Solutions for Statistical analysis of network data – Sheet 7

Recall that the Hamming distance between two adjacency matrices A_1 and A_2 is

$$d_1(A_1, A_2) = \frac{\|A_1 - A_2\|_1}{n^2},$$

and the Jaccard distance between A_1 and A_2 is

$$d_J(A_1, A_2) = \frac{\|A_1 - A_2\|_1}{\|A_1 + A_2\|_*},$$

where $\|.\|_*$ is the nuclear norm.

1. Substituting the two matrices into the formulae above, we get $d_1(A_1, A_2) = 4/36$ and $d_J(A_1, A_2) = 0.22$. The estimated edge density is $\rho_1 = 9/15$ for A_1 and $\rho_2 = 11/15$ for A_2.

2. Substituting the two matrices into the formulae, we get $d_1(A_1, A_2) = 4/36$ (similarly as in Problem 1) and $d_J(A_1, A_2) = 0.36$. The estimated edge density is $\rho_1 = 5/15$ for A_1 and $\rho_2 = 7/15$ for A_2; the relative change in edge density is higher in this case than in problem 1 , and this is reflected by the increased Jaccard distance value.

3. The exponential random graph formula:

$$\Pr\{\mathbf{A} = \mathbf{a}\} = \left(\frac{1}{\kappa}\right) \exp\left\{\sum_H \theta_H g_H(\mathbf{a})\right\}.$$

   (a) Requiring independence is equivalent to require that the above probability be of a product form in the terms of a. Any function g_H defined on any subgraph H other than a single edge is therefore excluded.

   (b) Let K_2 denote the complete graph on 2 nodes (that is, the edge graph). Then

$$\begin{aligned}\Pr\{\mathbf{A} = \mathbf{a}\} &= \left(\frac{1}{\kappa}\right) \exp\left\{\sum_H \theta_H g_H(\mathbf{a})\right\} \\ &= \left(\frac{1}{\kappa}\right) \exp\left\{\sum_{K_2} \theta_{K_2} g_{K_2}(\mathbf{a})\right\} \\ &= \left(\frac{1}{\kappa}\right) \exp\left\{\sum_{i>j} \theta_{K_2} g_{K_2}(a_{ij})\right\}.\end{aligned}$$

   (c) The Erdős-Rényi model can be recovered by choosing

$$\begin{aligned}g_{K_2}(a_{ij}) &= a_{ij}, \\ \theta_{K_2} &= \log \frac{p}{1-p}, \\ \kappa &= (1-p)^{-\binom{n}{2}}\end{aligned}$$