University of Konstanz
Algorithmics Group
V. Amati / J. Lerner

Network Modeling
Winter Term 2015/2016

Assignments $\mathcal{N}^{\underline{o}}$ 5

released: 25.11.2015 **due:** 30.11.2015

Task 1: Statistics at lower/upper bounds — 5 points

Let $\mathcal{G}$ be the set of all undirected, loopless graphs on n vertices and let

$$P_\theta(G) = \frac{1}{\kappa(\theta)} \cdot \exp\left(\theta_1 \cdot \mathrm{edges}(G) + \theta_2 \cdot \mathrm{triangles}(G)\right)$$

be the probability function of an ERGM on $\mathcal{G}$ with two statistics (number of edges and number of triangles) and associated "free" parameters θ_1 and θ_2 (that is, parameters that are yet to be estimated from a given graph).

Clearly the number of triangles of a given graph is a number between 0 and $\binom{n}{3}$ where these lower and upper bounds can also be attained. Your task is to answer the following questions.

(a) What happens if you estimate the maximum likelihood parameters from an observed graph that has 0 triangles?

(b) What happens if you estimate the maximum likelihood parameters from an observed graph that has $\binom{n}{3}$ triangles?

(c) In general, what happens if any statistic of the observed graph is at its lower or upper bound and you estimate the maximum likelihood parameters?

Hints: You might use the `ergm` function in R (and made-up networks) in order to get an idea of what happens in the above cases. (But you have to interpret and explain the output.)

Use the fact that the maximum likelihood parameters are exactly those that make the statistics of the observed graph equal to their expected values in the ERGM. Use the fact that in any ERGM the probability of any graph is larger than 0. What can you conclude about the expected number of triangles in relation to 0 and in relation to $\binom{n}{3}$?

Task 2: Simulating and estimating ERGMs (R) 15 points

Import the adjacency matrix of the Knecht friendship network observed at the third time point (file `net-3.csv`), the demographic characteristics (file `demographics.csv`), and the delinquency behaviour (file `delinquency.csv`) of the actors . Create a directed network object using the adjacency matrix. Add the gender of the pupils and the delinquency behaviour of the pupils at Wave 3 as attributes.

(a) Simulate 1000 networks from an estimated model specified by the statistics `edges`, `mutual`, and `nodematch("gender")`.

- (a.1) Calculate the mean values of these statistics of the simulated networks.
- (a.2) Compare these mean values with the observed value of the statistics.
- (a.3) Simulate again using the following parameter values: edges= -9, mutual=2, nodematch=1 (instead of the estimated ones). Again compare expected values with the observed ones.
- (a.4) Compute the goodness of fit of the estimated model and the one from (a.3). Describe in which aspects does the model fit well and in which not. (You can include the gof plots in your solutions PDF.)

(b) Estimate a model specified by `edges`, `mutual`, `nodematch("gender")` and `gwesp(alpha=0.1, fixed=TRUE)`.

- (b.1) Interpret the result.
- (b.2) Analyse the goodness of fit of this model and discuss it.

(c) Analyze the effect of the delinquency behaviour on friendship ties by including one network statistic in the model in (b) and estimating the new model.

(d) Which of these statistics are related to the "Running hypotheses" in the section "Running example: data, questions, and simple answers" in the lecture slides? Does the network support these hypotheses? Which of these hypotheses is/are not tested by this model? Can you specify a model that tests all 5 of the "Running hypotheses?"