INTERIM REPORT

# VARDAAN SHARMA

# S1882216

### GOALS

The goal of the project is to create a protein-protein interaction network from synaptic proteins found in Alzheimer’s disease proteomes, analyze the network, and compare it to a network of synaptic proteins identified from healthy synaptic proteomes.

### SO FAR

Data collection and base network creation.

Up till now, the proteins that were identified in 9 papers were extracted, and combined. This list was then mapped to their entrez ids for uniformity. The protein interactions of these proteins were identified from a list of known protein-protein interactions from three databases: BioGRID, INTACT and DIP. The complete synaptic network contains 1321 nodes and 3192 edges.

Preliminary network analysis.

The PPI network that was analyzed using the R package igraph with respect to nodes centrality measures including: Degree, Betweenness (Bet.), Closeness, Clustering Coefficent (CC), Page Rank (PR), Semi-Local centrality (SL), and mean shortest path (SP).

The largest connected component of the network (1243 nodes, 3177 edges) was identified and used as the main network.

The networks degree distribution was tested for evidence of scale-free structure, the goodness of fit of degree distribution was calculated using the R “PoweRlaw” package (version 0.50.0).

(Figure for scale free would be useful to show)

### FUTURE WORK

Clustering and Enrichment

The CDMSuite package will be used for detecting of community structure of the network. The different available clustering algorithms( Geodesic and Random Walk edge betweenness, Spectral Modularity ) in the package will be used and an algorithm will be chosen based on the cluster properties such as number of components etc. This clustered network will be tested for enrichment with Gene Ontology (GO) and disease terms using the same package.

Differentially expressed genes

The list of proteins differentially expressed in AD condition will be assembled from 9 papers in similar way.

Bridging proteins.

We will rank each of the proteins in the network according to their predicted importance for propagating signals within the network based on clustering results and centrality measures (degree). Bridging proteins are known to interact with many neighbors simultaneously, help organize function inside communities but also affect/influence other communities (beyond the ones they find themselves in) in the network. Our hypothesis is that differentially expressed genes would emerge in the ranking as topologically important.

Comparison

The clustered and enriched network will then be compared to the healthy synaptic network to identify its overlap with average synaptic proteome previously assembled in the lab. List of differentially expressed genes in AD will be mapped on the clusters to obtain functional and topological dependences.