On Scale-Free Prior Distributions and Their Applicability in Large-Scale Network Inference with Gaussian Graphical Models

Paul Sheridan, Takeshi Kamimura, and Hidetoshi Shimodaira

Tokyo Institute of Technology, Meguro-ku, Tokyo 152-8552, Japan
sherida6@is.titech.ac.jp

Abstract. This paper concerns the specification, and performance, of scale-free prior distributions with a view toward large-scale network inference from small-sample data sets. We devise three scale-free priors and implement them in the framework of Gaussian graphical models. Gaussian graphical models are used in gene network inference where high-throughput data describing a large number of variables with comparatively few samples are frequently analyzed by practitioners. And, although there is a consensus that many such networks are scale-free, the *modus operandi* is to assign a random network prior. Simulations demonstrate that the scale-free priors outperform the random network prior at recovering scale-free trees with degree exponents near 2, such as are characteristic of many real-world systems. On the other hand, the random network prior compares favorably at recovering scale-free trees characterized by larger degree exponents.

Keywords: Bayesian inference, complex networks, Gaussian graphical model, Markov chain Monte Carlo, prior distribution, scale-free, "small n, large p" problem, small-sample inference.

1 Motivation

Gaussian graphical models (GGMs) are commonly used to estimate a gene network from microarray data [17]. In this framework, p genes are represented by an undirected network $G = (V, E)$ where the node set $V = \{1, \ldots, p\}$ indexes the Gaussian random vector $X = (X_1, \ldots, X_p)$. A data matrix D of n microarray experiments is taken as a random sample from the multivariate Gaussian $X \sim N(\mu, \Sigma)$. The edge set E is defined by the conditional independence structure of X so that the edge $\{i, j\}$ is in E if, and only if, X_i and X_j are conditionally dependent given the remaining variables in X.

When Σ is nonsingular, conditional independence (a missing edge) between two variables X_i and X_j is equivalent to $\omega_{ij} = 0$ in the precision matrix $\Omega = \Sigma^{-1}$. Thus, model fitting over GGMs, known as *covariance selection*, amounts to identifying zero entries in Ω.

In classical GGM theory $n > p$ is necessary [3]; however, with genomic data it is frequently the case that $n \ll p$: the so called "small n, large p" problem.

J. Zhou (Ed.): Complex 2009, Part I, LNICST 4, pp. 110–117, 2009.

Many authors have taken to finding estimates for Ω when $n < p$ using either a full Bayasian approach [4] [7] [18], or via an empirical Bayes manner [10] [11].

Covariance selection, then, is accomplished either by heuristic searches, or by sampling the posterior distribution

$$\pi(G, \theta|D) \propto P(D|G)\pi(G|\theta)\pi(\theta) \tag{1}$$

over the space of undirected networks G on p nodes where $\pi(G|\theta)$ is a prior over networks that may depend on a set of parameters θ; $P(D|G)$ is the likelihood. To our knowledge, current methodologies assume G is sparse, and make inference on that basis. In particular, the approach to prior specification adopted in [7] assigns an inclusion probability $\beta = 2/(p-1)$ to each edge in G. This choice of β encourages sparse networks as the expected number of edges is p. In effect, $\pi(G|\theta)$ is the formula for the probability of a network under the random network model of [5], and we will refer to this as the *random prior*.

However, over the past decade, numerous examples of large-scale biological, technological, and sociological networks have been reported to be *scale-free*: that is, the *degree distribution* $p(k)$—the fraction of nodes in the network with degree k—closely follows a *power-law* $p(k) \propto k^{-\gamma}$ with exponent γ typically between 2 and 3 [9]. In particular, this property is thought to be a feature of gene networks [16] and $\gamma \approx 2.2$ has been verified for the known interactions in *S. cerevisiae* [6]. A further example comes from finance where it has been shown that cross-correlation between stock prices for companies on certain stock exchanges follows a power-law [15].

In this paper, we provide specifications for the prior $\pi(G|\theta)$ based on the formula for the probability of a network under three different scale-free models (Section 2). Our approach to employing scale-free network models in statistical inference to estimate a network G from data D is original insomuch as previous research has focused on estimating θ for a particular network, G. In Section 3, we give the results of a simulation study comparing these priors to the random prior at estimating scale-free trees from synthetic data. Finally, in Section 4 we muse on the practicality and possible future applications of our methodology.

2 Scale-Free Priors over Network Structures

In this section, we propose three scale-free assignments for $\pi(G|\theta)$. Each prior is defined by the formula for the probability of a network under a simple scale-free model. We selected the static model [8], the Poisson-growth (PG) model [13], and the proteome growth model [14].

Random network model: The random network, or Erdös-Rényi, model gives rise to a network G by connecting each pair of nodes in G with specified probability, β. Consequently, the probability of a network with $|E|$ edges is

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

where $\theta = (\beta)$ and $T = p(p-1)/2$ is the number of possible edges.

Static model: This model relies on *node fitness* as a generating mechanism, meaning that nodes in a network are assigned weights; edges are added to the network such that nodes with higher weight get more edges. Specifically, the static model is defined by $\theta = (\gamma, K)$ and generates a network as follows: Each node $1, \ldots, p$ is assigned a weight $P_i \propto i^{-1/(1-\gamma)}$ where $2 < \gamma < 3$ is the degree exponent. For $p \times K$ steps:

1. Select nodes i and j with probabilities P_i and P_j, respectively.
2. Connect i and j with an edge, unless they are already connected.

Unlike for the random network model, the probability of a network under any of the scale-free models depends on the order of the nodes in G. Therefore it is necessary to include the extra parameter $\sigma = (\sigma_1, \ldots, \sigma_p)$, a permutation of the nodes in V. Thus the posterior in Equation (1) becomes $\pi(G, \theta, \sigma|D)$ with accompanying prior $\pi(G|\theta, \sigma)$. The static model prior, as described in [8], is given by

$$\pi(G|\theta, \sigma) = e^{pK(1-M)} \prod_{e_{\sigma_i \sigma_j} \in G} \left(e^{2pK P_{\sigma_i} P_{\sigma_j}} - 1 \right)$$

where M is the sum of squares $\sum_{i=1}^{p} P_{\sigma_i}^2$.

PG model: The PG model is an offshoot of the Barabási-Albert (BA) model, which is based on two simple mechanisms: *growth*, where a network is built iteratively, over a series of steps $t = 1, \ldots, p$, by introducing a new node with m (fixed) edges at each step, and *preferential attachment* where the m edges are connected to exactly m nodes already in the system such that the probability a node of degree k gets an edge is proportional to $r(k) = k$, the *attachment function*. When $m = 1$, the BA model is known to have degree exponent $\gamma = 3$ [1].

In the PG model, m is assigned according to a Poisson random variable with parameter λ so that number of edges added at each step can vary. In addition, the attachment function is defined by $r(k) = k + a, k \geq 1$, and $r(0) = b$ where $a \geq -1$ is a small offset and $b = \geq 0$ is a threshold parameter. The PG model, then, is defined by $\theta(\lambda, a, b)$, and has been shown in [13] to follow a power-law with degree exponent ranging $\gamma > 2$. The formula for the associated prior is

$$P(G|\theta, \sigma) = \prod_{t=1}^{p-1} \left(\prod_{i=1}^{t} e^{-\lambda q_t(k_{\sigma_i, \sigma_t})} \frac{(\lambda q_t(k_{\sigma_i, t}))^{s_{\sigma_i, t}}}{s_{\sigma_i, t}!} \right).$$

where $k_{\sigma_i, t}$ is the degree of node σ_i at step t, q_t is the normalized attachment function at step t, and $s_{\sigma_i, t}$ is the number of edges connecting nodes σ_i and t.

Proteome growth: This model is based on growth and *node duplication*. In particular, At each step $t = 1, \ldots, p$, a node is selected from the network at random and duplicated so that the new node inherits its edge structure. Edges emanating from the new node are deleted with probability q; new edges are added between the new node and all other nodes in the system with a small probability β/t. Thus the model is specified by the parameter $\theta = (\beta, q)$, and have been show to exhibit scale-free-like properties [14].

The probability of a network under this model depends on the duplication history $\psi = \{\psi_2, \ldots, \psi_p\}$ in addition to σ, where ψ_t is the node in σ from which σ_t was duplicated. It follows that $\pi(G|\theta, \sigma, \phi)$ is simply the product of the edge inclusion probabilities, which are defined by

$$P(e_{\sigma_i,\sigma_t}) = \begin{cases} 1 - q(1 - \beta/t) & \text{if } \sigma_i \text{ is a neighbor of } \psi_t, \\ \beta/t & \text{otherwise.} \end{cases}$$

Unlike for the random network model, the probability of a network under any of the scale-free models depends on the order of the nodes in G. Therefore it is necessary to include the extra parameter $\sigma = (\sigma_1, \ldots, \sigma_p)$, a permutation of the nodes in V. Thus the posterior in Equation (1) becomes $\pi(G, \theta, \sigma|D)$ with accompanying prior $\pi(G|\theta, \sigma)$.

3 Simulation

We conduct a simulation study to compare the scale-free priors against the random prior at recovering a range of tree topologies from small-sample synthetic data sets.

Tree generation: We generated $p = 100$ node trees from given degree distributions using the stochastic algorithm described in [2], and Table 2 summarizes the trees generated in each case.

Data generation: Trees are used because it is simple to generate multivariate normal data satisfying their conditional independence structures. For each tree in Table 2 we generated ten small-sample data sets, each with $n = 10$ observations.

MCMC implementation: MCMC algorithms are commonly used for sampling from high-dimensional probability distributions such as those encountered in GGMs. We ran the ready-to-use MCMC software from [7] to explore the space of decomposable GGMs for each tree, under each simulated data set. To transition from a decomposable model G to another G' they add or delete an edge from G at random to obtain G'.

In order to accommodate our scale-free priors we modified their sampler to include both the node permutation σ and the duplication history ψ for the proteome growth model. We update the node permutation by choosing σ' in a "neighborhood" of σ. Specifically, we select a node $i \in \{1, \ldots, p\}$ at random, and

Table 1. A summary of scale-free network models used as the basis for prior distributions

Model	Mechanism	Parameters	γ	Ref.
Static	Node fitness	$\theta = (\gamma, K)$	$2 - 3$	[8]
PG	Pref. attach	$\theta = (\lambda, a, b)$	> 2.0	[13]
Prot. gr.	Node dup.	$\theta = (\beta, q)$	$2 - 3$	[14]

Table 2. A wide range of tree topologies were generated for used in the simulation: Generic trees having a binomial degree distribution, scale-free trees, and a star tree in which all nodes are connected to a central hub. The *two parameter model* of citeburda (with parameters α and β) was used as the model for generating most of the scale-free trees. γ is the value of the degree exponent as predicted by the generating model. APL stands for average path length, which, is smaller for more highly centralized trees.

	Model			
Topology	Name	Parameters	γ	APL
Generic	Erdös-Rényi	—	—	10.46
Generic	Two param.	$\alpha = 7.0, \beta = 1.5$	—	8.54
Scale-free	Two param.	$\alpha = 0.24, \beta = 3.0$	3.0	5.86
Scale-free	Two param.	$\alpha = 0.93, \beta = 2.5$	2.5	4.87
Scale-free	Barabási-Albert	$m = 1$	3.0	4.37
Scale-free	Two param.	$\alpha = 8.0, \beta = 2.0$	≈ 2.0	3.34
Scale-free	Two param.	$\alpha = 2.5, \beta = 2.5$	≈ 2.5	3.18
Scale-free	Two param.	$\alpha = 3.6, \beta = 2.2$	2.2	2.96
Scale-free	Two param.	$\alpha = 8.5, \beta = 2.1$	2.1	2.32
Star	—	—	—	1.98

proceed to transpose σ_i and σ_{i+1} to obtain σ'. In the case of the proteome growth model, the duplication history ψ is updated conditionally on σ' be rewiring the node represented by σ'_i to $\sigma'_i - 1$, if $\psi_{\sigma'_i} = \sigma'_i - 1$; otherwise σ'_i is rewired randomly to another node $\sigma'_j < \sigma'_i$. Additionally, we assigned the uniform distribution over each model parameter to compute $\pi(\theta)$, the prior distribution over the network model parameters.

We ran each MCMC chain for 5×10^6 steps after a burn-in of 10^5 steps. Each chain was started from the empty network, and for the scale-free priors with node σ and ψ were taken at random. We obtained similar results when starting from a variety of different initial conditions.

Results: Simulation results are summarized in Table 3. To estimate a network from a chain we took the $p - 1$ (i.e. the number of edges in a tree) edges of highest frequency over all networks in a chain. The model parameter estimate $\hat{\theta}$ was obtained by taking mean of the parameter values from a chain.

- The random network prior did quite well at recovering the generic trees as well as the scale-free trees with (large) $\gamma = 2.5$ to 3.0.
- The proteome growth prior exhibited poor performance overall. This can likely be attributed to model misspecification insomuch as this model may not be able to capture the tree structures that it was employed to estimate.
- The static and PG model priors outperformed the random prior at recovering the scale-free trees with underlying γ values near 2. This behavior likely reflects that as γ in a network tends to smaller values, the connectivity to the hub, the node with highest degree, tends to be higher, resulting in a more highly centralized network. In the pathological case of a star topology

Table 3. Results of the MCMC simulation ($p = 100$) with the numerical values in the table averaged over ten runs. PPV is the positive predictive value defined by $TP/(TP+FP)$, where TP stands for true positive, and FP for false negative.

	Generated Tree		Estimated Parameters		
Network Prior	Topology	γ	$\vec{\theta}$	$\hat{\gamma}$	PPV
Random	Generic	—	$\hat{\beta} = 0.015$	—	0.36
	Generic	—	$\hat{\beta} = 0.016$	—	0.35
	Scale-free	3.0	$\hat{\beta} = 0.016$	—	0.27
	Scale-free	2.5	$\hat{\beta} = 0.017$	—	0.28
	Scale-free	3.0	$\hat{\beta} = 0.016$	—	0.26
	Scale-free	≈ 2.0	$\hat{\beta} = 0.019$	—	0.19
	Scale-free	≈ 2.5	$\hat{\beta} = 0.016$	—	0.20
	Scale-free	2.2	$\hat{\beta} = 0.017$	—	0.12
	Scale-free	2.1	$\hat{\beta} = 0.017$	—	0.10
	Star	—	$\hat{\beta} = 0.018$	—	0.06
Static	Generic	—	$\hat{\gamma} = 2.82, \hat{K} = 0.63$	2.82	0.33
	Generic	—	$\hat{\gamma} = 2.80, \hat{K} = 0.70$	2.80	0.32
	Scale-free	3.0	$\hat{\gamma} = 2.70, \hat{K} = 0.76$	2.70	0.28
	Scale-free	2.5	$\hat{\gamma} = 2.58, \hat{K} = 0.90$	2.58	0.30
	Scale-free	3.0	$\hat{\gamma} = 2.76, \hat{K} = 0.70$	2.76	0.28
	Scale-free	≈ 2.0	$\hat{\gamma} = 2.39, \hat{K} = 1.18$	2.39	0.37
	Scale-free	≈ 2.5	$\hat{\gamma} = 2.54, \hat{K} = 0.88$	2.54	0.40
	Scale-free	2.2	$\hat{\gamma} = 2.25, \hat{K} = 1.27$	2.25	0.51
	Scale-free	2.1	$\hat{\gamma} = 2.30, \hat{K} = 1.16$	2.30	0.57
	Star	—	$\hat{\gamma} = 2.25, \hat{K} = 1.28$	2.25	0.78
PG	Generic	—	$\hat{\lambda} = 0.66, \hat{a} = -0.57, \hat{b} = 0.76$	2.88	0.27
	Generic	—	$\hat{\lambda} = 0.83, \hat{a} = -0.83, \hat{b} = 0.72$	2.56	0.17
	Scale-free	3.0	$\hat{\lambda} = 0.92, \hat{a} = -0.88, \hat{b} = 0.70$	2.51	0.23
	Scale-free	2.5	$\hat{\lambda} = 1.02, \hat{a} = -0.89, \hat{b} = 0.70$	2.51	0.28
	Scale-free	3.0	$\hat{\lambda} = 0.82, \hat{a} = -0.80, \hat{b} = 0.71$	2.58	0.22
	Scale-free	≈ 2.0	$\hat{\lambda} = 1.19, \hat{a} = -0.94, \hat{b} = 0.69$	2.50	0.41
	Scale-free	≈ 2.5	$\hat{\lambda} = 0.96, \hat{a} = -0.91, \hat{b} = 0.69$	2.49	0.45
	Scale-free	2.2	$\hat{\lambda} = 1.21, \hat{a} = -0.95, \hat{b} = 0.68$	2.49	0.62
	Scale-free	2.1	$\hat{\lambda} = 1.13, \hat{a} = -0.95, \hat{b} = 0.68$	2.48	0.62
	Star	—	$\hat{\lambda} = 1.20, \hat{a} = -0.95, \hat{b} = 0.70$	2.49	0.94
Prot. gr.	Generic	—	$\hat{\beta} = 0.64, \hat{q} = 0.91$	—	0.24
	Generic	—	$\hat{\beta} = 0.71, \hat{q} = 0.92$	—	0.22
	Scale-free	3.0	$\hat{\beta} = 0.72, \hat{q} = 0.92$	—	0.19
	Scale-free	2.5	$\hat{\beta} = 0.77, \hat{q} = 0.95$	—	0.23
	Scale-free	3.0	$\hat{\beta} = 0.68, \hat{q} = 0.92$	—	0.17
	Scale-free	≈ 2.0	$\hat{\beta} = 0.81, \hat{q} = 0.94$	—	0.02
	Scale-free	≈ 2.5	$\hat{\beta} = 0.72, \hat{q} = 0.92$	—	0.02
	Scale-free	2.2	$\hat{\beta} = 0.80, \hat{q} = 0.92$	—	0.10
	Scale-free	2.1	$\hat{\beta} = 0.75, \hat{q} = 0.94$	—	0.10
	Star	—	$\hat{\beta} = 0.83, \hat{q} = 0.94$	—	0.02

Note: For the proteome growth model, $\hat{\gamma}$ cannot be obtained analytically.

the scale-free priors perform exceptionally well in comparison to the random prior.
- The static model prior was able to produce was able to estimate the γ values with some accuracy, while the PG model prior did not discriminate between different degree exponents.

Out of the three scale-free prior distributions, the static model was the most accurate at estimating a network. In addition, it produced reasonable estimates for the degree exponent, γ.

4 Discussion

In summary, GGMs provide a framework for making inference about the conditional independence structure of a set a Gaussian variables when the number of observations are small compared to the number of variables. And this methodology, in various forms, is commonly applied in gene network inference. Our contribution in this paper has been to study the *a priori* inclusion of scale-free network topology into GGM inference via the specification of scale-free prior distributions. Our simulation study suggests that scale-free priors outperform the random prior at recovering scale-free trees, from small-sample data sets, with degree exponent γ near 2—a property thought to be typical of a wide variety of real-world networks.

We used trees, not networks, in our simulation in order to facilitate data generation. It would be beneficial to further this work by conducting a simulation over general network structures.

The implementation of [7], on which our results are based, can comfortably work for networks of a few hundred nodes; however, gene network inference can often involve networks with nodes numbering in the thousands. Therefore, scalability is a major issue, and integrating scale-free topological features using a different approach to GGM inference is of interest [10] [11].

References

1. Barabási, A.-L., Albert, R.: Emergence of scaling in random networks. Science 286, 509–512 (1999)
2. Burda, Z., Correia, J.D., Krzywicki, A.: Statistical Ensemble of Scale-Free Random Graphs. Phys. Rev. E 64, 046118+ (2001)
3. Dempster, A.P.: Covariance selection. Biometrics 28, 95–108 (1972)
4. Dorba, A., Hans, C., Jones, B., Nevins, J.R., West, M.: Sparse graphical models for exploring gene expression data. J. Multiv. Analysis 90, 196–212 (2004)
5. Erdös, P., Rényi, A.: On Random Graphs I. Pub. Math. Debrecen 6, 290–297 (1959)
6. Jeong, H., Mason, S.P., Barabasi, A.L., Oltvai, Z.N.: Lethality and Centrality in Protein Networks. Nature 411, 41–42 (2001)
7. Jones, B., Dobra, C., Carvalho, C., Hans, C., Carter, C., West, M.: Experiments in Stochastic Computation for High-Dimensional Graphical Models. Statistical Science 20(4), 388–400 (2005)

8. Lee, D.-S., Goh, H.-I., Kahng, B., Kim, D.: Scale-Free Random Graphs and Potts Model. Pramana J. Phys. 64, 1149–1159 (2005)
9. Newman, M., Barabási, A.-L., Watts, D.J.: The Structure and Dynamics of Networks. Princeton University Press, Princeton (2006)
10. Schäfer, J., Strimmer, K.: An Empirical Bayes Approach to Inferring Large-Scale Gene Association Networks. Bioinformatics 21, 754–764 (2005a)
11. Schäfer, J., Strimmer, K.: A Shrinkage Approach to Large-Scale Covariance Matrix Estimation and Implications for Functional Genomics. Stat. Appl. Genet. Mol. Biol. 4, article 32 (2005b)
12. Sheridan, P., Kamimura, T., Shimodaira, H.: Scale-Free Networks in Bayesian Inference with Applications to Bioinformatics. In: Proceedings of The International Workshop on Data-Mining and Statistical Science (DMSS 2007), Tokyo, pp. 1–16 (2007)
13. Sheridan, P., Yagahara, Y., Shimodaira, H.: A Preferential Attachment Model with Poisson Growth for Scale-Free Networks. Ann. Inst. Stat. Math. 60 (2008)
14. Solé, R.V., Pastor-Satorras, R., Smith, E., Kepler, T.B.: A Model of Large-Scale Proteome Evolution. Adv. Complex Systems 5, 43–54 (2002)
15. Vandewalle, N., Brisbois, F., Tordoir, X.: Non-Random Topology of Stock Markets. Quant. Finance 1, 372–374 (2001)
16. Wagner, A.: How the Global Structure of Protein Interaction Networks Evolve. Proc. R. Soc. B 270, 457–466 (2003)
17. Werhli, A.V.V., Grzegorczyk, M., Husmeier, D.: Comparative Evaluation of Reverse Engineering Gene Regulatory Networks with Relevance Networks, Graphical Gaussian Models and Bayesian Networks. Bioinformatics 22, 2523–2531 (2006)
18. Wong, K., Carter, C., Kohn, R.: Efficient Estimation of Covariance Selection Models. Biometrika 90, 809–830 (2004)