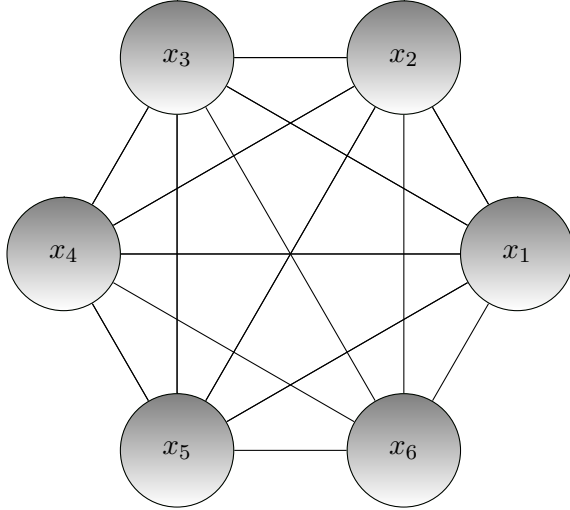


October 8, 2018

## 1 Gene network



- Each node corresponds to a gene
- Each edge corresponds to a pair of genes which are ‘co-expressed’

‘co-expressed’ meaning that their expression levels are highly correlated. The weights represent the strength of the connection between two genes.

True correlation between gene  $m, n$   $\rho_{mn}$  and correlation by observed data is  $r_{mn} = \text{corr}(x_m, x_n)$  by Fisher’s z-transformation  $w_{mn} = \text{arctanh}(r_{mn})$  approximately normal distribution .

$$w_{mn} \sim N(\text{arctanh}(\rho_{mn}), \frac{1}{N-3}) \quad (1)$$

Let  $E = \{e_{mn}\}$  be a set of true edges in the network.

The number of Gene  $G$  is larger and that most pairs are not co-expressed, and the number of possible edge  $K = G(G-1)/2$  is very larger than the number of true edges in the network  $|E|$

So in this gene network model we have some assumption

- This graphical model is undirected model because each node means correlation
- The network can be cycle??
- The network is sparse (  $|E| \ll K$  )

- The weight  $w_{mn}$  follows  $L_2N$  model

$L_2N$  model is a three component mixture model. First the 'null' component follow normal distribution with mean 0, which means majority of correlation between genes are 0. And the 'not null' components which mean strong positive/negative correlation follow log-normal distribution each.

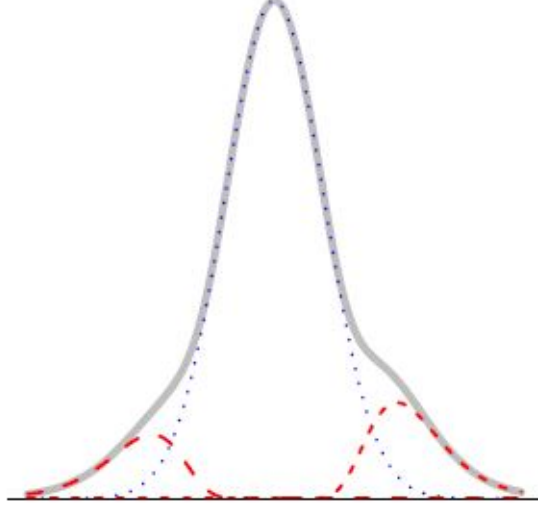


Figure 1:  $L_2N$  model

$$w_{mn}|e_{mn} \notin E \sim N(0, \sigma^2) \quad (2)$$

$$w_{mn}|[w_{mn} > 0, e_{mn} \in E] \sim \text{LogNormal}(\theta_1, \kappa_1^2) \quad (3)$$

$$-w_{mn}|[w_{mn} < 0, e_{mn} \in E] \sim \text{LogNormal}(\theta_2, \kappa_2^2) \quad (4)$$

$\sigma^2 = 1/(N - 3) + \sigma_0^2$ , where  $1/(N - 3)$  from Fisher's z-transformation and  $\sigma_0^2$  from random effect model.