EPFL

Course: MATH-448
Setter: Olhede

EXAMINATIONS

August 2020

MATH-448

Statistical analysis of network data

Ecole Polytechnique Federale de Lausanne

August 2020

MATH-448

Statistical analysis of network data

Date: Mock exam, 2020 Time: Mock exam, 3 h

All that can be used for this exam is a pen. No books, notes, summaries, formula collections or calculators are allowed. All questions should be answered.

All questions are marked out of a total of 1 point, bringing the total of the exam to a total of 5 points.

1. Assume we observe a network $\mathcal{G}$ on five nodes which has edges $(1,2)$, $(2,3)$, $(3,1)$, $(2,4)$, $(2,5)$ and $(3,4)$.

   (a) Plot this network and number the nodes in the diagram.

   (b) Please write down the adjacency matrix of this network.

   (c) Please write down the incidence matrix of this network.

   (d) Determine the degree of each node.

   (e) Determine the number of triangles in the network $\mathcal{G}$.

2. (i) Assume we observe a random graph $\mathcal{G}$ with n nodes whose $n \times n$ adjacency matrix A is generated according to the following mechanism. Assume we generate n independent uniform random variables ξ_i. Then for function $0 \leq g(x) \leq 1$ A is generated by

$$\Pr\{A_{ij} = 1\} = g(\xi_i)g(\xi_j), \quad 1 \leq i < j \leq n. \tag{1}$$

   (a) Write down the degree of node i in terms of the entries of A. Calculate the expectation of this degree.

   (b) Calculate the variance of d_i.

   (c) Determine the covariance of the ith degree d_i with the jth degree d_j.

   (ii) (a) Write down the definition of a finitely exchangeable graph model.

   (b) Show that the network model in part (a) is finitely exchangeable.

3. Consider a graph with adjacency matrix A on 5 nodes corresponding to

$$A = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}.$$

   (i) Form the $n \times n$ symmetric matrix of all the distances between the nodes in this network.

   (ii) Give the definition of and compute the closeness centrality (you do not have to simplify the expression).

   (ii) Give the definition of and compute the harmonic centrality (you do not have to simplify the expression).

4. Assume that we observe network $\mathcal{G}$ with adjacency matrix A that are realized using the Bernoulli distribution conditionally on a group membership. We will assume that it is generated by the K group stochastic blockmodel with parameters h_k (size of group k), group membership vector $z_i \in \{1, \ldots, K\}$ for $i = 1, \ldots, n$, as well as interaction matrix $\{\theta_{ab}\}$ for $1 \leq a \leq b \leq K$.

   (i) Please write down the likelihood of this model.

   (ii) Solve for θ_{ab} if z is known, this giving a maximum likelihood estimate conditionally on z. Form the profile likelihood by substituting in that estimate $\widehat{\theta}_{ab}(\widehat{z})$.

   (iii) Describe how to estimate z using spectral clustering.

5. Assume that we observe a network $\mathcal{G}$ with adjacency matrix A that are realized using the Bernoulli distribution conditionally on a group membership and degree structure π. We will assume that it is generated by the K group stochastic blockmodel with parameters h_k (size of group k), group membership vector $z_i \in \{1, \ldots, K\}$ for $i = 1, \ldots, n$, as well as interaction matrix $\{\theta_{ab}\}$ for $1 \leq a \leq b \leq K$, and that a degree correction is added on.

   (i) Describe how to estimate the parameters of the degree–corrected Stochastic BlockModel (SBM) using maximum likelihood.

   (ii) Calculate the expected degree of node i.

   (iii) Generalise the degree-corrected SBM into a directed model. Write down the probability distribution of the edges to form the likelihood of the observations.