Assignment $\mathcal{N}^{\underline{o}}$ 2

released: 24.10.2022 at 19:00 **due:** 09.11.2022 at 15:00

Task 1: ERGM distribution 4 points

Let $N = \{i, j, h\}$ be a set of three nodes with node attribute color v_i as $\{$ⓘ ⓙ ⓗ$\}$ and R an undirected relation.

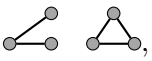
(1) Draw all the possible undirected networks that can be defined on N

(2) Consider the ERGM

$$P(X = x; \theta, v) = \frac{1}{\kappa(\theta)} \exp\Big\{\theta_1 S_1(x) + \theta_2 S_2(x, v) + \theta_3 S_3(x) + \theta_4 S_4(x)\Big\} ,$$

with $S_1(x)$ the number of edges, $S_2(x, v)$ the number of homophilous edges with respect to v, $S_3(x)$ the number of two-stars and $S_4(x)$ the number of triangles in the network x.
Give the mathematical formula for each network statistic. The formula shows how the network statistic is a function of the presence or absence of a tie x_{ij} and node attribute v_i variable.

(3) Compute the probability of each undirected network given that $\theta_1 = -\log(0.5)$, $\theta_2 = \log(2)$, $\theta_3 = \log(0.6)$ and $\theta_4 = -\log(0.8)$

(Two-star and triangle are defined by the local configurations  , respectively).

(4) Does the empty or full network have the highest probability under the network model described in (2) and (3)? If it is not the case, which is the network with the highest probability and why?

Task 2: Simulation from an ERGM **6 points**

The file `MHSim.R` contains the code to implement the Metropolis algorithm to simulate networks from the following ERGM:

$$P(X = x; \theta, v) = \frac{1}{\kappa} \exp\left\{\theta_1 \sum_{ij} x_{ij} + \theta_2 \sum_{i<j} x_{ij}x_{ji} + \theta_3 \sum_{ij} x_{ij}\mathbb{1}(v_i == v_j)\right\} .$$

with statistics the number of edges, reciprocal dyads, and homophily dyads on node attribute v. $\mathbb{1}(.)$ denotes the indicator function that takes value one when the argument is true and value zero otherwise.

(1) Some parts of the code are missing as denoted by the chunk code - - - MISSING - - -. Implement these in the R script, and include comments explaining what your code is doing.
(Please do not modify existing code even though more efficient solutions can be implemented.)

(2) The file `friend_net.Rda` contains an sna network object. A researcher suggested that plausible estimates of the parameters of the ERGM above for the friendship network are $\theta_1 = -2.76$, $\theta_2 = 0.68$ and $\theta_3 = 1.21$.

   i. Use the code developed in (1) to simulate friendship networks from the ERGM with parameters $\theta_1 = -2.76$, $\theta_2 = 0.68$ and $\theta_3 = 1.21$ using as node covariate the sex of the students coded as (1 = Girls, 0 = Boys). .
   (Use the functions `as.matrix()` and `get.node.attr()` to extract the adjacency matrix and sex covariate from the sna network object.)

   ii. Based on the simulations, do you think that the suggested values of the parameters are plausible estimates? Argue for your answer.

(3) Guess better estimates of θ_1, θ_2 and θ_3 based on the analysis in (2). Describe the procedure you used to obtain the guessed values.
(Please use the code and the analysis in (1), and (2). Obtaining better values using the `ergm` function is not considered a valid solution.)

Task 3: Estimation and interpretation of an ERGM **10 points**

The data set `friend_net.Rda` is a friendship network of high-school students. The network is directed as each student was asked to nominate other students as friends.
The data contains information of the network and the gender of the students (1 = Girls, 0 = Boys).

(1) Estimate an ERGM with an edge and a gender homophily parameter. Compute the conditional probability of observing a tie between two students i and j having the same gender and interpret the result.

(2) Specify an ERGM model to test simultaneously the following hypotheses:

   i. A tie is more likely between students when it reciprocates a friendship nomination (reciprocity).

   ii. A tie is more likely between students when it closes a transitive two-path (transitivity).

   iii. A tie is less likely between students when the sender has a higher out-degree (social activity).

   iv. A tie is more likely between students when the receiver has a higher in-degree (popularity).

   v. A tie is more likely between students with same gender (homophily).

(3) Estimate the ERGM specified in (2) and comment on the convergence of the algorithm.

(4) Evaluate the goodness of fit of the model according to four different criteria.

(5) Interpret the estimated parameters.

(6) Could you think of two other hypotheses that could be tested using ERGMs? State those hypotheses and provide the mathematical formula and the graphical representation of the effects that you need to include in the ERGM to test those hypotheses.

You are encouraged to work in groups of 3 or 4 people.
Please submit your solution (including R scripts in a file .R or .Rmd!) using moodle. Only one member of the group should submit the solution. Do not forget to report the names of all the group members in the documents you submit.