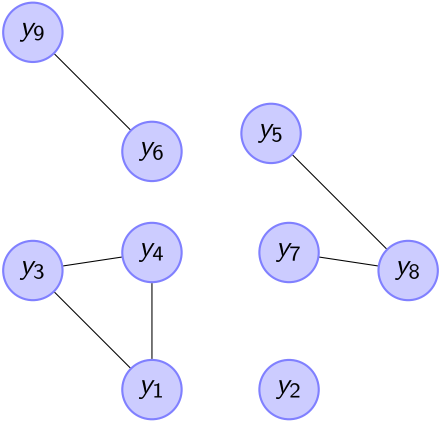
Undirected bayesian graph



There is an edge (i, j) if and only if $y_i \not\perp\!\!\!\perp y_j | y_{-ij}$

Setting and assumptions

We observe p gene expression levels across n patients. That is we have n observations $y_1, \ldots, y_n \in \mathbb{R}^p$ of a p dimensional vector. We assume that $y_1, \ldots, y_n \stackrel{i.i.d.}{\sim} \mathcal{N}_p(0, \Sigma)$ have a multivariate normal distribution.

- To estimate the graph we would like to know whether $y_i \not\perp\!\!\!\perp y_j | y_{-ij}$.
- Idea: We know that in the *precision matrix,* $\Omega = \Sigma^{-1}$ the (i, j)-th entry is zero only if $y_i \perp\!\!\!\perp y_j | y_{-ij}$
- How do we find the zero entries of Ω ?

Why should we use the Bayesian method ?

- We have access to the empirical precision matrix, we could derive statistical tests.
- This method becomes cumbersome when the number of parameters p becomes large.

If we consider Ω as being drawn from a prior distribution $p(\Omega)$ we can obtain a *posterior distrbution* $p(\Omega|X)$ of which the maximum is the *maximum à-posteriori* estimate Ω.

Spike and slab prior

The spike and slab prior for Ω helps us differentiate between zero and non-zero entries of Ω

$$
\begin{aligned}
y|\Omega &\sim N_p(0, \Omega^{-1}), \\
\omega_{ij}|\delta_{ij} &\sim \delta_{ij} N(0, v_1^2) + (1 - \delta_{ij}) N(0, v0^2) \text{for } i \neq j, \\
\omega_{ii} &\sim Exp(\lambda/2), \\
\delta_{ij}|\pi &\sim Bern(\pi), \\
\pi &\sim Beta(a, b).
\end{aligned}
$$

After a few manipulations we find that

$$
p(\Omega, \delta, \pi | y) \propto p(y|\Omega) p(\Omega|\delta) p(\delta|\pi) p(\pi)
$$