

# An Efficient Clustering algorithm for Discovering Protein Complexes and Functional Modules in Protein-Protein Interaction Networks

Bahar Sayoldin

**Short Abstract** — An emerging branch of biology is the study of proteins and how they influence biology. With complete genome sequencing, we can understand the expression, function and regulation of the proteins encoded by an organism. The full complement of proteins is called the proteome which has very wide application to explain how biological processes occur at the molecular level, how they differ in various cell types and causes disease. In general, proteome data provides a wide variety of studies on different protein properties with the purpose of understanding more details of structure, function, and control of biological system in health and disease. Protein interactions control many biological processes and can explain many physical traits. Therefore, identifying these protein-protein interactions (PPI) are important to systematically understanding their cellular role. These protein-protein interactions have different role in biology base on their composition, how they are dependent and life time of their association. The analysis of annotated protein help to discover that the protein in the same cell often interact with each other. Therefore, the function of unknown protein in the same cell can be discover based on their interaction with known protein. Furthermore, the interaction between two or more proteins can become inactive or destroy a protein, creating a new binding site. In addition, clustering algorithms can help locate strong factors that behave in the cell together. The goal is to find efficient clustering approaches to discover protein complexes or functional modules in protein-protein interactions

**Keywords** — Protein-Protein Interaction, Graph-based Clustering Techniques

## I. PURPOSE

Experimental and computational methods have increased the number of known or putative interactions in databases. Two main traditional approaches to study protein-protein interactions include Yeast Two-hybrid System and Mass Spectrometry[1]. There are also many other tools such as Microarray for discovery of the all possible protein-protein interactions. However, in some cases the data generated from these tools are erroneous. Therefore, developing a method to discover the accurate possible interaction and prediction is very useful. Some of technique use data-mining technique to identify pattern and the accuracy prediction of PPI such as Graph-based clustering methods [2]. The main purpose of clustering in protein interaction networks is to group the proteins into sets of clusters with more similarity within proteins in the same cluster than in different clusters [3]. In general, using graph-based clustering technique use for detecting protein complexes and functional modules in protein interaction networks. A

protein interaction graph can be weighted or unweighted. However, in most cases, protein interaction network is represented as an undirected graph  $G(V,E)$ . In this graph, vertices are proteins and edges are the interactions between proteins. The relationship between two proteins indicates by value 0 and 1 which represents the probability that this interaction is a true positive. The value of an interaction is 1 when the proteins interact with each other otherwise the value is 0. In an unweighted protein interaction graph, an assumption is made that this number is equal to 1 for all the edges of the graph. In general, the protein interaction graph provide a valuable tool that helps the better understanding of the functional organization of the proteome[4]. The purpose of this work would be finding clusters which are correspond to important components in proteinprotein interaction network. Theses component can be protein complex and functional modules. Protein Complexes are groups of protein that interaction with each other in the same place and forming a single multimolecular machine. On the other hand, functional models are groups of protein that interact with each other in a different time and place at the particular cellular. There are several benefits of clustering in PPI such as • Identify the PPI network and their related component • Find the function of each cluster based on the function of its member • Clarify of possible functions' of cluster members and comparing these functions with the function of other cluster members. The two main challenges to deal with this problem are finding a proper way to present a PPI in a graph and apply an efficient clustering algorithm on the graph.

## II. CONCLUSION

Propose an efficient clustering algorithm to discover the protein as cluster which help biologist predict cell behavior.

## REFERENCES

- [1] Lin, C., Cho, Y., Hwang, W., Pei, P., & Zhang, A. (n.d.). Clustering methods in protein-protein interaction network. *Department of Computer Science and Engineering State University of New York at Buffalo*, Retrieved from [alfred.cse.buffalo.edu/DBGROUP/bioinformatics](http://alfred.cse.buffalo.edu/DBGROUP/bioinformatics)
- [2] Wang, J., Li, M., & Deng, Y. (2012). Recent advances in clustering methods for protein interaction networks. *11*, Retrieved from <http://www.biomedcentral.com/1471-2164>
- [3] Spirin V, Mirny LA: Protein complexes and functional modules in molecular networks. *PNAS* 2003, 100:12123-12128.
- [4] Moschopoulos, C., Mavroudi, S., Likothanassis, S., & Kossida, S. (n.d.). Constructing and working with protein interaction graphs. *Department of Computer Engineering & Informatics, University of Patras, Rio, Patras, Greece*, Retrieved from [journal.embnet.org/index.php/embnetnews/article/view/14](http://journal.embnet.org/index.php/embnetnews/article/view/14)
- [5] Kossida, S. (2011). Noise reduction in protein-protein interaction graphs by the implementation of a novel weighting scheme. (10.1186/1471-2105-12-239), Retrieved from <http://www.biomedcentral.com/1471-2105/12/239>

<sup>0</sup>Acknowledgements:

<sup>1</sup>Department of Computer Science, New Mexico State University. E-mail: [sayoldin@nmsu.edu](mailto:sayoldin@nmsu.edu)