

# A Scale-Free Structure Prior for Graphical Models with Applications in Functional Genomics

Paul Sheridan\*, Takeshi Kamimura, Hidetoshi Shimodaira

Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Tokyo, Japan

## Abstract

The problem of reconstructing large-scale, gene regulatory networks from gene expression data has garnered considerable attention in bioinformatics over the past decade with the graphical modeling paradigm having emerged as a popular framework for inference. Analysis in a full Bayesian setting is contingent upon the assignment of a so-called structure prior—a probability distribution on networks, encoding *a priori* biological knowledge either in the form of supplemental data or high-level topological features. A key topological consideration is that a wide range of cellular networks are approximately scale-free, meaning that the fraction,  $P(k)$ , of nodes in a network with degree  $k$  is roughly described by a power-law  $P(k) \propto k^{-\gamma}$  with exponent  $\gamma$  between 2 and 3. The standard practice, however, is to utilize a random structure prior, which favors networks with binomially distributed degree distributions. In this paper, we introduce a scale-free structure prior for graphical models based on the formula for the probability of a network under a simple scale-free network model. Unlike the random structure prior, its scale-free counterpart requires a node labeling as a parameter. In order to use this prior for large-scale network inference, we design a novel Metropolis-Hastings sampler for graphical models that includes a node labeling as a state space variable. In a simulation study, we demonstrate that the scale-free structure prior outperforms the random structure prior at recovering scale-free networks while at the same time retains the ability to recover random networks. We then estimate a gene association network from gene expression data taken from a breast cancer tumor study, showing that scale-free structure prior recovers hubs, including the previously unknown hub SLC39A6, which is a zinc transporter that has been implicated with the spread of breast cancer to the lymph nodes. Our analysis of the breast cancer expression data underscores the value of the scale-free structure prior as an instrument to aid in the identification of candidate hub genes with the potential to direct the hypotheses of molecular biologists, and thus drive future experiments.

**Citation:** Sheridan P, Kamimura T, Shimodaira H (2010) A Scale-Free Structure Prior for Graphical Models with Applications in Functional Genomics. PLoS ONE 5(11): e13580. doi:10.1371/journal.pone.0013580

**Editor:** Joel S. Bader, Johns Hopkins University, United States of America

**Received:** February 17, 2010; **Accepted:** September 28, 2010; **Published:** November 5, 2010

**Copyright:** © 2010 Sheridan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This research was supported in part by Computationism as a Foundation for the Sciences, a program of the Japan Society for the Promotion of Science (JSPS) Global Centers for Excellence (COE) ([http://compview.titech.ac.jp/front-page-en/view?set\\_language=en](http://compview.titech.ac.jp/front-page-en/view?set_language=en)). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: sherida6@is.titech.ac.jp

## Introduction

### Gene Regulatory Networks and Gene Expression Data

Knowledge of the interactions among genes and gene products that occur within a cell is vital for understanding cellular behavior. These activities are largely a consequence of gene expression, the process whereby genes transcribe signature mRNA molecules that are translated into gene products of numerous kinds and functions. As it happens, genes do not express independently of one another; instead, their activities are coordinated in a complex system of control in which distinguished genes, called transcription factors, regulate the expression of other genes via their gene product proxies.

An undirected network  $G$  is a mathematical object consisting of a set of nodes and a set of unordered pairs of nodes called undirected edges. It differs from a directed network, which is also denoted by  $G$ , in that the latter is defined in terms of ordered pairs of nodes known as directed edges. Applying these straightforward abstractions to cellular processes has gained currency throughout the biosciences, so much so that a network mind-set has become a necessary precondition for thinking about systems of

gene regulatory interactions. For the purposes of this paper, a gene regulatory network is a directed network in which genes are identified with nodes and regulatory interactions with directed edges. From a purely statistical standpoint, it is best to regard a gene regulatory network as a convenient depiction of the true regulatory interactions of a system that, in reality, must be estimated from data.

Indeed, the network approach toward understanding gene regulatory systems only came to prominence in response to the advent of DNA microarray technology, which makes the profiling of mRNA expression levels for individual genes possible on a genome-wide scale. A typical experiment consists of a library of  $n$  expression profiles, each one a snapshot of the expression levels for  $p$  genes under a different experimental condition. The raw expression profile data is preprocessed and then arranged by row in an  $n \times p$  data matrix,  $D$ . In practice, not only is gene expression data notoriously noisy [1], but to make matters worse the number of samples is typically at least an order of magnitude smaller than the number of genes, that is,  $n \ll p$  (the “small  $n$ , large  $p$ ” problem), making the inference of regulatory interactions a challenging statistical problem [2].

There is now an extensive repertoire of algorithms available for the analysis of gene expression data, the majority of which are based on Boolean networks, differential equations, and graphical models [3]. Some approaches produce estimated gene regulatory networks that are directed networks, while others do not. In this paper, we work primarily with a variety of undirected graphical model known as gene association networks (GANs), in which undirected edges, called gene associations, correspond to certain statistical dependencies that are inferred from gene expression data. Therefore, in an effort to simplify the terminology, the terms “network” and “edge” will be used hereafter to mean undirected network and undirected edge, respectively. Although we will occasionally use the term “network” in a colloquial sense, such as in “network mind-set” or “network approach.” At any rate, the meaning should be clear from context.

## Graphical Models

Graphical models [4], [5], [6] are a suite of probabilistic models, widely used for estimating large-scale gene regulatory networks from gene expression data [7]. In this framework, genes are identified with the random variables of a multivariate distribution  $X$  with covariance matrix  $\Sigma$ , and each row of  $D$  is taken as a random sample from  $X$ . The conditional independence structure of  $X$  defines a network with the random variables as nodes and conditional dependencies latent between the random variables as directed or undirected edges; a diversity of models arise from the extent to which the dependencies are resolved [8].

Relevance networks comprise the simplest class of graphical model with absent edges corresponding to marginal independencies between the components of  $X$ . These networks have long been used in the analysis of genetic data [9]. But in terms of identifying regulatory interactions, relevance networks are bound to be misleading because marginal independence alone cannot discriminate among direct and indirect dependencies.

GANs provide a better alternative, circumventing this drawback by appealing to conditional independence as a criterion for edge exclusion. Gaussian graphical models (GGMs) are the gold standard. In a GGM, a pair of nodes do not share an edge when their underlying random variables from  $X$  are conditionally independent given all of the remaining random variables. However, GGMs too are not without disadvantages, as their estimation can be computationally intensive in a “small  $n$ , large  $p$ ” setting [10]. A class of GANs, bridging the gap between relevance networks and GGMs, has been advanced with this consideration in mind where absent edges are identified with lower order conditional independencies [11], [12], [13].

Lastly, Bayesian networks are a variety of graphical model founded on a more refined notion of conditional independence, conferring directionality to the edges; they are also well-established as a methodology for estimating gene regulatory networks [14].

## The Structure Prior

Inference within the graphical modeling paradigm amounts to an often painstaking exercise in covariance estimation and model selection. We defer a discussion on the problem of covariance estimation to the Methods section. That is because our interest pertains to model selection, which in a Bayesian setting is accomplished by sampling from the posterior distribution

$$P(G|D) \propto P(D|G) \times \pi(G) \quad (1)$$

over the appropriate space of networks using either heuristic searches or else Markov chain Monte Carlo (MCMC). The term

$P(D|G)$  is the likelihood and  $\pi(G)$  the structure prior, that is, a prior assigning a probability to each possible network.

The role of the structure prior is to direct inference toward graphical models consistent with biological prior knowledge, which may come in the form of *a priori* topological considerations or from *a posteriori* sources apart from the dataset. As far as the latter is concerned, previous research has concentrated on Bayesian networks [15], [16], [17], [18]. On the other hand, biologically-motivated topological assumptions are a consistent feature of graphical models tailored for genetic data. Heuristic search strategies often include implicit assumptions concerning network sparsity [19], [20], [21], [22]. In instances in which the structure prior is given explicit specification, standard practices include using a uniform prior capped at a small number of potential regulators per gene [23], or assigning it as a sparse random network [24], [25].

## Random and Scale-Free Networks

The theory of random networks was given its first systematic expression by Erdős and Rényi [26], [27]. According to the theory, a  $p$ -node random network is defined by an eponymous, generating algorithm — the Erdős-Rényi (ER) model — that works by connecting each pair of nodes in an initially empty network independently with probability  $\beta$ . This simple procedure gives rise to a probability distribution over the space of  $p$ -node networks, which is used to define the so-called random structure prior,  $\pi_r(G)$ . The degree distribution  $P(k)$  — the probability that a given node is connected to  $k$  other nodes — of a random network is binomially distributed according to  $\binom{p-1}{k} \beta^k (1-\beta)^{p-1-k}$  where the degree,  $k$ , of a node denotes the number of edges incident upon it. It follows, therefore, that degree in a random network has a strong central tendency, implying that the average degree of a random network is representative of the degree of a typical node.

Empirical studies, however, have firmly established that a wide variety of large-scale networks in nature, society, and technology exhibit heavy-tailed degree distributions that cannot be accounted for by random network theory [28], [29], [30]. This property is often approximately described by a power-law degree distribution,  $P(k) \propto k^{-\gamma}$ , over a large range of  $k$  with exponent  $\gamma$  typically between 2 and 3. A network that follows a power-law is called scale-free. It gets this name because the functional form of  $P(k)$  is retained under a scaling of the argument  $k$  by a constant factor  $\alpha$ :  $P(\alpha \times k) \propto P(k)$ . The scale-free property is thought to be a key organizational feature of cellular networks [31], and analyses suggest such an architecture for the gene regulatory networks of the model organisms *S. cerevisiae* [32] and *C. elegans* [33].

## Introducing a Scale-Free Structure Prior

Proposing a structure prior which incorporates the scale-free property is the topic of this paper. We define the scale-free structure prior,  $\pi_{sf}(G)$ , according to the probability of a network under a simple, scale-free network model. As for the underlying network model itself, a multitude of candidates have been proposed in the literature [34]. They fall into two broad categories: 1) growing models, where a network is generated via the successive addition of nodes and edges to a small seed network, and 2) non-growing models, where to a fixed number of nodes, pairs of nodes are chosen randomly and connected by edges.

The growing model approach employs a handful of simple universal mechanisms, thought to underpin disparate natural phenomena, to drive the stochastic evolution of networks toward power-laws. Preferential attachment is, perhaps, the best known

mechanism. The idea works something like this: the probability of attaching an edge from a newly added node to a node already in the system is roughly proportional to the degree of the old node. The Barabasi-Albert (BA) model [35] is the latter-day progenitor of a wide variety of preferential attachment models. The BA model generates a network via the successive addition of nodes and edges to a small seed network. At each step, a node is added to the system with a fixed number of emanating edges, which are subsequently preferential attached to the existing nodes. The resulting network follows a power-law with  $\gamma=3$  on average.

Preferential attachment is not considered to be the main driving force behind genome evolution; instead, gene (node) duplication and point mutations (edge dynamics) play the dominant role in shaping of gene regulatory networks [36]. The duplication model as formulated by [37] is such a network model, which in an analysis by [38] is suggested to approximately follow a power-law.

By contrast, in the non-growing approach, each node is assigned a fixed weight with the probability of a particular network depending on those weights. The ER model is an example of a non-growing model with uniform weights. Another non-growing model is the static model [39], which is a generalization of the ER model that has been shown to follow a power-law with  $\gamma$  tunable to any value greater than 2, depending on the specification of the model parameters; see Methods for details. We use the static model to define  $\pi_{sf}(G)$ . Indeed, this model is an appealing candidate for the purpose as the probability of a network is easy to compute compared with growing models of similar complexity. Moreover, the static model actually includes the ER model as a limiting case.

### A New Metropolis-Hastings Sampler for Networks

We implement an MCMC algorithm with  $\pi_{sf}(G)$  for GGMs adapted from [25], although it is important to point out that our methodology applies to graphical models in general. Reworking the algorithm is not simply a matter of plugging in a formula for  $\pi_{sf}(G)$  because it depends furthermore on a labeling of the nodes of  $G$ . Confronted with this complication, we design a novel Metropolis-Hastings sampler that solves the problem by including a node labeling,  $\sigma$ , which is defined in the Methods section, as a variable in the state space, thereby allowing it to be estimated.

### Summary of Contributions

In this paper, we advance a scale-free structure prior,  $\pi_{sf}(G)$ , for graphical models defined by the formula for the probability of a network under the static model. Our objective is to compare the performance of this prior with the commonly used random structure prior,  $\pi_r(G)$ , in the arena of simulation as well as with a real data example. We choose GGMs for this purpose, modifying the MCMC algorithm of [25] to include  $\pi_{sf}(G)$ . As mentioned above, one challenge of implementing  $\pi_{sf}(G)$  is that, unlike with  $\pi_r(G)$ , it requires a labeling of the nodes of  $G$ . We address this issue by introducing a Metropolis-Hastings sampler that includes the node labeling as a variable in the state space.

In a simulation study, we generate networks with given degree distributions together with Gaussian data in accordance with their implied conditional independence structures. As a case study we show that  $\pi_r(G)$  and  $\pi_{sf}(G)$  are equally effective at recovering a random network, but that  $\pi_r(G)$  is comparatively ineffective at recovering a scale-free network. In the full simulation study, we confirm that the aforementioned result holds, illustrating our main conclusion:  $\pi_{sf}(G)$  recovers random networks on an equal footing with the  $\pi_r(G)$ , yet surpasses it in recovering scale-free networks. Finally, we illustrate our methodology by analyzing a real gene expression dataset taken from a breast cancer tumor study by [40],

showing that in contrast with the random structure prior, the scale-free structure prior recovers hubs, including the estrogen regulator FOXA1 and the zinc transporter SLC39A6, which was previously unrecognized as a hub.

## Methods

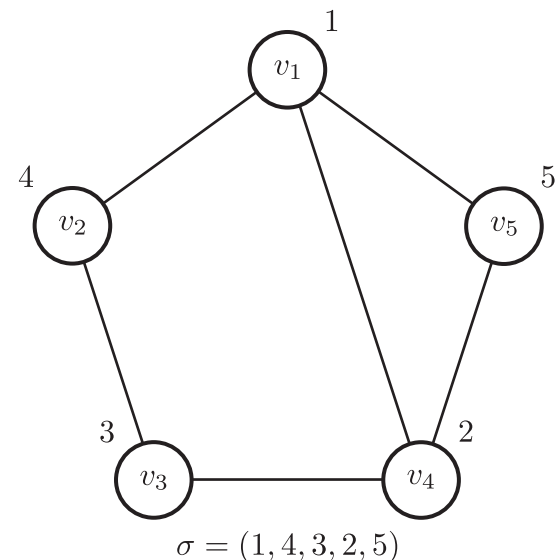
### Network Notation

The terms “network” and “graph” are used synonymously throughout this paper. An undirected network  $G=(V,E)$  is a mathematical object defined by a set of nodes  $V=\{v_1,v_2,\dots,v_p\}$  together with a set of undirected edges  $E$  consisting of unordered pairs  $\{v_i,v_j\}$  taken from  $V$ , provided that  $v_i\neq v_j$ . The set of all  $p$ -node, undirected networks is denoted by  $\mathcal{G}_p$ . A directed network is defined in an analogous manner, save that the elements of  $E$  are ordered pairs  $(v_i,v_j)$  called directed edges;  $v_i$  is called the parent and  $v_j$  the child.

It should be understood that a network refers to an undirected network, and likewise an edge is to be understood to mean an undirected edge. However, the following definitions are applicable to both undirected and directed networks. An empty network has no edges, that is  $E=\emptyset$ , while, in a complete network  $E$  is defined as the cross product  $V\times V$ . A subnetwork of  $G$  is a network whose node set  $V'$  is a subset of  $V$ , and whose edges are a subset of  $E$  restricted to  $V'$ . The subnetwork of  $G$  induced by a given subset of nodes  $V'\subseteq V$  is the subnetwork containing all edges from  $E$  that connect nodes in  $V'$ . Two nodes are said to be neighbors when they are connected by an edge. And, a network is itself connected when every pair of nodes is connected by a sequence of neighbors. Finally, a node labeling  $\sigma=(\sigma_1,\sigma_2,\dots,\sigma_p)$  is a permutation of the integers  $1,2,\dots,p$ , applied to the nodes of  $G$  so that each  $v_i\in V$  is represented by the integer  $\sigma_i$ ; see Figure 1. This node labeling is used later for defining the structure prior.

### Gaussian Graphical Models

In this section we sketch out the theory of GGMs essential to this paper. A detailed overview of the GGM estimation procedures outlined here is described in [25], while [8] is a good starting point



**Figure 1. An example of a node labeling.** A node labeling,  $\sigma=(3,4,2,1,5)$ , of the nodes of a network with  $p=5$  nodes so that, for instance,  $\sigma_2=4$  and  $\sigma_3=5$ .

doi:10.1371/journal.pone.0013580.g001

for understanding the niche they occupy in the larger context of graphical models.

Let  $X = (X_1, X_2, \dots, X_p)$  be a  $p$ -dimensional Gaussian random vector with zero-mean and positive definite covariance matrix,  $\Sigma$ . Two random variables  $X_i$  and  $X_j$  are not conditionally independent given the remaining variables in  $X$  if, and only if, there is a corresponding nonzero entry in the precision matrix,  $\Omega = \Sigma^{-1}$  [41]. The conditional independence structure of  $X$  can be represented by a network,  $G$ , where  $X_i$  is the value at node  $v_i$  and there is an edge between  $v_i$  and  $v_j$  when  $X_i$  and  $X_j$  are not conditionally independent. A GGM for  $X$  is the family of  $p$ -dimensional Gaussian distributions from  $N(0, \Sigma)$ , constrained by the structure of  $G$ .

Fitting a GGM to a given dataset — a task known as covariance selection — amounts to identifying zeros in the estimated precision matrix. In the classical setting when  $n > p$ , ensuring that  $\Sigma$  is positive definite, this is typically accomplished by inverting the estimated covariance matrix and then applying statistical tests to identify any entries significantly different from zero [4]. With genomic data, however, “small  $n$ , large  $p$ ” is the norm and, consequently,  $\Sigma$  will not generally be invertible.

This problem can be addressed in one of two ways. One way calls for restricting inference to pairwise independencies conditioned on fewer than all  $p - 2$  remaining random variables. A relevance network, for example, is constructed by estimating the pairwise correlations between all random variables, connecting any pair with correlation exceeding a specified cutoff value [9]. A related approach goes one step beyond a relevance network by estimating a GAN based on not only marginal but also first-order conditional independencies [11].

A more ambitious approach is to compute satisfactory small sample estimates for  $\Sigma$  and  $\Omega$  using Bayesian methods. Empirical Bayesian solutions are exemplified by shrinkage estimates [21] and sparsity encouraging lasso regression estimates [22]. Meanwhile, the full Bayesian scheme of [42] works by marginalizing over  $\Sigma$  to compute the likelihood term in (1), using a prior that constrains elements of the precision matrix to zero depending on  $G$ :

$$P(D|G) = \int_{\Sigma|G} P(D|\Sigma, G)P(\Sigma|G)d\Sigma. \tag{2}$$

The term  $P(D|\Sigma, G)$  is multivariate Gaussian, while the prior  $P(\Sigma|G)$  is hyper-inverse Wishart with hyperparameters  $\Phi$ , a positive definite dispersion matrix, and  $\delta > 0$ , a degrees of freedom parameter. Jones et. al [25] advise a small value for  $\delta$  as a reflection of ignorance, and take  $\Phi$  as the diagonal matrix  $\tau I$ , which assumes that the underlying Gaussian variables have common variance. A consequence of this assignment is that  $\delta$  can be used to specify  $\tau$  by making use of the fact that the marginal prior mode for each variance term is  $\text{var}(X_i) = \tau/(\delta + 1)$ .

GGM theory comes equipped with powerful techniques for computing the likelihood function when the underlying network is decomposable. Roughly speaking, a decomposable network can be broken down into distinguished subsets of nodes called maximal cliques. A clique is a subset of nodes whose induced subgraph is complete, and is called maximal when it is not contained within a larger complete subgraph. Computing the likelihood for a subset of  $X$  corresponding to a maximal clique is particularly tractable because the density is just an unrestricted multivariate Gaussian. Hence, when a network is decomposable, the evaluation of the likelihood term in (1) reduces to the computation of many likelihoods of smaller dimension [43]. We will return to these issues in the section on our MCMC implementation.

### The Static Model

A network model is a stochastic algorithm for generating networks that may depend on a vector of parameters,  $\theta = (\theta_1, \theta_2, \dots, \theta_M)$ . Associated with any model is a probability distribution, assigning a probability  $P(G|\theta, \sigma) \geq 0$  to each  $G \in \mathcal{G}_p$ , where  $\sigma$  is a node labeling of  $G$ .

The static network model [39] works by first assigning a weight  $w_i = \sigma_i^{-\mu} / \sum_{l=1}^p \sigma_l^{-\mu}$  to each node  $v_1, v_2, \dots, v_p$  where  $\mu$ , the Zipf exponent, is a tunable parameter in  $(0, 1)$ . To generate a network,  $G$ , the following step is repeated  $p \times K$  ( $p^{-\mu} \ll K \ll p^{1-\mu}$ ) times: select nodes  $v_i$  and  $v_j$  with probabilities  $w_i$  and  $w_j$  and connect them with an edge, unless  $v_i = v_j$  or  $v_i$  and  $v_j$  are already connected, in which case no edge is added to the network. The overall model parameter is  $\theta = (\mu, K)$ .

In order to work out the functional form of the degree distribution, it is enough to notice that, on average, each node acquires edges in proportion to its weight. Supposing for a moment that  $k_i$  denotes the degree of node  $v_i$ , we may write this as  $k_i \propto \sigma_i^{-\mu}$ . The probability distribution over the  $k_i$ 's is known as Zipf's law, and it has been shown to be equivalent to a power-law degree distribution with  $\gamma = 1 + 1/\mu$  [44]. It follows that the static model generates networks that follow a power-law with  $2 < \gamma < \infty$  depending on the choice of  $\mu$ . A rigorous derivation of the power-law appears in a comprehensive analysis of the static model by Lee et al. [45]. In the case when  $1/2 < \mu < 1$ , the exponent,  $\gamma$ , lies between 2 and 3, which is the most interesting range of values from the point of view of scale-free architecture. In contrast, for values of  $\mu < 1/2$ , which corresponds to  $\gamma > 3$ , the tail of  $P(k)$  is less pronounced. In the limit of  $\mu \rightarrow 0$ , or equivalently  $\gamma \rightarrow \infty$ , each weight  $w_1, \dots, w_p$  tends to  $1/p$ , resulting in the ER model with edge inclusion probability  $\beta = 1 - (1 - 2/p^2)^{pK}$ . To be clear, the static model actually includes the ER model as a special case.

A formula for the probability of a network is provided in the same analysis. The probability that nodes  $v_i$  and  $v_j$  are connected in the final network is  $1 - (1 - 2w_i w_j)^{pK}$ , which is well-approximated by  $e^{-2pKw_i w_j}$  when  $p$  is large. The probability of a network, then, is given by overall product of the edge inclusion probabilities

$$P(G|\theta, \sigma) = \prod_{\{v_i, v_j\} \in E} \left(1 - e^{-2pKw_i w_j}\right) \prod_{\{v_i, v_j\} \notin E} e^{-2pKw_i w_j}, \tag{3}$$

assuming independence.

### A Scale-Free Structure Prior

The structure prior is generically defined as

$$\pi(G) = \frac{1}{p!} \sum_{\sigma} \int_{\theta} P(G|\theta, \sigma) \pi(\theta) d\theta \tag{4}$$

where  $P(G|\theta, \sigma)$  is the probability of a  $p$ -node network,  $G$ , under a certain network model given a vector of parameters,  $\theta$ , and a node labeling,  $\sigma$ ; the summation is over all permutations of  $\sigma$ . It is obvious from the definition that  $\theta$  and  $\sigma$  are hyperparameters and must be dealt with accordingly. In the case of  $\sigma$ , each one of the  $p!$  possible node labeling assignments is allotted uniform weight. In our work, we additionally impose that  $\pi(\theta)$  is a uniform prior, leaving the details to be described below within the contexts of specific network models.

The simplest means of dealing with uncertainty about graphical structures is to assign uniform weight either to each  $G \in \mathcal{G}_p$ , that is,  $\pi(G) = 1/|\mathcal{G}_p|$ , or to the subspace of decomposable networks [43]. This approach is in fact a special case of the probability of a network under the ER network model when  $\beta = 1/2$ . A related

approach uses a prior that distributes probability mass uniformly according to the number of edges as opposed to individual networks [42]. More recently, the ER model has been explicitly employed as a structure prior [24], [25]. The random structure prior,  $\pi_r(G)$ , is formally defined by

$$P(G|\theta) = \beta^{|E|}(1-\beta)^{T-|E|} \tag{5}$$

where  $\theta = (\beta)$  and  $T = \binom{p}{2}$  is the number of possible edges. A node labeling would be superfluous due to symmetry. To foster sparsity  $\beta$  may be fixed at  $2/(p-1)$  so that the expected number of edges comes out to be  $p$  when (5) is taken over all networks. Although strictly speaking the value will be somewhat lower in the decomposable case [25]. The approach taken in this paper goes one step further by simply taking  $\pi(\beta)$  as uniform over the unit interval.

As explained above, the static model with parameter  $\theta = (\mu, K)$  is a generalization of the ER model that is accommodating to scale-free topologies. We define the the scale-free structure prior,  $\pi_{sf}(G)$ , according to the probability of a network under (3). The static model has two parameters,  $\theta = (\mu, K)$ , and they are not exactly independent as the domain of  $K$  is a function of  $\mu$ . This means that the prior  $\pi(\mu, K)$  must actually be treated as the product of  $\pi(\mu)$  and  $\pi(K|\mu)$ . We take each term to be uniform over its respective domain.

### MCMC Implementation

MCMC algorithms are commonly used for sampling from high-dimensional probability distributions such as those encountered in modern bioinformatic applications [46], [47], [48]. In this section, we describe a Metropolis-Hastings sampling scheme for updating the state variables  $G$ ,  $\theta$ , and  $\sigma$ . Our main interest is in inference for the posterior  $P(G|D)$ . We take the approach of estimating the target distribution

$$P(G|D, \theta, \sigma) \propto P(D|G)P(G|\theta, \sigma)\pi(\theta)\pi(\sigma) \tag{6}$$

with  $\pi(\sigma) = 1/p!$  and then marginalize over  $\theta$  and  $\sigma$  to obtain  $P(G|D)$ . In the process, any  $\theta_m$  ( $m = 1, \dots, M$ ) or  $\sigma_i$  ( $i = 1, \dots, p$ ) can be estimated from a histogram of values, constructed from an MCMC chain. While methodology for sampling from  $P(G|D, \theta) \propto P(D|G)P(G|\theta)\pi(\theta)$  is well-established for GGMs [43], [42], [25], the concept of including  $\sigma$  as a state space variable is new to our work. In principle, it is possible to marginalize over all permutations of  $\sigma$  at each step, that is,  $P(G|\theta) = (1/p!) \sum_{\sigma} P(G|\theta, \sigma)$ . This approach, however, quickly becomes unfeasible as the number of nodes becomes large. What is more, very few assignments for  $\sigma$  actually capture scale-free network structure, making the marginalization difficult to estimate by random sampling. Instead, we include  $\sigma$  in the MCMC directly. We describe a Metropolis-Hastings sampler for  $\sigma$  below, and provide an implementation in C computer code for decomposable GGMs, built largely on the work of [25].

### Metropolis-Hastings Sampler

**Updating  $G$ .** The space of decomposable graphs can be traversed by adding or deleting a single edge in the transition from a current network,  $G$ , to a proposed network,  $G'$  [49]. In an arrangement of this sort,  $G$  and  $G'$  will have nearly identical maximal cliques, leading to extensive cancellation in the likelihood ratio  $P(D|G)/P(D|G')$  [42]. This coupled with the closed form expressions for (2) in the decomposable case, results in

considerable computational savings in comparison with the same computations for non-decomposable models. However, in the transition from  $G$  to  $G'$ , special care is required to preserve decomposability. To that end, a theorem of [43] provides easily verifiable, necessary and sufficient conditions to determine whether or not a network is decomposable. In their implementation [25], a transition is accomplished by first deciding to either add or delete an edge to  $G$  by the flip of a coin. Next the appropriate move is made at random to obtain  $G'$  as shown in Figure 2. If  $G'$  happens to be non-decomposable, then it is rejected outright.

**Updating  $\theta$ .** Each hyperparameter  $\theta_m$  ( $m = 1, \dots, M$ ) is updated as follows: select a value for  $\theta_m$  uniformly from  $(\theta_m - \epsilon, \theta_m + \epsilon)$  for a given step size  $\epsilon$ , rejecting when  $\theta_m$  falls outside its domain  $(\theta_{mmin}, \theta_{mmax})$ .

**Updating  $\sigma$ .** In order to obtain  $\sigma'$  we select an integer  $h \in \{1, \dots, p-1\}$  at random, find nodes  $v_i$  and  $v_j$  such that  $\sigma_i = h$  and  $\sigma_j = h+1$ , and then exchange the values of  $\sigma_i$  and  $\sigma_j$ ; see Figure 2.

### Network and Parameter Estimation

**Estimating  $G$ .** An MCMC sample of the posterior (1) becomes increasingly threadbare as the number of variables grows, so much so that the frequency of a network in a chain is an inadequate approximation to its true probability, even for problems of moderate dimension. So too for the maximum posterior network — the single most probable network in a chain — unless its probability mass dominates a possibly multi-modal landscape, comprising a near-infinity of alternative models, its status as a representative estimator is questionable [50]. This is even more important in our implementation, as we carry the model parameters through the computation. Alternatively, a more representative estimator can be pursued by exploiting marginal probabilities of edge inclusion, which do reflect posterior density. We took our estimated network to be the network of all edges in the sample with marginal probability greater than  $c$ , which we denote by  $\hat{G}_\pi^c$ ; the subscript  $\pi$  denotes the structure prior.

**Estimating  $\theta$ .** Let  $\theta_m^{(r)}$  denote the  $r$ 'th value of  $\theta_m$  ( $m = 1, \dots, M$ ) in an MCMC chain of length  $R$ .  $\theta_m$  is estimated by averaging over the values in an MCMC sample so that  $\hat{\theta}_m = \sum_{r=1}^R \theta_m^{(r)}$ .

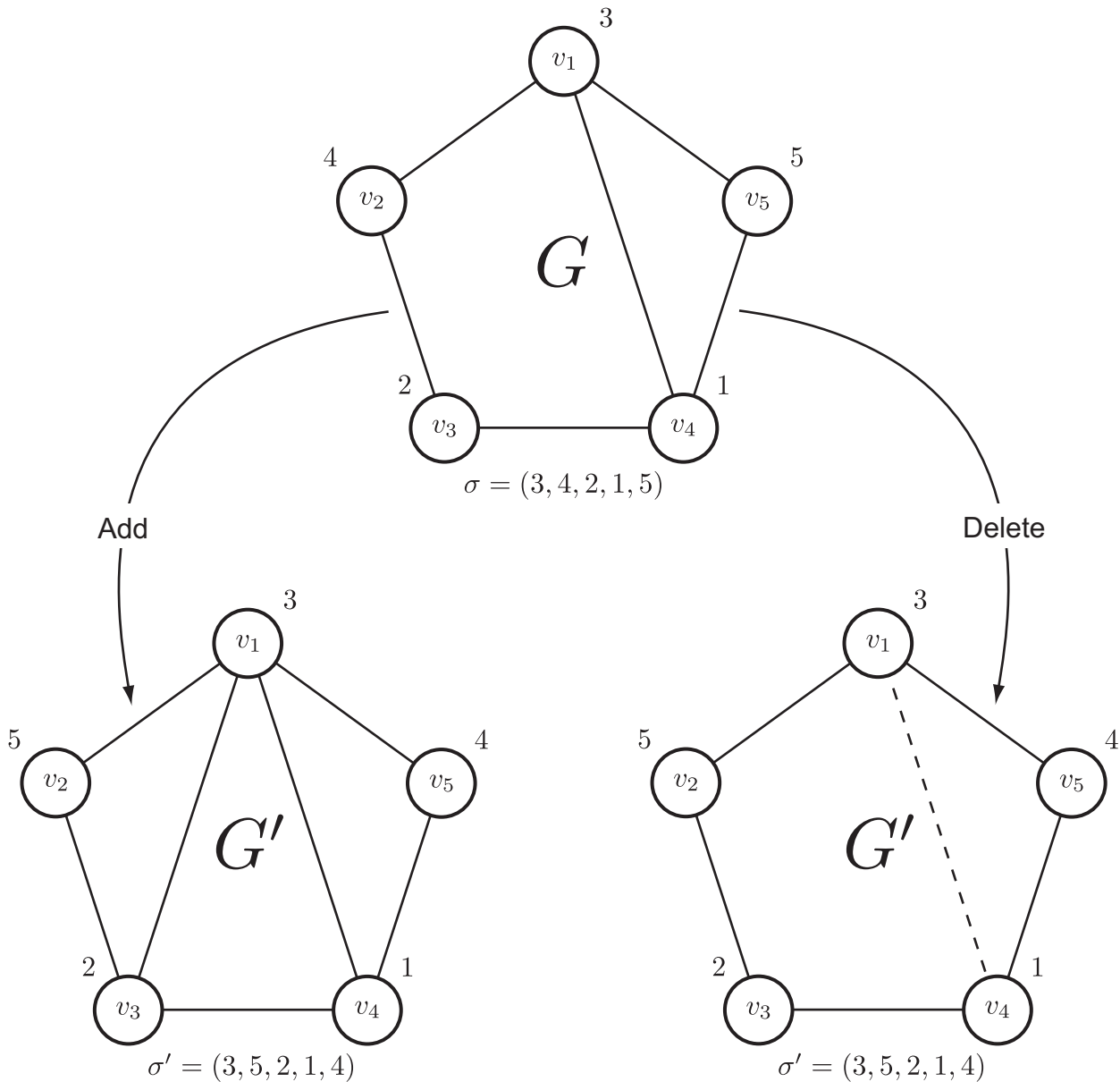
## Results

### Simulation Design

We carried out a simulation study in order to evaluate the relative performance of the random and scale-free structure priors. In our experiments, we generated trees invested with a variety degree distributions that can be thought of as falling along a spectrum ranging from binomial to scale-free on through to more extreme heavy-tail forms, called crumple trees, culminating finally with a star tree. For each tree, we generated multivariate Gaussian data under the assumption that a tree represents the true underlying conditional independence structure of a GGM. We then ran our Metropolis-Hastings sampler for both structure priors in an effort to recover each true tree from the data.

**Data generation.** In order to simulate trees we more or less relied on the stochastic algorithm of [51]. Their approach rests on specifying a formula for the degree distribution,  $P(k)$ , for a  $p$ -node connected tree. Then, roughly speaking, they use MCMC to draw a tree that is maximally random under  $P(k)$ .

The reason for restricting our simulation to trees is that data satisfying their implied conditional independence structures can be generated by a simple iterative procedure. With this end in mind,



**Figure 2. An example of the Metropolis-Hastings transition step.** The current network,  $G$ , with node labeling  $\sigma = (3, 4, 2, 1, 5)$  is updated to a proposed network,  $G'$ , by adding/deleting a single edge to/from  $G$  at random. This picture shows two possible ways for  $G$  to be updated. In one instance, a new edge is added between  $v_1$  and  $v_3$  to obtain  $G'$ , while the other  $G'$  is obtained by deleting the edge between  $v_1$  and  $v_4$ . As for the proposed node labeling, an integer  $h$ , in this example  $h=4$ , is selected randomly from  $\{1, 2, 3, 4\}$ . From there the node labels  $\sigma_2=h=4$  and  $\sigma_5=h+1=5$  are swapped to get  $\sigma' = (3, 5, 2, 1, 4)$ . doi:10.1371/journal.pone.0013580.g002

it is convenient to imagine the edges as being directed according to index so that an edge from  $v_i$  to  $v_j$  implies that  $i < j$ . The procedure begins with simulating  $X_1$ , which is identified with node  $v_1$ , as a standard normal random variable,  $Z_1$ . Next, any  $X_j$  corresponding to a child of  $v_1$  is simulated as  $X_j = (X_1 + Z_j) / \sqrt{2}$ . The step  $X_j = (X_i + Z_j) / \sqrt{2}$  is then repeated from parent  $v_i$  to child  $v_j$  until all nodes have been reached. The scaling factor,  $\sqrt{2}$ , ensures that each  $X_i \in X$  has unit variance.

**Performance measures.** Let  $TP$  (true positive) denote the number of edges correctly identified by the estimated network with  $FP$  (false positive),  $FN$  (false negative), and  $TN$  (true negative) defined similarly. The positive predictive value,  $PPV = TP / (TP + FP)$ , and the sensitivity,  $Se = TP / (TP + FN)$ ,

are reported for each estimated network. While it is often customary to include specificity,  $TN / (FP + TN)$ , along with  $PPV$  and  $Se$ , its conspicuous absence here is for good reason. Since GANs are sparse,  $TN$  is sure to be very large in comparison to  $FP$ . As a result, even a moderate change to  $FP$  will have little influence on the specificity, making this an unsuitable measure of performance.

**Simulated Example**

This section serves as a prelude to an extensive simulation study, illustrating our methodology by means of a simple example. Specifically, we set  $p=250$  and generated a binomial tree and a scale-free tree with  $\gamma=2.3$ , and then simulated an  $n=75$



observation dataset from each. In each case, we attempted to recover the true tree from the scaled dataset using our Metropolis-Hastings sampler implemented with 1)  $\pi_r(G)$ , and 2)  $\pi_{sf}(G)$ .

For each chain, the Metropolis-Hastings sampler was run for  $10^7$  steps after a burn-in of  $10^6$ , starting from the empty network and identity node labeling. The value for the step size,  $\epsilon$ , required for updating the hyperparameters was set to 0.05 with  $\mu_{min}=0.01$ . As for the hyper-inverse Wishart parameters, we choose  $\delta=3$  which fixes  $\tau$  at 4 since the data was standardized. The values of the hyperparameters were recorded at every 100<sup>th</sup> step after burn-in. The runtime for the Metropolis-Hastings sampler with  $\pi_r(G)$  on a dual 1.8 GHz PowerPC G5 processor was 5.92hrs for the binomial tree and 5.67hrs for the scale-free tree. The corresponding runtimes with  $\pi_{sf}(G)$  were 15.20hrs and 12.93hrs.

The results of the case study are shown in Table 1. In this experiment, our expectation that  $\pi_{sf}(G)$  will recover the scale-free tree more accurately than  $\pi_r(G)$  is confirmed. It should also be noted that  $\pi_{sf}(G)$  was able to recover a reasonable value for the scale-free exponent, too. Not to mention that it recovered the binomial tree on par with  $\pi_r(G)$ , thereby allaying the potential drawback that it would infer a heavy-tailed network, even from binomial data. Remember, this can be explained by the rather large value of  $\hat{\gamma}$ . Recall that when  $\gamma$  is large,  $\pi_{sf}(G)$  actually approximates  $\pi_r(G)$ . And although it may seem odd that  $\pi_{sf}(G)$  fared slightly better on the binomial tree, the disparity falls within the boundaries of sampling variation. More precisely, we ran the Metropolis-Hastings sampler 10 times for each structure prior, starting each run from a different random seed, and found that the standard deviation of the sensitivity was 0.02 in each case. Finally, we ran the uniform structure prior on both trees, but decided against including the results in Table 1 due to very poor performance.

## Extended Simulation

Table 2 contains the results of our main simulation. In the previous section, we focused on two particular trees: one binomial, the other scale-free. This time we generated 25 trees ( $p=250$ ) under each model listed in the table together with accompanying datasets of  $n=75$  observations. The models listed as scale-free, not including the BA model, and the crumpled one were generated from a two-parameter family of distributions [51]. The parameter setting for generating the crumpled trees was  $\alpha=8$  and  $\gamma=2$ . The simulation settings used for each MCMC run are identical to those of the case study. Finally, the values of  $PPV$ ,  $Se$ , and  $\hat{\gamma}$  reported in the table are averaged over the 25 chains. The simulation was run on the supercomputer, Tsubame [52]. The system has a total of 639 Sun Fire  $\times 4600$  nodes. Each node has 8 AMD Opteron Dual Core model 880, 2.4GHz cpus.

Just as with the simulated example,  $\pi_{sf}(G)$  recovers the binomial trees equally as well as  $\pi_r(G)$ . In fact, the  $PPV$  agreed

to two decimal places, while the  $Se$  was actually a little higher under the scale-free structure prior. This slight discrepancy can be accounted for by noting that the standard deviation of  $Se$  was 0.04 for both priors. Also as expected, the more heavy-tailed the underlying trees become, the more  $\pi_{sf}(G)$  outperforms  $\pi_r(G)$ . The difference becomes huge in the extreme case of a star tree. Moreover,  $\pi_{sf}(G)$  demonstrated the ability to roughly recover the scale-free exponent of the underlying tree.

## Real Data Example

We demonstrate our methodology on a subset of the gene expression data from a breast cancer study by [40] that was originally analyzed in [25]. The dataset (Dataset S1) consists of expression profiles for  $p=150$  genes related to the estrogen receptor gene ESR1 (also known as ER-alpha) derived from  $n=49$  tumor samples. This gene is an estrogen-activated transcription factor key to the proliferation of cancerous cells that is found to be overexpressed in luminal type A and B breast cancers. The overall level of ESR1 expression is higher in type A than in type B with the former correlating with better prognosis [53].

The Metropolis-Hastings sampler was run on the standardized data with both the random structure prior and its scale-free counterpart, yielding the corresponding GANs  $\hat{G}_{\pi_r}^c$  and  $\hat{G}_{\pi_{sf}}^c$ . For comparison's sake, the edge inclusion threshold,  $c$ , was tuned for each run so that the resulting GAN comprised exactly 150 edges; the value of  $c$  is 0.370 for  $\pi_r(G)$  and 0.747 for  $\pi_{sf}(G)$ . In both cases, the Metropolis-Hastings sampler was started from the empty network with identity node labeling and  $11 \times 10^6$  iterations were run with the first  $10^6$  discarded as burn-in. The hyperparameter assignments were identical to those of the simulated examples. The runtime on a dual 1.8 GHz PowerPC G5 processor was 2.70hrs with  $\pi_r(G)$  and 11.93hrs with  $\pi_{sf}(G)$ .

At this stage, comparing the performance of the scale-free structure prior in a broader context is of key importance. To this end, we used the software packages ARACNE [54], [55] and BANJO [23] to analyze the gene expression data as well. ARACNE constructs a relevance network based on estimated mutual information between all pairs genes, but there is a twist. After a relevance network is inferred by connecting any pair of genes with mutual information greater than a certain cutoff value, some edges suspected to represent indirect interactions are eliminated using the data processing inequality principle. We chose the cutoff value to be 0.2735 so that the number of estimated edges was 150, while all other program arguments were set at their default values. The code itself was run in a matter of minutes. BANJO, on the other hand, constructs a Bayesian network from discrete data using a heuristic search strategy to explore the space of  $p$ -node, directed networks without cycles. Each network happened upon in the search is ranked using a Bayesian Dirichlet equivalence scoring metric. We discretized the

**Table 1.** Case study.

Topology	True $\gamma$	Random Structure Prior		Scale-Free Structure Prior		$\hat{\gamma} \pm s.d.$
		$PPV$	$Se$	$PPV$	$Se$	
Binomial	—	0.96	0.59	0.96	0.60	$19.74 \pm 3.79$
Scale-Free	2.3	0.88	0.54	0.90	0.79	$2.18 \pm 0.04$

Summary of the networks estimated using  $\pi_r(G)$  (left) and  $\pi_{sf}(G)$  (right) when the true tree topology is binomial in the one case and scale-free with  $\gamma=2.3$  in the other.  $PPV$  is the positive predictive value and  $Se$  the sensitivity, which is computed in respect to the the number of edges correctly identified by the estimated network. The value  $\hat{\gamma}$  is the estimated scale-free exponent, obtained by averaging over every 100<sup>th</sup> value in the MCMC chain.

doi:10.1371/journal.pone.0013580.t001

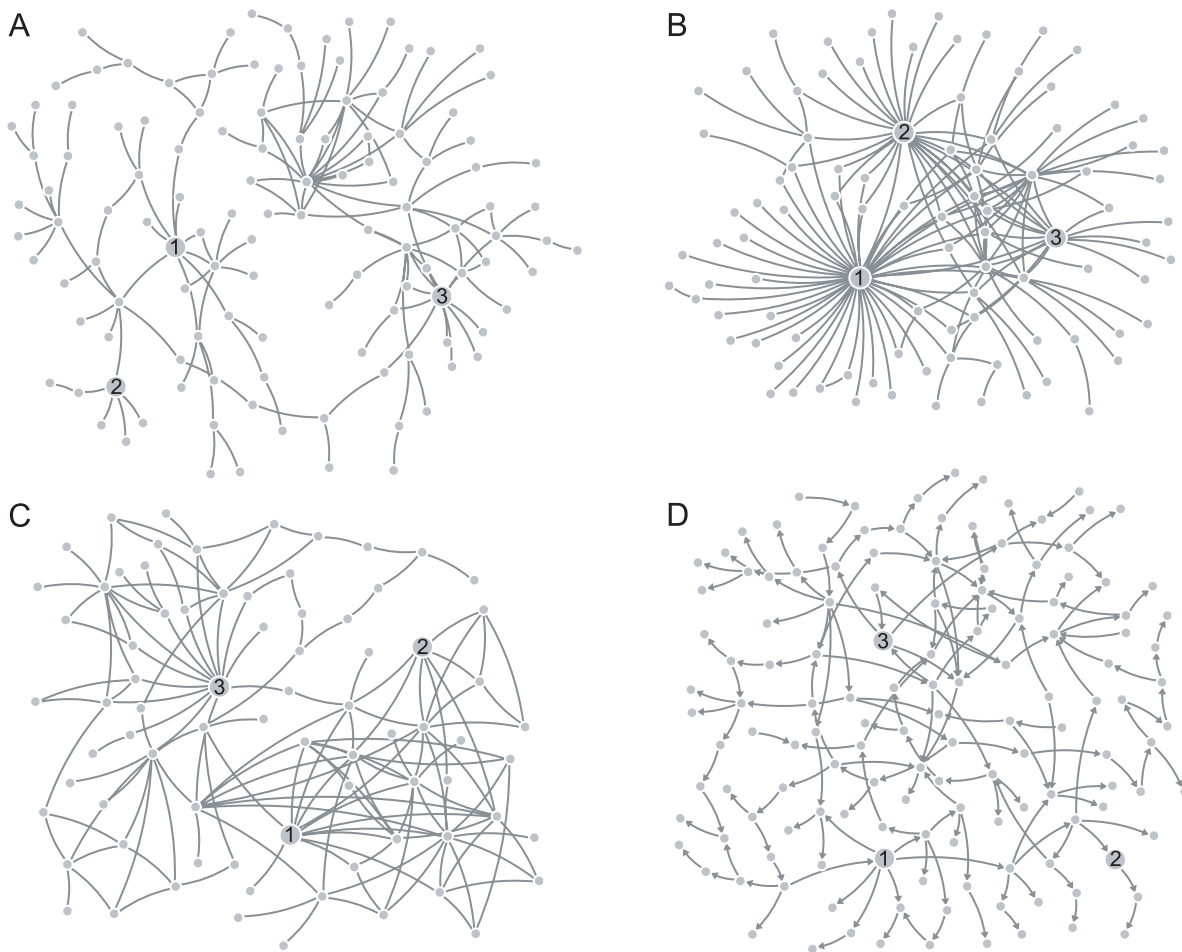
**Table 2.** Full simulation.

Topology	True $\gamma$	Random Structure Prior		Scale-Free Structure Prior		$\hat{\gamma}$
		PPV	Se	PPV	Se	
Binomial	—	0.96	0.59	0.96	0.60	18.70
Scale-Free (BA)	3.0	0.95	0.58	0.94	0.65	2.49
Scale-Free	2.5	0.93	0.56	0.93	0.62	2.35
Scale-Free	2.3	0.86	0.49	0.90	0.71	2.18
Crumple	—	0.76	0.38	0.90	0.80	2.10
Star	—	0.63	0.30	0.90	0.86	2.10

Summary of the networks estimated using  $\pi_r(G)$  (left) and  $\pi_{sf}(G)$  (right) for a variety of topologies. A total of 25 trees of were generated for each kind of topology; each has  $p = 250$  nodes with an accompanying  $n = 75$  observation dataset. PPV is the positive predictive value and Se the sensitivity, which are computed according the the number of edges correctly identified by the estimated network.  $\hat{\gamma}$  is the estimated scale-free exponent. The values in the table are averaged over the 25 MCMC runs. doi:10.1371/journal.pone.0013580.t002

data into three categories and limited the number of parents any given node may have to 10. Once again, all other program arguments were set at their default values. The estimated network, which was found to have 162 directed edges, was the highest scoring network after running BANJO for 1hr.

Figures 3(A) and (B) show the GANs estimated with  $\hat{G}_{\pi_r}^c$  and  $\hat{G}_{\pi_{sf}}^c$ . The latter exhibits clear hubs, supporting the view that a gene regulatory network consists of a small minority of hub genes with the vast majority of genes engaged in a small number of interactions. By contrast, the topology of  $\hat{G}_{\pi_r}^c$  is relatively



**Figure 3. Estimated networks for the breast cancer expression dataset.** (A) The gene association network estimated with random structure prior,  $\pi_r(G)$ , and (B) the gene association network estimated with the scale-free structure prior,  $\pi_{sf}(G)$ . (C) The relevance network estimated with ARACNE, and (D) the Bayesian network estimated with BANJO. The labeled nodes are the largest hubs as identified by  $\pi_{sf}(G)$  (1:FOXA1, 2:SLC39A6, 3:E2F3). For ease of visualization, only the largest connected subnetworks are displayed. doi:10.1371/journal.pone.0013580.g003



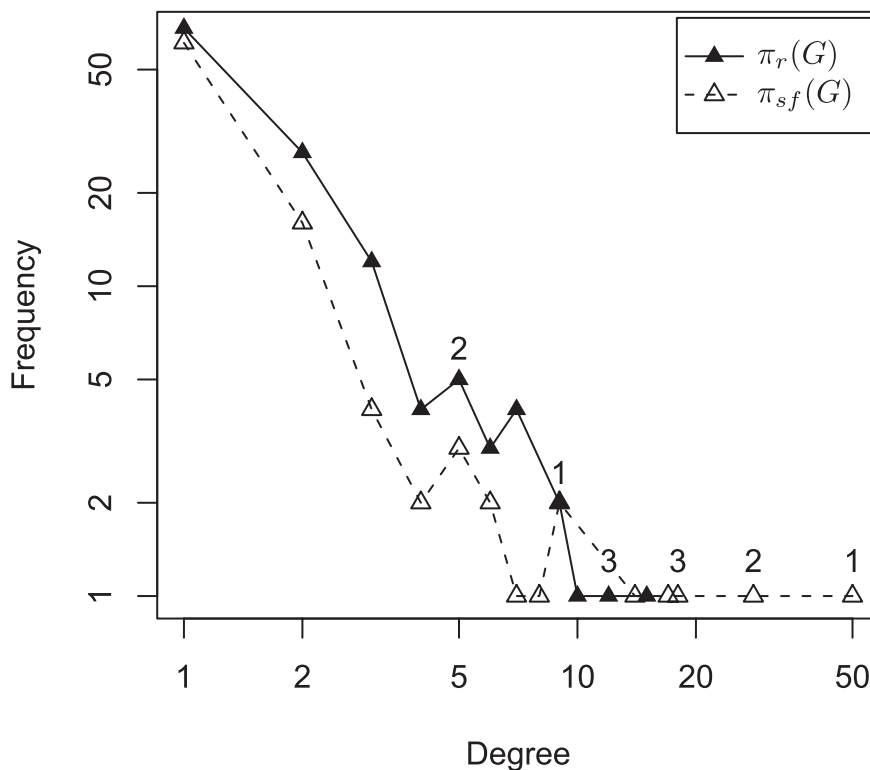
decentralized with no single gene dominating the network. Additionally, the estimated value of exponent  $\gamma$  in the static model was 2.28, in line with findings in the literature for gene regulatory networks [31]. Turning now to Figures 3(C) and (D), it is interesting to see that the topology of the relevance network echoes that of the GAN inferred using the scale-free structure prior. The same can be said for the Bayesian network and the random structure prior GAN. Of course, a more exquisite experimental technique is the only sure-fire way to validate the individual regulatory interactions suggested by these graphical models. These results, however, are telltale in one respect. In a study comparing different reconstruction methods on simulated data [56], it was reported that BANJO performs well only when  $n \gg p$ , while ARACNE shows good performance even when  $n \ll p$ .

The topological dissimilarity between the two GANs is again made evident by a visual inspection of their degree distributions, plotted in Figure 4. The most abundantly connected node in  $\hat{G}_{\pi_r}^c$  has degree 15, whereas  $\hat{G}_{\pi_{sf}}^c$  contains four nodes with degree exceeding this value; the largest hubs correspond to the genes FOXA1 (HNF-3A), SLC39A6 (LIV-1), and E2F3 (KIAA0075) and have degree 50, 28, and 18, respectively. The main hub FOXA1 is a forkhead box family transcription factor that is necessary for optimum expression of roughly half of all ESR1-regulated genes [57]. In a recent study [58], it was found that FOXA1 is expressed predominantly in luminal type A carcinomas, making it a potential marker of good prognosis. Previously unrecognized as a hub, SLC39A6 functions as a zinc transporter, and was identified in [59] to be highly expressed in ESR1-positive tumors as well as showing a highly significant association with the spread of breast cancer to the lymph nodes. Meanwhile, E2F3 is a transcription factor that has been shown to regulate numerous genes involved in cell cycle progression [60].

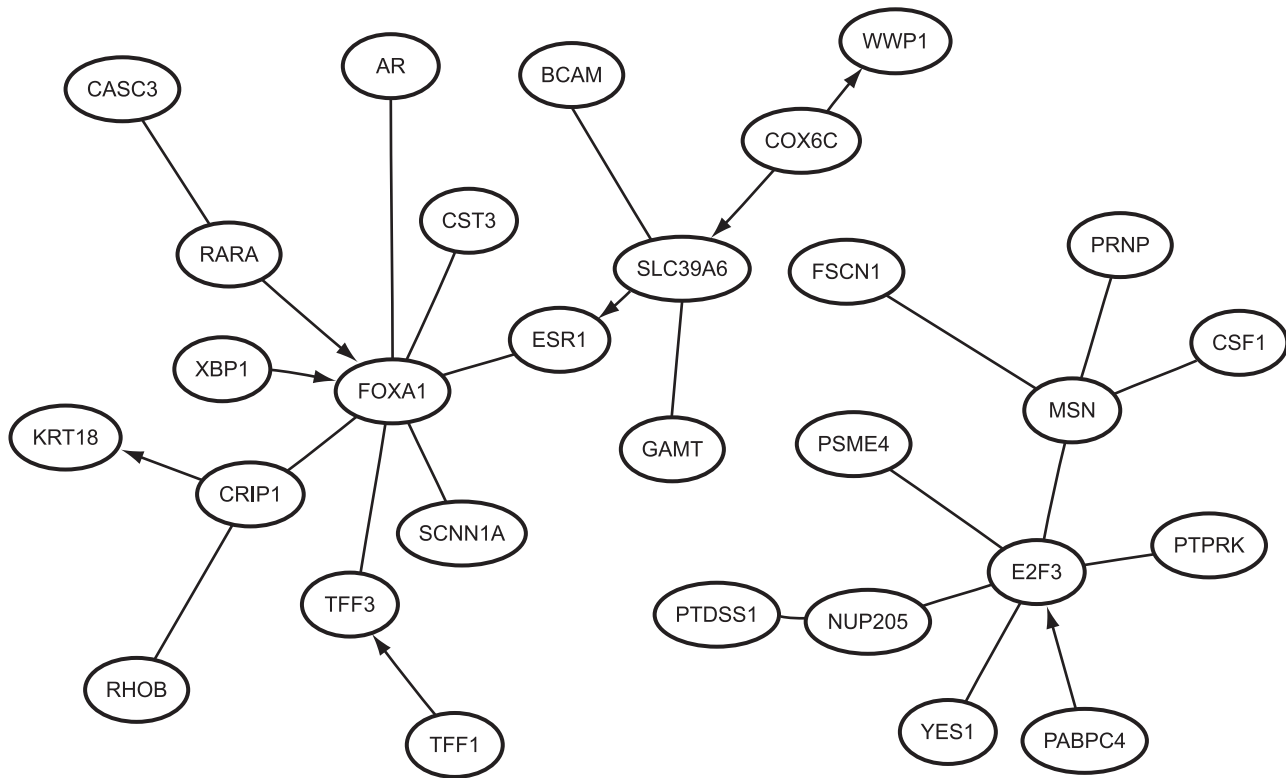
Finally, both GANs agreed with the relevance network on some established regulatory interactions as can be seen in Figure 5. For instance, FOXA1 is connected to AR (androgen receptor), which is known to regulate estrogen receptor expression [61]. FOXA1 has also been shown to play a direct role in the transcription of the TFF1 (pS2) gene [62], and our work agrees with [24] on the role of TFF3 (ITF) as an intermediary. By contrast, the Bayesian network agreed on very few of these interactions. Part of the explanation is likely to rest in using the maximum posterior network as the estimated network. As we drew attention to in the section Network and Parameter Estimation, a single network of high posterior probability may be a less representative estimator than an network consisting of edges that occur with high frequency in an MCMC chain. Another possible contributing factor is that the number of observations was insufficient for BANJO, but what is also unclear is the extent to which discretizing the expression data affected the quality of the inference.

## Discussion

The main purpose of this paper has been to introduce a scale-free structure prior,  $\pi_{sf}(\mathcal{G})$ , for graphical models with a view toward the inference of large-scale GANs from datasets consisting of few observations,  $n$ , for a comparatively large number of variables,  $p$ . It is important to point out that the true network need not follow a power-law in order for the scale-free prior to be applicable; rather,  $\pi_{sf}(\mathcal{G})$  is a convenient distribution that can account for heavy-tailed degree distributions — a crucial limitation of the random structure prior. That said, we have shown in simulated examples that  $\pi_{sf}(\mathcal{G})$  performs markedly better than the random structure prior at recovering networks characterized by heavy-tailed degree distributions. What is more,



**Figure 4. Degree distributions of the gene association networks.** The labels 1, 2, and 3 indicate the locations of the genes FOXA1, SLC39A6, and E2F3, respectively. The plots are on a log – log scale. doi:10.1371/journal.pone.0013580.g004



**Figure 5. Gene interactions identified by all methods.** A subnetwork of the gene interactions involved with the estrogen receptor gene, ESR1, that were commonly identified by the random structure prior,  $\pi_r(G)$ , the scale-free structure prior,  $\pi_{sf}(G)$ , and the relevance network (estimated with ARACNE). The directed edges indicate the gene interactions on which the Bayesian network (estimated with BANJO) agreed with the other three methods.

doi:10.1371/journal.pone.0013580.g005

$\pi_{sf}(G)$  proved versatile enough to recover random networks on par with  $\pi_r(G)$  itself. Above all, our analysis of the breast cancer expression data illustrates the practical value of the scale-free structure prior as an instrument to aid in the identification of candidate hub genes with the potential to direct the hypotheses of molecular biologists, and thus drive future experiments.

A node labeling  $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_p)$ , that is, a permutation of the integers  $1, 2, \dots, p$  applied to the nodes of  $G$  so that each  $v_i \in V$  is represented by the integer  $\sigma_i$ , is an essential prerequisite for any MCMC implementation of the scale-free structure prior. The reason is that the scale-free network model underlying  $\pi_{sf}(G)$ , or any other scale-free network model for that matter, is so elaborate that the nodes are not interchangeable in regard to computing the probability of  $G$ . And, although easily overshadowed by  $\pi_{sf}(G)$  itself, our new Metropolis-Hastings sampler for  $\sigma$  is an innovative contribution in its own right. Our sampler uses a simple pair swapping strategy for updating  $\sigma$ , and one future topic of research is to investigate the comparative performance of more ingenious update schemes. More research is also required in order to assess how accurately  $\sigma$  can be estimated.

We take pains to point out that while our implementation is for GGMs, the methodology described here applies to graphical models more generally. For instance,  $\pi_{sf}(G)$  could be applied crudely to Bayesian network inference by simply ignoring edge directionality, or else the underlying static model could be modified to have directed edges. The latter approach raises an interesting consideration: in gene regulatory networks, according to the prevailing wisdom [31], it is actually only the out-degree distribution that follows a power-law. By contrast, the in-degree of

a node is usually small and its distribution is better approximated by a sort of restricted exponential function. While this distinction gets blurred when inference is conducted with undirected graphical models, Bayesian networks provide an obvious incentive for taking it into account. Indeed, Bayesian networks may prove to be a more promising area of application because they currently are able to handle much larger networks than GGMs [63].

Although the static model is not biologically motivated, it is a defensible choice as an underlying model for  $\pi_{sf}(G)$  on the grounds that it is a simple model with the potential to describe any network topology; not to mention that it includes the ER model as a limiting case. But there is more, implementing a structure prior based on a growing network model poses some added difficulties because not only will the probability of a network depend on the choice of seed network, but evaluating  $P(G|\theta, \sigma)$  will result in a greater expenditure of computational resources as the edge inclusion probabilities depend on the order in which they were added to the network.

All the same, we implemented two other scale-free structure priors based on growing models; one on the Poisson-growth, preferential attachment model [64], and another on the biologically meaningful duplication model. In the former case, we were able to get away with using a single node as the seed network, and we found that while this prior recovered heavy-tailed networks as well as  $\pi_{sf}(G)$ , yet it understandably struggled to accurately recover random networks. Meanwhile, the duplication model based structure prior was highly sensitive to the choice of seed network in addition to being unstable due to the complexity of the model. One future avenue of research is to adapt these

models, or the MCMC implementation, to be more applicable for use as a prior distributions. The primary motivation for doing so is that the model parameters have biological meaning, and their estimation could prove of independent interest.

The estimation of network model parameters has been an incidental aspect of our work; however, it is related to the quite different problem of fitting network models to known biological networks. Likelihood and likelihood-free methods have been developed [65], [66] in order to fit a hybrid preferential attachment/duplication and divergence model to some protein-protein interaction networks, obtaining estimates of the model parameters. These methodologies assume that the ordering of the nodes in time, that is  $\sigma$ , is known, but in most cases this information is unknown. In the future, our Metropolis-Hastings sampler could very well be applied to this problem.

Software is available from the corresponding author upon request.

## References

- Husmeier D (2003) Sensitivity and specificity of inferring genetic regulatory interactions from microarray experiments with dynamic bayesian networks. *Bioinformatics* 19: 2271–2282.
- West M (2003) Bayesian data regression models in the “large  $p$ , small  $n$ ” paradigm Oxford University Press. pp 723–732.
- Schlitt T, Brazma A (2007) Current approaches to gene regulatory network modelling. *BMC Bioinformatics* 8: S9+.
- Whittaker J (1990) Graphical models in applied multivariate statistics. New York: John Wiley & Sons.
- Lauritzen SL (1996) Graphical models. Oxford Statistical Science Series. New York: Oxford University Press.
- Edwards D (2000) Introduction to graphical modelling. New York: Springer-Verlag.
- Grzegorzczak M, Husmeier D, Werhli AV (2008) Reverse engineering gene regulatory networks with various machine learning methods. In: Emmert-Streib F, Dehmer M, eds. *Analysis of Microarray Data*. Weinheim: Wiley-VCH. pp 101–142.
- Markowitz F, Spang R (2007) Inferring cellular networks - a review. *BMC Bioinformatics* 8: S5+.
- Butte AJ, Tamayo P, Slonim D, Golub TR, Kohane IS (2000) Discovering functional relationships between rna expression and chemotherapeutic susceptibility using relevance networks. *Proc Natl Acad Sci U S A* 97: 12182–12186.
- Schäfer J, Strimmer K (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology* 4.
- Wille A, Bhlmann P (2006) Low-order conditional independence graphs for inferring genetic networks. *Statistical Applications in Genetics and Molecular Biology* 5: 1.
- Castelo R, Roverato A, Chickering M (2006) A robust procedure for gaussian graphical model search from microarray data with  $p$  larger than  $n$ . *Journal of Machine Learning Research* 7: 2006.
- Magwene PM, Kim J (2004) Estimating genomic coexpression networks using first-order conditional independence. *Genome Biology* 5: R100.
- Pe’er D, Nachman I, Lital M, Friedman N (2000) Using bayesian networks to analyze expression data. *Journal of Computational Biology* 7: 601–620.
- Imoto S, Higuchi T, Goto T, Tashiro K, Kuhara S, et al. (2003) Combining microarrays and biological knowledge for estimating gene networks via bayesian networks. In *CSB ’03: Proceedings of the IEEE Computer Society conference on Bioinformatics*. Washington, DC, USA: IEEE Computer Society. pp 104–113.
- Tamada Y, Kim S, Bannai H, Imoto S, Tashiro K, et al. (2003) Estimating gene networks from gene expression data by combining bayesian network model with promoter element detection. *Bioinformatics* 19 Suppl 2: 11227–11236.
- Barnard A, Hartemink AJ (2005) Informative structure priors: joint learning of dynamic regulatory networks from multiple types of data. In: *PSB ’05: Pacific Symposium on Biocomputing World Scientific*. pp 459–470.
- Werhli AV, Husmeier D (2007) Reconstructing gene regulatory networks with bayesian networks by combining expression data with multiple sources of prior knowledge. *Statistical applications in genetics and molecular biology* 6.
- Pe’er D, Regev A, Tanay A (2002) Minreg: inferring an active regulator set. *Bioinformatics* 18 Suppl 1: S258–S267.
- Segal E, Shapira M, Regev A, Pe’er D, Botstein D, et al. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature Genetics* 34: 166–176.
- Schäfer J, Strimmer K (2005) An empirical bayes approach to inferring large-scale gene association networks. *Bioinformatics* 21: 754–764.
- Meinshausen N, Bhlmann P (2006) High dimensional graphs and variable selection with the lasso. *Annals of Statistics* 34: 1436–1462.
- Yu J, Smith VA, Wang PP, Hartemink AJ, Jarvis ED (2004) Advances to bayesian network inference for generating causal networks from observational biological data. *Bioinformatics (Oxford, England)* 20: 3594–3603.
- Dobra A, Hans C, Jones B, Nevins JR, West M (2004) Sparse graphical models for exploring gene expression data. *Journal of Multivariate Analysis* 90: 196–212.
- Jones B, Carvalho CM, Dobra A, Hans C, Carter C, et al. (2005) Experiments in stochastic computation for high-dimensional graphical models. *Statistical Science* 20: 388–400.
- Erdős P, Rényi A (1959) On random graphs. *Publicationes Mathematicae* 6: 290–297.
- Erdős P, Rényi A (1960) On the evolution of random graphs. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences* 5: 17–61.
- Albert R, Barabási AL (2002) Statistical mechanics of complex networks. *Reviews of Modern Physics* 74.
- Dorogovtsev SN, Mendes JFF (2002) Evolution of networks. *Advances in Physics* 51: 1070–1187.
- Newman MEJ (2003) The structure and function of complex networks. *Siam Review* 45: 167–256.
- Albert R (2005) Scale-free networks in cell biology. *J Cell Sci* 118: 4947–4957.
- Farkas I, Jeong H, Vicsek T, Barabasi, Oltvai ZN (2003) The topology of the transcription regulatory network in the yeast, *saccharomyces cerevisiae*. *Physica A: Statistical Mechanics and its Applications* 318: 601–612.
- Kim SK, Lund J, Kiraly M, Duke K, Jiang M, et al. (2001) A gene expression map for *caenorhabditis elegans*. *Science* 293: 2087–2092.
- Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang D (2006) Complex networks: Structure and dynamics. *Physics Reports* 424: 175–308.
- Barabasi AL, Albert R (1999) Emergence of scaling in random networks. *Science* 286: 509–512.
- Ohno S (1970) Evolution by gene duplication G. Allen and Unwin; Springer.
- Solé R, Pastor-Satorras R, Smith E, Kepler T (2002) A model of large-scale proteome evolution. *Advances in Complex Systems* 5: 43–54.
- Pastor-Satorras R, Smith E, Solé RV (2003) Evolving protein interaction networks through gene duplication. *J Theor Biol* 222: 199–210.
- Goh KI, Kahng B, Kim D (2001) Universal behavior of load distribution in scale-free networks. *Phys Rev Lett* 87.
- West M, Blanchette C, Dressman H, Huang E, Ishida S, et al. (2001) Predicting the clinical status of human breast cancer by using gene expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 98: 11462–11467.
- Dempster AP (1972) Covariance selection. *Biometrics* 28: 95–108.
- Wong K, Carter C, Kohn R (2003) Efficient estimation of covariance selection models. *Biometrika* 90: 809–830.
- Guidici P, Green PJ (1999) Decomposable graphical model determination. *Biometrika* 86: 785–801.
- Adamic LA, Huberman BA (2002) Zipf’s law and the internet. *Glottometrics* 3: 143–150.
- Lee DS, Goh KI, Khang B, Kim D (2005) Scale-free random graphs and potts model. *Pramana Journal of Physics* 64: 1149–1159.
- Rannala B, Yang Z (1996) Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *J Mol Evol* 43: 304–311.
- Yang Z, Rannala B (1997) Bayesian phylogenetic inference using dna sequences: a markov chain monte carlo method. *Mol Biol Evol* 14: 717–724.
- Ronquist F, Huelsenbeck JP (2003) Mrbayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
- Frydenberg M, Lauritzen SL (1989) Decomposition of maximum likelihood in mixed interaction models. *Biometrika* 76: 539–555.

## Supporting Information

**Dataset S1** This file contains the gene expression data that we analyzed in our paper.

Found at: doi:10.1371/journal.pone.0013580.s001 (0.05 MB TXT)

## Acknowledgments

The authors kindly thank Mariko Okada-Hatakeyama and Seiya Imoto for fruitful discussions as well as Beatrix Jones for fielding our various inquiries concerning the gene expression dataset. Also special thanks go to the editor, the two anonymous referees, and Brent Hagen for their most constructive comments and suggestions.

## Author Contributions

Conceived and designed the experiments: PS TK HS. Performed the experiments: PS. Analyzed the data: PS. Wrote the paper: PS HS.

50. Carvalho LE, Lawrence CE (2008) Centroid estimation in discrete high-dimensional spaces with applications in biology. *Proceedings of the National Academy of Sciences* 105: 3209–3214.
51. Burda Z, Correia JD, Krzywicki A (2001) Statistical ensemble of scale-free random graphs. *Phys Rev E* 64: 046118.
52. Matsuoka S (2008) *The road to TSUBAME and beyond*, Boca Raton, FL: Chapman & Hall/CRC, chapter 14. *Computational Science Series*. pp 289–310.
53. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America* 98: 10869–10874.
54. Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, et al. (2005) Reverse engineering of regulatory networks in human b cells. *Nature Genetics* 37: 382–390.
55. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, et al. (2006) Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC bioinformatics* 7 Suppl 1.
56. Bansal M, Belcastro V, Ambesi-Impiombato A, di Bernardo D (2007) How to infer gene networks from expression profiles. *Molecular Systems Biology* 3.
57. Carroll JS, Brown M (2006) Estrogen receptor target gene: an evolving concept. *Molecular Endocrinology* 20: 1707–1714.
58. Nakshatri H, Badve S (2007) Foxa1 as a therapeutic target for breast cancer. *Expert Opinions on Therapeutic Targets* 11: 507–514.
59. Taylor KM (2008) A distinct role in breast cancer for two liv-1 family zinc transporters. *Biochemical Society Transactions* 36 (Pt6): 1247–1251.
60. Giangrande P, Hallstrom T, Tunyaplin C, Calame K, Nevins J (2003) Identification of e-box factor tfe3 as a functional partner for the e2f3 transcription factor. *Molecular and Cellular Biology* 23: 3707–3720.
61. Sahlin L, Norstedt G, Eriksson H (1994) Androgen regulation of the insulin-like growth factor 1 and the estrogen receptor in rat uterus and liver. *Journal of Steroid Biology* 51: 57–66.
62. Beck S, Sommer P, dos Santos Silva E, Blin N, Gött P (1999) Hepatocyte nuclear factor 3 (winged helix domain) activates trefoil factor gene tff1 through a binding motif adjacent to the tata box. *DNA Cell Biology* 18: 157–164.
63. Hartemink AJ (2005) Reverse engineering gene regulatory networks. *Nature Biotechnology* 23: 554–555.
64. Sheridan P, Yagahara Y, Shimodaira H (2008) A preferential attachment model with poisson growth for scale-free networks. *Annals of the Institute of Statistical Mathematics* 60: 747–761.
65. Wiuf C, Brammier M, Hagberg O, Stumpf MP (2006) A likelihood approach to analysis of network data. *Proceedings of the National Academy of Sciences of the United States of America* 103: 7566–7570.
66. Ratmann O, Jørgensen O, Hinkley T, Stumpf M, Richardson S, et al. (2007) Using likelihood-free inference to compare evolutionary dynamics of the protein networks of *h. pylori* and *p. falciparum*. *PLoS Computational Biology* 3: 2266–2278.