AC297r: Milestone 3 Summary Report

Victor Avram, Sergio Miguel Moya Jimenez, M. Eagon Meng, Wenhan Zhang

For milestone 3, we explored the use of graph neural networks. Specifically we

used and modified a graph autoencoder architecture proposed by Hasibi et al 2020.

The graph autoencoder model works on both the expression data and a network

representation of the network data in the format of an adjacency matrix (nodes

correspond to genes and edges correspond to relationships between genes). Through

simultaneous training on both the network and node features (i.e. the expression

profiles per gene), the model is capable of creating predictions for unknown node

feature values.

A maize protein-protein interaction (PPI) network from http://comp-sysbio.org/

ppim/ was used as the network representation of the expression data. This allows

us to incorporate new, post-transcriptional information into the model. As well, this

network allows us to realize causal link between nodes as opposed to correlative links

between nodes when using a network derived from our expression data. While only

11,399 genes overlapped between those found in the PPI network and those found in

our data, the combination of these two data sources has the potential to increase the

performance of the graph autoencoder model.

We ran the graph autoencoder with the default settings and are currently

iterating over different versions of the model. We are modifying both the underlying

architecture (e.g. making the neural network deeper given that it is very shallow) as

well as other hyperparameters.