Assignment № 1

released: 23.10.2023 at 19:00   **due:** 08.11.2023 at 12:00

Task 1: MR-QAP logit regression   **10 points**

The data set `friend_net_v2.Rda` contains a friendship network of 31 high-school students. The network is directed as each student was asked to nominate other students as friends.
The following information are provided:

- `friend_net`: a network object containing friendship nominations

- `attributes`: a data frame containting the following attributes of students:

   - *id*: the identifier of the student
   - *sex*: self-reported gender (0 = girls, 1 = boys)
   - *activity*: the type of extracurricular activity that the student participates in (1 = hockey, 2 = chess, 3 = football)
   - *smoke*: whether the student reported having used any tobacco products during the past three months (0 = none, 1 = at least once)

(1) Import the data. Build a QAP to test if a friendship nomination is more likely between a pair of students with same gender. (Use the functions `as.matrix()` and `get.node.attr()` to extract the adjacency matrix and sex covariate from the `sna` network object.)

(2) Add to the model in (1) variables to test the following hypotheses simultaneously:

   i. Boys are more likely to send friendship nominations than girls

   ii. Smokers are more likely to receive friendship nominations than non-smokers.

   iii. A friendship nomination is more likely between a pair of students pariticipating in the same activity.

   Argue for the definition of the variables. When several operationalizations are possible choose one of them.

(3) Estimate the model specified in (2). Interpret the coefficients of the model and determine whether the data support the hypotheses listed in (2).

(4) Could you think of another hypothesis that could be tested using QAPs? State your hypothesis and provide the corresponding statistic.

(5) Test the hypothesis formulated in (4) by adding the corresponding variable in the MRQAP specified in (3). Comment on the results.

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:

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

with statistics the number of edges, reciprocal dyads, and homophily dyads on gender v. $1(.)$ 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) With the data from `friend_net`, a member of your research team 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 gender of the students.

    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 your 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**

Now we want to analyze the high-school friendship network in `friend_net.Rda` using ERGM.

(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) Add variables to the ERGM specified in (1) 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).

(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.

Task 4: Comparing ERGM and MR-QAP   4 points

Building on the analysis in the previous tasks, we want to compare the results from MQ-QAP and ERGM.

(1) Replicate the hypotheses in Task 1(2) using ERGM, with and without the structural terms we specified in Task 4 (2). Comment on the similarity and difference of the results using ERGM compared with those using MRQAP. (You can set the `main.method` to "Stochastic-Approximation" in `control.ergm` to enhance convergence.)

(2) Could you think of another hypothesis that could be tested using ERGMs? State your hypothesis and provide the mathematical formula and the graphical representation of the effect that you need to include in the ERGM to test the hypothesis.

You are encouraged to work in groups of 3 or 4 people.
Please submit your solution (including R scripts in a zip file with .R and .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.