# Bayesian inference for multiple Gaussian graphical models with application to metabolic association networks

Siew Li Linda Tan

(Joint with Ajay Jasra, Maria De Iorio and Timothy Ebbels)

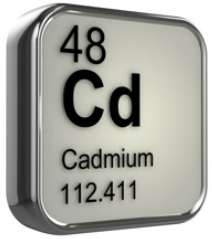
National University of Singapore

# Introduction

- Profiling of metabolites (products of metabolic reactions) is important for understanding biological systems. Metabolite levels are often correlated – analyzing these associations provide information about metabolic relationships and the physiological state of a system.

**Urinary metabolic data**

- Acquired using $^{1}$H NMR spectroscopy for 127 individuals living close to a smelter in Bristol, UK, that produces airborne cadmium (very toxic metal, acute exposure poses health risks).




- Study correlation structure of 22 urinary metabolites in response to Cd exposure using Gaussian graphical models (GGMs).
- Analysis of such metabolic association networks can point to differences in the underlying biological reactions caused by Cd exposure.

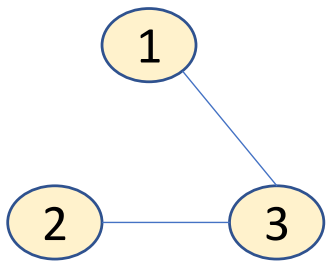
# Introduction (Gaussian graphical models)

- Gaussian graphical models (GGMs) provide an important tool for studying the dependence structure among a set of random variables.
- Assuming the variables have a joint Gaussian distribution, a zero in the precision matrix indicates conditional independence between associated variables.

**Precision matrix**

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 3 \end{pmatrix}$$

**Underlying graph**




**Nodes denote variables.**
**Edges denote conditional dependencies.**

- To investigate effect of different experimental conditions on the dependence relationships among variables, multiple GGMs (one for each condition) have to be estimated. Joint inference encourage sharing of information across graphs and allow for common structure.

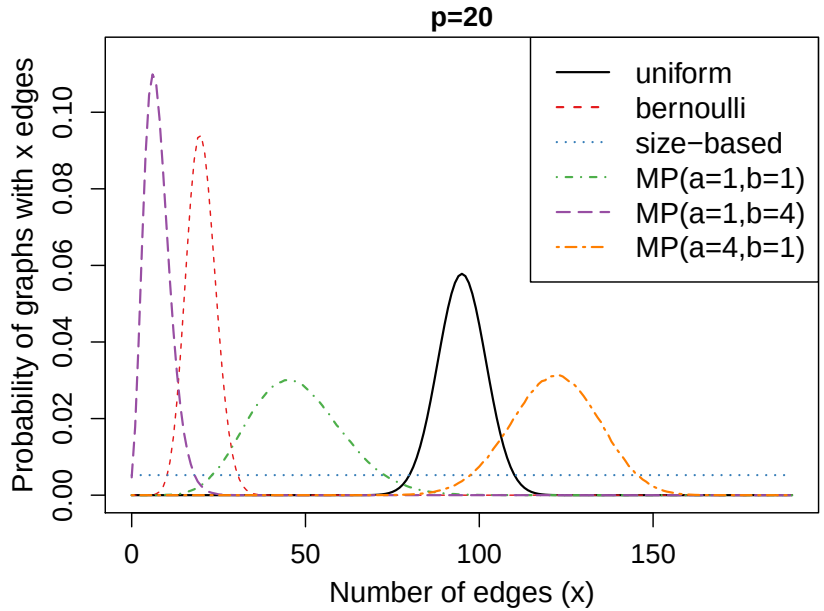
# Contributions

- Bayesian inference for GGMs using G-Wishart prior (Roverato, 2002). The G-Wishart is the family of conjugate distributions for the precision matrix, where entries corresponding to missing edges in the underlying graph are constrained to be zero.
- Consider the unrestricted graph space (allow non-decomposable graphs) and propose using the multiplicative model (Chung and Lu, 2002) as a prior on graphs for estimating GGMs. This prior is extended to multiple GGMs via logistic regression.
- Develop a sequential Monte Carlo algorithm to obtain joint posterior inference for multiple GGMs.
- Illustrate proposed methods using simulated data and the urinary metabolic dataset.

# Priors for graphs

Given $p$ nodes, there are $2^r$ possible graphs, where $r = \binom{p}{2}$ is the total number of possible edges.

- Uniform prior: Each graph has equal probability of arising.
- Erdős-Rényi model: Every edge has probability $\alpha$ of being included.
  P(graph with $x$ edges) $= \alpha^x(1-\alpha)^{r-x}$.
  Reduces to uniform prior when $\alpha = 0.5$.
- Bernoulli prior (Jones et al. 2005):
  $\alpha = \frac{2}{p-1}$, expected no. of edges is $p$.




- Place Beta$(a, b)$ prior on $\alpha$ (Carvalho and Scott, 2009).
- Size-based prior (Armstrong et al. 2009): Every size $0, \ldots, r$, has equal probability and every graph of the same size has equal probability.
  P(graph with $x$ edges) $= \frac{1}{r+1}\binom{r}{x}^{-1}$. Controls number of spurious edges.

# Multiplicative prior

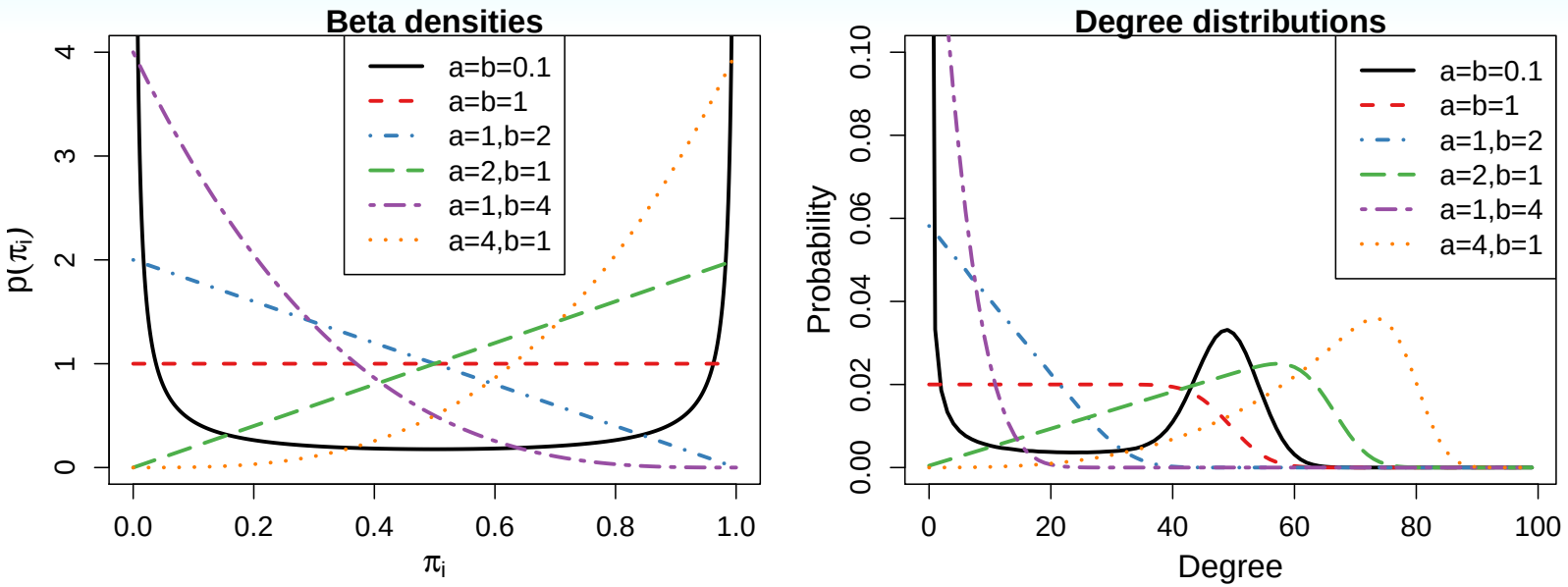
- Let $G = (V, E)$ be a simple graph with vertex set $V = \{1, 2, \ldots, p\}$ and edge set $E \subseteq \{(i, j) \in V \times V : i < j\}$. The adjacency matrix $A = [A_{ij}]$ of $G$ is a symmetric binary matrix where $A_{ij}$ is 1 if an edge is present between nodes $i$ and $j$, and 0 otherwise.
- Chung and Lu (2002):
  - Expected degree sequence: $\{d_1, \ldots, d_p\}$, $\mathsf{P}(A_{ij} = 1) \propto d_i d_j$.
  - Capture more diverse degree distributions and real-world networks.
- Multiplicative model:

$$A_{ij} \overset{\text{indep}}{\sim} \text{Bernoulli}(\pi_i \pi_j) \text{ for } 1 \leq i < j \leq p, \text{ where } 0 \leq \pi_i \leq 1.$$

$$\pi_i \overset{\text{indep}}{\sim} \text{Beta}(a, b) \text{ for } i = 1, \ldots, p, \quad a, b > 0.$$

  - $\pi_i$ reflects activity level (connectivity) of node $i$.

# Degree distributions of multiplicative prior




We derive some degree and clustering properties for the multiplicative prior.

- Degree distributions of a variety of shapes (e.g. right-skewed, U-shaped) can be obtained by varying $a$ and $b$.
- The multiplicative model can accommodate greater variation in the degree distribution than the Erdős-Rényi model by $\mathcal{O}(p^2)$ given the same mean degree.

# Investigating power-law

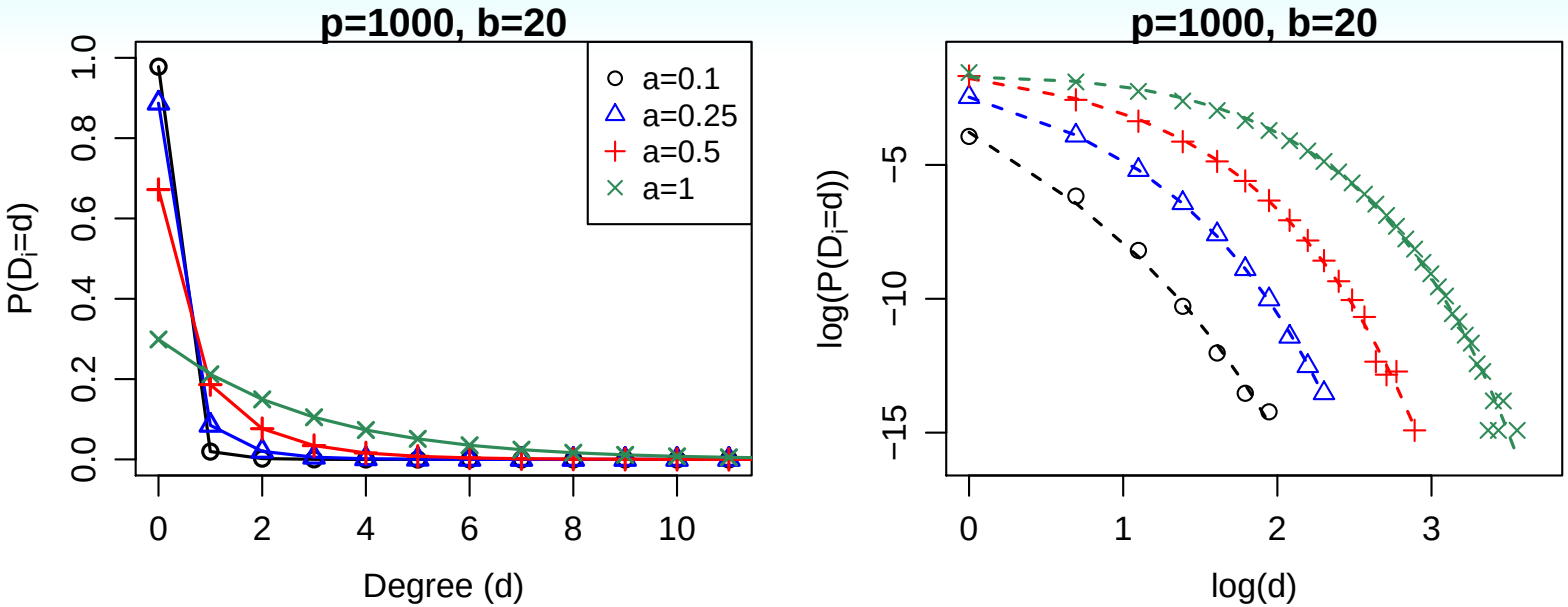



Figure: Left: degree distributions. Right: relationship between $\log P(D_i = d)$ and $\log d$ should be a straight line if power-law is satisfied.

- Degree distributions of scale-free networks follow power-law:

$$\mathrm{P}(D_i = d) \propto d^{-\gamma} \text{ where } \gamma > 0.$$

- The multiplicative prior comes close to but does not quite induce power-law networks as the right tail is not sufficiently heavy. However, the points are well fitted by a power law with an exponential cutoff, $P(D_i = d) \propto d^{-\gamma} \mathrm{e}^{-\tau d}$.

# Multiple Gaussian graphical model

- **Data**: $H$ observations of $p$ variables which fall into $K$ groups.
  - $y_h = (y_{h1}, \ldots, y_{hp})$ denotes the $h^{\text{th}}$ observation.
  - $S_k$ is a set containing indices of observations which belong to group $k$ with $H = \sum_{k=1}^{K} |S_k|$.
- Assume

$$y_h | \Omega_k \sim \mathsf{N}(0, \Omega_k^{-1}) \text{ for } h \in S_k,$$

  where $\Omega_k$ is a precision matrix.
- Let $G_k = (V, E_k)$ be a simple graph with vertex set $V = \{1, 2, \ldots, p\}$ and edge set $E_k \subseteq \{(i,j) \in V \times V : i < j\}$. Node $i \in V$ represents the $i$th variable and each edge $(i,j) \in E_k$ corresponds to $\Omega_{k,ij} \neq 0$.

$$\begin{array}{c} y_{hi} \text{ and } y_{hj} \text{ are conditionally independent} \\ \text{in } G_k \text{ given the rest of the variables in } y_h \end{array} \Leftrightarrow \begin{array}{c} (i,j) \notin E_k \\ (\Omega_{k,ij} = 0) \end{array}$$

# G-Wishart prior

- Conjugate prior for $\Omega_k$ is $G$-Wishart distribution, $W_{G_k}(\delta_k, D_k)$, which has density,

$$p(\Omega_k|G_k) = \frac{1}{I_{G_k}(\delta_k, D_k)}|\Omega_k|^{(\delta_k-2)/2}\exp\left\{-\frac{1}{2}\text{tr}(\Omega_k D_k)\right\}.$$

  $\Omega_k$ is constrained to the cone $P_{G_k}$ of positive definite matrices with entries equal to zero for all $(i,j) \notin E_k$ and $I_{G_k}(\delta_k, D_k)$ is a normalizing constant.

- $G$-Wishart distribution reduces to Wishart distribution when $G_k$ is complete (normalizing constant can then be evaluated analytically).

# Priors over graphs: $K = 1$

- We use the multiplicative model to assign prior probabilities to graphs. Let $A_k = [A_{k,ij}]$ be adjacency matrix of $G_k$.

$$A_{k,ij} | \pi_{k,i}\pi_{k,j} \overset{\text{indep}}{\sim} \text{Bernoulli}(\pi_{k,i}\pi_{k,j}),$$

  where $0 \leq \pi_{k,i} \leq 1$.

- $K = 1$: $G_1 = G$ and let $\pi_{1,i} = \pi_i \sim \text{Beta}(a, b)$. Then

$$\begin{aligned} p(G|a,b) &= \int p(G|\pi)p(\pi|a,b)\, d\pi \\ &= \frac{1}{B(a,b)^p} \int \prod_{i,j:\, i<j} (1-\pi_i\pi_j)^{(1-A_{ij})} \prod_{i=1}^{p} \pi_i^{(a+d_i-1)}(1-\pi_i)^{(b-1)}\, d\pi, \end{aligned}$$

  where $d_i$ is the degree of node $i$.

# Priors over graphs: $K > 1$

- Propose a joint prior for $G_1, \ldots, G_K$, which may depend on graph specific covariates. Let $x_k = (x_{k1}, \ldots, x_{kQ})^T$ be a vector of covariates for $G_k$, $\beta_i = (\beta_{i1}, \ldots, \beta_{iQ})$ be a vector of coefficients for node $i$ and

$$\pi_{k,i} = \frac{\exp(\beta_i^T x_k)}{1 + \exp(\beta_i^T x_k)}, \qquad \beta_{iq} | \sigma_q^2 \overset{\text{indep}}{\sim} \mathsf{N}(0, \sigma_q^2).$$

- Let $x = (x_1, \ldots, x_K)$, $\beta = (\beta_1^T, \ldots, \beta_p^T)^T$ and $\sigma^2 = (\sigma_1^2, \ldots, \sigma_Q^2)$. Then

$$p(G_1, \ldots, G_K | x, \sigma^2) = \int p(\beta | \sigma^2) \prod_{k=1}^{K} p(G_k | x_k, \beta) d\beta$$

$$= \int \prod_{i=1}^{p} \prod_{q=1}^{Q} \left\{ \frac{1}{\sqrt{2\pi\sigma_q^2}} \mathrm{e}^{-\frac{\beta_{iq}^2}{2\sigma_q^2}} \right\} \prod_{k=1}^{K} \left\{ \prod_{i=1}^{p} \pi_{k,i}^{d_{k,i}} \prod_{i<j} (1 - \pi_{k,i}\pi_{k,j})^{1 - A_{k,ij}} \right\} d\beta,$$

  where $d_{k,i}$ is the degree of node $i$ in $G_k$.

- $p(G|a,b)$ or $p(G_1, \ldots, G_K | x, \sigma^2)$ are estimated efficiently using Laplace approximation.

# Priors over graphs for urinary metabolic data

- Let $x_k$ include an intercept and an indicator for level of Cd exposure (1 if above median, 0 otherwise).

| | $G_1$ | $G_2$ |
|---|---|---|
| $x_k$ | $(1,0)$ | $(1,1)$ |
| | correlation structure of group with Cd exposure $\leq$median | correlation structure of group with Cd exposure $>$median |
| $\pi_{k,i}$ | $\frac{1}{1+\exp(-\beta_{i1})}$ | $\frac{1}{1+\exp(-\beta_{i1}-\beta_{i2})}$ |

- $\beta_{i1}$ determines popularity of node $i$ in $G_1$ while $\beta_{i2}$ controls the difference in popularity of node $i$ between $G_1$ and $G_2$. As $|\beta_{i2}| \uparrow$, difference in connectivity of node $i$ between $G_1$ and $G_2 \uparrow$.
- We can also add a third covariate for gender (1 if male, 0 for female). Then $K = 4$, $x_1 = (1,0,0)$, $x_2 = (1,0,1)$, $x_3 = (1,1,0)$ and $x_4 = (1,1,1)$.

# Priors over graphs (illustration)

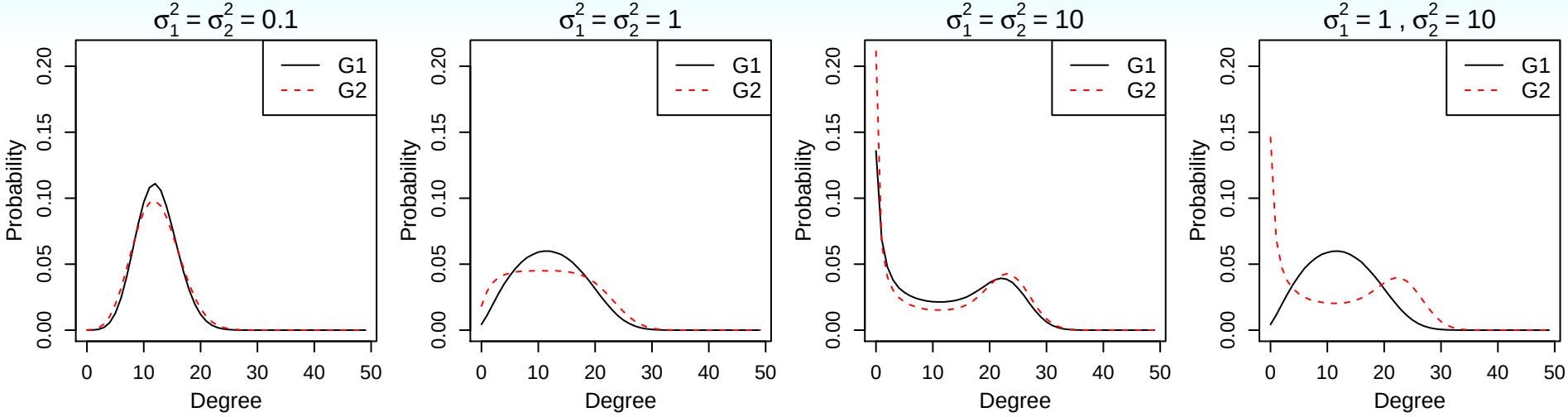



Figure: Prior degree distributions of $G_1$ and $G_2$ for $p = 50$. Covariates for $G_1$ and $G_2$ are (1,0) and (1,1) respectively.

- When $\sigma_1^2 = \sigma_2^2 = 0.1$, $\pi_{1,i} \approx \pi_{2,i} \approx 0.5$. Degree distribution resembles that of Erdős-Rényi model where each edge is formed with probability 0.25.
- When $\sigma_1^2$ is large, connectivity of each node tends to 0 or 1. Thus the degree distribution resembles the case where each $\pi_i \sim \text{Beta}(0.1, 0.1)$.
- Distinction between degree distribution of $G_1$ and $G_2$ becomes greater as $\sigma_2^2$ increases.

# Posterior distributions

- Let $y = (y_1, \ldots, y_H)$. For $K > 1$, the joint distribution of the model is

$$p(y, \Omega_1, \ldots, \Omega_K, G_1, \ldots, G_K, \beta | x, \sigma^2)$$
$$= p(\beta|\sigma^2) \prod_{k=1}^{K} \left\{ p(G_k|x_k, \beta) p(\Omega_k|G_k) \prod_{h \in S_k} p(y_h|\Omega_k) \right\}.$$

  Integrating out $\Omega_k$ and $\beta$, $p(G_1, \ldots, G_K|y, x, \sigma^2)$

$$\propto p(G_1, \ldots, G_K|x, \sigma^2) \prod_{k=1}^{K} \frac{I_{G_k}(\delta_k + |S_k|, D_k + \sum_{h \in S_k} y_h y_h^T)}{I_{G_k}(\delta_k, D_k)}.$$

- When $K = 1$, the dependence on $x$ and $\sigma^2$ is replaced by $a$ and $b$:

$$p(G|y, a, b) \propto p(G|a, b) I_G(\delta + H, D_k + \sum_{h=1}^{H} y_h y_h^T) / I_G(\delta, D).$$

# Posterior inference

- For any $G$, let $(\mathcal{P}_1, \ldots, \mathcal{P}_L)$ be a perfect sequence of the prime components of $G$ and $(\mathcal{S}_2, \ldots, \mathcal{S}_L)$ be the separators. Then

$$I_G(\delta, D) = \prod_{l=1}^{L} I_{G_{\mathcal{P}_l}}(\delta, D) / \prod_{l=2}^{L} I_{G_{\mathcal{S}_l}}(\delta, D),$$

  where $G_{\mathcal{P}_l}$ and $G_{\mathcal{S}_l}$ denote subgraphs induced by $\mathcal{P}_l$ and $\mathcal{S}_l$ respectively.

- The separators are complete and $I_{G_{\mathcal{S}_l}}(\delta, D)$ can be evaluated exactly.
- $I_{G_{\mathcal{P}_l}}(\delta, D)$ can be evaluated exactly if $\mathcal{P}_l$ is complete. Otherwise, estimate using
  - $\delta$ is small: Monte Carlo method (Atay-Kayis and Massam, 2005)
  - $\delta$ is large: Laplace approximation (Lenkoski and Massam, 2005)
- This procedure is feasible for small graphs ($p \leq 22$). When $p$ is large, techniques that avoid evaluating prior normalizing constants (and jointly explore the space of graphs and precision matrices) must be used.

# Sequential Monte Carlo (SMC) sampler

- Suppose we wish to sample from a complex target $\lambda(x)$.

$$\text{(Start)} \quad \underbrace{\lambda_1}_{\substack{\text{distribution easy} \\ \text{to sample from}}} \quad , \quad \underbrace{\lambda_2, \ \ldots, \ \lambda_{T-1}}_{\substack{\text{intermediate} \\ \text{distributions}}}, \quad \underbrace{\lambda_T = \lambda}_{\substack{\text{target} \\ \text{distribution}}} \quad \text{(End)}$$

- Let $N$ denote the number of samples and $\mathcal{N} = 1, \ldots, N$.
- At any time $t$, maintain weighted samples $\{W_t^{(n)}, X_t^{(n)} | n = 1, \ldots N\}$ and use these particles to generate samples from the target distribution at the next time point using sequential importance sampling.
- Motivation: It is easier to move the particles from one target to the next if $\lambda_t$ is close to $\lambda_{t+1}$.

# SMC sampler for GGMs

- Aim: sample from $p(G_1, \ldots, G_K|y) \propto \gamma(G_1, \ldots, G_K|y)$.
- Construct the sequence of target densities:

$$p(G_1, \ldots, G_K|y)^{\phi_1},\ p(G_1, \ldots, G_K|y)^{\phi_2},\ \ldots,\ p(G_1, \ldots, G_K|y)^{\phi_T},$$

  where $0 < \phi_1 < \phi_2 < \ldots \phi_T = 1$ are user-specified temperatures.
- At $t$, maintain samples $\{W_t^{(n)}, (G_1, \ldots, G_K)_t^{(n)} | n \in \mathcal{N}\}$ approximating $p(G_1, \ldots, G_K|y)^{\phi_t} \propto \gamma(G_1, \ldots, G_K)^{\phi_t}$.
- **Initialization** Sample each edge in $G_k$ independently with probability 0.5 for $k = 1, \ldots, K$ and repeat $N$ times. The weight of each sample is

$$w_1^{(n)} = \gamma((G_1, \ldots, G_K)_1^{(n)})^{\phi_1} 2^{rK},$$

  where $r = p(p-1)/2$.

# SMC Algorithm for multiple GGMs

- **Transition** Increase the temperature from $\phi_{t-1}$ to $\phi_t$ at time $t \geq 2$,

$$w_t^{(n)} = w_{(t-1)}^{(n)} \gamma((G_1, \ldots, G_K)_{t-1}^{(n)})^{\phi_t - \phi_{t-1}}.$$

- **Resampling** To prevent degeneracy of the particle approximation, resample when effective sample size, ESS $= \{\sum_{n=1}^{N} (W_t^{(n)})^2\}^{-1} < N/3$.
- **MCMC steps** Perform a small number of MCMC steps with invariant distribution $p_t(G_1, \ldots, G_K|y)$ to improve mixing. Generate candidates for each sample by flipping a small number, $M$, of randomly selected edges for each $G_k$.
- The SMC algorithm is easily parallelizable as computation of weights and the MCMC steps can be performed independently for the $N$ samples.

# SMC Algorithm for multiple GGMs

At $t = 1$,

- Draw $(G_1, \ldots, G_K)_1^{(n)}$ uniformly at random from the joint graphical space for $n \in \mathcal{N}$.
- Compute weights $\{w_1^{(n)}\}$ and normalized weights $\{W_1^{(n)}\}$.

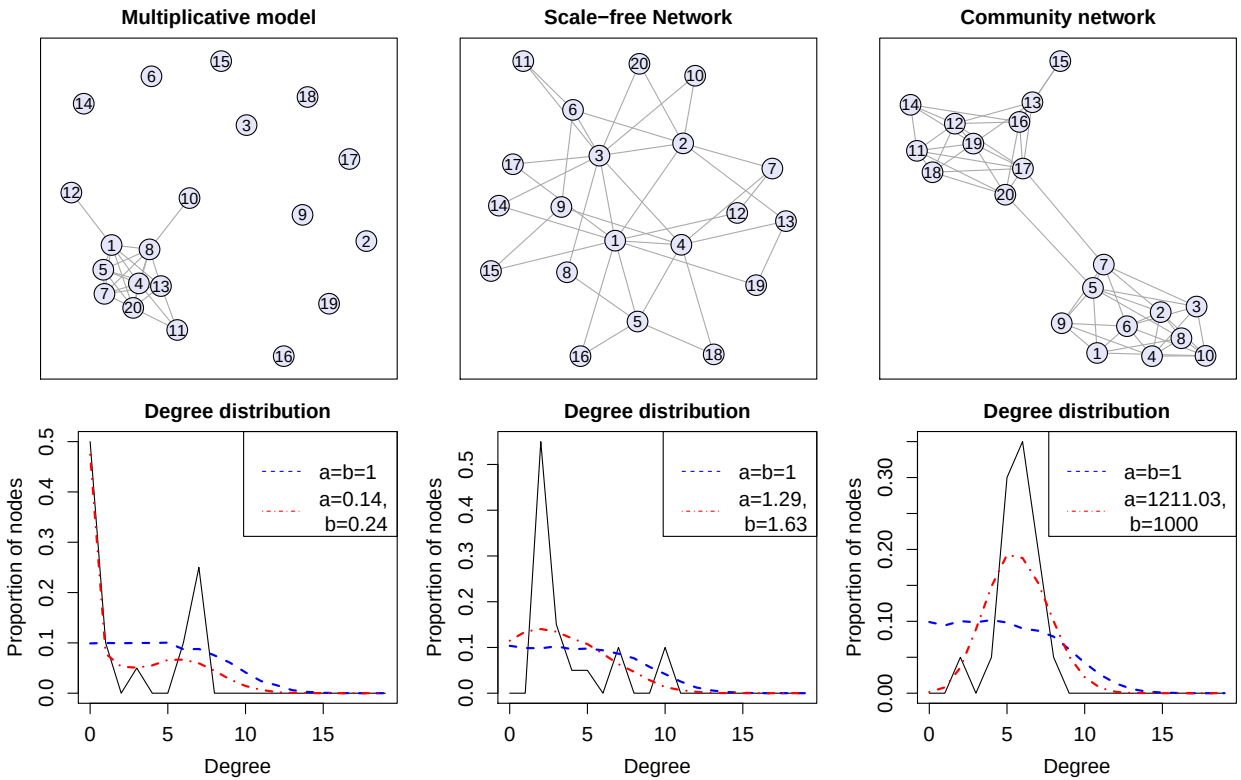
For $t = 2, \ldots, T$,

- Update weights $\{w_t^{(n)}\}$ and normalized weights $\{W_t^{(n)}\}$.
- If ESS $< N_{\text{threshold}}$, resample particles and set $W_t^{(n)} = 1/N$ for $n \in \mathcal{N}$.
- For $n \in \mathcal{N}$,
  - Randomly select $M$ edges from the set of all possible edges..
  - Set $(G_1, \ldots, G_K)_t^{(n)} = (G_1, \ldots, G_K)_{t-1}^{(n)}$.
  - For $m = 1, \ldots, M$, $k = 1, \ldots, K$, let $(G_1, \ldots, G_K)_c^{(n)}$ be sample candidate obtained from $(G_1, \ldots, G_K)_t^{(n)}$ by flipping the $m$th selected edge in $G_k$. Accept sample candidate with probability $\min\left[\{\gamma((G_1, \ldots, G_K)_c^{(n)})/\gamma((G_1, \ldots, G_K)_t^{(n)})\}^{\phi_t}, 1\right]$.

# Simulations

- Compare multiplicative prior with uniform and size-based prior.
- Set $K = 1$, $p = 20$ nodes. Generate data from three types of networks:

1. Multiplicative model: Generate $\pi_i \overset{\text{i.i.d.}}{\sim} \text{Beta}(0.1, 0.1)$, $A_{i,j} \sim \text{Bern}(\pi_i \pi_j)$.
2. Scale-free network: Generate using Barabási-Albert (BA) model.
3. Community network: Divide $p$ nodes into two equal-size groups.
   Assume within and across group interaction rates of 0.6 and 0.02.



# Simulations

- Ten datasets are simulated from the GGM by setting $H = 100$. The underlying network is regarded as the "true" graph.
- For multiplicative prior, consider (1) $a = b = 1$, (2) find $a$, $b$ such that shape of prior degree distribution resembles that of true graph.
- SMC sampler: set $N = 500$, $M = 3$, $\{\phi_t\} = (0.01, 0.02, \ldots, 1)$, $T = 100$. Average CPU time for each experiment is $(6.4 \pm 0.5)$ hours.
- Using the weighted samples, we compute the posterior probability of occurrence of each edge and summarize results using the area under ROC curves (AUC).

# Simulations

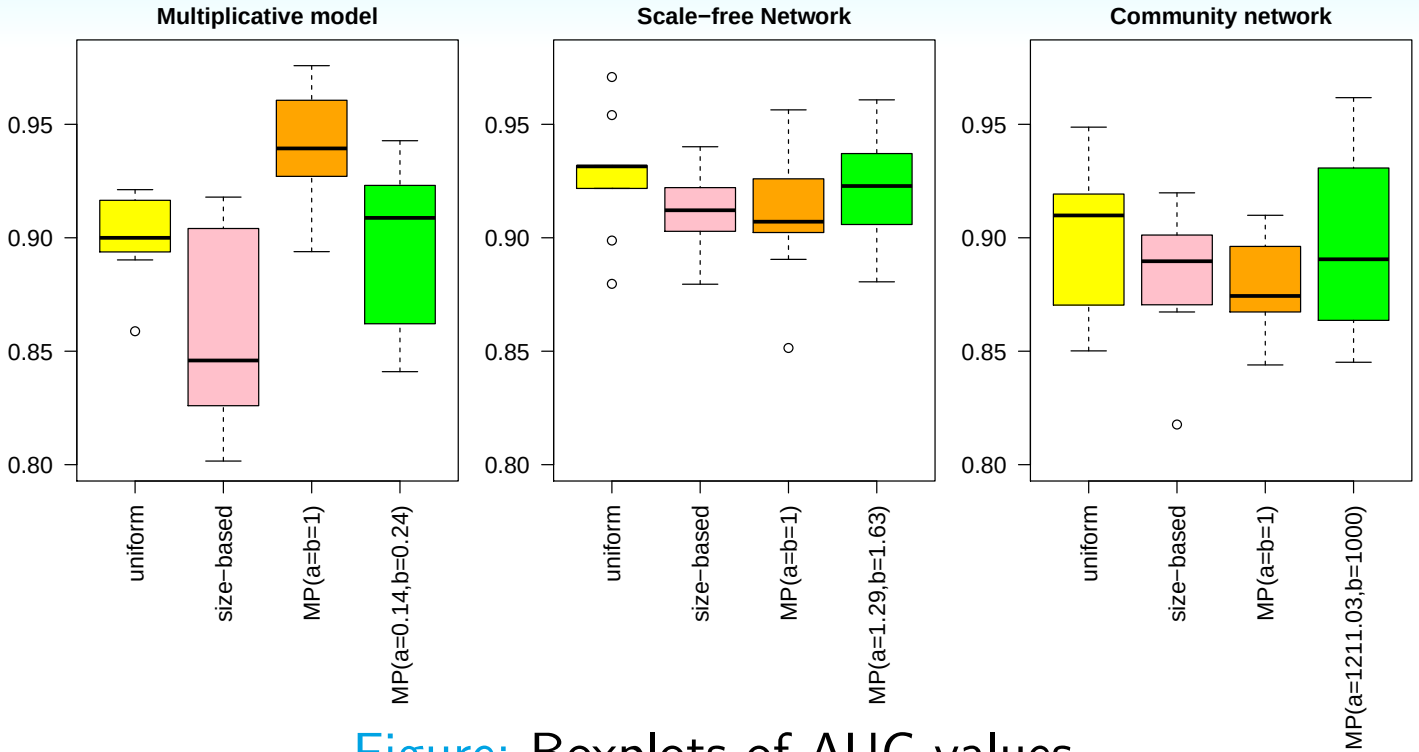


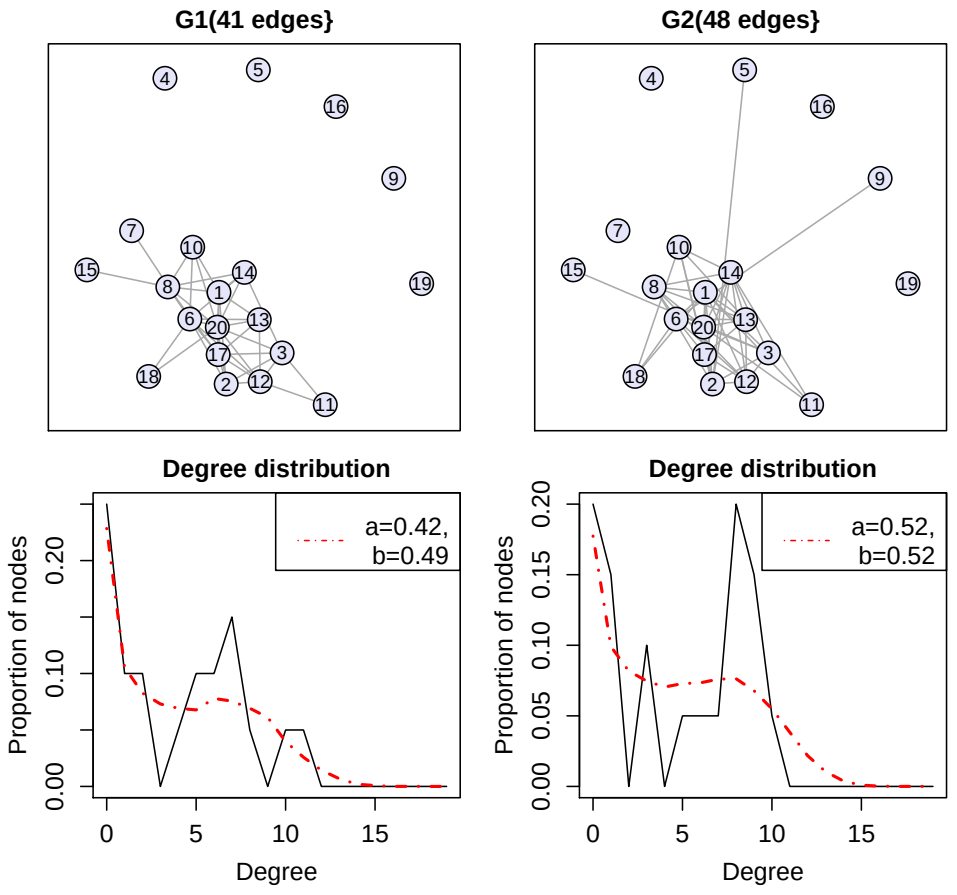

Figure: Boxplots of AUC values.

- The multiplicative priors performed better than the uniform and size-based priors for data simulated from the multiplicative model.
- For data simulated from scale-free and community networks, performance of different priors are quite similar. For these networks, the multiplicative prior performed better if hyperparameters were tuned to match the degree distribution of the true graph.

# Simulations

- Investigate ability of multiplicative prior to borrow information across graphs when the nodes have similar connectivity.
- Simulate 10 datasets each with $H = 100$ observations, $p = 20$ variables, $K = 2$ groups, $x_1 = (1, 0)$ and $x_2 = (1, 1)$.

- Assume 50 observations in each group. The graphs are generated from the multiplicative model setting $\sigma_1^2 = 10$ and $\sigma_2^2 = 0.01$.
- Connectivity of nodes in $G_1$ may vary over a wide range while connectivity of each node in $G_1$ and $G_2$ are similar.

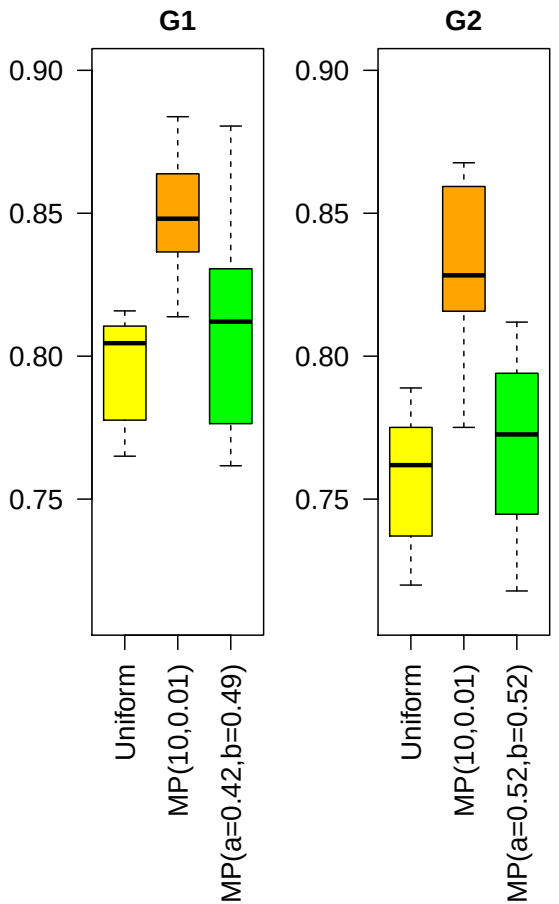


# Simulations

Compare

1. uniform prior (equal prior probability for each pair of graphs).
2. joint multiplicative prior with $\sigma_1^2 = 10$, $\sigma_2^2 = 0.01$
3. independent multiplicative priors for $G_1$ and $G_2$ with hyperparameters matched to degree distributions of true graphs.

- Using SMC algorithm (same settings), average CPU time for joint prior $(K = 2)$ is $(12.4 \pm 0.5)$ hours and for the independent multiplicative priors $(K = 1)$ is $(6.5 \pm 0.4)$ hours.
- The joint multiplicative prior performs better than the uniform prior and independent multiplicative priors indicating the ability of the multiplicative prior to encourage similarity in connectivity of nodes across graphs.




Boxplots of AUC values

# Urinary metabolic data

- Investigate correlation structure of $p = 22$ urinary metabolites in response to Cd exposure. Two analyses:
  1. Consider the individuals as a homogeneous group.
  2. Divide into two groups: $S_1$ (control group, Cd exposure $\leq$median) and $S_2$ (diseased group, Cd exposure $>$median).
- Use `R` package `GeneNet` to obtain fast shrinkage estimators of partial correlation in the network. The degree distributions provide a basis for determining appropriate hyperparameters for the multiplicative prior.
- The observations of each variable are first standardized. For SMC sampler, set $N = 500$, $M = 5$, $\{\phi_t\} = (0.005, 0.01, \ldots, 1)$, $T = 200$. Average CPU time for each experiment is $(24.7 \pm 3.0)$ hours for $K = 1$ and $(48.0 \pm 7.5)$ hours for $K = 2$.

# Case: $K = 1$

- Compare multiplicative prior ($a = b = 1$ and $a = b = 0.1$), size-based prior and uniform prior.
- Fit GGM and compute posterior probability of occurrence of each edge. The graphs obtained under multiplicative and size-based priors are similar and much sparser than that of the uniform prior.

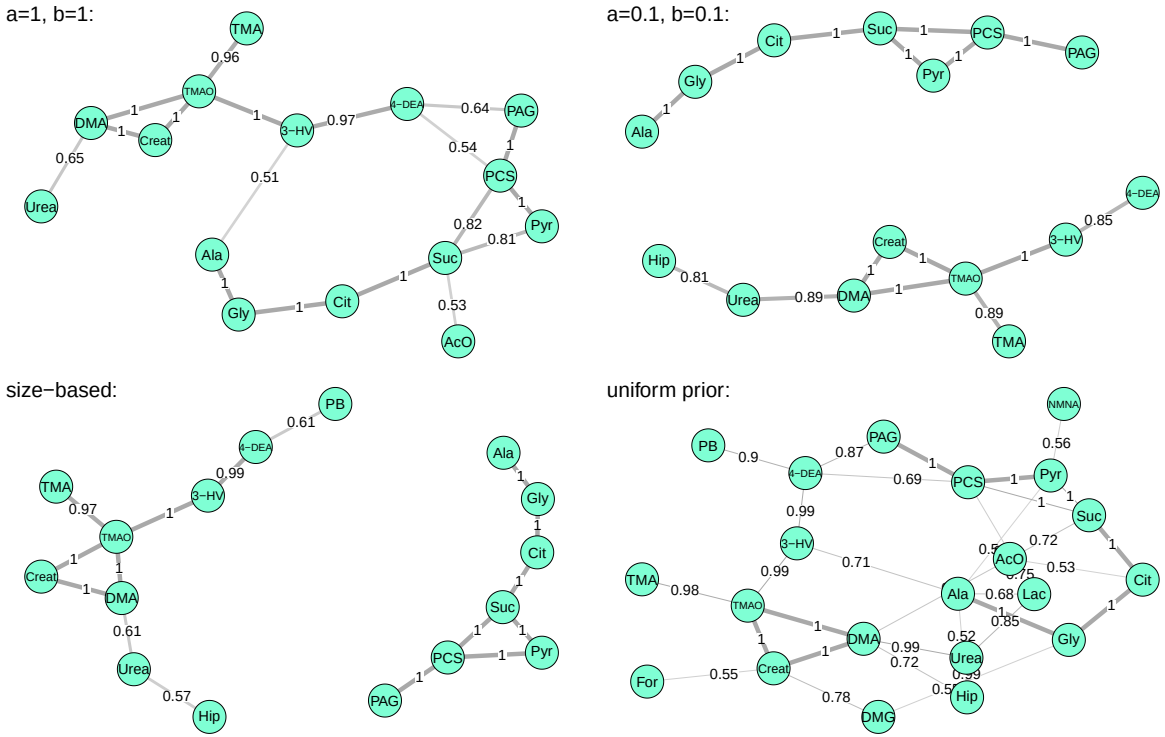



Figure: Only edges with posterior weights greater than 0.5 are shown. Width of each edge is proportional to its posterior probability.

# Case: $K = 1$

- For the multiplicative prior, we can obtain uncertainty measures of the tendency of each node to form connections with other nodes.
- $a = b = 0.1$ places too much prior weight at 0 and 1. $a = b = 1$ provides a better fit.

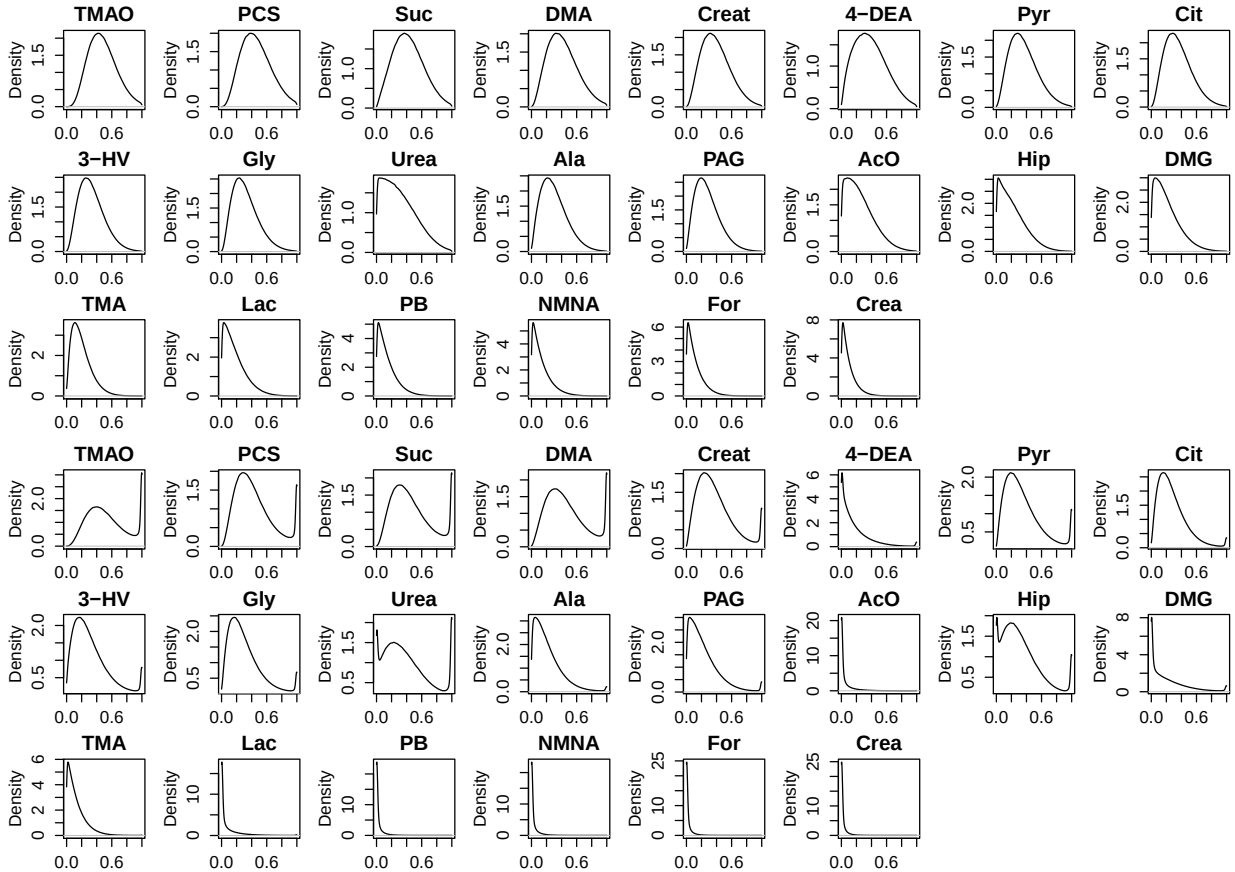



Figure: Posterior distribution of $\pi_i$ of each metabolite under multiplicative prior with $a = b = 1$ (first 3 rows) and $a = b = 0.1$ (last 3 rows).

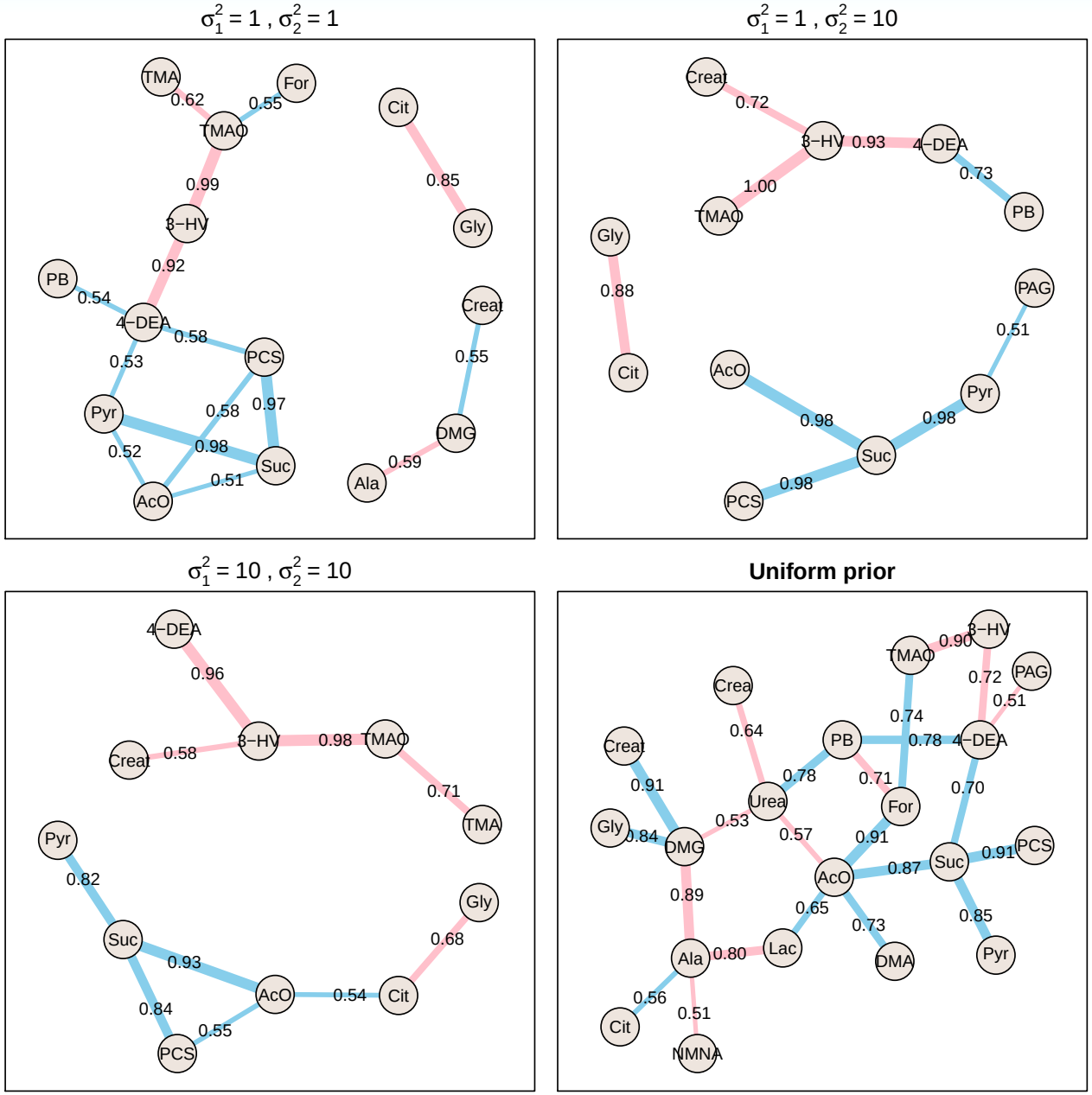
# Case: $K = 2$

- Difference in graphical structure between $G_1$ (Cd exposure $\leq$medium) and $G_2$ (Cd exposure $>$medium) is of interest.
- Consider four priors: multiplicative model with
  1. $\sigma_1^2 = \sigma_2^2 = 1$ (suggested by degree distributions from GeneNet)
  2. $\sigma_1^2 = 1$ and $\sigma_2^2 = 10$ (allow structure of $G_2$ to vary more from $G_1$)
  3. $\sigma_1^2 = \sigma_2^2 = 10$ (assumes connectivities are closer to 0 or 1)

  and the uniform prior.
- To compare $G_1$ and $G_2$, we construct differential networks which display only edges likely to appear in one graph but not the other. Let $\rho_{ij}^k$ be the posterior marginal probability of inclusion of edge $(i, j)$ in $G_k$. We estimate $\rho_{ij}^k$ as the proportion of SMC samples for which $(i, j)$ was included in $G_k$ and consider an edge as differential if $|\rho_{ij}^1 - \rho_{ij}^2| > \kappa$.

# Differential network for different priors

- Here we set $\kappa = 0.5$.
- Edges in blue are likely to appear in $G_1$ but not in $G_2$ and pink edges are likely to appear in $G_2$ but not in $G_1$. The labels indicate the estimate of $|\rho_{ij}^1 - \rho_{ij}^2|$ for each edge.



# Case: $K = 2$

- The full network ($K = 1$) and the differential network ($K = 2$) show similar topological characteristics corresponding to subgraphs of metabolites.
- $K = 2$: different prior hyperparameters lead to different levels of shrinkage, but there is similarity in biological interpretation – both figures show three different sub-graphs linking metabolites with shared origin.
  - A group of organic acids (Suc, Pyr, AcO, PCS) are connected, sometimes also with PAG.
  - The second group contains TMA and TMAO (part of choline metabolism), plus 3-HV and 4-DEA (products of amino acid catabolism). Choline is an essential nutrient metabolised in the liver. Due to its long biological half-life, Cd accumulates in human tissues, especially the liver and kidney, so this observation may point towards a possible mechanistic connection.

# Case: $K = 2$

  - The third group links Cit and Gly (closely associated in central carbon metabolism). A strong correlation between Cd and Cit was found by Ellis et al. (2012), while Valcarcel et al. (2011) found a significant deregulation of the dependency network associated with dimethylglycine, a biproduct of the synthesis of Gly from choline.

- It is plausible that several of the metabolites found in the full and differential networks are involved in pathways disregulated due to Cd exposure.
- Interpretation of dependency networks is difficult as metabolite associations derive from a variety of factors but they give us a novel view of the data not exposed in conventional analyses, and may help to generate new hypotheses to be investigated by future biochemical experiments.

# Conclusion

- We propose using the multiplicative random graph model as a prior on the graphical space of GGMs.
  - Encourage sparsity or particular degree structures, when such prior knowledge is available.
  - Accommodate a wider range of degree structures than the Erdős-Rényi model, e.g. right-skewed or U-shaped degree distributions, by varying the hyperparameters.
- Illustrate how this prior can be applied to both single and multiple GGMs
- Develop SMC sampler for posterior inference (stable and parallelizable).
- Multiplicative prior yields rich posterior inference, enabling a study of the connectivity of each node and how the propensity to connect varies across different experimental conditions.