A STUDY OF THE PROTEIN-PROTEIN INTERACTION NETWORK OF CAENORHABDITIS ELEGANS

Under the guidance of Dr. Krishnan Rajkumar

Jha Aditya Dilip, En. No. 19/11/EC/018

What is a Graph?

A graph G is an ordered pair $G = (V, E)$ where, V is a set of nodes or vertices and

$E \subseteq \{(x,y) \mid x,y \in V \text{ and } x \neq y\}$ is a set of edges.

Graphs or Networks are used to model many phenomena in the real world from physical and biological to social and informational systems.

PROTEIN PROTEIN INTERACTION

Interaction between proteins in the cell makes life possible.

Interactions occur between defined binding regions in the proteins.

Have a particular biological meaning.

Why?

Can distinguish between normal and diseased state.

Identify important proteins responsible for a particular disease.

Assign putative roles to uncharacterised proteins

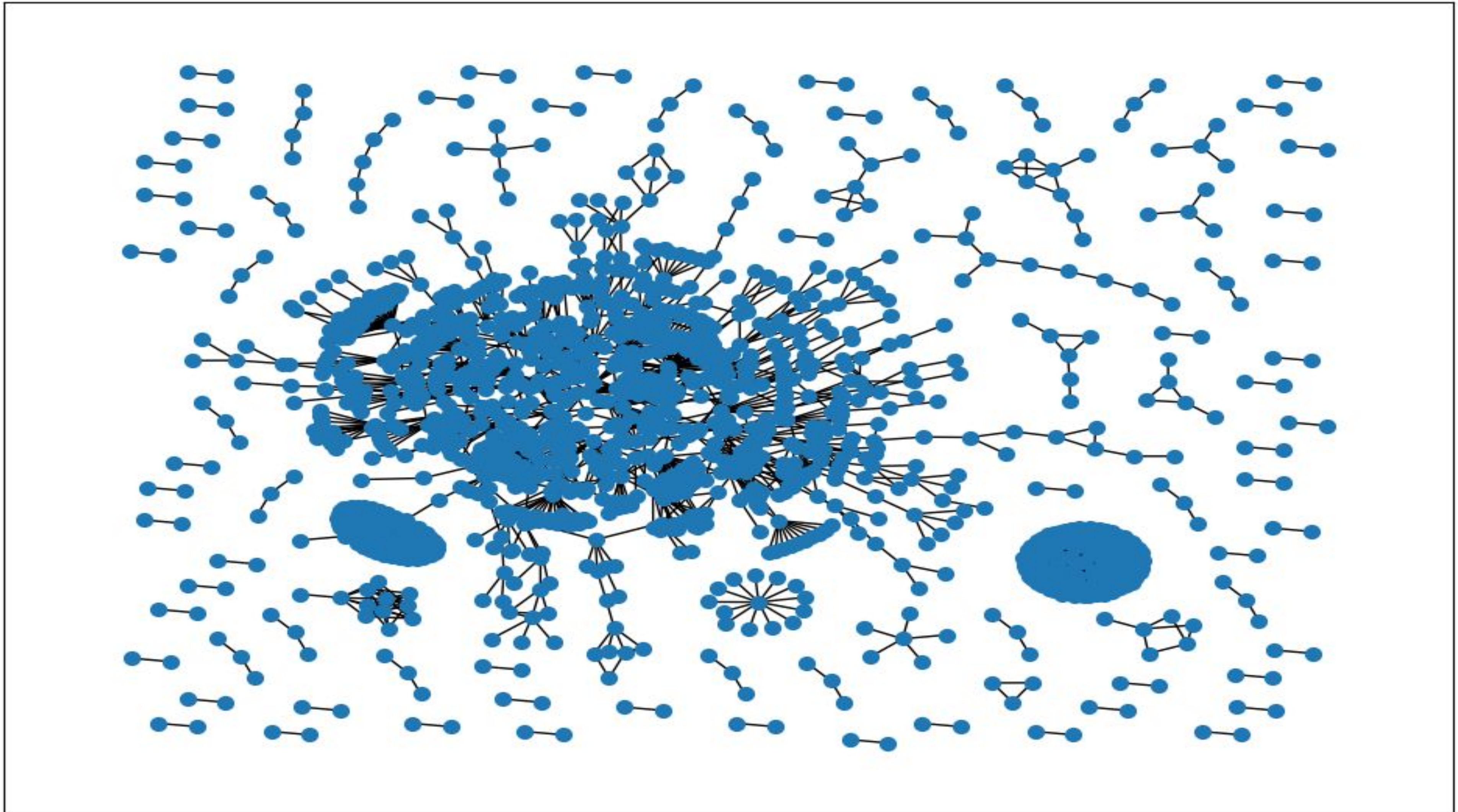
Get information about signalling pathways

Data

Protein protein interaction data of Caenorhabditis elegans(C. elegans).

Wormnet v3, a network-assisted hypothesis generating server for C. elegans.

1648 inferred links amongst 1387 proteins



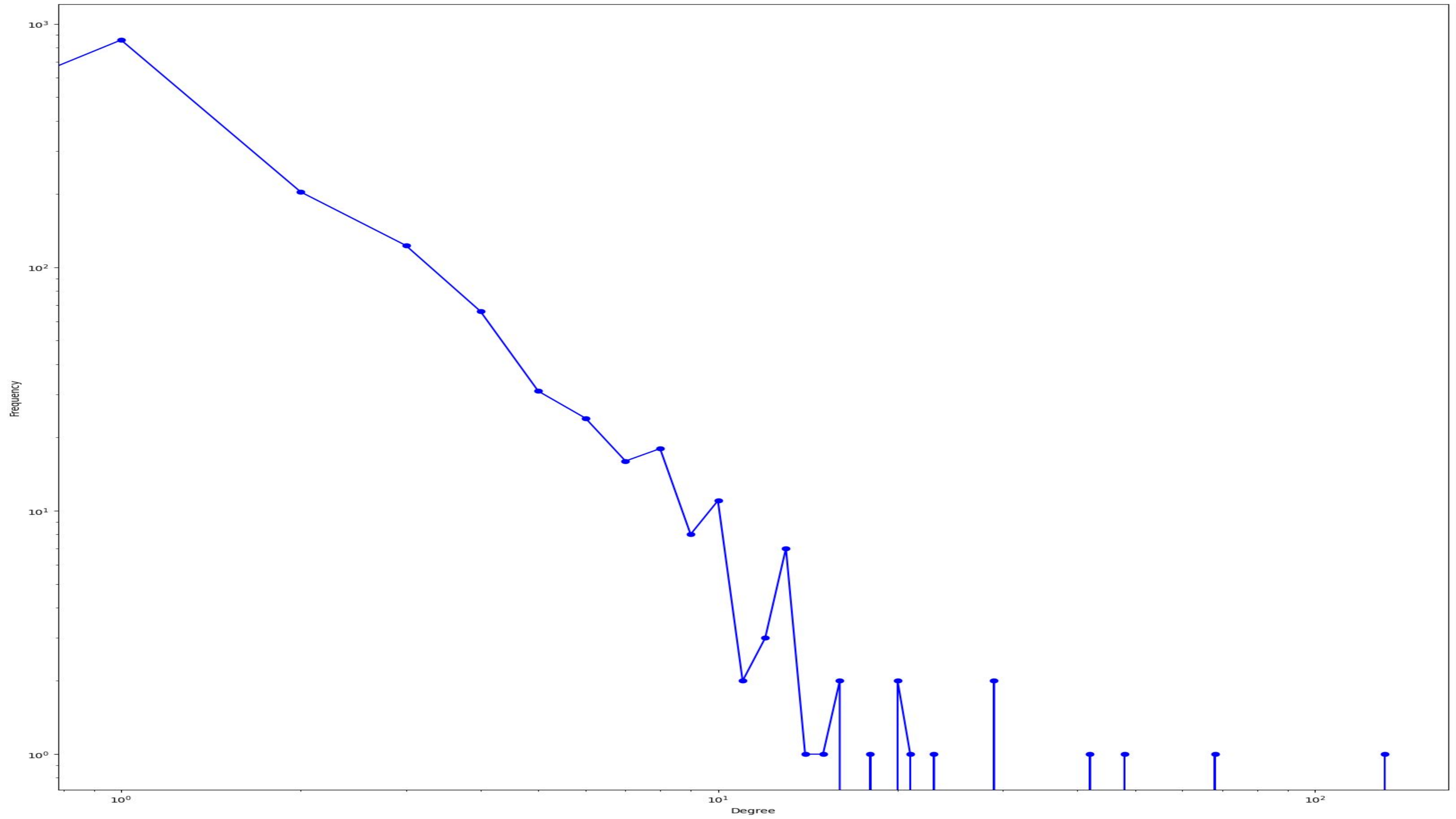
Degree Distribution

Majority of the nodes have less degree and few have more.

Presence of Hubs can be deduced in the network.

Reduces the chance of failure if it occurs randomly

Leaves the network vulnerable to failures of few nodes



Frequency
10^3
10^2
10^1
10^0
10^0
10^1
10^2
Degree

Clustering

Clustering co-efficient value = 0.076

Less cohesion between nodes in the network

Suggests reduced tendency to form functional modules

Proteins involved in many distinct biological processes

Diameter

Shortest path between the two most furthest nodes in the network.

Diameter is 22, main component nodes = 993 nodes

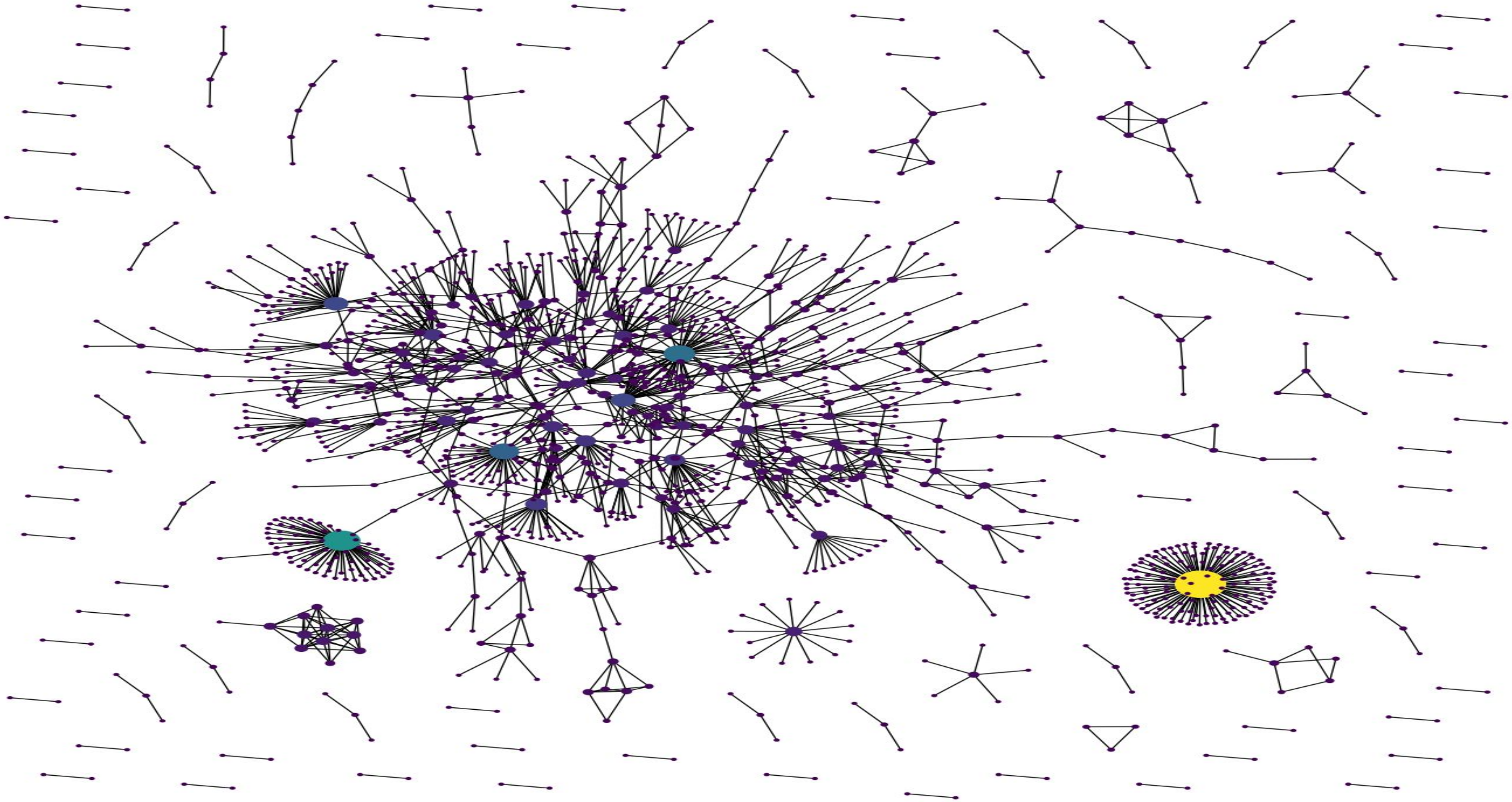
Even at the local level, nodes do not exhibit small worldness

Centrality

Centrality is a measure of how important a given node is in the network.

As the concept of importance varies across different applications so does the type of centrality.

Degree Centrality



Betweenness Centrality

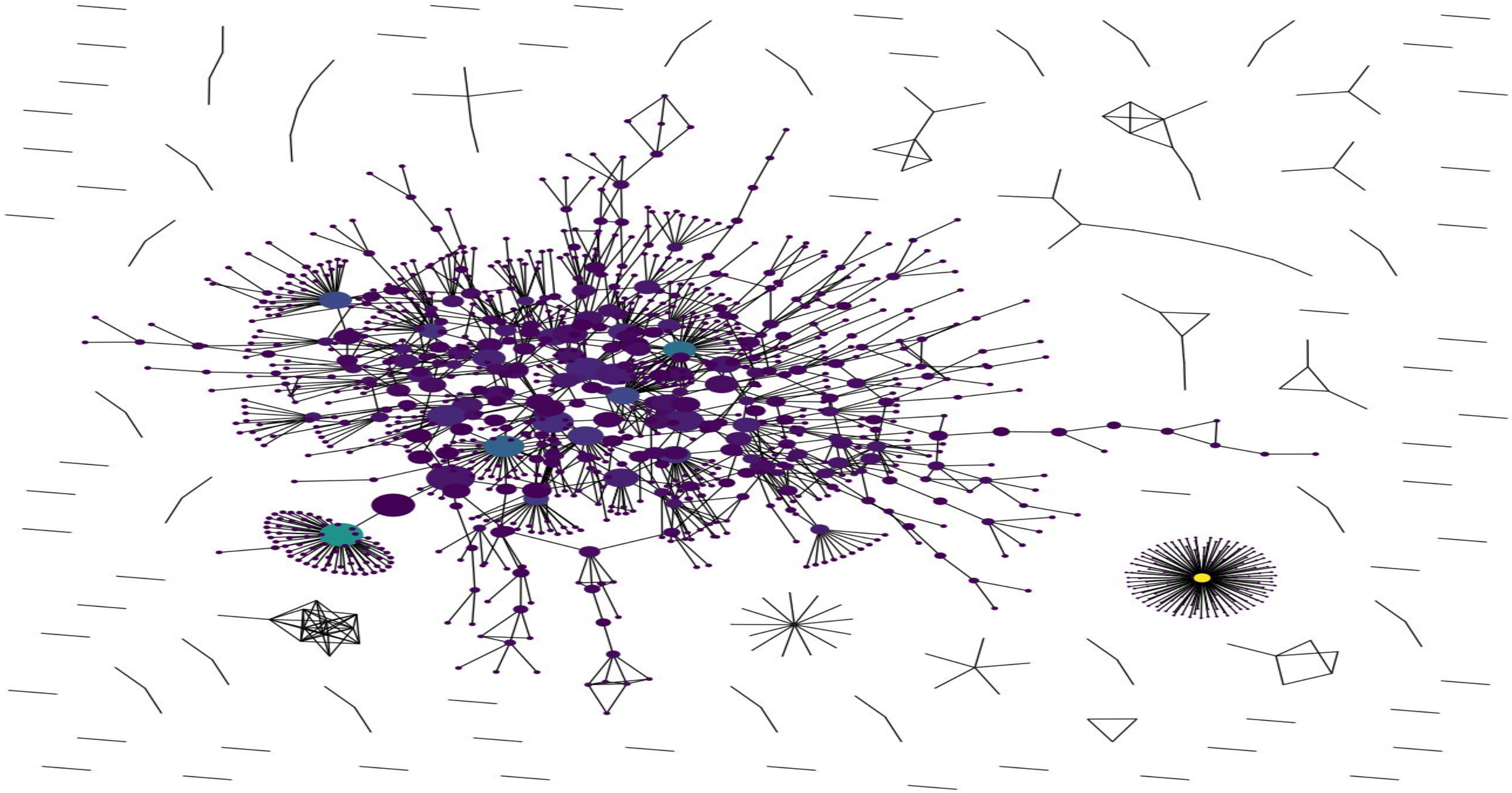
Any two pair of nodes have at least one shortest path.

This centrality measures the number of shortest paths that passes through a node

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Apt measure when the flow is through shortest path.

Betweenness



Implications of Centrality

Nodes with high betweenness centrality are located close to each other

This is because most nodes are connector nodes

These protein yield significant influence on the network

Information doesn't flow through hubs but a wide range of nodes

Modularity

Nodes more strongly connected to some set nodes(groups or communities) than with others.

Modularity is the fraction of the edges that fall within the given groups minus the expected fraction if edges were distributed at random.

$$Q = \frac{1}{2m} \sum_{i,j} (A_{ij} - \frac{k_i k_j}{2m}) \delta(c_i, c_j)$$

Finding Communities

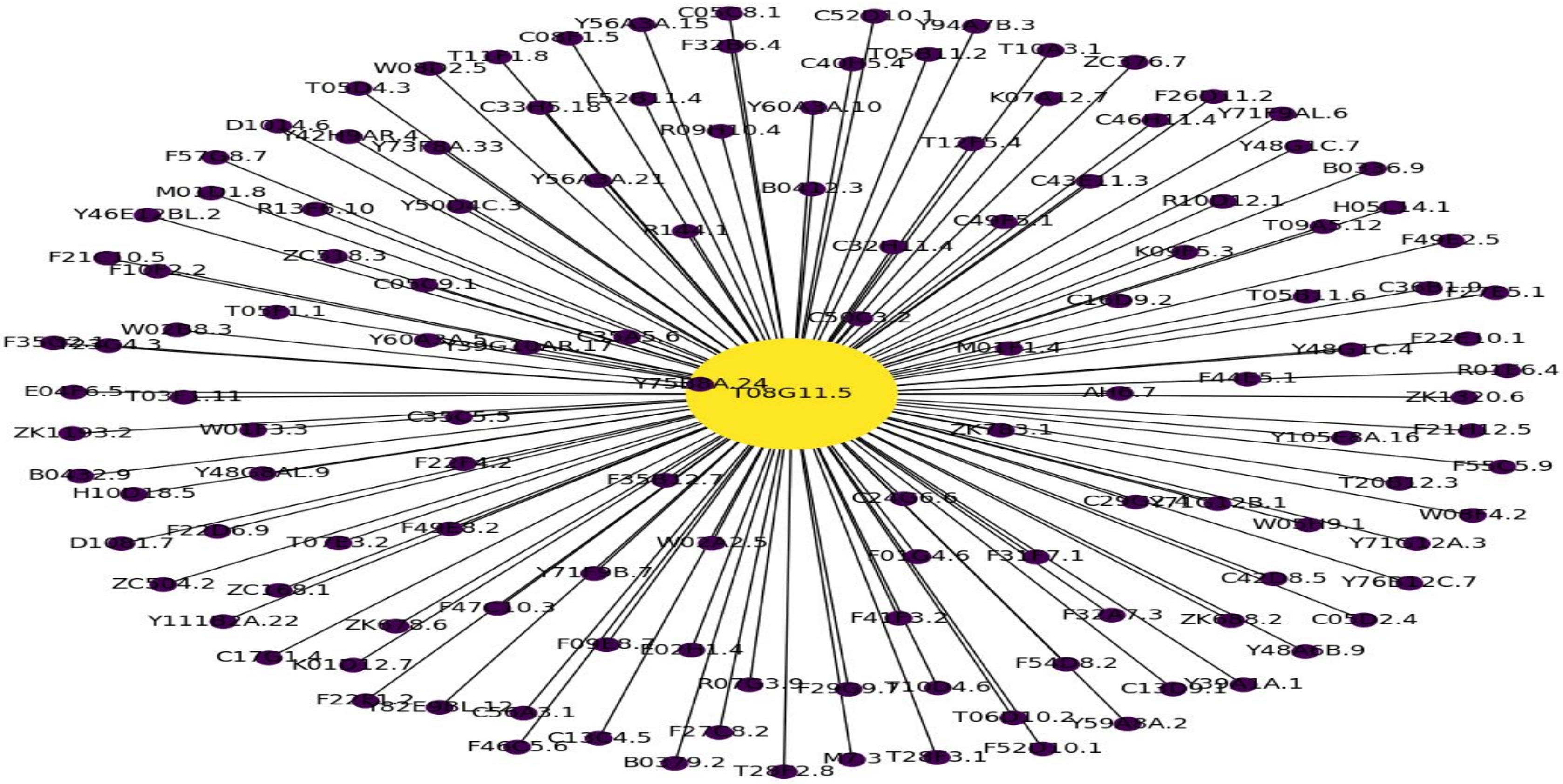
Using the measure of modularity, we find communities in the network by a greedy approach towards maximizing modularity.

These communities give us functional units

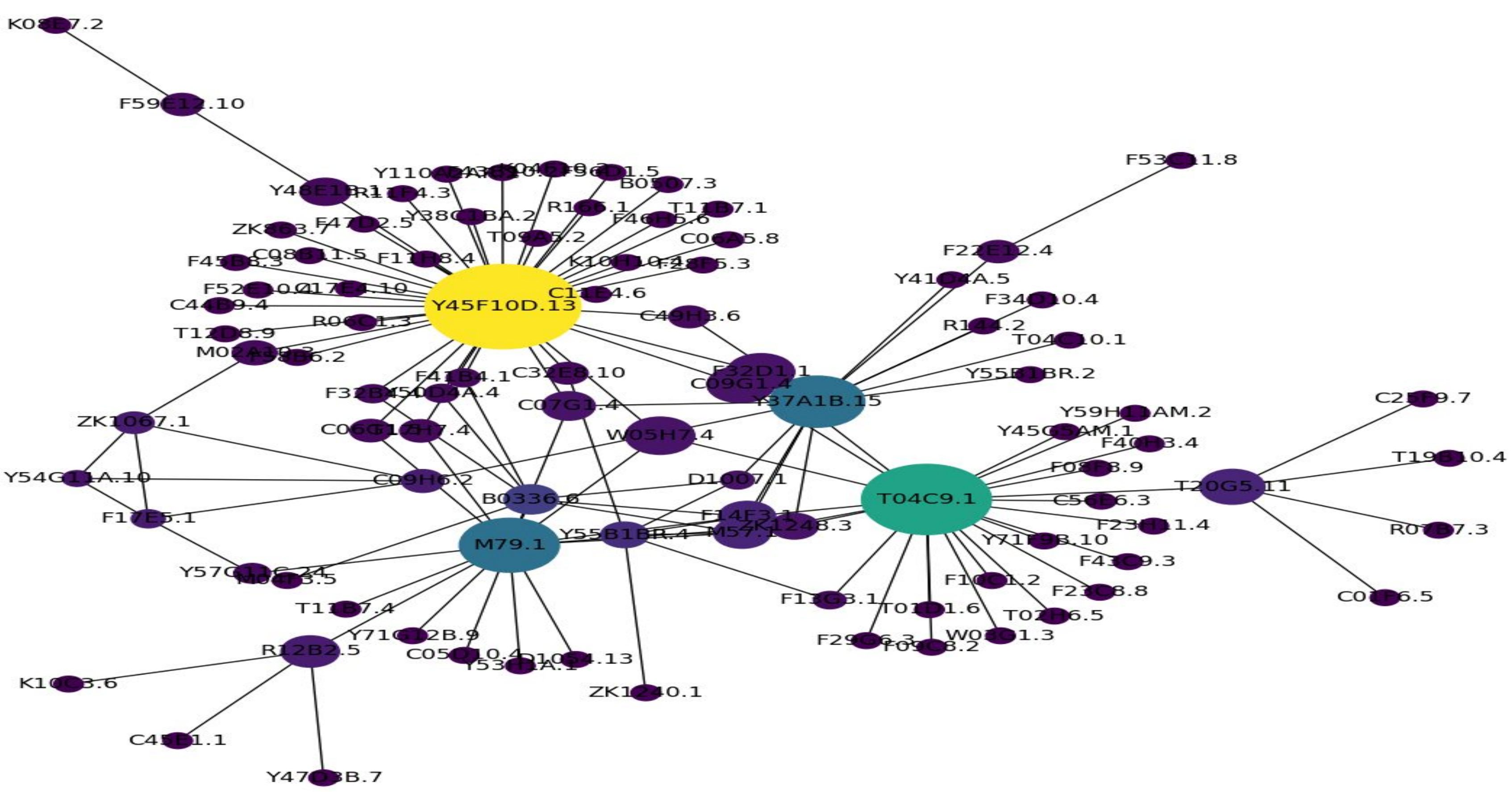
Hubs are distributed throughout different and unconnected communities

This solves the vulnerability of having hubs to some extent

Largest Community



Second Largest Community



Backbone

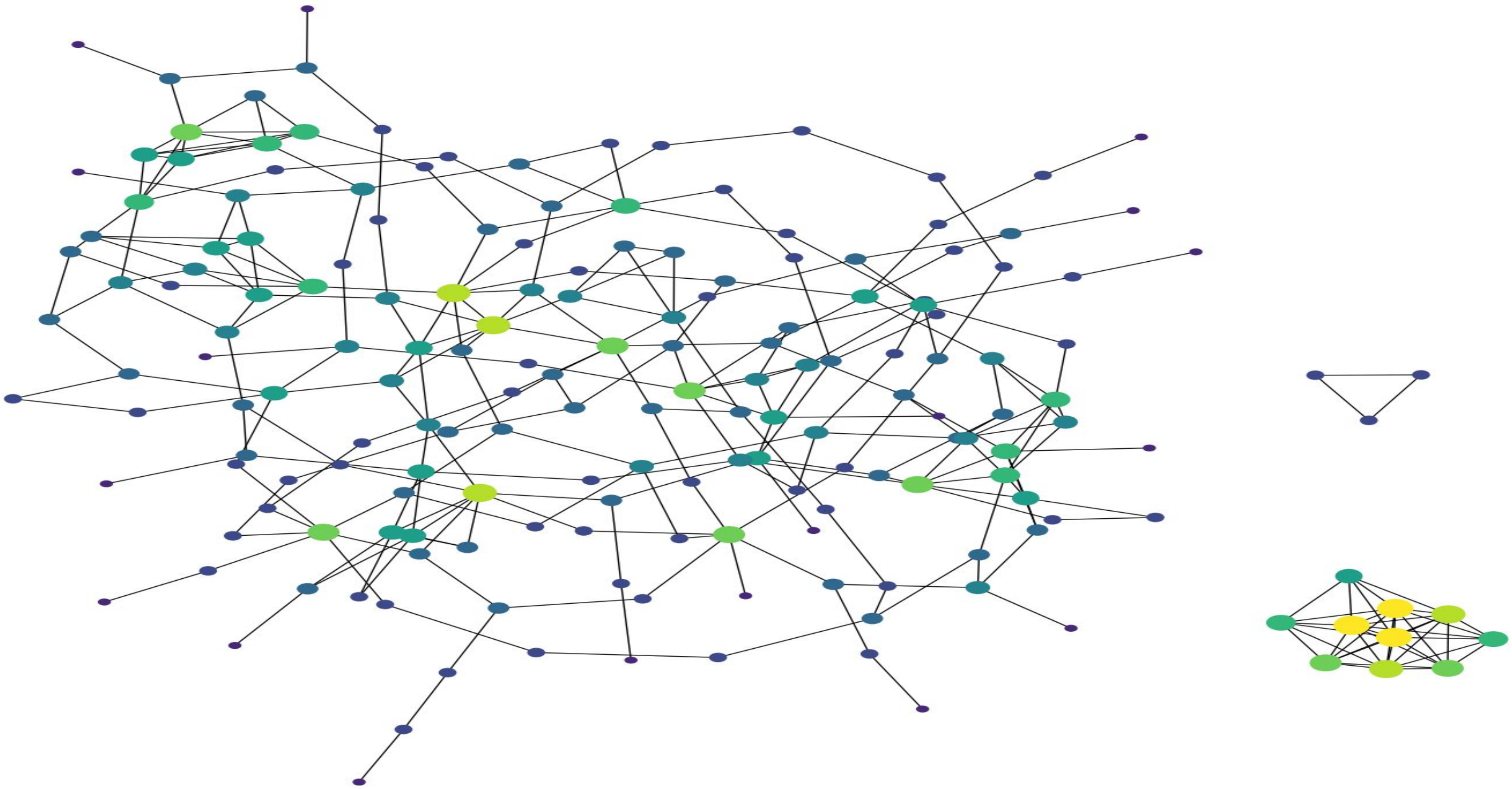
Subgraph extracted by taking the most important nodes

The nodes here are more connected to each other

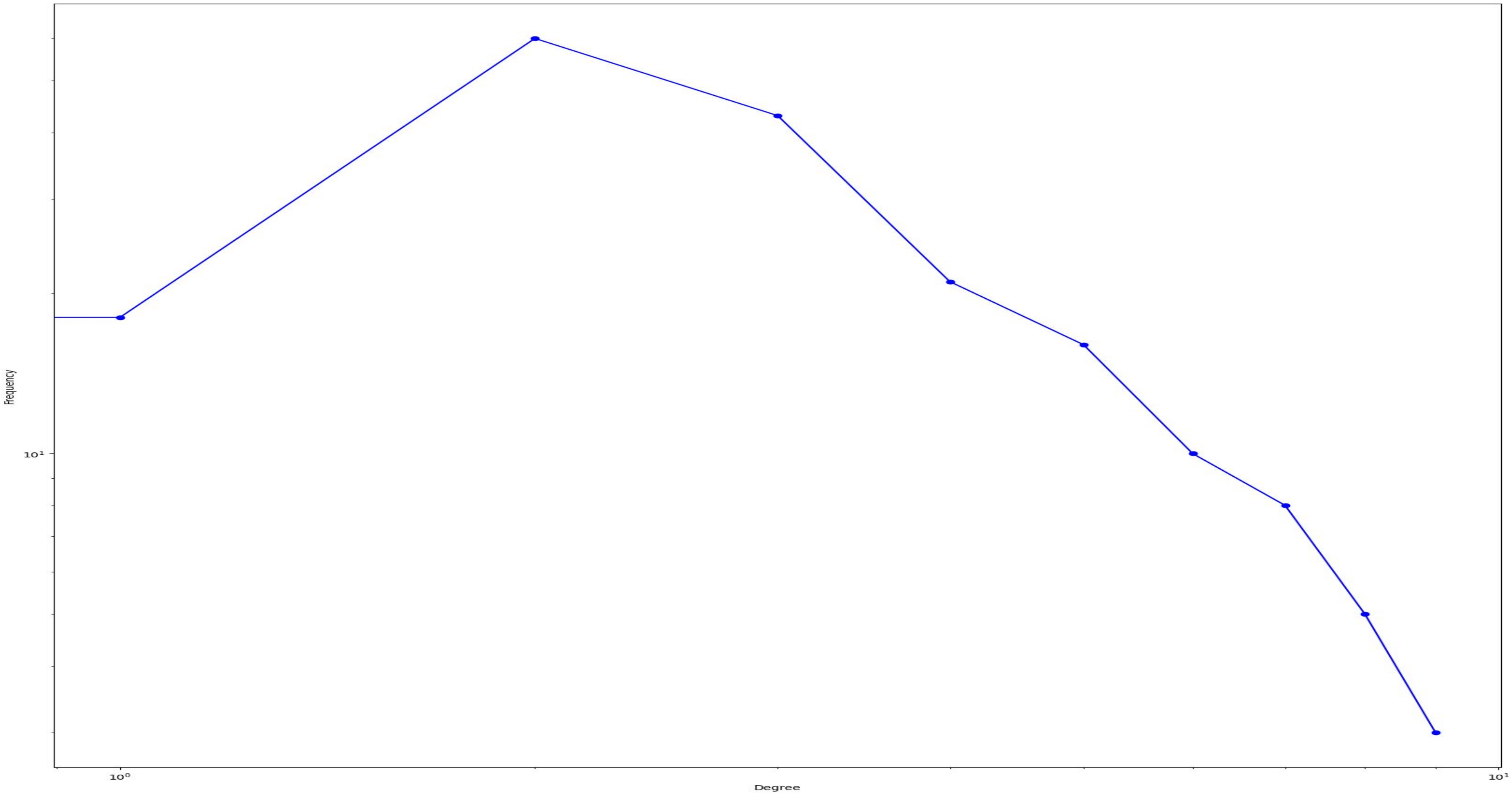
Tendency to form hubs is less

Small worldness to some extent

Backbone



Degree Distribution of Backbone



Conclusion

Existence of hubs proteins in the network

Hubs are distributed in different communities

Existence of many connector proteins that connect different communities

Important proteins in the network are connected more cohesively