Abstract

Protein-Protein Interaction (PPI) networks are believed to be important sources of information related to biological processes and complex metabolic functions of the cell. Protein complex is a physical aggregation of several proteins via molecular interaction with each other at the same location and time. Cellular processes are typically carried out by protein complexes and functional modules. Identifying protein complexes are of great importance for understanding cellular organization and functions of organisms.

In a PPI network, high density modules correspond to protein complexes. Many protein complexes are not dense sub-graphs and neglect many peripheral proteins that connect to the core protein clusters with few links. The underlying topology reveals that protein complexes are densely connected within themselves but sparsely connected with the rest of the PPI network.

In most of the cases, similarity of two proteins are measured based on their topological structure. Since functional information of gene product is stored as Gene Ontology, so along with the topological structure the information stored in gene ontology can be used to find similarity between two proteins. In this work, a method is proposed to find protein complexes in a PPI network based on topological structure which incorporates GO information in the clustering process. Our method has been biologically validated by GO semantic similarity score and co-localization score and satisfactory results have been obtained. We have also compared our method with some well known methods and obtained better results in terms of various parameters.

Keywords: Protein-Protein interaction network, Protein complex, Gene Ontology, normalized term overlap, modularity, objective function, precision, recall, F-measure.