

345b Transcriptional Regulatory Network Reconstruction Via Integer Linear Programming

Joao M. S. Natali and Jose M. Pinto

Much effort has been recently dedicated to the identification of genome-wide transcription regulatory networks by means of comprehensive high-throughput experiments that are capable of capturing the systemic behavior of the transcription coordination phenomenon. However, such data have limited use if not interpreted by a coherent mathematical framework capable of deciphering the underlying mechanisms and interconnections within a global regulatory network. A fundamental advantage of Mathematical Programming methods with respect to statistical data treatment lies on the fact that the former rely on model-based decision making algorithms that require minimal human interference and, therefore, yield an unbiased treatment and extraction of information from large scale experimental results.

The present work comprises the development of Linear Programming (LP) and Integer Linear Programming (ILP) approaches to model and analyze the gene regulatory network of *S. cerevisiae*, centered on a logic inference based representation of regulatory events and on a direct evaluation of experimental quantitative results found in literature (Lee et al., *Science* 298, 799, 2002; Harbison et al., *Nature* 431, 99, 2004). The former approach employs a network representation of regulatory pathways based on a bipartite graph formed by the complete set of Transcription Factors (TF) considered, linked to the entire set of genes from the studied organism. It further uses a quantification of regulatory signals as fluxes through this structure and the final problem is posed as a Minimum Cost Network Flow problem, for which many efficient solution procedures have been developed.

The approach based on ILP has the important advantages of being able to represent logical events in an unambiguous way, and of explicitly accounting for relationships between modeled variables, that can be expressed as sets of logical propositions. For that purpose, disjunctive programming approaches are used to represent transcriptional events synchronized by mutually exclusive regulatory elements such as direct interactions and expression of TFs. Possible interactions between TFs and genes are placed by the model in one of three categories: activation, repression, or the lack of a sensible interaction. From this definition, sets of constraints representing regulatory elements are defined for TF-Gene pairs that are found to interact in each of the defined possibilities. Microarray experiments data is used to define a set of physiological conditions upon which the expression profiles obtained will be based. The objective function is, then, defined in a way to minimize the existence of false TF-Gene interactions weighted by the probability that similar location analysis experimental results would be obtained from random values, for each of the pairs. Moreover, a high coherence between the expression predictions and microarrays experimental data is sought, which composes the second objective term.

Our preliminary results show that gene regulatory networks can effectively be modeled by linear relationships between appropriate variables resulting in LP and ILP problems. We successfully determined sets of regulatory interactions that concisely describe the influence of the expressed TF's on the coordination of genomic-scale transcription regulation. Furthermore, our method based on ILP is able to discern between activation and repression interactions, an important task that current statistical methods can not do. Further work is being dedicated to improving the applicability of the model and the computational methods applied for its solution.