

Assignment \mathcal{N}^o 2

released: 25.10.2021 at 18:30 **due:** 10.11.2021 at 12:00

Task 1: Simulation from ERGM

10 points

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

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

with statistics the number of edges and reciprocal dyads.

- (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 `communication.csv` contains an observed adjacency matrix. A researcher suggested that plausible estimates of the parameters of the ERGM above for the communication network are $\theta_1 = -0.6$ and $\theta_2 = 1.5$.
 - i. Use the code developed in (1) to simulate communication networks from the ERGM with parameters $\theta_1 = -0.6$ and $\theta_2 = 1.5$.
 - 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 and θ_2 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 2: Estimation and interpretation of ERGM

10 points

The data set `lawyeradvice.csv` is a network of advice ties between 36 lawyers who were partners at a law firm spread over three cities. The network is directed. Senders reported asking for advice, and receivers are the lawyers who

gave the advice to sender. The adjacency matrix contains node IDs in the first column and first row.

The data set `lawyerattr.csv` contains information on the gender of each individual ('gender'), the city in which each of the employees work ('office_location'), and their duration of the employment at the firm, in years ('seniority'). The row order of these data corresponds to the order of the adjacency matrix, with the first row containing the variable names.

- (1) Estimate an ERGM with edge and gender homophily parameters. Compute the conditional probability of observing a tie between two employees i and j which have the same gender. Interpret the result.
- (2) Specify an ERGM model to test simultaneously the following hypotheses:
 - i. There is evidence for reciprocity
 - ii. There is evidence against transitivity
 - iii. Ties between lawyers working in the same city are more likely.
 - iv. Ties between lawyers of the same gender are more likely.
 - v. Lawyers who have been at the firm for longer ask fewer others for advice.
 - vi. Lawyers who have been at the firm longer are more likely to be asked for advice.
 - vii. Lawyers who have been at the firm for a similar amount of time are more likely to ask each other for advice.
- (3) Estimate the ERGM specified in (2) and comment on the convergence of the algorithm.
- (4) Evaluate the goodness of fit of the model.
- (5) Interpret the estimated model.
- (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 .R or .Rmd file!) 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.