# Assignment № 1

**released:** 14.10.2024 at 19:00 **due:** 30.10.2024 at 12:00

The data set `Krackhardt-High-Tech_Multiplex_Social.zip` contains the multiplex social network of the managers of a high-tech company [1]. The network is composed of three layers, that are stored as edges list in the file `Krackhardt-High-Tech_multiplex.edges`:

layer ID 1: A directed binary network of *advice-seeking* between colleagues. There is a tie from $i$ to $j$ if $i$ ask for advice to $j$.

layer ID 2: A directed binary network of *friendship* nomination. There is a tie from $i$ to $j$ if $i$ nominated $j$ as a friend.

layer ID 3: A directed binary network of *reporting* relationships in the organization. There is a tie from $i$ to $j$ if $i$ reports to $j$ as part of $i$'s functions in the organization.

Additional information on the managers is provided in the file `Krackhardt-High-Tech_multiplex_attributes.csv`:

`ID`: the unique identifier of the manager.

`age`: the age in years of the manager.

`tenure`: the number of months that the manager has been part of the company.

`level`: the hierarchical level of the manager in the organization. 1 = CEO, 2 = Vice President, 3 = manager.

`department`: the department that the manager belongs to in the company. Coded 1,2,3,4 with the CEO in department 0, i.e., not in a department.

[1] Available here and originally used in: Krackhardt, David. "Cognitive social structures." *Social Networks* 9.2 (1987): 104-134

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

(1) Import the data. Build a QAP to test if friendship and advice relations correlate. Use at least 5,000 permutations for reporting the results.

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

   i. A friendship nomination is more likely between a pair of managers within the same deparment.

   ii. Senior managers are less likely to nominate friends.

   iii. A friendship nomination is more likely between a pair of managers of a similar age.

   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 MR-QAP 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:

$$P(X = x; \theta) = \frac{1}{\kappa} \exp\left\{\theta_1 \sum_{ij} x_{ij} + \theta_2 \sum_{i<j} x_{ij} x_{ji} + \theta_3 \sum_i \binom{x_{+i}}{2}\right\} .$$

with statistics the number of edges, reciprocal dyads and 2-istar. Where $\kappa$ is the normalizing constant, $x_{ij}$ is the element of the adjacency matrix, $x_{+i}$ is the indegree for node $i$, and $\theta_1$, $\theta_2$ and $\theta_3$ are the parameters of the model. The 2-istar statistic count the number of distinct incoming-2-stars in the network, where it is defined to be a node $j$ and a set of 2 different nodes $\{i, k\}$ such that the direct ties $(i, j)$ and $(k, j)$ exist. In principle, the in-2-star parameter models the heterogeneity in a graph with respect to popularity.

(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) A member of your research team suggested that plausible estimates of the parameters of the ERGM above for the advice network are $\theta_1 = -2.76$, $\theta_2 = 0.68$ and $\theta_3 = 0.05$.

i. Use the code developed in (1) to simulate advice networks from the ERGM with parameters $\theta_1 = -2.76$, $\theta_2 = 0.68$ and $\theta_3 = 0.05$. .

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 $\theta_1$, $\theta_2$ and $\theta_3$ 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**

Now we want to analyze the friendship nominations using ERGM.

(1) Estimate an ERGM with the edges and a deparment homophily parameters. Compute the conditional probability of observing a tie between two managers $i$ and $j$ belonging to the same department and interpret the result.

(2) Add effects to the ERGM specified in (1) to test simultaneously the following hypotheses:

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

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

iii. A tie is more likely 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 auxiliary statistics. Comment on the results.

(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 3.2. Comment on the similarity and difference of the results using ERGM compared with those using MR-QAP.

(2) Could you think of another hypothesis that could be tested using ERGMs, but not with MR-QAP? 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.

**Submission instructions:** You are encouraged to work in groups of 3 or 4 people. Each group should submit their solution via Moodle, with one member designated to submit on behalf of the group.

Please ensure that:

- All group members' names are listed on the documents submitted.
- The submission includes a **single PDF** file that contains all essential information, including code, results, plots, and written explanations, as this PDF will be the primary document for grading. We suggest using Rmarkdown or Quarto to create the PDF document as these tools simplify the process to create a single document.
- Any accompanying R scripts (`.R`, `.Rmd`, or `.qmd`) should be zipped and included with the submission; these files will be referenced only if additional verification of computations is needed.

Thank you for following these guidelines to ensure a smooth grading process!