

BMI 826. Computational network biology. Homework 2

Due Nov 27th, 2018

Instructions

The goal of this homework is to review some concepts from the dynamic network section and network clustering. The homework is due in class on Nov 27th.

Problem 1 (10 points). Defining concepts

For each of the terms below provide a brief description of what they mean.

- Dynamic Bayesian network
- Markov chain Monte Carlo
- Input/output HMM
- Graph clustering
- Scale free distribution
- Network motifs
- Multi-task clustering
- Modularity
- Simulated annealing
- Graph Laplacian

Problem 2 (10 points). Context specific networks

Suppose you are given genome-wide expression samples from two different conditions and a static protein-protein interaction network. In each condition there are $n = 5$ different biological replicates of the genome-wide expression levels and you are measuring $m = 5,000$ genes. Given the protein-protein interaction network and this gene expression data your goal is to identify components of the network that are more active in one condition versus another.

- (a) Given a subnetwork describe how you would *score* it based on its relative activity in one condition versus another.

- (b) Describe a strategy to search for subnetworks with a high score.
- (c) Describe how you would extend this approach to more than two conditions. Clearly define what the score of a subnetwork would capture in this case.
- (d) How will you change your answers to the above questions if you had to use *Graph clustering* as part of your solution?

Problem 3 (10 points). Context-specific and dynamics in networks

Suppose you have a few dozen samples for each of five different but related disease conditions. You are interested in identifying at the molecular level what are the differences and similarities of the networks between these conditions. Describe an approach to solve this problem which is different from the solution in Problem 2. State any assumptions you need to make. Compare your solution in this problem to that in Problem 2 to highlight the key advantages and disadvantages of the two approaches.

Problem 4 (10 points). Network dynamics

Suppose you are given a time series experiment of gene expression profile measurements and noisy and incomplete interaction network. Describe two approaches based on probabilistic graphical models that leverage both the expression data and the interaction network to learn a dynamic network. State what the random variables of the model are, what the dependencies mean and what the parameters of the model are. Feel free to use any formal notation to precisely describe each model. Describe briefly how you will learn each model from the data. Describe the strengths and weaknesses of each approach. Finally, describe how you would assess the quality of your learned model.

Problem 5 (10 points). Graph clustering

- (a) For each of the three algorithms below for graph clustering, describe what each algorithm takes as input, what output it produces and what it does: (i) Girvan Newman clustering, (ii) Hierarchical agglomerative clustering, (iii) Spectral clustering.
- (b) Compare and contrast these three algorithms in terms of their strengths and weaknesses.
- (c) Describe two measures you can use to assess the quality of your clusters. You can assume that you can leverage domain-specific information to evaluate your clusters.