T-61.5110 Modeling biological networks

Exam, February 18, 2014

You are NOT allowed to use calculators or any other additional equipments/material in the exam. Please write your answers in English. Please write carefully.

1. a) Compute the average node degree statistic for the network shown in Fig. 1 (a). (3 points)

   b) The distribution of the average node degree for the 5-node Bernoulli random network model with parameter $p = 0.5$ is shown in Fig. 1 (b). Given a significance level of $\alpha = 0.05$, is the network in Fig. 1 (a) statistically significantly different from this Bernoulli random network model? (3 points)

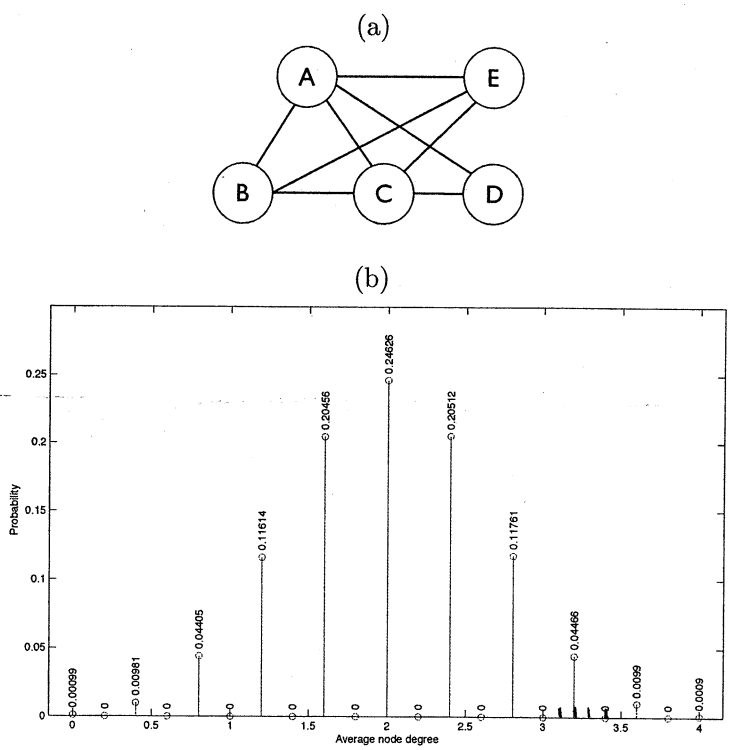

Figure 1: (a) An undirected biological network. (b) The distribution of the average node degree for 5-node Bernoulli network model with $p = 0.5$.

2. Consider the simple model of Michaelis-Menten enzyme kinetics

$$\begin{aligned} S + E &\rightarrow SE \\ SE &\rightarrow S + E \\ SE &\rightarrow P + E \end{aligned}$$

   a) Represent it mathematically as a Petri net, $N = (P, T, Pre, Post, M)$, assuming that there are currently 100 molecules of substrate S, 20 molecules of the enzyme E, and no molecules of the substrate-enzyme complex SE or the product P. (2 points)

b) If the first reaction occurs 20 times, the second 10, and the last 5, what will be the new state of the system? (2 points)

c) Can you identify any T-invariants for this system? If you can, identify one T-invariant. (2 points)

3. Assume you are given a PSFM model θ for a DNA binding protein A. Assume the width of the PSFM (and thereby the width of the binding site/motif) is n.

   a) You are given two DNA sequences, a wild-type $X = (x_1 x_2 x_3 \dots x_n)$ and a mutant $X^* = (x_1 x_2^* x_3 \dots x_n)$, where $x_2 \neq x_2^*$. Describe how you can computationally decide which one of the sequences, X or X^*, is more likely to bind protein A. (3 points)

   b) You are given a DNA sequence $Z = (z_1 z_2 \dots z_\ell)$, where $\ell > n$. Explain how you can assess the statistical significance that Z contains a binding site for protein A. (3 points)

4. Explain how the maximum likelihood principle and the steepest descent (i.e., gradient descent) numerical optimization method can be applied to estimate parameters of ordinary differential equation (ODE) models, given a parametric ODE model and time-series measurements. You may assume additive Gaussian noise. In particular, explain the steps specific for the application to ODE model, including calculation of the likelihood, numerical derivation etc. (i.e., a general explanation of the steepest descent will not be enough). (6 points)

5. Approaches to model biological networks can be categorized into qualitative models (e.g. Boolean networks or discrete-valued Bayesian networks) and quantitative models (e.g. ODEs and coupled chemical reaction networks). Discuss the benefits and drawbacks of both qualitative and quantitative network models. Discuss e.g.:

   - accuracy of models, i.e., how realistic a network model is
   - computational complexity/difficulty to simulate and analyze a model
   - computational aspects of constructing a network model, including network structure selection and parameter estimation
   - what kind of measurement data is needed for network structure selection and parameter estimation
   - applicability in practical work

   Use Boolean *or* Bayesian networks as an example of qualitative approaches, and use ODEs *or* coupled chemical reaction networks as an example of quantitative approaches. (6 points)