Progression Modeling of Dynamic Biological Systems

Saptarshi Pyne (*Roll No.* 146101005)
Supervisor: Dr. Ashish Anand
Department of Computer Science and Engineering
Indian Institute of Technology Guwahati

July 19, 2016

Problem Formulation: Biological processes are manifestations of interplay between numerous variables in dynamic biological systems. The state of such a system can be defined as a function of the states of the system variables. Dynamic systems progress in time through transitions from one state to another. Progression modeling of such systems help us to identify the key variables or modules (highly interdependent group of variables) responsible for a particular state transition. It is substantially applied in biomedical research to find out the variables critical for transition from a healthy stage to a disease stage. The same is true for pure life science research where progression modeling sheds light on how natural developmental processes take place in living systems. In order to distinguish the key variables that regulate the rest of the variables to cause a system-wide state transition, we need to decipher the interdependencies between all the variables. The data-driven approach is to monitor changes in their values across time and under different conditions; then analyze the acquired data to learn how values of these variables change under different conditions and how changes in the value of one variable influences that of the others. Following this approach, one class of methods is proposed that models temporal progression of a system as a collection of time-varying networks where nodes and edges of each network represent the system variables and their interdependencies, respectively. The differences in the network topologies and the edge parameters between two consecutive networks represent the changes in interdependencies among the system variables during that time period. For example, a variable whose value remains unchanged during a state transition, can play key regulatory role during another transition. Once the networks are inferred, an array of network analysis methods are applied on them to identify the nodes and paths (sequence of edges) crucial for critical state transitions. Effective application of the data-driven network inference methods depend principally on what type of data is available. In this report, a comprehensive range of input data settings is chosen based on availability and compatibility with the contemporary applications. For each data setting, the state-of-the-art methods are critically reviewed along with their historical perspectives. It in turn provides a focused range of research challenges. Finally, a subset of these challenges are chosen as the author's PhD objective based on feasibility and timeliness of the potential contribution. The objective is divided into three subproblems. **The first subproblem** is to develop a novel Joint Network Inference (JNI) algorithm for inferring Gene Regulatory Networks (GRNs)

from time-course genome-scale gene expression data of cohort studies. **The second subproblem** is to extend the inference power of the proposed algorithm to infer multiple time-varying networks for each individual. **The final subproblem** is to extend the capability of the proposed algorithm even further so that it can utilize prior knowledge and other data types (like - proteomic data and clinical data) to generate time-varying heterogeneous cellular signaling networks.