Exercises for Statistical analysis of network data – Sheet 2

1. Assume we draw observations from a configuration model with

$$\pi_i = \frac{\theta_n}{i^\gamma}, \quad i = 1, \ldots, n.$$

Calculate the degrees from the realized adjacency matrix A via

$$d_i = \sum_{j \neq i} A_{ij}.$$

(a) We find that

$$\begin{aligned} \mathrm{E}(d_i) &= \sum_{l \neq i} \pi_i \pi_l \\ &= \pi_i \sum_{l \neq i} \pi_l \\ &= \theta_n^2 \frac{1}{i^\gamma} \sum_{l \neq i} \frac{1}{l^\gamma} = \theta_n^2 \frac{1}{i^\gamma} \begin{cases} \zeta(p) & \text{if} \quad p > 1 \\ \log(n)(1+o(1)) & \text{if} \quad p = 0 \\ \frac{n^{1-p}}{1-p}(1+o(1)) & \text{if} \quad 0 < p < 1 \end{cases}. \end{aligned} \tag{1}$$

(b) The variance of d_i is found from determining the second moment. We have

$$\begin{aligned} \mathrm{E}(d_i^2) &= \sum_{l \neq i} \sum_{k \neq i} \mathrm{E} A_{il} A_{ik} \\ &= \sum_{l \neq i} \sum_{k \neq i} \left(\mathrm{E} A_{il} \mathrm{E} A_{ik} (1 - \delta_{lk}) + \mathrm{E} A_{il} \delta_{lk} \right) \\ &= \sum_{l \neq i} \sum_{k \neq i} \left(\pi_i^2 \pi_l \pi_k (1 - \delta_{lk}) + \pi_i \pi_l \delta_{lk} \right) && (2) \\ &= \theta_n^4 \frac{1}{i^{2\gamma}} \sum_{l \neq i} \sum_{k \neq i,l} \frac{1}{(il)^\gamma} + \theta_n^2 \frac{1}{i^\gamma} \sum_{l \neq i} \frac{1}{l^\gamma} && (3) \end{aligned}$$

and then the standard formula of the variance can be applied.

$$\begin{aligned} \mathrm{Var}\{d_i\} &= \mathrm{E}(d_i^2) - \mathrm{E}^2(d_i) \\ &= \theta_n^4 \frac{1}{i^{2\gamma}} \sum_{l \neq i} \sum_{k \neq i,l} \frac{1}{(kl)^\gamma} + \theta_n^2 \frac{1}{i^\gamma} \sum_{l \neq i} \frac{1}{l^\gamma} - \theta_n^2 \frac{1}{i^\gamma} \sum_{l \neq i} \frac{1}{l^\gamma} \theta_n^2 \frac{1}{i^\gamma} \sum_{l' \neq i} \frac{1}{(l')^\gamma} && (4) \\ &= \theta_n^2 \frac{1}{i^\gamma} \sum_{l \neq i} \frac{1}{l^\gamma} - \theta_n^4 \frac{1}{i^{2\gamma}} \sum_{l \neq i} \frac{1}{l^{2\gamma}} && (5) \\ &\asymp \theta_n^2 \frac{1}{i^\gamma} \begin{cases} \zeta(p) & \text{if} \quad p > 1 \\ \log(n)(1+o(1)) & \text{if} \quad p = 0 \\ \frac{n^{1-p}}{1-p}(1+o(1)) & \text{if} \quad 0 < p < 1 \end{cases}, && (6) \end{aligned}$$

unless $\gamma = 0$ or $\gamma > 1$ which we do not allow.

(c) The dispersion is the ratio of the mean to the variance. By comparing the mean and variance we see this is the case if $0 < \gamma < 1$.

(d) The covariance can be computed like the variance from first principles: We have

$$\begin{aligned} \mathrm{E}(d_i d_j) &= \sum_{l \neq i} \sum_{k \neq j} \mathrm{E} A_{il} A_{jk} \\ &= \sum_{l \neq i} \sum_{k \neq j} \left(\mathrm{E} A_{il} \mathrm{E} A_{jk} (1 - \delta_{lk} \delta_{ij} - \delta_{ik} \delta_{lj}) + \mathrm{E} A_{il} A_{jk} (\delta_{lk} \delta_{ij} + \delta_{ik} \delta_{lj}) \right) \end{aligned}$$

and then the standard formula of the co-variance can be applied, and simplified like before.

2. Show that the degree–corrected stochastic blockmodel is exchangeable.

   We note from class that we have defined a connection probability matrix Θ which has entries θ_{ab} for $1 \leq a < b \leq k$. Then with 1-d function $g(x)$ we draw

$$A_{ij} | z_i,\, z_j, \xi_i, \xi_j = \text{Bernoulli}(\theta_{z_i z_j} + g(\xi_i) g(\xi_j)), \quad 1 \leq j < i \leq n. \tag{7}$$

   We note that the distribution of z_i and ξ_i is unchanged when we apply a permutation to i and so deduce the result.

3. Calculate the expected degree of node i from the random dot product graph.

   We note that the random dot product graph takes the form

$$\mathrm{E}\left\{A_{ij} \,|\, \boldsymbol{\Xi}\right\} = \rho_n \cdot \boldsymbol{\xi}_i^T \boldsymbol{\xi}_j,$$

   where the latent position of node i, namely $\boldsymbol{\xi}_i$ is generated by probability density function $f(\boldsymbol{\xi})$ iid. We therefore find that

$$\begin{aligned} \mathrm{E} d_i &= \sum_{j \neq i} \mathrm{E} \rho_n \cdot \boldsymbol{\xi}_i^T \boldsymbol{\xi}_j \\ &= \rho_n \sum_{j \neq i} \mathrm{E} \boldsymbol{\xi}_i^T \mathrm{E} \boldsymbol{\xi}_j. \end{aligned} \tag{8}$$

4. Reformulate the stochastic block model as a random dot product graph. How do we select d the latent dimension of the RDGP relative to k the number of blocks.

   We make a discrete measure. Note that a SBM has θ_{ab} as the connection probability if one node is in group a and the other in group b. We say an RDPG with latent positions $\boldsymbol{\xi}$ is an SBM with k blocks if the number of distinct rows in $\boldsymbol{\Xi}$ is k, denoted $\boldsymbol{\Xi}(1), \ldots, \boldsymbol{\Xi}(k)$. In this case, we define the block membership function $\tau : t\ \{1, \ldots, n\} \mapsto \{1, \ldots, K\}$ be a function such that $\tau(i) = \tau(j)$ if and only if $\boldsymbol{\Xi}(i) = \boldsymbol{\Xi}(j)$.

5. What is the size of the automorphism group of an edge? This is 2. We can put either of the nodes at either place.

6. What is the size of the automorphism group of a triangle? This is 6. We can put any node in any of the three slots, then any of the two for the remaining two nodes, this yields 3! which is 6.