

Gene Network inference for high-dimensional problems

A. Mohammadi and E. Wit

March 28, 2013

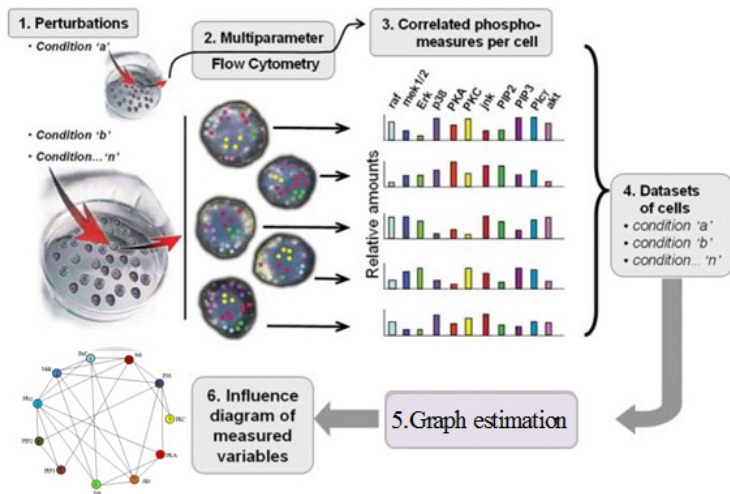


university of
 groningen

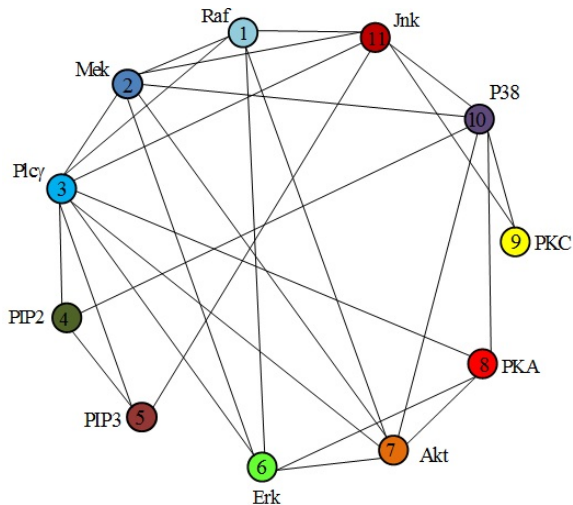
faculty of mathematics
 and natural sciences

MOTIVATION

Flow cytometry data with 11 proteins from Sachs et al. (2005)



RESULT FOR CELL SIGNALING DATA



PROBLEM IN BAYESIAN GRAPH ESTIMATION

$$p(G|data) = \frac{p(G)p(data|G)}{\sum_{G \in \mathcal{G}} p(G)p(data|G)}$$

Trans-dimensional MCMC in general

- ▶ Reversible-jump MCMC
- ▶ Birth-death MCMC

Our solution

- ▶ We proposed birth-death MCMC method for undirected graph estimation
- ▶ Implement to R : BDgraph package

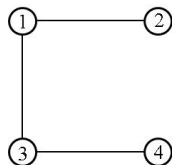
Gaussian graphical model

Respect to graph $G = (V, E)$ as

$$\mathcal{M}_G = \left\{ \mathcal{N}_p(0, \Sigma) \mid K = \Sigma^{-1} \text{ is positive definite based on } G \right\}$$

Pairwise Markov property

$$X_i \perp X_j \mid X_{V \setminus \{i,j\}} \Leftrightarrow k_{ij} = 0,$$



$$K = \begin{bmatrix} k_{11} & k_{12} & k_{13} & 0 \\ & k_{22} & 0 & 0 \\ & & k_{33} & k_{34} \\ & & & k_{44} \end{bmatrix}$$

BIRTH-DEATH PROCESS

- ▶ Spatial birth-death process: Preston (1976)



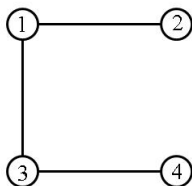
- ▶ Birth-death MCMC: Stephens (2000) in mixture models



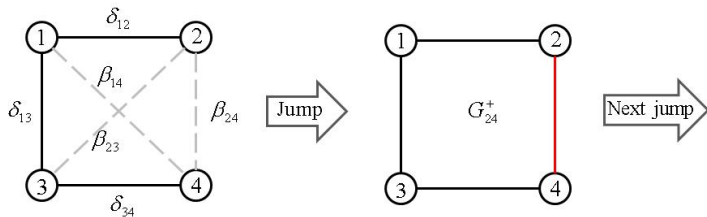
Birth-death process in GGM

- ▶ Adding new edge in birth and deleting edge in death time

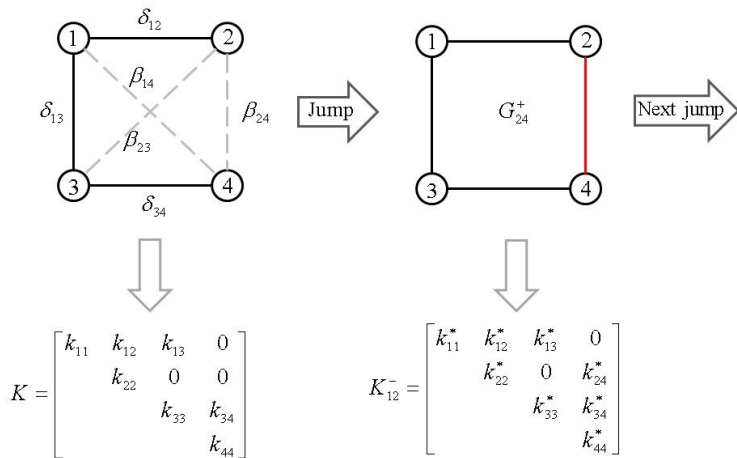
SIMPLE CASE



SIMPLE CASE



SIMPLE CASE



CANVERGENCY

Preston (1976): Backward Kolmogorov

Under Balance condition, process converges to unique stationary distribution.

Mohammadi and Wit (2013): BDMCMC in GGM

Stationary distribution = Posterior distribution of (G,K)

So, relative sojourn time in graph $G = p(G|data)$

PROPOSED BDMCMC ALGORITHM

Step 1: (a). Calculate birth and death rates

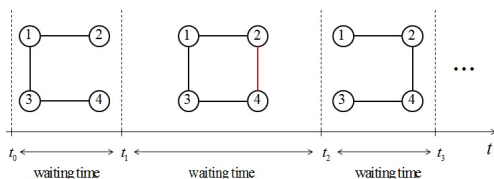
$$\beta_{\xi}(K) = \lambda_b, \quad \text{new link } \xi = (i, j)$$

$$\delta_{\xi}(K) = \frac{b_{\xi}(k_{\xi})p(G_{\xi}^{-}, K_{\xi}^{-} | \mathbf{x})}{p(G, K | \mathbf{x})} \lambda_b, \quad \text{existing link } \xi = (i, j)$$

(b). Calculate waiting time,

(c). Simulate type of jump, birth or death

Step 2: Sampling new precision matrix: K_{ξ}^{+} or K_{ξ}^{-}



PROPOSED PRIOR DISTRIBUTIONS

Prior for graph

- ▶ Discrete Uniform
- ▶ Truncated Poisson according to number of links

Prior for precision matrix

- ▶ G-Wishart: $W_G(b, D)$

$$p(K|G) \propto |K|^{(b-2)/2} \exp \left\{ -\frac{1}{2} \text{tr}(DK) \right\}$$

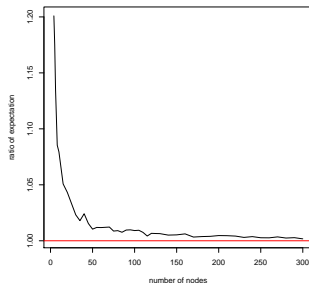
$$I_G(b, D) = \int_{\mathbb{P}_G} |K|^{(b-2)/2} \exp \left\{ -\frac{1}{2} \text{tr}(DK) \right\} dK$$

COMPUTING DEATH RATES

$$\delta_{\xi}(K) = \frac{I_G(b, D)}{I_{G_{\xi}^{-}}(b, D)} \left(\frac{|K^{-\xi}|}{|K|} \right)^{(b^*-2)/2} \exp \left\{ -\text{tr}(D^*(K^{-\xi} - K))/2 \right\} \gamma_b$$

$$\frac{I_G(b, D)}{I_{G^{-\xi}}(b, D)} = 2\sqrt{\pi} t_{ii} t_{jj} \frac{\Gamma((b+\nu_i)/2)}{\Gamma((b+\nu_i-1)/2)} \frac{E_G[f_T(\psi^{\nu})]}{E_{G^{-\xi}}[f_T(\psi^{\nu})]}$$

Plot for ratio of normalizing constants



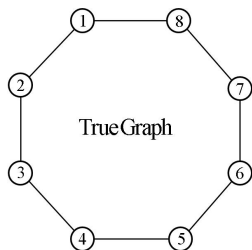
R PACKAGE

BDgraph package

- ▶ Graph estimation for high-dimensional cases
- ▶ Graph estimation for low-dimensional cases

SIMULATION: 8 NODES

$$\mathcal{M}_G = \{ \mathcal{N}_8(0, \Sigma) | K = \Sigma^{-1} \in \mathbb{P}_G \}$$

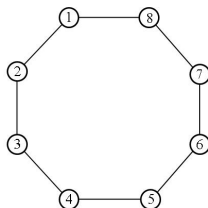
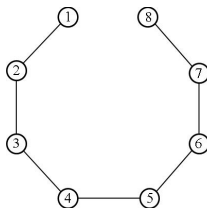


$$K = \begin{bmatrix} 1 & .5 & 0 & 0 & 0 & 0 & 0 & .4 \\ & 1 & .5 & 0 & 0 & 0 & 0 & 0 \\ & & 1 & .5 & 0 & 0 & 0 & 0 \\ & & & 1 & .5 & 0 & 0 & 0 \\ & & & & 1 & .5 & 0 & 0 \\ & & & & & 1 & .5 & 0 \\ & & & & & & 1 & .5 \\ & & & & & & & 1 \end{bmatrix}$$

SOME RESULT

Effect of Sample size

Number of data	20	30	40	60	80	100
$p(\text{true graph} \mid \text{data})$	0.018	0.067	0.121	0.2	0.22	0.35
false positive	0	0	0	0	0	0
false negative	1	0	0	0	0	0



SIMULATION: CIRCLE GRAPH WITH 120 NODES

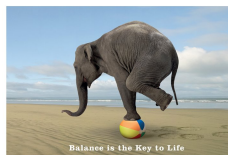
$$\mathcal{M}_G = \left\{ \mathcal{N}_{120}(0, \Sigma) \mid K = \Sigma^{-1} \in \mathbb{P}_G \right\},$$

- ▶ $n = 2000 \ll 7260$
- ▶ Priors: $K \sim W_G(3, I_{120})$ and $G \sim TU(\text{all possible graphs})$
- ▶ 10000 iterations and 5000 iterations as burn-in

Result

- ▶ Time: 4 hours
- ▶ $p(\text{true graph} \mid \text{data}) = 0.09$ which is most probable graph

SUMMARY



MOHAMMADI, A. AND E. C. WIT (2013) *Gaussian graphical model determination based on birth-death MCMC inference*, arXiv preprint arXiv:1210.5371v4