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

Network Modeling
Winter Term 2015/2016

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

released: 18.11.2015 **due:** 23.11.2015

Task 1: Dependency in ERGM — 5 points

Let $\mathcal{G}$ be the set of undirected, loopless graphs with $n = 3$ vertices and consider an exponential random graph model $(\mathcal{G}, P)$ with only one statistic, namely $t(G)$ (the number of triangles) with associated parameter value $\log(3)$.

Compute the edge probability of a dyad and the conditional edge probability, given that there is an edge on another dyad. Are dyads independent or not in this model?

Task 2: Inappropriate Sampling from ERGM — 5 points

Provide a (preferably simple) example of an ERGM $(\mathcal{G}, P)$ which demonstrates that the following algorithm returns graphs G with probability **different from** $P(G)$.

Algorithm 1: incorrect ERGM sampling

Data : edge set E
Output: random graph $G = (\{1, \ldots, n\}, E)$

$E \leftarrow \emptyset$
foreach $e \in D = \{d_1, \ldots, d_{\binom{n}{2}}\}$ **do**
    randomly add e to E with probability

$$\frac{P(V, E \cup \{e\})}{P(V, E) + P(V, E \cup \{e\})}$$

Hints: use a small model in which you have dependency among dyads. Show that the algorithm turns the very first dyad d_1 into an edge with a probability that is different from the probability $P(d_1 \in E)$ defined by the ERGM. You might reuse probability-calculations from the lecture or from another task.

Task 3: Interpreting ERGM parameters — 10 points

Download the *Preprocessed Knecht Classroom Data* from `http://algo.uni-konstanz.de/lehre/ws15/nm/local/data/data.html`.

(a) Import the adjacency matrix of the network observed at the third time point (file `net-3.csv`) and the demographic characteristics of the actors (file `demographics.csv`)

- (a.1) Symmetrize the adjacency matrix using the function `symmetrize` in the R package **sna**. Use the ``strong'' rule.
- (a.2) Create a `network` object using the symmetrized matrix. Check that the network is undirected and add the gender of the pupils as an attribute.
- (a.3) Estimate an ERGM model specified only by the number of edges and interpret the result.
- (a.4) Estimate an ERGM model specified by the number of: edges, triangles, 2-stars, 3-stars and the same gender covariate (`nodematch`). Interpret the result.

(b) The data are directed. During the lecture, several statistics were introduced for undirected ties. Analogous statistics exist for directed ties. Therefore, we can estimate ERGMs also for directed data.

- (b.1) Create a directed `network` object using the adjacency matrix observed at the third time point. Check that the network is directed and add the gender of the pupils as an attribute.
- (b.2) Estimate an ERGM specified by the following statistics: number of edges, reciprocal dyads, and homophilous dyads with respect to gender. Interpret the result.
- (b.3) Estimate an ERGM specified by the following statistics: number of edges, reciprocal dyads, and a statistic accounting for whether girls have more ties than boys (`nodefactor`). Interpret the result and explain the difference to the result from b.2.

(b.4) In the model from b.3, set `base=0` in the `nodefactor` statistic. Interpret the result. Now, additionally, drop `edges` from the model. Interpret the result and compare it to the one obtained in b.3.

(b.5) Estimate an ERGM specified by the following statistics: number of reciprocal dyads and the term `nodemix(''gender'')`. Discuss the results – comparing them to findings from models estimated in b.2, b.3, b.4. How could you specify a model with better-interpretable parameters? Estimate such a model and discuss the results.

(b.6) Estimate an ERGM specified by the following statistics: number of edges, reciprocal dyads, homophilous dyads with respect to gender and transitive triplets. Explain the output.

N.b.: To find the name and definition of the effects check the help `?ergm.terms`