

Exercises for Statistical analysis of network data-Sheet 4

1. Starting from the model of an exchangeable network draw a latent variable ξ_i iid uniform. Then draw realisations of

$$A_{ij} \mid \boldsymbol{\xi} = \text{Bernoulli} (g(\xi_i) g(\xi_j)) .$$

This is the exchangeable version of the Chung-Lu model. Calculate the expected number of copies of a path on three nodes and two edges.

2. Assume that

$$A_{ij} \mid \Xi = \text{Bernoulli} (\boldsymbol{\xi}_i^T \boldsymbol{\xi}_j)$$

where (ξ_i) is iid from some k -variate distribution. Calculate the mean and variance of d_i .

3. Now assume that we draw realisations from the sparse exchangeable model of Bollobas and Riordan (2007):

$$A_{ij} \mid \boldsymbol{\xi} = \text{Bernoulli} (\rho_n f(\xi_i, \xi_j)) .$$

Repeat the calculations of problem 1 with this new framework. If $\rho_n \rightarrow 0$ as $n \rightarrow \infty$ what is the limit of the expected number? Starting from this we argue that unless we define the homomorphism count more carefully, we will just get degenerate properties of the count. We always assume $\|f\|_1 = 1$.

4. We noted the distribution of $\hat{\rho}$ for the Erdos-Renyi model. Construct a confidence interval for ρ based on this distribution.
5. Assume that $\rho = 1/n^{1.5}$. Determine the mean square error (variance plus bias square) in estimating ρ by the proportion of present edges. Compare with the mean square error of estimating ρ as zero. Which is preferred and why?

6. Now assume that

$$A_{ij} \mid \boldsymbol{\pi} \sim \text{Bernoulli} (\pi_i \pi_j)$$

where $\boldsymbol{\pi}$ is a deterministic vector. Determine the mean, variance and covariance of d_i and d_j if we draw realizations from that model.