_{ij}^{(t)}| \leq C_2 \tag{30}$$

Put simply, each term in the covariance matrix $\Sigma^{(t)}$ *changes slowly* over time. In other words, as we make more measurements with smaller time gaps between consecutive measurements, the difference between the inferred networks at consecutive time points will decrease. This is indeed the case for most biological systems.

### 4.1.1 Inference of dynamic interactions under smooth variation

As mentioned in Section 2.2, there are scenarios where it is advantageous to discretize the measurements to some set $\mathscr{X}$. A common choice is $\mathscr{X} = \{-1, 1\}$, which corresponds to genes being downregulated or upregulated. Song et. al. [58, 1] have studied inference of time-varying discrete graphical models where the gene expression are discretized to $\mathscr{X} = \{-1, 1\}$. They extended the local neighbourhood search procedure [64, 50] to handle inference of dynamic interaction networks under smoothness conditions in (29)-(30). This is achieved by introducing a weighted negative log likelihood analogous to (25)

$$\tilde{\ell}^{(t)}(W_{\backslash i}) = -\frac{1}{C}\sum_{k=1}^{n} w^{(t)}(k) \log P(X_i^{(k)} | X_{\backslash i}^{(k)}, W_{\backslash i}) \tag{31}$$

where $C = \sum_{k=1}^{n} w^{(t)}(k)$ is the scaling term, and $w^{(t)}(k)$ is given by a symmetric non-negative kernel $K$ as described in (24). The modified optimization problem for the $i^{th}$ row at time $t$ is given by

$$W_{\backslash i}^{(t)} = \arg \min_{W_{\backslash i} \in \mathbb{R}^{p-1}} \left( \tilde{\ell}^{(t)}(W_{\backslash i}) + \lambda \|W_{\backslash i}\|_{\ell_1} \right) \tag{32}$$

The above optimization is same as in (18) except for the weighted log likelihood term $\tilde{\ell}^{(t)}(W_{\backslash i})$. Consequently, the network is constructed independently at each time $t$ as described in 3.1.

They use this approach to study the evolution of gene regulatory network in *Drosophila Melanogaster*, the common fruit fly, over its developmental cycle based on 66 gene expression measurements collected in [5]. The expression measurements can be catgorized into four stages, i.e. embryonic (1-30 time point), larval (31-40 time point), pupal (41-58 time point) and adult stages (59-66 time point). In order to verify the biological findings of the method, they focus on three groups of genes consisting of $25-30$ genes that are known to be functionally implicated in

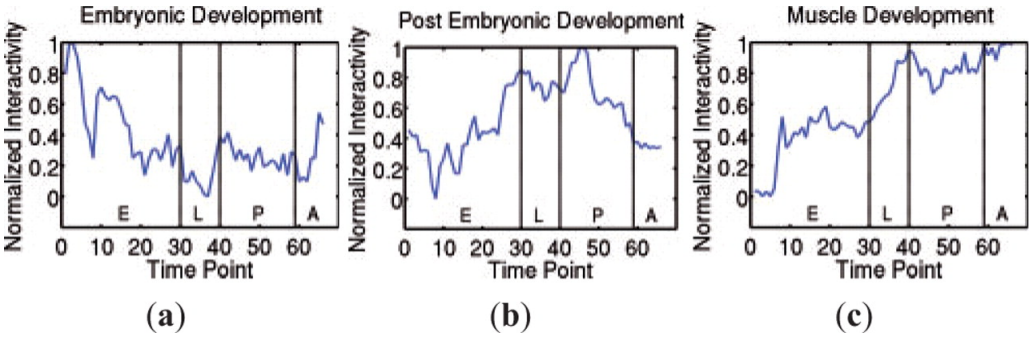



**Fig. 3** Interactivity of three groups of genes related to $(a)$ embryonic development; $(b)$ post-embryonic development; and $c$ muscle development. It shows the interactivity of three groups of genes during the different developmental cycles, showing that each group of genes is more active during its development stage. Thus, the time-varying networks inferred using [58] are consistent with known biological findings. (Reprinted with permission from [58])

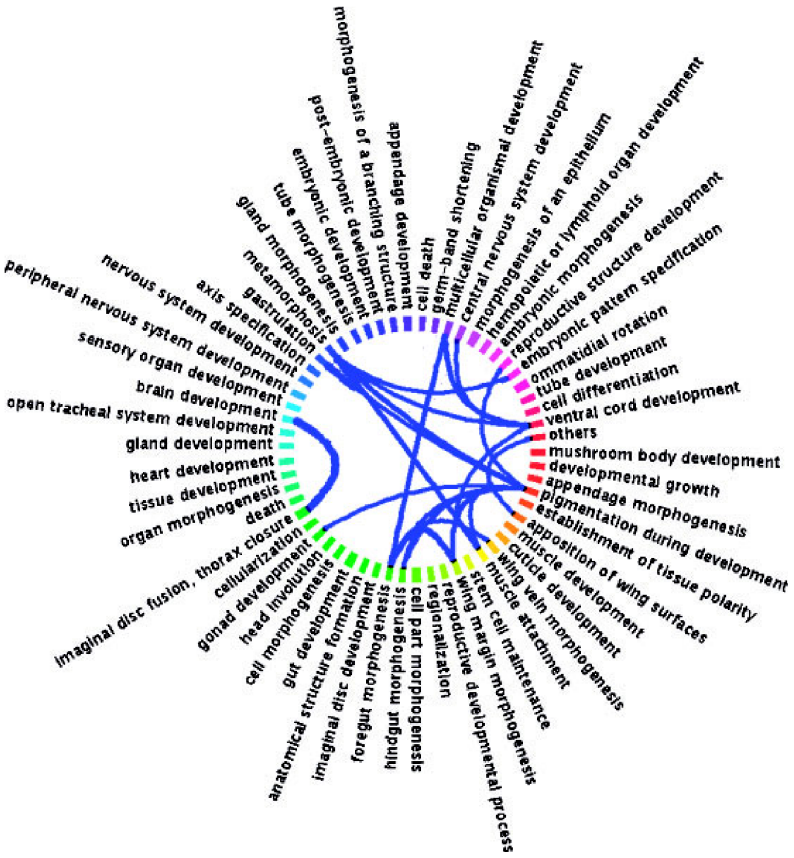
different developmental stages, namely, embryonic development, post-embryonic development and muscle development. They measure the interactivity of the group in the interaction networks found using their method, and observe that each group of genes has more non-zero interactions during their development stage. Figure 3 shows the interactivity of three groups of genes during the different developmental cycles, showing that each group of genes is more active during its development stage.

They also investigate the interactions between genes from different functional groups. Figure 4 shows the connectivity between different functional groups. This yields fresh insight into interactions between functional groups. For example, the method reveals high activity between genes related to metamorphosis, wing margin morphogenesis, wing vein morphogenesis and apposition of wing surfaces during early embryonic stage (Figures 4 $(b)$,$(c)$). Such interaction is typically visible during the transition from pupa stage to adult stage (Figures 4 $(r)$,$(s)$) when wing development occurs [5]. This behaviour could potentially indicate diverse functionality of these genes.

Ahmed and Xing [1] extended the above approach to handle sharp structural changes such as sudden rewiring of a gene network in response to a stimulus. They use the approach to study the evolution of regulatory network in *Drosophila Melanogaster* consisting of 4028 genes at 66 different time points over its life cycle.

## *4.2 Model-based methods for modeling interaction dynamics*

Methods discussed in Section 4.1 do not make any assumptions on the structure of the network beyond sparsity (few non-zero interactions) and smoothness temporal variation (nearby time instants have similar interaction profile). However, more




(*a*)

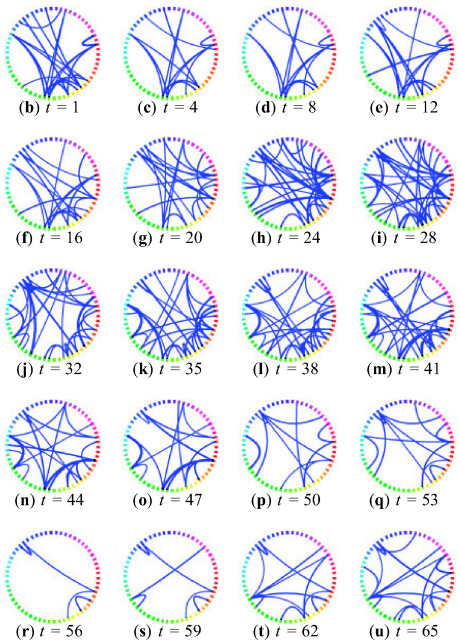



**Fig. 4** (*a*) Average network between functional groups obtained using [58]. Each color patch denotes an ontological group, and the position of the ontological groups remains the same from (*a*) to (*u*). The annotation in the outer rim indicates the function of each group. (Adapted with permission from [58])

information is available in case which can be incorporated into the analysis. For example, the yeast (static) interaction network is known with high degree of confidence based on perturbation studies [63, 35]. It has also been observed that even though the interaction networks are highly dynamic in nature, they feature a number of building blocks i.e. motifs and subgraphs that recur over time [33]. Finally, there is extensive information available about the functional roles performed by the genes [46, 6]. Even though it is clear that this information could potentially aid in reconstruction of dynamic interaction networks; integration poses a challenge, and it is only recently that methods [29, 36] have been explored that leverage this information to aid in dynamic interaction network reconstruction.

### 4.2.1 Inference of interaction dynamics based on recurring motifs

Guo et. al. [29] studied a model-based approach that leverages information about recurring subgraphs that appear in the interaction networks at different times. Their approach considered the interaction $W_{ij}^{(t)}$ between genes $i$ and $j$ at time $t$ to be either absent or present with some strength $W_{ij}$ i.e.

$$W_{ij}^{(t)} = \begin{cases} W_{ij} & \text{if } E_{ij}^{(t)} = 1 \\ 0 & \text{otherwise} \end{cases} \tag{33}$$

They modeled the evolution of the interaction network under the Markov assumption i.e. the interaction network $E^{(t)}$ at time $t$ depends only on the interaction network $E^{(t-1)}$ at previous time instant $(t-1)$,

$$P(E^{(1)}, E^{(2)}, \ldots, E^{(n)}) = P(E^{(1)}) \prod_{t=2}^{n} P(E^{(t)} | E^{(t-1)}) \tag{34}$$

The transition probability $P(E^{(t)}|E^{(t-1)})$ is specified in terms of simple features $(\Psi_f)$ that measure some global statistic extracted from the interaction network,

$$P(E^{(t)}|E^{(t-1)}) = \frac{1}{Z(\theta, E^{(t-1)})} \exp\left( \sum_{f=1}^{F} \theta_f \Psi_f(E^{(t)}, E^{(t-1)}) \right) \tag{35}$$

Examples of simple features are “density”, “stability” and “transitivity” given by

$$\Psi_1(E^{(t)}, E^{(t-1)}) = \sum_{ij} E_{ij}^{(t)} \qquad \text{(density)} \tag{36}$$

$$\Psi_2(E^{(t)}, E^{(t-1)}) = \sum_{ij} \mathbb{I}(E_{ij}^{(t)} = E_{ij}^{(t-1)}) \qquad \text{(stability)} \tag{37}$$

$$\Psi_3(E^{(t)}, E^{(t-1)}) = \sum_{ijk} \frac{E_{ij}^{(t)} E_{ik}^{(t-1)} E_{kj}^{(t-1)}}{\sum_{ij} E_{ik}^{(t-1)} E_{kj}^{(t-1)}} \qquad \text{(transitivity)} \tag{38}$$

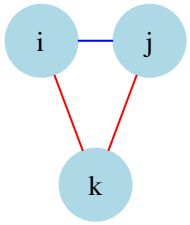



**Fig. 5** Example of transitivity feature. The edges $(i,k)$ and $(j,k)$ present in $E^{(t-1)}$ at time $(t-1)$ are shown in red, while edge $(i,j)$ in $E^{(t)}$ at time $t$ are shown in blue. Here $E^{(t)}_{ij}E^{(t-1)}_{ik}E^{(t-1)}_{kj}=1$.

where $\mathbb{I}(\cdot)$ denotes the indicator function i.e. $\mathbb{I}(E^{(t)}_{ij}=E^{(t-1)}_{ij})$ is one if edge $(i,j)$ is present in the interaction networks $E^{(t)}$ and $E^{(t-1)}$ at times $t$ and $(t-1)$ respectively, and zero otherwise.

The density feature $\Psi_1(E^{(t)},E^{(t-1)})$ counts the number of interactions in the network $E^{(t)}$ at time $t$. For example, if the coefficient $\theta_1$ corresponding to the density feature in (35) is negative, this favors sparse interaction network $E^{(t)}$ at time $t$. The stability feature $\Psi_2(E^{(t)},E^{(t-1)})$ counts the number of interactions that are preserved across the two network $E^{(t)}$ and $E^{(t-1)}$. If the coefficient $\theta_2$ corresponding to the stability feature in (35) is positive, this discourages large number of changes between $E^{(t)}$ and $E^{(t-1)}$. As an extreme case, if $\theta_2$ is extremely high, we recover a static network.

The transitivity feature $\Psi_3(E^{(t)},E^{(t-1)})$ measures the fraction of genes $i$ and $j$ that are connected to a common gene $k$ at time $(t-1)$, and also which have an edge $(i,j)$ at time $t$. Figure 5 shows the case when $E^{(t)}_{ij}E^{(t-1)}_{ik}E^{(t-1)}_{kj}=1$. If $\theta_3$ corresponding to the transitivity feature in (35) is positive, this means a common gene interacting with two genes (which may or may not be interacting with each other) is likely to induce interactions among the two possibly non-interacting genes. In other words, this favors the triangle motif in the network.

Other motifs e.g. stars, cycles, etc. can similarly be favored or discouraged in the network structure by careful choice of features and weight function $\theta$. This class of models are known as Exponential Random Graph Models (ERGM) (see [51] for an introductory survey).

Guo et. al. [29] use a sampling based approach to infer the time varying networks $E^{(t)}$ based on gene expression measurements assuming that the interaction dynamics are Markovian as in (34). They analyze the muscle development subnetwork of *Drosophila Melanogaster* (fruit fly) consisting of 11 genes during four stages of the life cycle, including embryonic, larval, pupal and first 30 days of adulthood. Their results agree closely with known interactions, and suggest hitherto unknown linkages which can be investigated further experimentally.

### 4.2.2 Inference of interaction dynamics based on functional roles of genes

In a recent work, Jethava et. al. [36] studied the dynamics of the interaction network in terms of the functional roles performed the genes. Their approach assumes Markovian dependence between interaction strengths $W^{(t)}$ and $W^{(t-1)}$ at times $t$ and $(t-1)$ respectively, i.e.

$$P(W^{(t)}|W^{(t-1)},\ldots,W^{(1)}) = P(W^{(t)}|W^{(t-1)}) \tag{39}$$

They modeled the time-varying interaction between two genes as governed by the functions performed by the two genes. In effect, the dynamics of interaction $W_{ij}^{(t)}$ between any two genes $i$ and $j$ is conditionally independent of other interactions in the network conditioned on the roles. This assumption allows tractable inference in moderate and large networks (few thousand genes). A Bayesian approach is used to model network sparsity with selection of appropriate priors.

Most genes perform multiple functions at different times, each contributing to some vital requirement in the life cycle. Their model assumes that at each time, the interaction between two genes depends on the *active functional roles* of the two genes at that time. Thus, the overall network structure at each time depends on the different processes happening in the organism at that time.

The model infers the latent time-varying interactions using functional information of genes. This is used to infer the interaction network in *S. Cerevisiae* (baker's yeast) at different time points with varying nutrient availability. Figure 6 shows an example of the inferred network at time $t = 4.1$ hrs. The network changes due to external stimulus (change in nutrient availability). The model successfully captures sharp change in the network (the interaction strengths get inverted) in response to critical change in nutrient availability from Carbon-rich environment to Nitrogen-rich environment.

## 5 Integrative analysis from multiple sources

The availability of data from multiple sources such as protein-DNA binding, protein-protein interactions (PPI), miRNA-mRNA interactions, time series gene expressions under genetic perturbations etc. has led to a new challenge in network inference based on multiple sources. In many cases, the additional data belongs to a single time point under a single condition e.g. protein-DNA binding, miRNA-RNA interations, etc. (see [27], Section 3.3 for a discussion of methods integrating static with time series data).

A new scenario has arisen where one wishes to combine several time series data to infer the related network at different time points. For example, if time series gene expressions are measured under the same experimental conditions for several different strains of an organism with minor genetic variation; the inferred networks for the different strains should be largely similar. Alternatively, if one is measuring

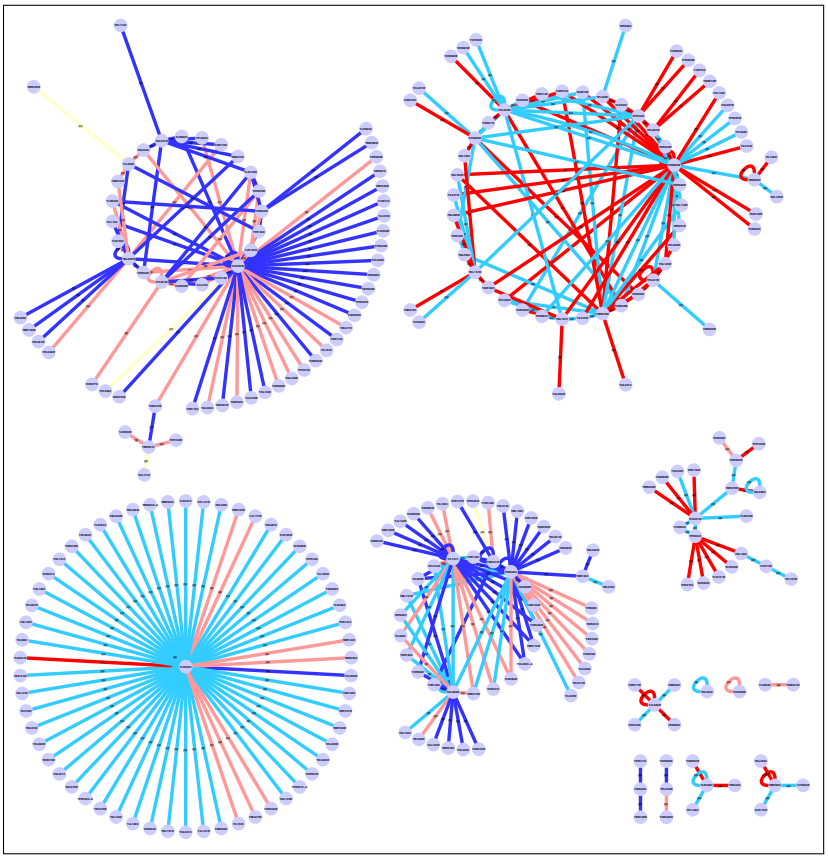


**Fig. 6** Time varying interaction strengths between the genes at time $t = 4.1$ hrs in [36] in experiment with gradual change in nutrient availability. The edge colors denote their interaction strength, which was classified as strong repressing (red), low repressing (pink), no effect (yellow), low inducing (light blue) and strong inducing (dark blue). (Reprinted with permission)

time series expression data using different methods or under different experimental conditions, some of the interactions might be different while the remaining network structure is largely preserved.

From a computational perspective, one models such time series data as being non i.i.d. which are generated from different but closely related interaction networks. For Gaussian Graphical Models, this corresponds to multiple graphical models that share the same variables and a large part of the dependence structure. Guo et. al. [30] recently investigated the joint estimation of multiple graphical models under the assumption that the underlying network structure is largely preserved across the multiple data sources. In their method, no additional assumption is made beyond the large common substructure across the different data sources. We discuss this further in Section 5.1.

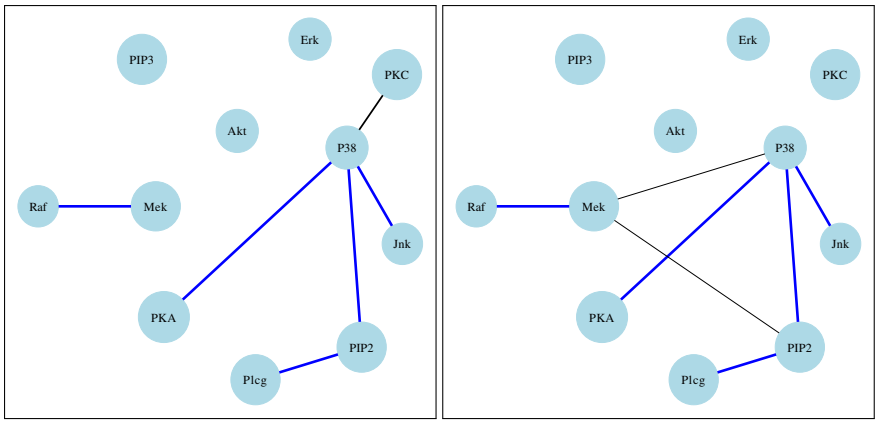



**Fig. 7** Example of sparse networks sharing a large common substructure. The common substructure is highlighted in blue. Such networks can be extracted from different data sources or data from multiple experiments using approach outlined in [30].

In many cases, we might have further information about the variation in the network structure between different data sources. For example, if one of the sources corresponds to an time series data with some genetic perturbation such as gene knockout; the network structure would largely vary in a close neighbourhood of the knocked out gene. Jethava et. al [36] explore this for dynamic network reconstruction in yeast based on data from several strains of yeast having genetic perturbations. This is discussed further in Section 5.2.

## *5.1 Data integration based on common network substructure*

Guo et. al. [30] investigate joint estimation of network structure for multiple graphical models. Their approach assumes the underlying model structure (conditional independencies) is largely preserved across the different networks.

In order to model this, they model the interaction strength $w_{ij}^{[k]}$ between nodes $i$ and $j$ in $k^{th}$ network as a product of two terms, namely, a term $\theta_{ij}$ which is common across all networks and $\gamma_{ij}^{[k]}$ which is different across the networks arising from different data sources i.e.

$$w_{ij}^{[k]} = \theta_{ij}\gamma_{ij}^{[k]} \tag{40}$$

If $\theta_{ij}$ is zero, then all the networks have $i$ and $j$ conditionally independent i.e. no edge is present between nodes $i$ and $j$ across all networks. However, if $\theta_{ij}$ is not zero, some of the networks can still have $\gamma_{ij}^{[k]} = 0$ while other networks can have $\gamma_{ij}^{(k')} \neq 0$ yielding different network structure. In order to ensure common substructure, they use $\ell_1$ regularization based approach by introducing sparsity constraint on $\theta$ and $\gamma^{[k]}$

i.e.

$$\arg\min_{\Gamma^{[k]},\Theta} \sum_{k=1}^{K} \left(\mathrm{Tr}(S^{[k]}W^{[k]}) - \log\det W^{[k]}\right) + \eta_1 \|\Theta\|_{\ell_1} + \eta_2 \sum_{k=1}^{K} \|\Gamma^{[k]}\|_{\ell_1} \tag{41}$$

The parameters $\eta_1$ controls the degree of commonality in the network. A high value of $\eta_1$ promotes common substructure across the different networks. The parameter $\eta_2$ controls the degree of sparsity in the networks. The resulting optimization is solved using the GLASSO software as a subroutine.

This approach allows systematic integration of data from different sources in order to obtain sparse networks with large common substructure. In order to extend this procedure to the case of dynamic networks, one can use the weighted empirical covariance $\hat{S}^{(t),[k]}$ in network $k$ at time $t$ based on kernel reweighting as discussed in Section 4.1. Figure 7 shows an example of two sparse networks with a large common substructure.

## *5.2 Integration of time series data under genetic perturbations*

Jethava et. al. [36] studied the problem of dynamic network reconstruction in *S. Cerevisiae* from multiple experiments with genetic perturbation i.e. where one or two genes have been knocked out. Their approach combines the network perturbation effect into dynamic network reconstruction under the assumption that the network changes drastically near the perturbations (genes knocked out); while subnetworks not related to the functional roles performed by the knocked out gene are minimally impacted.

This is modeled by considering the interaction strength $w^{(t),[k]}$ in perturbed network $k$ at time $t$ as a product of two terms, namely, a base interaction strength $w_{ij}^{(t)}$ and a edge damping coefficient $\gamma_{ij}^{[k]}$.

$$w_{ij}^{(t),[k]} = w_{ij}^{(t)} \gamma_{ij}^{[k]} \tag{42}$$

The damping coefficient $\gamma_{ij}^{[k]}$ of edge $(i,j)$ in network $k$ depends on the distance of nodes $i$ and $j$ from the genes knocked out in the perturbed network $k$ i.e.

$$\gamma_{ij}^{[k]} = (1-\gamma_i^{[k]})(1-\gamma_j^{[k]}) \tag{43}$$

Since a base network is known for Yeast with high degree of confidence, this is used to compute the node damping $\gamma_i^{[k]}$ by diffusing the effect of the gene knockout through the network i.e. if a gene $i$ is knocked out in network $k$, then $\gamma_i^{[k]}$ is 1; otherwise, $\gamma_i^{[k]}$ is computed by averaging the damping coefficients $\gamma_j^{[k]}$ for all genes $j$ which interact with gene $i$.

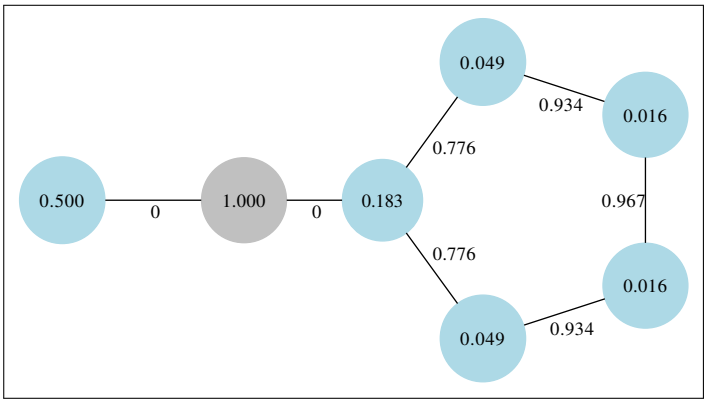



**Fig. 8** Example of damping with one gene knocked out with damping coefficient $\beta = 0.5$. The knockout gene is indicated in gray; and the node and edge damping coefficients are shown. The node damping coefficients quickly decrease as one goes away from the knockout gene. Consequently, the impact of knockout decreases (edge damping close to 1) for interactions far from the knockout point.

$$\gamma_i^{[k]} = \begin{cases} 1 & \text{if gene } i \text{ is knocked out} \\ \frac{\beta}{d(i)} \sum_{j \in N(i)} \gamma_j^{[k]} & \text{otherwise} \end{cases} \tag{44}$$

The parameter $\beta$ controls the range of perturbation effects. A small value of $\beta$ mean that the effect of gene knockout is limited to close neighbours only; while a large value of $\beta$ means the pertubation effects are long-ranged. Figure 8 shows an example of damping coefficients computed for a network with single gene knockout. The edges next to the knockout gene are impacted the most, while interactions far away from the knockout point are treated as being the same across all networks.

This approach decouples the inference procedure from the effect of network perturbations - while allowing incorporation of data from multiple experiments into reconstrution of the dynamic networks. Consequently, one can use perturbation studies in concert with time series data.

## 6 Software

### GLASSO

Graphical lasso (GLASSO) is a popular software written in R and Matlab, for estimating sparse inverse covariance matrix using lasso ($\ell_1$) penalty. This can be used to find a sparse static interaction network based on microarray expression data. The software is available at `http://www-stat.stanford.edy/~tibs/glasso/`.

### KELLER

KELLER is a software in Matlab for estimating time-varying regulatory networks based on time series gene expression data using $\ell_1$ regularization approach. It assumes that the interaction network changes smoothly over time i.e. the network between consecutive observation times are very similar structurally. The software is available at `http://cogito-b.ml.cmu.edu/keller/`.

### TESLA

TESLA is a software in Matlab for estimating time-varying networks based on node observations using $\ell_1$ regularization approach. This can be used to find dynamic interaction network (different at different time-points) based on microarray expression measurements. It detects sharp changes such as sudden rewiring of the network in response to external stimulus. The software is available at `http://www.sailing.cs.cmu.edu/tesla/index.html`.

### NETGEM

NETGEM is a software in Matlab for estimating time-varying interaction network based on microarray expression data using a Bayesian approach. It models the network dynamics contingent on the functional roles performed by interacting genes. It incorporates time series data with perturbation analysis to improve network reconstuction by combining time series data from several perturbed networks. The software is available at `http://www.cse.chalmers.se/~jethava/netgem.html`.

## 7 Discussion

This survey discusses recent methods for reconstruction of time-varying networks based on time series gene expression data. This problem is ill-posed due to high-dimensional data i.e. number of variables $p$ is much larger than number of observations $n$. Additional assumptions on the network structure as well as the temporal dynamics governing network evolution are required in order to facilitate reconstruction of time-varying interaction networks.

A popular assumption on the network structure is the sparsity of interaction network i.e. each gene interacts with at most few other genes. This agrees closely with domain knowledge and yields biologically plausible networks. This network sparsity is imposed by using $\ell_1$ regularization in optimization-based methods, or an sparsity inducing prior in Bayesian approach.

The underlying causes governing network evolution in time-varying interaction networks are not well-understood. A number of simplifying assumptions have been made in order to model different aspects of the network evolution including smooth variation i.e. the underlying network changes slowly over time, piece-wise constant with sharp changes, Markovian dynamics. The reconstructions using the different methods have been shown to yield biologically plausible networks. Further, the reconstructed networks often predict transient interactions which may be experimentally verified; leading to a deeper understanding of biological processes. A clear understanding of network evolution is yet to emerge; and this is an exciting direction for future research.

We also discuss methods which allow systematic analysis of time-series data corresponding to multiple related networks. These methods allow network reconstruction based on multiple data sources e.g. gene interaction networks, miRNA-mRNA interactions, protein-protein interactions (PPI); as well as multiple experiments with genetic perturbations. Such approaches allow better network reconstruction by combining information from several experiments. This is becoming increasingly relevant with growth in publicly available -omics data due to recent advances in sequencing methods.

## References

1. Ahmed, A., Xing, E.: Recovering time-varying networks of dependencies in social and biological studies. Proceedings of the National Academy of Sciences **106**(29), 11,878–11,883 (2009)
2. Alon, U.: An introduction to systems biology: design principles of biological circuits, vol. 10. CRC press (2007)
3. Ambroise, C., Chiquet, J., Matias, C.: Inferring sparse gaussian graphical models with latent structure. Electronic Journal of Statistics **3**, 205–238 (2009)
4. Androulakis, I., Yang, E., Almon, R.: Analysis of time-series gene expression data: Methods, challenges, and opportunities. Annu. Rev. Biomed. Eng. **9**, 205–228 (2007)
5. Arbeitman, M., Furlong, E., Imam, F., Johnson, E., Null, B., Baker, B., Krasnow, M., Scott, M., Davis, R., White, K.: Gene expression during the life cycle of drosophila melanogaster. Science **297**(5590), 2270–2275 (2002)
6. Ashburner, M., Ball, C., Blake, J., Botstein, D., Butler, H., Cherry, J., Davis, A., Dolinski, K., Dwight, S., Eppig, J., et al.: Gene ontology: tool for the unification of biology. Nature genetics **25**(1), 25 (2000)
7. Banerjee, O., El Ghaoui, L., d'Aspremont, A.: Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data. The Journal of Machine Learning Research **9**, 485–516 (2008)
8. Barabási, A., Oltvai, Z.: Network biology: understanding the cell's functional organization. Nature Reviews Genetics **5**(2), 101–113 (2004)
9. Barabasi, L., Gulbahce, N., Loscalso, J.: Network medicine: A network–based approach to human disease. Nature Reviews Genetics **12**, 56–68 (2011)
10. Boyd, S., Vandenberghe, L.: Convex optimization. Cambridge Univ Pr (2004)
11. Buhl, S.: On the existence of maximum likelihood estimators for graphical gaussian models. Scandinavian Journal of Statistics pp. 263–270 (1993)
12. Bühlmann, P., Van De Geer, S.: Statistics for High-Dimensional Data: Methods, Theory and Applications. Springer-Verlag New York Inc (2011)

13. Candes, E., Tao, T.: The dantzig selector: statistical estimation when p is much larger than n. The Annals of Statistics **35**(6), 2313–2351 (2007)
14. Carroll, S.: Evolution at two levels: on genes and form. PLoS biology **3**(7), e245 (2005)
15. Cipollina, C., van den Brink, J., Daran-Lapujade, P., Pronk, J., Porro, D., de Winde, J.: Saccharomyces cerevisiae sfp1: at the crossroads of central metabolism and ribosome biogenesis. Microbiology **154**(6), 1686–1699 (2008)
16. Clarke, R., Ressom, H., Wang, A., Xuan, J., Liu, M., Gehan, E., Wang, Y.: The properties of high-dimensional data spaces: implications for exploring gene and protein expression data. Nature Reviews Cancer **8**(1), 37–49 (2008)
17. Davidson, E.: Genomic regulatory systems: development and evolution. Academic Pr (2001)
18. Dempster, A.: Covariance selection. Biometrics pp. 157–175 (1972)
19. Donoho, D.: High-dimensional data analysis: The curses and blessings of dimensionality. AMS Math Challenges Lecture pp. 1–32 (2000)
20. Donoho, D.: Compressed sensing. Information Theory, IEEE Transactions on **52**(4), 1289–1306 (2006)
21. Duchi, J., Shalev-Shwartz, S., Singer, Y., Chandra, T.: Efficient projections onto the l 1-ball for learning in high dimensions. In: Proceedings of the 25th international conference on Machine learning, pp. 272–279. ACM (2008)
22. Ernst, J., Nau, G., Bar-Joseph, Z.: Clustering short time series gene expression data. Bioinformatics **21**(suppl 1), i159–i168 (2005)
23. Friedman, J., Hastie, T., Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso. Biostatistics **9**(3), 432–441 (2008)
24. Friedman, J., Hastie, T., Tibshirani, R.: The elements of statistical learning, 2 edn. Springer Series in Statistics (2009)
25. Friedman, N., Linial, M., Nachman, I., Pe'er, D.: Using bayesian networks to analyze expression data. Journal of computational biology **7**(3-4), 601–620 (2000)
26. Fu, W., Song, L., Xing, E.: Dynamic mixed membership blockmodel for evolving networks. In: Proceedings of the 26th Annual International Conference on Machine Learning, pp. 329–336. ACM (2009)
27. Gitter, A., Lu, Y., Bar-Joseph, Z.: Computational methods for analyzing dynamic regulatory networks. Methods in molecular biology (Clifton, NJ) **674**, 419 (2010)
28. Glass, L., Kaplan, D.: Time series analysis of complex dynamics in physiology and medicine. Medical progress through technology **19**, 115–115 (1993)
29. Guo, F., Hanneke, S., Fu, W., Xing, E.: Recovering temporally rewiring networks: A model-based approach. In: Proceedings of the 24th international conference on Machine learning, pp. 321–328. ACM (2007)
30. Guo, J., Levina, E., Michailidis, G., Zhu, J.: Joint estimation of multiple graphical models. Biometrika **98**(1), 1–15 (2011)
31. Hartemink, A., et al.: Reverse engineering gene regulatory networks. Nature biotechnology **23**(5), 554–555 (2005)
32. de Hoon, M., Imoto, S., Miyano, S.: Inferring gene regulatory networks from time-ordered gene expression data using differential equations. In: Discovery science, pp. 283–288. Springer (2002)
33. Hu, H., Yan, X., Huang, Y., Han, J., Zhou, X.: Mining coherent dense subgraphs across massive biological networks for functional discovery. Bioinformatics **21**(suppl 1), i213–i221 (2005)
34. Ideker, T., Sharan, R.: Protein networks in disease. Genome Res. **18**, 644–652 (2008)
35. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., Sakaki, Y.: A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proceedings of the National Academy of Sciences **98**(8), 4569 (2001)
36. Jethava, V., Bhattacharyya, C., Dubhashi, D., Vemuri, G.: Netgem: Network embedded temporal generative model for gene expression data. BMC bioinformatics **12**(1), 327 (2011)
37. Kim, S., Imoto, S., Miyano, S.: Dynamic bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data. Biosystems **75**(1), 57–65 (2004)

38. Koh, K., Kim, S., Boyd, S.: An interior-point method for large-scale l1-regularized logistic regression. Journal of Machine learning research **8**(8), 1519–1555 (2007)
39. Koller, D., Friedman, N.: Probabilistic graphical models: principles and techniques. The MIT Press (2009)
40. Lam, C., Fan, J.: Sparsistency and rates of convergence in large covariance matrix estimation. Annals of statistics **37**(6B), 4254 (2009)
41. Lauritzen, S.: Graphical models, vol. 17. Oxford University Press, USA (1996)
42. Lin, C., Weng, R., Keerthi, S.: Trust region newton method for logistic regression. The Journal of Machine Learning Research **9**, 627–650 (2008)
43. Luscombe, N., Babu, M., Yu, H., Snyder, M., Teichmann, S., Gerstein, M.: Genomic analysis of regulatory network dynamics reveals large topological changes. Nature **431**(7006), 308–312 (2004)
44. Ma, S., Gong, Q., Bohnert, H.: An arabidopsis gene network based on the graphical gaussian model. Genome research **17**(11), 1614–1625 (2007)
45. Meinshausen, N., Bühlmann, P.: High-dimensional graphs and variable selection with the lasso. The Annals of Statistics **34**(3), 1436–1462 (2006)
46. Mewes, H., Frishman, D., Gruber, C., Geier, B., Haase, D., Kaps, A., Lemcke, K., Mannhaupt, G., Pfeiffer, F., Schüller, C., et al.: Mips: a database for genomes and protein sequences. Nucleic Acids Research **28**(1), 37–40 (2000)
47. Parisi, G., Shankar, R.: Statistical field theory. Physics Today **41**, 110 (1988)
48. Perrin, B., Ralaivola, L., Mazurie, A., Bottani, S., Mallet, J., dAlche Buc, F.: Gene networks inference using dynamic bayesian networks. Bioinformatics **19**(suppl 2), ii138–ii148 (2003)
49. Peer, D., Regev, A., Elidan, G., Friedman, N.: Inferring subnetworks from perturbed expression profiles. Bioinformatics **17**(suppl 1), S215–S224 (2001)
50. Ravikumar, P., Wainwright, M., Lafferty, J.: High-dimensional ising model selection using 1-regularized logistic regression. The Annals of Statistics **38**(3), 1287–1319 (2010)
51. Robins, G., Pattison, P., Kalish, Y., Lusher, D.: An introduction to exponential random graph $p^*$ models for social networks. Social networks **29**(2), 173–191 (2007)
52. Rothman, A., Bickel, P., Levina, E., Zhu, J.: Sparse permutation invariant covariance estimation. Electronic Journal of Statistics **2**, 494–515 (2008)
53. Sachs, K., Perez, O., Pe'er, D., Lauffenburger, D., Nolan, G.: Causal protein-signaling networks derived from multiparameter single-cell data. Science's STKE **308**(5721), 523 (2005)
54. Schadt, E.: Molecular networks as sensors and drivers of common human diseases. Nature **416**, 218–223 (2009)
55. Schäfer, J., Strimmer, K.: An empirical bayes approach to inferring large-scale gene association networks. Bioinformatics **21**(6), 754–764 (2005)
56. Schliep, A., Schönhuth, A., Steinhoff, C.: Using hidden markov models to analyze gene expression time course data. Bioinformatics **19**(suppl 1), i255–i263 (2003)
57. Shermin, A., Orgun, M.: Using dynamic bayesian networks to infer gene regulatory networks from expression profiles. In: Proceedings of the 2009 ACM symposium on Applied Computing, pp. 799–803. ACM (2009)
58. Song, L., Kolar, M., Xing, E.: Keller: estimating time-varying interactions between genes. Bioinformatics **25**(12), i128–i136 (2009)
59. Soranzo, N., Bianconi, G., Altafini, C.: Comparing association network algorithms for reverse engineering of large-scale gene regulatory networks: synthetic versus real data. Bioinformatics **23**(13), 1640–1647 (2007)
60. Speed, T., Kiiveri, H.: Gaussian markov distributions over finite graphs. The Annals of Statistics **14**(1), 138–150 (1986)
61. Tegner, J., Yeung, M., Hasty, J., Collins, J.: Reverse engineering gene networks: integrating genetic perturbations with dynamical modeling. Proceedings of the National Academy of Sciences **100**(10), 5944 (2003)
62. Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological) pp. 267–288 (1996)

63. Uetz, P., Giot, L., Cagney, G., Mansfield, T., Judson, R., Knight, J., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al.: A comprehensive analysis of protein–protein interactions in saccharomyces cerevisiae. Nature **403**(6770), 623–627 (2000)
64. Wainwright, M., Ravikumar, P., Lafferty, J.: High-dimensional graphical model selection using l˜ 1-regularized logistic regression. Advances in neural information processing systems **19**, 1465 (2007)
65. Werhli, A., Grzegorczyk, M., Husmeier, D.: Comparative evaluation of reverse engineering gene regulatory networks with relevance networks, graphical gaussian models and bayesian networks. Bioinformatics **22**(20), 2523–2531 (2006)
66. Wille, A., Zimmermann, P., Vranová, E., Fürholz, A., Laule, O., Bleuler, S., Hennig, L., Prelic, A., Von Rohr, P., Thiele, L., et al.: Sparse graphical gaussian modeling of the isoprenoid gene network in arabidopsis thaliana. Genome Biol **5**(11), R92 (2004)
67. Workman, C., Mak, H., McCuine, S., Tagne, J., Agarwal, M., Ozier, O., Begley, T., Samson, L., Ideker, T.: A systems approach to mapping dna damage response pathways. Science's STKE **312**(5776), 1054 (2006)
68. Yeang, C., Mak, H., McCuine, S., Workman, C., Jaakkola, T., Ideker, T.: Validation and refinement of gene-regulatory pathways on a network of physical interactions. Genome Biology **6**(7), R62 (2005)
69. Yeung, M., Tegnér, J., Collins, J.: Reverse engineering gene networks using singular value decomposition and robust regression. Proceedings of the National Academy of Sciences **99**(9), 6163 (2002)
70. Yuan, M., Lin, Y.: Model selection and estimation in the gaussian graphical model. Biometrika **94**(1), 19–35 (2007)
71. Zhou, S., Lafferty, J., Wasserman, L.: Time varying undirected graphs. Machine Learning **80**(2), 295–319 (2010)
72. Zou, M., Conzen, S.: A new dynamic bayesian network (dbn) approach for identifying gene regulatory networks from time course microarray data. Bioinformatics **21**(1), 71–79 (2005)