

Master thesis/project/summer work in bioinformatics:

Inference and visualization of molecular networks in trees

Keywords: machine learning, networks, web-based visualization

Forest trees are a renewable source of raw material not only for paper production, but also for energy. However, to get higher productivity from forests we need to acquire basic understanding of important processes such as development and growth. To this end, experimentalists collect huge amount of molecular data from aspens trees using transcriptomics, proteomics and metabolomics platforms. The goal of this project is to use computational methods to explain this experimental data in terms of network models that describe interactions between genes, proteins and metabolites, and the underlying regulatory logics hard-wired in the trees' DNA. These models will become important platforms from which experimentalist can obtain understanding, overview and hints to the direction of future experiments.

Network induction. We induce different types of networks. *Co-expression networks* simply treat genes, protein and metabolites as nodes and add edges between nodes with similar profiles in data. *Transcriptional networks* predict how regulatory proteins (transcription factors) regulate the expression of genes. Here we treat transcription factors and transcriptional modules (many co-expressed genes) as nodes and add edges between nodes were we predict that one, or a combination of, transcription factors regulate a module. *Metabolic networks* predict which catalytic proteins (enzymes) that catalyze which metabolites. Here enzymes and metabolites are nodes and edges signify catalytic reactions.

To infer network models from data we use *machine learning* techniques. These are used to predict interactions between genes/proteins/metabolites from examples. Algorithms must be developed to handle the time complexity of the combinatorial problems involved in searching for the best regulatory mechanisms.

Visualization. We will build web-based visualization tools that biologist can use to browse networks, click on nodes to get more information about genes/proteins/metabolites, color certain nodes of interest, etc.

Contact. Projects can be tailor-made to your competence and interest. Contact me and we will discuss the specifics of your project!

Torgeir R. Hvidsten, Assistant professor in bioinformatics at Umeå Plant Science Centre.

Email: torgeir.hvidsten@plantphys.umu.se

Phone: 090-786 5348

Website: <http://www.trhvidsten.com/>