Exercises for Statistical analysis of network data – Sheet 8

1. The probability of edge formation in the latent class model is

$$\Pr\{A_{ij} = 1 | z_i, z_j, \theta_{ab}\} = \theta_{z_i z_j},$$

with $\theta_a b$ being the edge probability between a node with $z_i = a$ and $z_j = b$. This can be recovered from the general expressions (Hoff 2007)

$$\begin{aligned}
&\Pr\{A_{ij} = 1 | z_i, z_j, x_{ij}, \theta\} = \mathrm{E}(A_{ij}) = p_{ij}, \\
&g(p_{ij}) = \eta_{ij}, \\
&\eta_{ij} = \gamma + \beta^T x_{ij} + \alpha((z_i, z_j)
\end{aligned}$$

by choosing the identity link $g(\mu) = \mu$, $\gamma = 0$ and $\alpha(z_i, z_j) = \theta_{z_i z_j}$.

2. Define the latent variable z_i as an indicator vector of length k taking the value 1 at index a if node i is of class a and 0 everywhere else.. Define the $K \times K$ matrix Λ to be $\Lambda_{ij} = \theta_{ab}$ if node i belongs to block a and node j belongs to block b. Then $\Pr\{A_{ij} = 1\} = z_i^T \Lambda z_j$, which is a special case of the eigenmodel.

3. Less trivial: see Hoff (2018), Modeling homophily and stochastic equivalence in symmetric relational data, arXiv:0711.1146v1.

4. Define the latent variable z_i as an indicator vector of length k taking the value 1 at index a if node i is of class a and 0 everywhere else, and let Θ denote the block parameter matrix, with elements θ_{ab}.

$$\Pr\{\boldsymbol{A} = \boldsymbol{a} | \boldsymbol{z}, \Theta\} = \prod_{i<j} \prod_{a,b} \left[\theta_{ab}^{a_{ij}} (1 - \theta_{ab})^{1-a_{ij}}\right]^{z_{ia} z_{jb}},$$

from which the log-likelihood is

$$\ell(\Theta; \boldsymbol{a}, \boldsymbol{z}) = \sum_{i<j} \sum_{a,b} z_{ia} z_{jb} \left\{a_{ij} \log \theta_{ab} + (1 - a_{ij}) \log(1 - \theta_{ab})\right\}.$$

Assuming the indicator variables known, equating the derivatives of the log-likelihood with respect to θ_{ab} with zero, we obtain

$$\hat{\theta}_{ab} = \frac{\sum_{i<j} z_{ia} z_{jb} a_{ij}}{h}.$$

5. Letting the latent variables z_i be defined on any reasonable metric space with a distance metric $d(z_i, z_j)$, and specifying a threshold r,

$$\Pr\{A_{ij} = 1\} = \begin{cases} 1 & \text{if} \quad d(z_i, z_j) < r \\ 0 & \text{if} \quad d(z_i, z_j) \geq r. \end{cases}$$

6. Assume that the Poisson distributions of the edge weights have a different parameter according to which blocks their endpoints belong to. Denote this by the matrix Λ whose elements are λ_{ab}. The probability of a certain configuration $\boldsymbol{a}$ of the edge weight matrix which summarises the information in an observed graph, is now

$$\Pr\{\boldsymbol{A} = \boldsymbol{a} | \boldsymbol{z}, \theta_{ab}\} = \prod_{i<j} \prod_{a,b} \left[\frac{\lambda_{ab}^{a_{ij}} \exp(-\lambda_{ab})}{a_{ij}!}\right]^{z_{ia} z_{jb}},$$

from which the log-likelihood is now

$$\ell(\Lambda; \boldsymbol{a}, \boldsymbol{z}) = \sum_{i<j} \sum_{a,b} z_{ia} z_{jb} \left\{a_{ij} \log \lambda_{ab} - \lambda_{ab} - \log a_{ij}!\right\}.$$

Introducing the notations as in the lecture

$$\bar{A}_{ab}(\boldsymbol{z}) = \frac{2}{h_{ab}} \sum_{i<j} z_{ia} z_{jb} a_{ij},$$
$$h_{ab}(\boldsymbol{z}) = \sum_{i<j} z_{ia} z_{jb},$$

we obtain the log-likelihood form

$$\ell(\Lambda; \boldsymbol{a}, \boldsymbol{z}) = \sum_{a,b} \left\{ h_{ab}(\boldsymbol{z}) \bar{A}_{ab}(\boldsymbol{z}) \log \lambda_{ab} - h_{ab}(\boldsymbol{z}) \lambda_{ab} - c \right\},$$

where c is constant with respect to the model parameters. The partial derivatives with respect to λ_{ab} are

$$\frac{\partial \ell(\Lambda; \boldsymbol{a}, \boldsymbol{z})}{\partial \lambda_{ab}} = h_{ab}(\boldsymbol{z}) \bar{A}_{ab}(\boldsymbol{z}) \lambda_{ab}^{-1} - h_{ab}(\boldsymbol{z}),$$

from which the estimator of λ_{ab} is

$$\hat{\lambda}_{ab} = \bar{A}_{ab}(\boldsymbol{z}).$$

Profiling this would consist of taking all possible configurations of $\boldsymbol{z}$ and computing the log-likelihood value at each. The final estimator would then be the one belonging to the $\hat{\boldsymbol{z}}$ which maximises the log-likelihood.