

## Bioinformatics III

Prof. Dr. Volkhard Helms

Thorsten Will, Maryam Nazarieh, **Duy Nguyen**, Daria Gaidar  
Winter Semester 2016/2017

Saarland University  
Chair of Computational Biology

### Exercise Sheet 7

**Due: 16.12.2016 13:15**

Submit your solutions on paper, hand-written or printed at the *beginning* of the lecture or in building E21, Room 3.03. Alternatively you may send an email with a single PDF attachment. If possible, please include source code listings. Additionally hand in all source code via mail to [duy.nguyen@bioinformatik.uni-saarland.de](mailto:duy.nguyen@bioinformatik.uni-saarland.de).

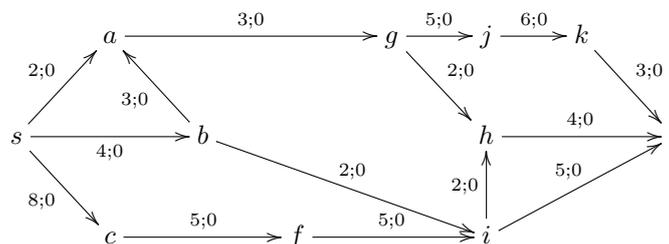
## FFEK Algorithm and DREAM Challenge

### Exercise 7.1: FFEK Algorithm (40 points)

Apply the Ford, Fulkerson, Edmonds, and Karp (FFEK) algorithm explained in the lecture to determine the s-t-cut and the capacity of the network given below.

For each iteration, give the indices of the nodes, the resulting  $f$ -augmenting path with its capacity, and the updated  $val(f)$ . Sketch the newly found  $f$ -augmenting paths. Also update the currents through the arcs.

If you find multiple possible paths from s to t with the same length, then choose the one with the highest  $\Delta Q$ .



### Exercise 7.2: GRN (DREAM Challenge) (60 points)

Apply one of the three models introduced in the lecture (Noise, Linear and Sigmoidal) to predict the directed unsigned GRN topology of E.coli from steady state and time series gene expression data. The target network is of size 10 genes without self-regulatory interactions.

Download the gene expression dataset and use the following:

- *heterozygous.tsv* contains the steady state levels for the wild-type and the heterozygous knock-down strains for each gene. Thus, for a network of size 10 there are 11 experiments (wild-type plus knock-down of every gene).
- *null-mutants.tv* contains the steady state levels for the wild-type and the null-mutant strains for each gene. Thus, for a network of size 10 there are 11 experiments (wild-type plus knock-down of every gene).
- *trajectories.tsv* contains time courses of the network recovering from several external perturbations. For the network of size 10, you have 4 perturbations (each one with 21 time points).

Finally, the expected output to be submitted should be a ranked list of regulatory link predictions ordered according to the significance of each prediction.

For example: G1 G2 score

Where G1 and G2 are two different genes (no self-interactions). Links are directed: the gene in the first column regulates the gene in the second column. (If both G1 regulates G2 and G2 regulates G1, then both lines should be included). Score is between 0 and 1 and indicates the confidence level you set to this link prediction (sample output file is also attached).

Have fun!