

Self-supervised completion of Gene Regulatory Networks using Variational Graph Autoencoders

Marco Stock

August 26, 2021

Abstract

Genes are interacting in complex network structures. The computational reconstruction of such Gene Regulatory Networks (GRN) in the representation of graphs from sequencing data, remains a very important and challenging task. While individual gene interactions may be experimentally detected and validated by, e.g., knockout experiments, this approach lacks the scalability to infer large interaction networks. The recent research progress on Graph Neural Networks (GNN) enabled their successful application in several problems, such as in the protein folding predictor AlphaFold 2. Here we use a Variational Graph Autoencoder (VGAE) to complete partly known GRNs borrowing information from RNA sequencing data sets. This self-supervised machine learning method uses a given incomplete GRN to predict missing gene interactions. The predicted gene interactions can then be validated experimentally. The approach is suitable for both bulk and single cell RNA-sequencing data combined with partly known ground truth interaction networks. In our ongoing work, the first version of the model is applied to different data sets to get an unbiased performance estimate of the predictions and it is moreover benchmarked against other supervised methods of gene interaction inference.