

CS760 Project Proposal:

1. The members of the team are **Stewart Kerr, Andrew Maule, Noah Stafford, Harry Zhou, Nathaniel Schleif**.

2. The main task to be addressed by this project is to discover the network of interactions between genes. We will do this by analyzing mRNA expression data derived from RNA-seq experiments. Specifically, we will use a dynamic bayesian network to model the gene interactions. Our principal task will be to implement dynamic bayesian network learning algorithms using python that accept preprocessed RNA-seq experimental data. Rather than making the focus of our project RNA-seq analysis, we will use already analyzed experiments. Two example datasets we have identified is a yeast dataset published by Gasch *et al* 2017 looking at salt stress and an *Arabidopsis* paper (Klepikova *et al.* 2015) that looked at the development of the apical meristem.

3. The experiments we plan to run deal mostly with the different types of data that could be analyzed. One big comparison would be the difference in inference capability in perturbation studies (like the salt stress of yeast) versus development studies (such as the apical meristem study of *Arabidopsis*). Assessment of these created networks will mostly rely on looking at their ability to reconstitute known 'benchmark' networks. Sensitivity/specificity analysis is one avenue for assessing algorithm performance, by generating a confusion matrix between the predicted edges representing interactions and the actual benchmark. Benchmarks can be generated by combining multiple data sources, such as previously inferred networks, tentative interactions discovered through techniques such as CHIP-seq and DAP-seq (establishes protein-DNA interactions), and traditional functional biology experiments using gene perturbation. A further experiment that will be done if there is sufficient time will be to integrate prior-knowledge into the network.

4. Harry, Noah, and Stewart will principally be working on the actual implementation of a dynamic bayesian network. Collaboration on the project will occur through GitHub with communication mediated by GroupMe. Andrew and Nathaniel will be more focused on the results interpretation end of things given our biological knowledge.

5. **April 5th:** Have the structure of the program diagrammed out and GitHub repository created

April 12th: Specific experiments settled on and data formatted to be accepted by program.

April 22nd: Program completed and the beginning of experiment running.

April 29th: Experiments complete and interpretation and paper writing begins.

May 6th: Turn in paper.