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

Network Modeling
Winter Term 2013/2014

# Assignments № 5

**released:** 27.11.2013 **due:** 03.12.2013, 12:00h

## Task 1: Interpreting ERGM Parameters — 6 points

Up till now, we were only dealing with undirected networks. However, we can also model directed networks with an ERGM. For example, consider the following hypothetical case. We did a survey in a company and asked the employees with whom it is important to communicate to get work completed effectively. Further we gathered some attributes of the employees. The binary attribute *seniority* indicates whether the employee was a senior or not. The *projects* attribute is a continuous variable denoting the number of projects the employee was involved in, i.e. how experienced is the employee. The *office* attribute is a categorical variable showing in which office the employee is located. We are now interested in understanding the structure of the informal communication ties. So we try to model the gathered informations with an ERGM. After a lot of brainstorming we come up with several parameters that might be important to include in the model and run our estimations. The output is depicted in the table on the last page.

(1) Which of the parameters are significant?

After our estimations we send the network data to the boss of the company. Since he doesn't know how to interpret the table, he sends us several hypotheses what he thinks played a role in the network formation:

(1) *"Employees tended to reciprocate communication."*

(2) *"There were popular employees with whom many others have to communicate to effectively complete tasks."*

(3) *"There was a tendency towards a hierarchical path closure."*

(4) *"Experience in projects affected the tendency to communicate."*

(5) *"Employees tended to communicate with others of similar experience."*

Which of these hypotheses are true and which are false? Give an explanation using the data from the table.

**Task 2: Degenerated Model** **4 points**

Consider the following ERGM with the edge statistic $m(G)$ and the triangle statistic $t(G)$:

$$P(G) = \frac{1}{\kappa} \exp\left(\eta m(G) + \tau t(G)\right)$$

(1) For which parameter settings do you expect degeneracy, i.e. when will you end up in an almost empty network or in an almost complete network, independent of the network you start the simulation with?

Now let's assume $\eta < 0$ and $\tau > 0$.

(2) How does the density of the starting network affect the resulting network of the simulation? Explain which processes lead to this outcome.

**Task 3: R: Modeling an ERGM by yourself** **10 points**

In this task you are on your own. Load the *faux.magnolia.high* network data, from the ergm package with the command *data("faux.magnolia.high")*. The network describes friendship among students of a highschool.

Before you begin modeling these data, it is good to have a general handle on their nature. Which attributes are included in the data? What are structural properties of the network (e.g. number of triangles and edges)? You could also try to plot the network and set the vertex colors to different attributes to get an idea, if there is a tendency to form ties between students with the same attribute. If the plotting takes to long you can also just look at the mixingmatrices.

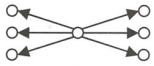
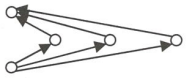
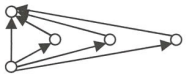
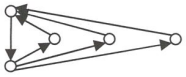
(1) Try to come up with some hypotheses for the tie formation process.

(2) Define a model that includes some structural statistics and statistics that can prove your hypotheses right or wrong.

(3) Estimate the model and explain your outcome. Which parameters are significant? How do you interpret the parameters?

(4) If the estimation does not converge, exclude the paramteres you expect to cause troubles. Or is there a possibility to circumvent the problems?

(5) As a last step, try to sample from your model and compare it to the original network. Do you observe any differences or even degeneracy?

| Network effect | | Estimate (SE) |
|---|---|---|
| **Purely structural effects (endogenous)** | | |
| Arc |  | −1.96 (0.73) |
| Reciprocity |  | 2.88 (0.46) |
| Popularity (in-degree) |  | −0.27 (0.32) |
| Activity (out-degree) |  | −0.34 (0.34) |
| Simple 2-path[3] |  | −0.06 (0.08) |
| Multiple 2-paths |  | −0.06 (0.09) |
| Transitivity (transitive path closure of multiple 2-paths) |  | 1.22 (0.19) |
| Cyclic closure (cyclic closure of multiple 2-paths) |  | −0.37 (0.17) |
| **Actor relation effects (exogenous)** (black nodes indicates actor with attribute) | | |
| Sender (seniority) |  | −0.56 (0.29) |
| Sender (projects) |  | 0.01 (0.02) |
| Receiver (seniority) |  | 0.08 (0.23) |
| Receiver (projects) |  | −0.02 (0.02) |
| [4] Homophily (seniority) |  | 0.64 (0.26) |
| Heterophily (projects) |  | −0.08 (0.02) |
| Homophily (office) |  | −0.01 (0.17) |
| **Covariate network (exogenous)** | | |
| Advice entrainment (covariate arc) |  | 1.76 (0.30) |