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

Network Modeling
Winter Term 2014/2015

Assignments $\mathcal{N}^{\underline{o}}$ 4

released: 19.11.2014 **due:** 25.11.2014 at 12:00h

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

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

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: Estimating ERGMs **15 points**

Import the adjacency matrix of the 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 network object using the adjacency matrix and name the object `netwdir`. Check that the network is directed and add the gender of the pupils and the delinquency behaviour of the pupils at wave 3 as attributes (SEE Assigment 3, Task 3).

(a) Simulate 1000 networks from the model including `edges`, `mutual`, `nodematch("gender")` statistics.

  (a.1) Calculate the mean values of the statistics of the simulated networks.

  (a.2) Compare these mean values with the observed value of the statistics.

  (a.3) What happens to the simulated networks if you use the following coefficient values: edges= -9, mutual=2, nodematch=1.

(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.

  (b.3) Compare this model with the one obtained from point (b.3) in Assignment 3. What can you conclude?

(c) The model in (b) reveals that there is homphily with respect to gender. Specify a model that allows to distinguish the effect of homophily of friendship within boys and girls.

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

(e) Can you related some of the statistics used in this task to the hypotheses in the 13^{th} (16 of 253) slide of the lecture? Does the network support these hypotheses?

Hint: inspect the code in the updated script at `http://www.inf.uni-konstanz.de/algo/lehre/ws14/nm/local/data/NAwithR.R`