Notes for network study group

Jiaxin Hu

February 4, 2022

Main presentation: The main presentation given by Rogini and Wan from Northeastern University tackles to the problem on how to obtain gene expression network from gene correlation matrix. After the preprocessing to clean the outliers based on previous works, they apply two method to obtain the gene networks:

1. Hard thresholding: generate the edges between the genes with correlation over 99-th (or 90-th) percentiles;

2. Adaptive marginal likelihood fliter (MLF) test: original MLF test assumes a null hypothesis where the number of connection between two genes follows a binomial distribution with parameter related to the degree of the genes. They propose an adaptive MLF using normal approximation due to the correlation matrix, and generate the network by the test results.

They characterise the networks by several statistics including degree correlation, average degree, the longest path, and e.t.c.. They show that the sparsity property and the characterization of the networks generated by these two methods are different.

Discussions:

- What kind of network is desirable?

  (I did not hear a clear conclusion about this question. But it is definitely true to check out the interpretation of the results.)

- Why we want to study the modules (clusters) of genes?

  There are two reasons that we want to study the the modules of genes: (1) genes may express together; (2) gene expression modules may be related to the cell type and the cell interaction. For example, cells in immune system may produce a particular protein if there is an external object in the fish. The module of the gene expressions related to that protein may reflect the cell type and action.

  It is also interesting to study whether the modules of genes will change with different environments and other features.

- Checking gene ontology may be helpful to interpret or assess the gene module results.

- Gene expression data is usually over-dispersed. We may not assume the gene expression data follows Poisson distribution, and we may not fully rely on the Pearson correlation coefficient.