**Identifying highly discriminative interaction subnetworks from seed link communities**

James A. Eddy, Bobak Hadidi, Nathan D. Price

**Abstract**

We present an extension to DIRAC that offers two advantageous features: (i) instead of *a priori* defined database networks, the method instead operates on specific regions (subnetworks) within protein-protein interaction (PPI) networks generated by high-throughput data; (ii) based on the links between genes in the PPI network, the method is able to adaptively grow and shrink networks to identify more discriminative signatures for classification.

**Introduction**

**Methods**

Methods overview:

1. Protein-protein interaction networks constructed from interaction databases
2. Link communities clustering used to define subnetworks
3. DIRAC used to identify most variably expressed subnetworks; top k subnetworks are seeds for adaptive search (SFFS)
4. Seed networks adaptively modified with SFFS to identify more accurate classifier subnetworks

*PPI network construction*

Protein-protein interaction (PPI) information is obtained from public, manually curated databases including the Biological General Repository for Interaction Datasets (BioGRID), the Human Protein Reference Database (HPRD), and the Database of Interacting Proteins (DIP), and used to construct PPI networks. These PPI networks serve as inputs to the computational analyses used here.

*Link communities clustering*

Link Communities is a hierarchical clustering method based on the similarity of links rather than nodes in the graph [63]. In the resulting dendrogram, nodes may occupy multiple agglomerations due to their links; this is important because it allows genes to be grouped into multiple subnetworks.

*DIRAC classification*

[brief description here]

*Adaptive subnetwork modification*

The initial pool of networks used as the search space DIRAC may not yield the best signatures for distinguishing between two phenotypes. Networks defined *a priori* in pathway databases—regardless of the level of curation—may not be representative of functioning sets of genes in a particular phenotype, and therefore would be unlikely to serve as accurate classifiers. While interaction networks represent mechanistic links between genes, they may also not be specific to the phenotypes being studied; furthermore, subnetworks defined according to network structure may not necessarily correspond to informative differential expression of genes. We are using a sequential forward floating search algorithm [64] to grow and shrink subnetworks identified as variably expressed, in an effort to improve classification accuracy.

**Results and Discussion**