

Assignment 3

for lecture "Bioinformatics III" WS 08/09



Return before lecture on Nov. 20, 2008 or by email to

p.walter@bioinformatik.uni-saarland.de

until **Nov. 21**. This assignment will be discussed in the tutorial on Nov. 24, 2008, room15,
building E1 3

1. Comparison of scale-free networks and random networks (40 points)

- Consider the following values describing the properties of a network: $P(k)$ (average degree of nodes), $C(k)$ (average cluster coefficient of all nodes with k vertices) and l (average length of the shortest path between all pairs of nodes in the network). Which of these values are independent of the number of nodes in random and scale-free networks?
- Give a definition for the following terms: small world property, ultra-small-world property,
- Implement a tool that
 - determines the cluster coefficient $C(k)$ of each node in a given network
 - averages the cluster coefficients of the nodes of the same degree
- What is the meaning of the cluster coefficient? What is the difference between the two networks that differ in their average cluster coefficients?

2. Clusters in networks (40 points)

Consider the destruction of a network by removal of nodes. Is the split of a connected graph into components already related to nodes with high degree (=hubs) or can you give a counter example? Explain your answer?

- From the interactions listed in the "Biomolecular Interaction Network Database" (BIND) for man, *Saccharomyces cerevisiae* and *E. coli* (id: 562), determine the histogram of cluster sizes $P(C(k))$, the size of the largest cluster N_{max} and the average cluster coefficient $\langle C \rangle$ (see (1)).

Hint: You can start from the tools you wrote in the last assignment, but do not limit the interactions to proteins and small-molecules this time.

Hint: Be careful to exclude self-interactions, i.e., where both interaction partners are identical, or you will run into an infinite loop during the cluster identification.

- To check the stability of these biological networks against directed attacks take the interaction network of the *Saccharomyces cerevisiae* and determine (the labels of) the 200 nodes with the highest degrees. Compare the size of the largest cluster N_{max} and the number of clusters N_{cl} of the original network to networks, where you delete the 10, 20, 50, 100, or 200 nodes with the highest degrees and also to networks, where you randomly delete the same numbers of nodes. Does the network behave as you were told in the lecture? Explain your answer.

Hint: you don't have to give the list of the labels of the 200 nodes with the highest degrees.

- Repeat part (b) for a scale-free network according to Barabási–Albert with 2500 nodes.

3. Network theory

(20 points)

a) In the lecture you learnt the definition of a edge-cut (κ_e) and vertex-cut(κ_v). Give an example of a graph G satisfying the following conditions:

- $\kappa_v(G) = \kappa_e(G) = \delta_{\min}(G)$
- $\kappa_v(G) < \kappa_e(G) = \delta_{\min}(G)$
- $\kappa_v(G) = \kappa_e(G) < \delta_{\min}(G)$
- $\kappa_v(G) < \kappa_e(G) < \delta_{\min}(G)$

Contact: Peter Walter, phone -3613, email: p.walter@bioinformatik.uni-saarland.de