

Abstract

Within microorganisms the Transcriptional Regulatory Network (TRN) plays an important role in maintaining cellular homeostasis under changing environmental conditions. Therefore understanding the structure and dynamics of this network is fundamental for understanding and ultimately predicting organism behavior. With the emergence of the microarray technology genome wide data has become available that provides snapshots of the activity of the TRN. An important computational challenge is to infer or reverse-engineer the structure and dynamics of this TRN from available data.

The computational problem of inferring TRNs from gene expression data is however underdetermined: multiple equivalent solutions exist that each explain the data equally well. Ensemble methods provide an elegant way for dealing with this problem of underdetermination by considering multiple equivalent solutions and by reinforcing those solutions that are repeatedly retrieved. In this thesis we present different ensemble strategies to improve upon and extend the scope of existing methods to infer the TRN from gene expression data.

In a first part we focus on module detection or the detection of sets of coexpressed genes from gene expression data. In particular we develop an ensemble strategy for existing query-based biclustering methods in order to extend their application to input sets that are heterogeneous in their expression profiles. As such the method can be used to interrogate gene expression compendia for experimentally derived gene lists, as is illustrated on an *Escherichia coli* and *Salmonella Typhimurium* case study.

In a second part we focus on inference of the TRN itself. Here, we first present Stochastic LeMoNe. This method uses a stochastic

Abstract

optimization approach to output multiple equivalent outcomes of the network inference problem. By using ensemble averaging we demonstrate that both module detection and inference of the transcriptional program can be improved. Further we illustrate that by making certain assumptions on the inference problem, Stochastic LeMoNe is biased towards making correct predictions for only subparts of the TRN. Building upon this observation, we categorized existing network inference methods according to their conceptual differences and illustrated how these differences result in distinct methods highlighting different parts of the TRN.