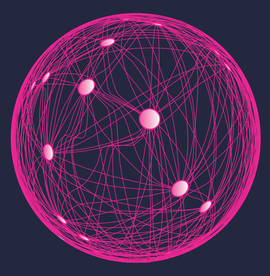


A Scale-free Prior over Graph Structures for Bayesian Inference of Gene Networks

Takeshi Kamimura
kamimur1@is.titech.ac.jp

Hidetoshi Shimodaira
shimo@is.titech.ac.jp

Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, Ookayama, Meguro, Tokyo 152-8552, Japan.

Keywords: Scale-free random graphs, Power law, Bayesian Inference, Gaussian graphical model, Markov chain Monte Carlo, microarray, gene network

I. Introduction

In recent years, a large amount of gene expression data has been collected and estimating a gene network has become one of the central topics in the field of bioinformatics. Several methodologies have been proposed for constructing a gene network based on gene expression data and Gaussian graphical model is also one of the effective methods. When we look at the method from a Bayesian perspective, questions of the nature and consistency of prior probability specification (prior probabilities over graphical structure etc) have yet to be definitively determined, though a lot of ideas have been suggested [2, 4].

Recent studies of networks such as the Internet or World Wide Web have revealed that the probability that a node of these networks has k edges, or equivalently k adjacent nodes, follows a power law $P(k) \propto k^{-\gamma}$ over a large range of k, with an exponent γ that ranges between 1 and 3 depending on the system. Such netwoks are called *scale free* and this property is suggested to be appropriate for biological netwoks as well [6].

In this study, we propose a new prior based on this property of "real-world" networks. This method is applied to *S. cerevisiae* gene expression data [1].

This work is supported in part by Grant KAKENHI -17700276 from MEXT of Japan.

II. Gaussian Graphical Model

Graphical models provide representations of the conditional independence structure of a multivariate distribution as well as access to efficient algorithms for computaion of conditional and marginal densities. Multivariate Gaussian graphical models are defined in terms of Markov properties, i.e., conditional independences associated with the underlying graph. Thus, model selection can be performed by testing these conditional independences, which are equivalent to specified zeros among certain (partial) correlation coefficients. The graph G consists of a set of nodes V and a set of edges E. Two nodes v_i and v_j are conditionally independent given the remaining variables if, and only if, $\{v_i, v_j\} \notin E$. The details of Gaussian graphical model are described in [2].

Formal inference is inherently structured by composition; from a Bayesian perspective, we are interested in posterior distributions

$$P(G|Y) \propto P(Y|G)P(G).$$

For the first term $P(Y|G)$, we refered to [2] and, in our study, we propose the way to constitute a new prior $P(G)$.

Fig. 1

Fig. 2

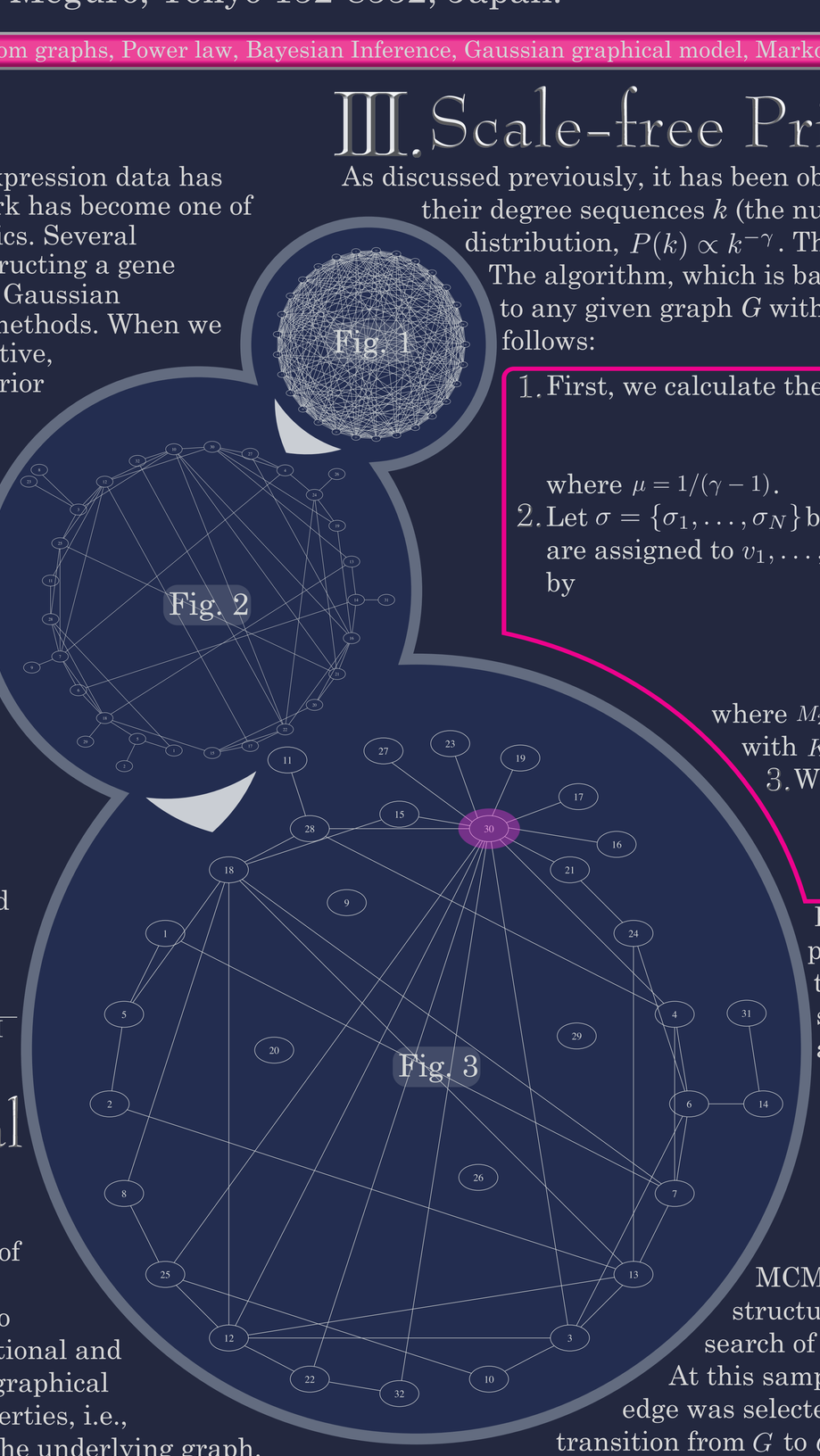


Fig. 3

III. Scale-free Priors over Graphs

As discussed previously, it has been observed that many biological networks share global properties and their degree sequences k (the number of edges per node) often follow a long-tailed power-law distribution, $P(k) \propto k^{-\gamma}$. Thus, we would like to construct the prior based on this property.

The algorithm, which is based on the model introduced in [3, 5], to assign a prior probability to any given graph G with a fixed set of nodes ($V = \{v_1, \ldots, v_N\}$) can be expressed as follows:

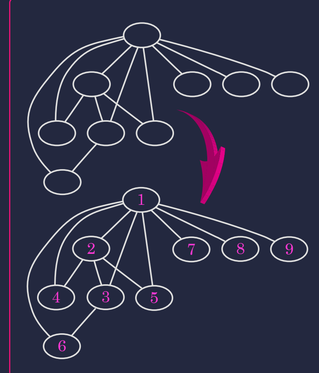
1. First, we calculate the the following numbers for $i = 1, \ldots, N$,
$$P_i = \frac{i^{-\mu}}{\sum_{j=1}^{N} j^{-\mu}} \approx \frac{1-\mu}{N^{1-\mu}} i^{-\mu}$$
where $\mu = 1/(\gamma - 1)$.
2. Let $\sigma = \{\sigma_1, \ldots, \sigma_N\}$ be a permutation of $\{1, \ldots, N\}$. For a given permutaion, $\sigma_1, \ldots, \sigma_N$ are assigned to $v_1, \ldots, v_N$, respectively, and the conditional probability of G is defined by
$$\begin{aligned} P(G|\sigma) &= \prod_{\{v_i,v_j\}\in E} (1 - e^{-2NKP_{\sigma_i}P_{\sigma_j}}) \prod_{\{v_i,v_j\}\notin E} e^{-2NKP_{\sigma_i}P_{\sigma_j}} \\ &= e^{-NK(1-M_2)} \prod_{\{v_i,v_j\}\in E} (e^{2NKP_{\sigma_i}P_{\sigma_j}} - 1) \end{aligned}$$
where $M_2 \equiv \sum_{i=1}^{N} P_i^2$ and we can select K on the condition that $K_l \ll K \ll K_u$ with $K_l \sim N^{-\mu}$ and $K_u \sim N^{1-\mu}$.
3. We define $P(G)$ by averaging $P(G|\sigma)$ over all permutaions
$$P(G) = \frac{1}{N!} \sum_{\sigma:\ \text{all permutations}} P(G|\sigma)$$

However, as the number of nodes gets larger, the number of permutations dramatically increases. A possible approximation is to generate randomly σ for B times (say B = 10,000) in the summation of Step 3, but we consider rather better approximation methods in Section IV.

IV. Markov Chain Monte Carlo Algorithm

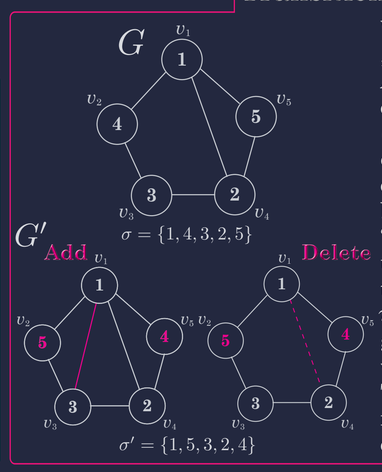
MCMC is a much used tool for exploring the space of graphical structures. We implemented the Metropolis-Hastings sampler for a search of not only decomposable but also non-decomposable graph space. At this sampler, the choice to add or delete an edge was made, and then an edge was selected at random from those appropriate for that type of move. The transition from G to G' and approximate methods for calculating the prior probability of G are as follows:

Transition type-1



At each MCMC step, we change the graph structure by adding or deleting an edge randomly and an approximation for calculating the prior probability of G is to calculate $P(G|\hat{\sigma})$ based on the permutaion $\hat{\sigma}$ that maximizes $P(G|\sigma)$ instead of averaging $P(G|\sigma)$; the more edges a node has, the smaller number i we assign to the node, and we define $P(G)$ proportional to $P(G|\hat{\sigma})$.

Transition type-2



We define $P(G, \sigma) = P(G|\sigma)P(\sigma)$, where $P(\sigma) = 1/N!$.
At each MCMC iteration, we change the graph structure from G to G' by adding or deleting an edge randomly and we also change σ to σ' at the same time by choosing σ' in a "neighborhood" of σ. We have used the following definition of the neighborhood: $\tau'_i = \tau_{i+1}$, $\tau'_{i+1} = \tau_i$ for randomly generated i in $\{1, \ldots, N-1\}$, where $\tau_j = i$ for $\sigma_i = j$.
Then the Metropolis-Hastings ratio for the prior part can be calculated with $P(G', \sigma')/P(G, \sigma)$.

V. A Numerical Example

We applied the new prior to the *S. cerevisiae* gene expression data. We focused on 32 genes which are arranged in the right table. The Metropolis-Hastings was run for 1,000,000 steps (we used Trasition type-1 until 10,000 steps for fast convergence to the stationary distribution and Transition type-2 for the rest) and we took γ = 2.2 and $K = \frac{K_l + K_u}{2}$. Figure 1, Figure 2 and Figure 3 are the resulting networks using different priors and they had the highest log posterior probabilities in each chain.

- Figure 1: Estimated gene network using the uniform prior over all graph structures. This network is very dense and the number of edges a node has is almost uniform, which is inconsistent with the biological observations [6, 7].
- Figure 2: Estimated gene network using a Bernoulli prior on each edge inclusion probability. This approach to prior specification penalizes only the number of edges, so the estimated network is sparser, but the number of edges per node is almost uniform and it is inconsistent with the biological observations [6, 7].
- Figure 3: Estimated gene network using the proposed scale-free prior. It shows that the estimated network based on scale-free priors is sparser and it has hubs (ex. node 30), which is consistent with the biological observations [6, 7].

	gene name		gene name
1	RAD51	17	MCM1
2	CLN1	18	ACE2
3	CLB2	19	GAT3
4	BUD9	20	ACA1
5	TSL1	21	KRE33
6	JIP1	22	RCO1
7	EGT2	23	RFX1
8	SWI5	24	SFL1
9	SPO16	25	SIP3
10	FKH2	26	SMK1
11	MBP1	27	UGA3
12	SWI6	28	UME6
13	NDD1	29	WAR1
14	STE12	30	YER184C
15	SWI4	31	YGR067C
16	FKH1	32	YRR1

References

[1] Aach, J., Rindone, W. and Church, GM., Systematic management and analysis of yeast gene expression data. *Genome Res.*, 10:431-435, 2000.

[2] Beatrix, J., Carlos, C., Adrian, D., Chris, H., Chris, C. and Mike, W., Experiments in Stochastic Computation for High-Dimensional Graphical Models. *SAMSI Technical Report*, 2004-1, 2004.

[3] Goh, K. -I., Kahng, B. and Kim, D. Universal Behavior of Load Distribution in Scale-free Networks. *Phys. Rev. Lett.*, 87:278701, 2001.

[4] Imoto, S., Higuchi, T., Goto, T., Tashiro, K., Kuhara, S. and Miyano, S. Combining microarrays and biological knowledge for estimating gene networks via Bayesian networks. *Journal of Bioinformatics and Computational Biology*, 2:77–98, 2003.

[5] Lee, D. -S., Goh, K. -I., Kahng, B. and Kim, D. Scale-free random graphs and Potts model. *Pramana-J. Phys.*, 64:1149–1159, 2005.

[6] Newman, M, E, J. The Structure and Function of Complex Networks. *SIAM Rev.*, 45:167–256, 2003.

[7] Vera van Noort, Berend Snel and Martijn A. Huynen. The yeast coexpression network has a small-world, scale-free architecture and can be explained by simple model. *EMBO reports.*, 5:280-284, 2004.