

Homework 3: Due Dec 15th 2016 in class

Prof. Sushmita Roy

Instructions

This homework will refresh some concepts from the network dynamics and network alignment section.

Problem 1 (10 points).

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. Given the protein-protein interaction network and this gene expression data describe an algorithm to identify parts of the network that are differentially active between two conditions? Explain the choice of your algorithm.

Problem 2 (10 points)

Suppose you are given a collection of a large number of time courses and a noisy physical interaction network. Describe an approach that uses this network and expression to learn a dynamic network. What types of predictions can you make with this dynamic network?

Problem 3 (10 points)

Recall the JACS object in the MATISSE paper. Describe what the JACS object captures and the key aspects of the JACS score. Describe how the JACS object and its associated score is different from “Active subgraphs” and its score. Describe the search algorithms for finding JACS and Active subgraphs. Give one strength and weakness of each search algorithm.

Problem 4 (10 points)

Suppose you had noisy protein-protein interaction networks from two species. Describe a local network alignment algorithm to find conserved subnetworks.

Problem 5 (10 points)

Compare and contrast the pairwise Global Network Alignment approach of the FUSE method with that of IsoRank. How does each approach handle global alignment across more than two species? Highlight two strengths and weaknesses of each approach.

Problem 6 (10 points)

Suppose you are given protein-protein interaction networks, gene expression data for two species. Assume that the gene expression data in each species is measuring global gene expression levels in 10 different conditions. Assume further that each species has thousands of genes, so you do not have sufficient samples to infer a regulatory network but you could infer co-expressed gene modules. Finally, assume that you have the sequence orthology between the genes of each species, that is, you know which gene in one species corresponds to which gene(s), if any, in the other species. Using the approaches we have seen in class, describe an approach to find conserved gene modules across species such that the gene modules are also connected on the respective protein-protein interaction network.

Problem 7 (10 points)

Consider the question in problem 6. Suppose we now have more than two species. Describe the key extensions to your approach in (6) to find a conserved and co-expressed gene modules.

Problem 8 (10 points)

You are given a weighted interaction network. Suppose you knew that the nodes interact with nodes from the same community and there are k such communities. Describe an algorithm to find these communities. Explain your choice. How will you assess the quality of the community structure you have inferred.

Problem 9 (10 points)

Consider the question in problem 8. Suppose we did not know how many communities there are and rather there might be a hierarchically structure to communities in the graph. Describe an approach to find these hierarchically organized communities.

Problem 10 (10 points)

Recall that JACs uses protein-protein interaction networks as a “constraint” graph. Briefly describe what the role of the “constraint” graph is. Describe a spectral approach to find objects similar to JACs.