**Solutions 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) Calculate the expectation of $d_i$. You may use the result that for large $n$ (Gradshteyn + Ryzkin 0.233, and an integral squeezing argument):

$$\sum_{j=1}^{n} \frac{1}{j^p} \approx \begin{cases} \zeta(p) & \text{if} \quad p > 1 \\ \log(n)(1+o(1)) & \text{if} \quad p = 1 \\ \frac{n^{1-p}}{1-p}(1+o(1)) & \text{if} \quad 0 < p < 1 \end{cases}.$$

*Solution:We find that*

$$\mathrm{E}\,(d_i) = \sum_{l \neq i} \pi_i \pi_l = \pi_i \sum_{l \neq i} \pi_l$$

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

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

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

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

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

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

(c) Calculate the dispersion of $d_i$. Are there instances when the dispersion approaches unity ?
*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) What is the covariance of $d_i$ and $d_j$ if $i \neq j$ ?
*The covariance can be computed like the variance from first principles: We have*

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

*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.
*Solution: The probability of an edge between two nodes $i$ and $j$ in the DC-SBM is given by*

$$A_{ij} \mid z_i, z_j, \xi_i, \xi_j = \text{Bernoulli}\left(\theta_{z_i z_j} \cdot g\left(\xi_i\right) g\left(\xi_j\right)\right)$$

*where $z_i$ and $z_j$ are the community labels for nodes $i$ and $j, \xi_i$ and $\xi_j$ are their degree-correction factors, and $g(x)$ is a function that modifies the connection probability based on the degree correction. To check exchangeability, we need to determine if permuting the node indices changes the overall distribution of the adjacency matrix. Both the community labels $z_i$ and the degree-correction variables $\xi_i$ are drawn independently for each node, and identically within their own class (i.e. hte $\xi_i$'s are iid and the $z_i$'s are iid). When we permute the node indices, the sets of block labels and degree-correction parameters remain identically distributed, meaning that the connection probabilities between nodes are not affected by this permutation. Since the edge probabilities depend only on the values of $z_i, z_j, \xi_i, \xi_j$, and these values are distributed independently of the node indices, permuting the indices does not change the joint distribution of the edges. Therefore, the adjacency matrix has the same distribution after any permutation of the node labels, demonstrating that the DC-SBM is exchangeable.*

3. Calculate the expected degree of node $i$ from the random dot product graph.
*Solution: We note that the random dot product graph takes the form*

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

*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*

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

4. Reformulate the stochastic block model as a random dot product graph. How do we select the latent dimension, call it $d$, of the RDGP relative to $k$ the number of blocks.
*Solution: We make a discrete measure. Note that a SBM has $\theta_{ab}$ 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 $\Xi$ is $k$ , denoted $\Xi(1), \ldots, \Xi(k)$. 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 $\Xi(i) = \boldsymbol{\Xi}(j)$*

5. What is the size of the automorphism group of an edge ?
*Solution: 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 ?
*Solution: 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.*