Network Edge Prediction Strategy Analysis

Evan Momen

Introduction

- Protein-protein interactions (PPIs) studied in yeast species (Saccharomyces cerevisiae - baker's yeast)
- Various PPIs modeled as a network
  - Nodes = proteins
  - Edges = interactions
- Yields important information regarding protein complexes and their involvement in biological pathways

Problem

- Accurate modeling of PPIs as a network is difficult
  - PPI Core network contains 2708 proteins
- Bad edges may be present, important ones may be missing
- Use computational means (statistics and machine learning) to improve network modeling accuracy

Why yeast?

- Network improvements are easily verifiable
- Gene ontology (GO)
  - Major bioinformatics initiative for creating a computational representation of genes and gene products
  - Three domains:
    - Cellular component (CC)
    - Molecular function (MF)
    - Biological process (BP)
- New network should have greater agreement with GO domains

Previous research

- Improvements to PPI Core network via 'random walk with resistance' (RWS) algorithm
- Measures node similarity based on nodes' distances to all other nodes in the network
- Some edges added, some removed. Overall number unchanged.
- Improvements verified through GO

Current project

- Continuation of previous research
- Stage 1
  - Greater transformations on the network, changing:
    - Number of network edges
    - Degree of network nodes
    - Network quality (as measured through box plots and IQR's)
- Stage 2
  - Use Deep Learning to optimize the above parameters

Technology

- Matlab
- Python3
  - Numpy
  - Pandas (?)

Questions?