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

Network Modeling
Winter Term 2015/2016

Assignments $\mathcal{N}^{\underline{o}}$ 3

released: 11.11.2015 **due:** 16.11.2015

For this exercise sheet, assume (as in Sheet02) a partition of the node set V into 'girls' and 'boys': $V = V_1 \uplus V_2$, and consider the set $\mathcal{G}$ of all *directed*, loopless graphs with vertex set V.

For the following tasks you have to experimentally estimate parameters. In some cases this is hard or even impossible. Don't spend too much time in trying but, in any case, report what you did and what is your best result.

Hint: Implementing Task 3 solves the other two when you set the unused parameters to zero.

Task 1: (R) Incrementally-defined Reciprocity Model 7 points

Let $p \in (0, 1)$ be a real number between zero and one and let p_r be a real number such that $0 < p + p_r < 1$.

(1) Implement in R a random graph generator for the following incrementally defined model.
   (a) Iterate once over all dyads (u, v) with $u \neq v$;
   (b) Check whether the reverse dyad (v, u) is an edge;
      i. If no: add the edge (u, v) with probability p;
      ii. If yes: add the edge (u, v) with probability $p + p_r$;

(2) Implement in R a function that estimates the expected number of edges and the expected number of mutual edges (that is: reciprocated edges) for the above model in the following way:
   (a) Sample a given number N of graphs from the model;
   (b) Compute the numbers of edges / mutual edges for each graph;

(c) Take the average (mean) as an estimate for the expectation;

(3) Try to find (more or less systematically) the parameters p and p_r such that the expected number of edges and mutual edges is (almost) equal to the observed number of edges and mutual edges in the 3rd wave of the (preprocessed) Knecht Classroom Data.

Task 2: (R) Incrementally-defined Triangle Model **7 points**

Let $p \in (0, 1)$ be a real number between zero and one and let p_t be a real number such that $0 < p + p_t < 1$.

(1) Implement in R a random graph generator for the following incrementally defined model.

(a) Iterate once over all dyads (u, v) with $u \neq v$;

(b) Check whether u and v have a common friend. (We say that u and v have a common friend if there is any third vertex w such that $(w, u) \in E$ and $(w, v) \in E$.)

i. If no: add the edge (u, v) with probability p;

ii. If yes: add the edge (u, v) with probability $p + p_t$;

(2) Implement in R a function that estimates the expected number of edges and the expected number of edges whose endpoints have a common friend for the above model in the following way:

(a) Sample a given number N of graphs from the model;

(b) Compute the two numbers for each graph;

(c) Take the average (mean) as an estimate for the expectation;

(3) Try to find (more or less systematically) the parameters p and p_t such that the expected number of edges and edges whose endpoints have a common friend is (almost) equal to the observed numbers in the 3rd wave of the (preprocessed) Knecht Classroom Data.

Task 3: (R) Putting it all together **6 points**

Recall the "homophily model" defined in Task 1 in Assignment 2.

(1) Implement in R a random graph generator for the incrementally defined model with the parameters

   - p (the baseline edge probability)
   - p_h (homophily parameter)
   - p_r (reciprocity parameter)
   - p_t (triangle parameter)

   by combining the network effects from the homophily model and the two models above. When deciding about the edge probability of the next dyad check for 'same-gender', 'reciprocated dyad' and 'common friend'; add up the probabilities for all conditions satisfied by that dyad (take the baseline probability p only once).

(2) Implement in R a function that estimates the expected number of edges, the expected number of same-gender edges, the expected number of mutual edges, and the expected number of edges whose endpoints have a common friend by sampling a given number of graphs.

(3) Try to find (more or less systematically) the four parameters such that these four expected numbers are (almost) equal to the observed numbers in the 3rd wave of the (preprocessed) Knecht Classroom Data. Start with the parameter estimates for p_r and p_t from Tasks 1 and 2 and the homophily parameter p_h computed in Task 1 in Assignment 2; what are the resulting expected numbers?