



INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI
SHORT ABSTRACT OF THESIS

Name of the Student : Saptarshi Pyne

Roll Number : 146101005

Programme of Study : Ph.D.

Thesis Title: Improving Reconstruction Efficiency of Time-varying Gene Regulatory Networks

Name of Thesis Supervisor(s) : Dr. Ashish Anand

Thesis Submitted to the Department/ Center : Department of Computer Science and Engineering

Date of completion of Thesis Viva-Voce Exam : June 2, 2020

Key words for description of Thesis Work : Computational Biology, Bioinformatics, Probabilistic Graphical Models, Bayesian Networks, Dynamic Bayesian Networks, Gene Regulatory Networks, Transcriptional Regulatory Networks, Gene Regulation, Network Reconstruction, Network Inference, Gene Expression

SHORT ABSTRACT

Motivation: Flow of regulatory-information between genes are often modelled as a ‘time-varying gene regulatory network’. Here, nodes represent genes. A directed edge from gene X to gene Y at time-interval T is interpreted as ‘X regulates Y at time-interval T’. A widely-used approach is to reverse-engineer or ‘reconstruct’ the network from time-series ‘gene expression data’ using computational algorithms called ‘time-varying gene regulatory network reconstruction algorithms’. **Problem Formulation:** We apply the existing algorithms on three benchmark datasets. It is found that an algorithm named ‘ARTIVA’ provides state-of-the-art correctness. However, ARTIVA requires months for processing large-scale datasets with thousands of variables. Hence, the objective of the thesis is set to designing algorithms that can process large-scale datasets in acceptable time, while remaining competitive to ARTIVA in correctness. **Contributions:** Towards that objective, the thesis makes four contributions. In the first contribution, an algorithm named ‘TGS’ is proposed. It can process large-scale datasets in minutes. However, its correctness is not competitive to that of ARTIVA. In the second contribution, we propose an algorithm named ‘TGS-Plus’ (TGS+). It is as fast as TGS, and competitive to ARTIVA in correctness. Nonetheless, TGS+’s main-memory requirement grows exponentially at $(2^n * (n + 1))$, where n is the number of genes in a dataset. As a result, it tends to encounter memory-related issues for large-scale datasets. In the third contribution, an algorithm named ‘TGS-Lite+’ is proposed. It preserves the same time-complexity and correctness as those of TGS+. At the same time, TGS-Lite+’s main-memory requirement grows linearly at $(2n + 2)$ only. Nevertheless, we observe that TGS-Lite+ is unable to capture ‘transient’ edges – the edges that remain active for a small number of time intervals, yet, can have cascading effects. For this reason, capturing transient edges is critical to understand the underlying interactions. In the fourth contribution, we propose an algorithm named ‘TGS-T-Lite+’. It is able to capture twice as many edges as that of TGS-Lite+ for larger benchmark-datasets. **Conclusion:** In conclusion, this thesis advances the state of ‘time-varying gene regulatory network reconstruction’ algorithms in terms of time-efficiency, memory-efficiency, and correctness. It also discusses future opportunities.