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

Network Modeling
Winter Term 2013/2014

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

released: 20.11.2012 **due:** 26.11.2012, 12:00h

Task 1: Inappropriate Sampling from ERGMs **4 points**

So far, algorithms for sampling from $G(n, p)$ or the preferential attachment model decided about the inclusion/exclusion of edges one after the other.

Provide a (preferably simple) example which demonstrates that a corresponding strategy does not work out for the ERGM class in general, i. e. starting with the empty edge set $E = \emptyset$ and sequentially adding edge e to E with probability $\frac{P(V,E\cup\{e\})}{P(V,E)+P(V,E\cup\{e\})}$ yields incorrect outcomes — you are allowed to reuse probability calculations from the lecture.

Task 2: Appropriate Sampling from ERGMs **6 points**

Let $\mathcal{G}$ the set of undirected, loopless graphs with $n = 3$ vertices and consider the exponential random graph model $(\mathcal{G}, P)$ containing only the *number of two-stars* statistic $g_1 = s_2$ with parameter $\theta_1 = \ln 2$.

According to the Gibbs sampling strategy defined in the lecture, specify the transition probabilities π in a Markov chain on $\mathcal{G}$ with unique stationary distribution P.
(Hint: You don't have to calculate the transition probabilities for all pairs of graphs explicitly. Many transition probabilities are 0 and for isomorphic graphs they are equivalent.)

Task 3: R: power iteration, Gibbs, gof **10 points**

Power iteration and Gibbs sampling

(1) Implement the power iteration as a function. The input parameters should be a transition matrix π and an initial distribution P_0. Use

the transition matrix π from task 2 and set P_0 to $[1, 0, 0, 0, 0, 0, 0, 0]$, so we start the iteration from the empty graph. In each step i reevaluate the current convergence parameter ε in the following way: $\varepsilon = \left\|P^{(i)} - P^{(i-1)}\right\|_2$. Run the power iteration as long as ε is greater than $9 \cdot 10^{-6}$ and store the total number of needed iteration steps T. Report a Matrix where row i should contain the distribution $P^{(i)}$ in iteration step i and the value of ε. *(This should be a Matrix with T rows and 9 columns.)*. Compare $P^{(T)}$ with the stationary distribution you calculated in Task 2.
(Matrix Vector Multiplication is done with %% in R)*

(2) Implement a simple Gibbs sampling for the ERGM from Task 2. Starting like in (1) with the empty graph, perform T sampling steps and store in each step, on which graph you were. This should be done with a Matrix A in the following way: $A[i, j] = 1$ if you were on graph j in step i. So A has T rows and 8 columns and after finishing the Gibbs sampling there should be exactly one entry in each row that is one. Perform this Gibbs Sampling 1000, sucht that at the end the entry $A[i, j]$ equals the number of times you visited graph j in step i. Derive from the Marix A the probability distribution of the graphs. Discuss differences/similarities to the probability distribution you obtained in (1). *(The best way to solve this task is to just use the transition matrix π and the sample function of R with appropriate parameters.)*

Goodness of fit

(1) Recall the greys anatomy network from the last sheet. Create two *ergm* models. The first one should use the statistics *nodematch("sex")*, *edges*, *degree(1)* and the second one *nodematch("sex")*, *edges*, *absdiff("birthyear")*. Apply the *gof* function to both models. Which model fits the original data better in terms of degree distribution, edgewise shared partner and geodesic distances? Explain why.