Name: Zhuoqing Fang
PI Name: Gary Peltz

Email: fangzq@stanford.edu
PI email: gpeltz@stanford.edu

Graph Neural Network for Genetic Discovery from Genome Wide Association Studies

Zhuoqing Fang and Gary Peltz

Department of Anesthesia, Pain and Perioperative Medicine,
Stanford University School of Medicine, Stanford CA 94305

Since mouse and human GWAS identify many gene candidates, but only one (or a few) are causative, the ability to prioritize candidate genes is critical for genetic discovery. While many gene prioritization methods have been developed, relatively few connect genes with phenotypes using published literature. Here, we build a large gene-phenotype knowledge graph that is based upon analysis of 29M Pubmed papers using nature language processing technology. We then develop a novel heterogeneous graph convolutional network-based gene prioritization method (GNNphrank) that embeds a gene and phenotype (represented by MeSH terms) knowledge graph, which also utilizes information obtained from a protein-protein interaction network and from protein structural features. Our GNN model utilizes gene-phenotype associations to assess the probability of whether a candidate gene is connected to a phenotype. The model was trained in an end-to-end manner, without manually crafted features or pre-defined data fusion rules. We show that GNNphrank produces markedly improved results relative to those obtained with a simple linear feedforward neural network model, and we demonstrate that it works well when murine or human GWAS results are assess with evidence from the published literature.