International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-2, December 2018

An Interactive Tool for Protein - Protein Interaction Network Visualization

Varsha H. Patil, Swati K. Bhavsar

Abstract: ***In scientific research, visualization of protein- protein interaction network is an important aspect. Protein- protein interaction contains two or more proteins binding together to form an interaction network. This interaction network is formed with nodes as proteins and interaction between them as edges. In this paper, an algorithm Graphviz as an interactive visualization tool for protein- protein interaction network visualization has been discussed. The proposed algorithm is tested on PPI network datasets. Experimental results clearly illustrates that it works efficiently for visualization of PPI densed graphs.***

Index Terms: Interaction, Protein, PPI, Visualization

I. INTRODUCTION

Biological information is managed with the help of computer and information technology. Its application is known as bioinformatics. Bioinformatics is the application of computer and information technology to the management of protein -protein interaction is a molecular clocking between proteins occurs in a cell [1]. PPI networks are used in many medical applications for predicting the risk of metastasis and to classify breast cancer into sub types [2].

PPI network is represented as a graph, where nodes represent proteins and the edges are interactions between proteins in a network. This network of graph is represented as either directed graph or undirected graph. In undirected graph model of PPI, if protein X interacts with Y, then protein Y interacts with X [3]. Also the number of edges connected to a particular protein node exhibits the degree of a particular node. And furthermore its statistical model is represented with degree distribution. These models need to be properly visualized and it leads to the protein network graph visualization. Hence visualization of graph is very popular in the domain of information visualization [4]. The graphs are fundamental structural representation of information which is formed by interconnection of nodes and edges. Visualization of this graph is the visual representation of the nodes and edges of a graph. Visualization tools are an essential layer to identify and analyze insights from connected data. Graph visualization is useful because of many reasons like-

1. Less time for assimilate information: As the human brain processes visual information much faster than written one. Visual data always ensures better understanding and reduces the time to action.
2. Better understanding of problem: one can achieve better understanding of problem by visualizing pattern and contests. Graph visualization not only visualizes relationships but also assist to understand the contest of data.
3. Effective form of Communication: Visual representations are more effective medium to share finding and offer more instinctive way to understand the data.
4. Easy to use: Any user without special technical and programming skills can interact with graphs visualization.

The paper is organized as follows. Section 2 presents related works for visualization techniques, Section 3 presents methodology and core ideas of proposed algorithm. Experimental set up is presented in section 4. Section 5 discusses results. Finally, section 6 summarizes conclusion.

II. RELATED WORK

Over the last decade, significant development has been seen in development of protein-protein interaction (PPI) visualization system. The currently available system for graph visualization is mainly focused on small graphs [2, 15]. Some of the techniques are working good for large graph visualization but works only for static graph. These techniques are application specific and only applicable to graphs having specific structural properties. Shannon et al. proposed visualization tool known as Cytoscape. The system provides visualization and relationships between nodes and interaction in user friendly manner [5]. Cline et al used cytoscape for combining a PPI network for gene expression profiling [6]. Breitkreuts et al. proposed Ospery visualization tool which provides graphical representation of gene function in color coded form [7].Royer et al. developed additional plugins for special purposes in visualization of protein-protein interaction network [8] termed a power graphs. These power graphs are having additional feature of being compressibility and applicability of all types of networks such as protein interaction, regulatory networks or homology networks [8]. Pavlopoulos and O'Donogheu et al. come up with visualization tool known as Arena3D. It contains layers of data in 3D space.By using different clustering algorithms; data is clustered and placed on different layers [9-11]. Freeman et al. proposed a tool for layout, visualization and clustering of large scale networks known as BioLayout Express 3D[12]. The Markov Clustering Algorithm (MCL) was used for clustering analysis. Pavlopoulos et al. introduced JClust and Medusa tool for cluster analysis algorithms [13, 16].

Manuscript published on 30 December 2018.
***Correspondence Author(s)
Dr. Varsha H. Patil, Department of Computer Engineering, Matoshri College of Engineering & Research Centre (MCERC), Savitribai Phule Pune University, Pune (Maharashtra), India.
Ms. Swati K. Bhavsar, Department of Computer Engineering, Matoshri College of Engineering & Research Centre (MCERC), Savitribai Phule Pune University, Pune (Maharashtra), India.

By study of all these visualization techniques, it has been observed that if size of graph is too large, leads to major challenge in efficient graph visualization. It can cause various problem named as-

i. Algorithmic Complexity: As the graph size is too large, algorithms take more time for its processing. Graph algorithms are NP Complete or NP hard [17-19].
ii. Display Clutter: Graph contains thousands of elements and more. When the size of data grows, graph becomes visually confusing. Data is so large it could succeed the number of display pixels [17].
iii. Readability: Human perception able to visualize only small graphs including finding the node and links between nodes [17].
iv. Navigation: Navigating large information spaces, such as graphs with thousands of nodes suffers from small display problems of viewing a large space [17].

Along with these problems, there are some major constraints to be understood while designing visualization system are- user constraints, data constraints and aesthetic constraints.

User constraints: It contains important aspect of interactive tasks the user is carrying out while visual representation. Compatibility of visualization and user: User is working on some scientific research project and his required visualization system may vary according to application. Therefore user preference while evaluating visual representation is more important and it leads to compatibility of visualization tool and user.

Compatibility of visualization and interaction techniques: user is interested in exploration of graph structure in depth. Proper exploration with easy searching of required data is important aspect. It can be achieved by compatibility of visualization and interaction techniques.

Compatibility of different visualization techniques: The main focus of graph analysis is not only on structure of graph but on the internal relationship among vertices and edges. This technique is known as Needle Grid [10].

Data constraints: Data available for visual representation in some of the forms as-

Static Data vs. Dynamic Updates: Static data can be visualized efficiently by using various graph representation layouts like node, space, 3D and matrix layout. But for dynamic data, it not works much efficiently as original layout is disturbed [12].

Size and Density of the Graph: for densed graphs visualization, it suffers from major problem called as "screen bottleneck", and it occurs when the number of nodes to display at once exceeds the number of available pixels on the screen [1,17].

Aesthetic constraints: The hardest problem when visualizing graphs is to find an appropriate layout in the 2D- or 3D-space. Some of the most prominent aesthetical criteria for a desirable graph layout are:

- Less number of edge crossings
- Small area of drawing
- Less number of bends along the edges
- Small but uniform edge lengths

Hence, there is need to find the solution for considering all the constraints and develop an efficient visualization tool for a graph.

All the techniques discussed in section 2 were applicable to the original graphs. But sometimes the graph size is too large and cannot be loaded into limited memory, so need to partition these large graphs into smaller ones and visualize it separately to retrieve important information from it. For better representation and illustration of graphs and partitions (sub-graphs) generated after partitioning graph G by using Cutset Computing and Partitioning Algorithm(CPA), graph visualization tool is required. Existing tools for graph visualization are application specific and they required input data in specific format only. To overcome this limitation, visualization tool is developed using vis java script library [14].

III. METHODOLOGY

Input dataset is given as input to pre processing unit. Preprocessing generates output as graph $G(V, E)$, where $V=\{v_1,v_2,v_3..\}$ and $E=\{e|\ e= (v_i, v_j,e_{ij})\}$, where v_i, v_j and l_{ij} are source vertex, destination vertex and label of edge e.

Algorithm 1: Protein –Protein Interaction network Visualization

Input: Graph $G\ (V, E)$
Output: Graph G is visu alized as a network where nodes represent vertices in a graph and connectivity between two vertices represented as an edge.
PP_nodes_array represents vertices of network graph as vertex id as a label
PP_edge_array represents edge between two vertices
1. Initialize *PP_nodes_array*
2. Initialize *PP_ edges_ array*
3. Create an empty dataset.html file
for each vertex v_i in Graph G
for each vertex v_j in graph G
If v_i and v_j are adjacent vertices then
Update *PP_nodes_array* by using label of v_i
End if
Update *PP_* edge_array by using edge connectivity from v_i to v_j.
5. Write *PP_nodes_array* and *PP_edge_array* to dataset.html file

IV. EXPERIMENTAL SETUP

Proposed Algorithm has been implemented and tested on the 1GHz, single core CPU: 512 MB RAM. Performance of "Graphviz: Graph Visualization Algorithm" is tested on standard protein- protein interaction network dataset as full net.

V. RESULTS

Full net protein- protein interaction dataset contains 20,095 vertices and 25964 edges. It is partitioned into 10 clusters using CPA clustering algorithm. Every cluster is visualized by developed visualization tool. Results obtained after visualization of first 8 clusters are as shown in figure 2(1) to 2(8).

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

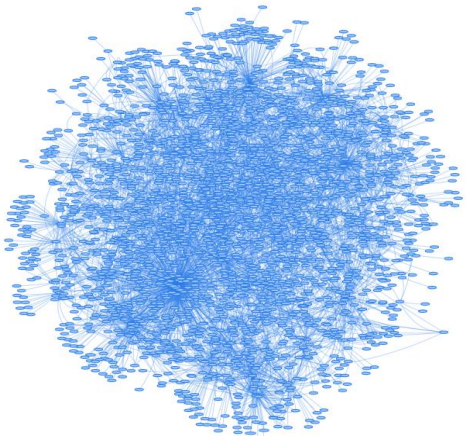

Fig 2 (1): Sub-graph 1

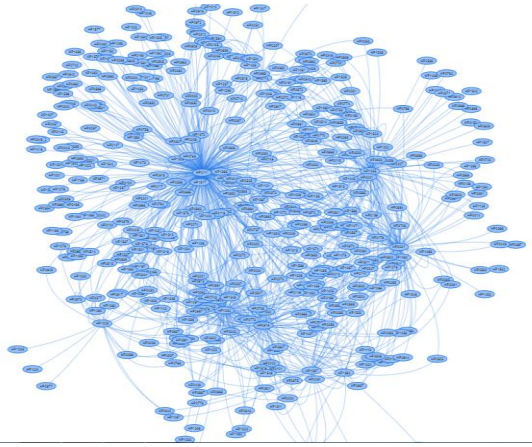

Fig 2 (2): Sub-graph 2

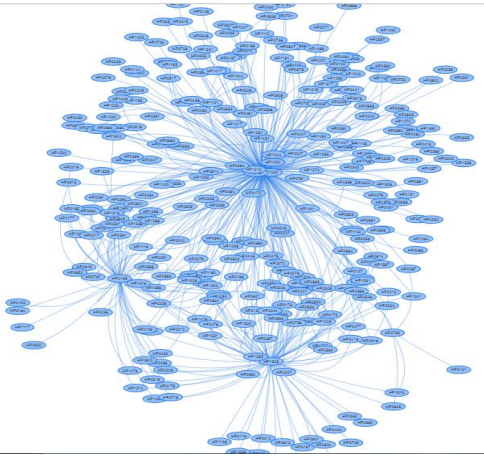

Fig 2 (3): Sub-graph 3

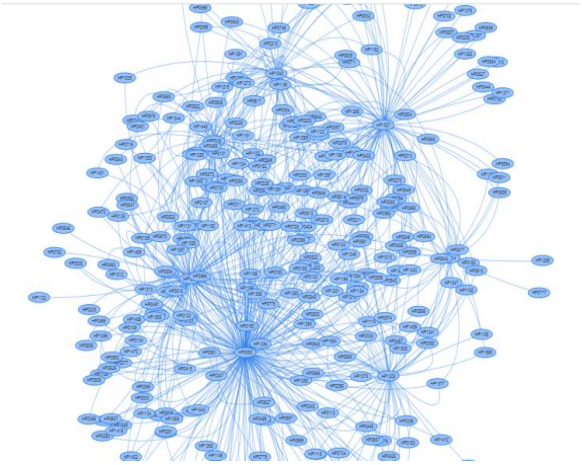

Fig 2 (4): Sub-graph 4

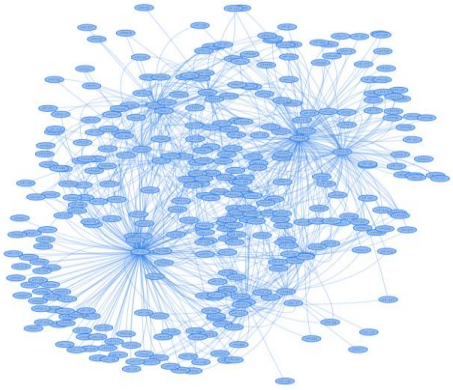

Fig 2 (5): Sub-graph 5

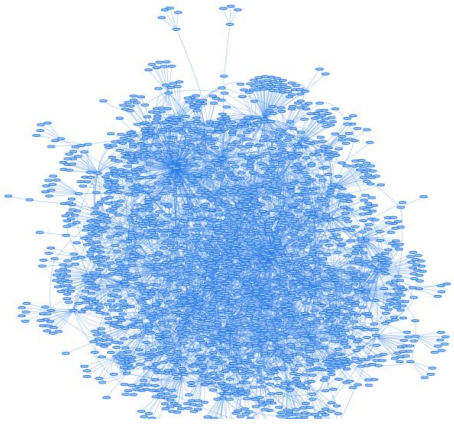

Fig 2 (6): Sub-graph 6

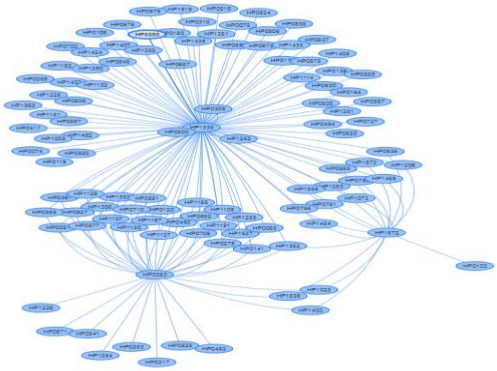

Fig 2 (7): Sub-graph 7

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

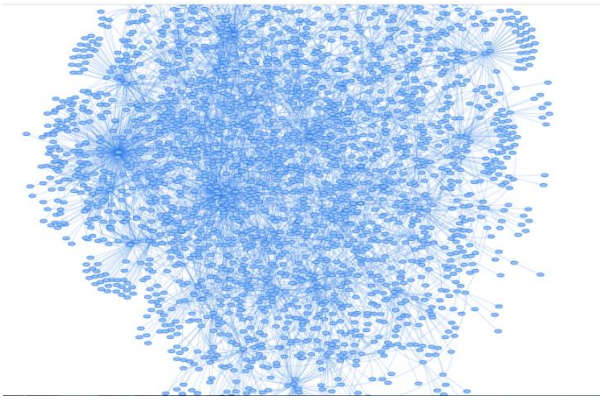

Fig 2 (8): Sub-graph 8

VI. CONCLUSIONS

The central problem addressed in this paper is visualization of protein- protein interaction network graph using an efficient visualization tools. Graphviz: An Interactive Visualization Tool for PPI densed graph has been implemented. It is useful to visualize any PPI densed graph represented in vertex edge form. The performance of the algorithm is evaluated using standard datasets as input which are considered as benchmarks for assessment of graph visualization algorithms. It is observed that Visualization tool is capable of efficient dynamic visualization of 128 sub-graphs obtained after partitioning a large protein- protein interaction network. Visualization tool efficiently visualizes protein- protein interaction network. This visualization tool will be useful for visualizing many complex data sets. The tool avoids clutters of edges and efficiently utilizes display area, provide better and quick visualization of a graph with vertex labels and edge labels.

REFERENCES

1. G Fiona Browne, Huiru Zheng, HaiyingWang, and Francisco Azuaje, “From Experimental Approaches to Computational Techniques: A Review on the Prediction of Protein-Protein Interactions”, Advances in Artificial Intelligence, Hindawi, vol. 2010, pp. 1-15.
2. Soon J., Yoon B. J. & Doufherty E. R., “ Identification of Diagnostic Sub network Markers for Cancer in Human Protein-Protein Interaction Network”, BMC Bioinformatics, 2008, 11:8
3. Gentleman R. & Huber W. , “Making the most of high- throughput protein- interaction data”, Genome Biology8(10), 2007, pp 112-115
4. Hans-Jorg Schulz, Heidrun Schumann, “Visualizing Graphs - A Generalized View”, Proceedings of the Information Visualization (IV’06), © IEEE-7695-2602-0/06, 2006, pp 1-8
5. Shannon, P. , “Cytoscape: A Software Environment for Integrated Models of Bimolecular Interaction Networks”, Genome Res 2003, 13(11), pp. 2498-2504
6. Cline, M.S., “Integration of Biological Networks and Gene Expression Data Using Cytoscape”, Nature Protocols 2, 2007, pp. 2366 - 2382.
7. Breitkreutz, B.J., Stark, C. & Tyers, M. Osprey, “A Network Visualization System”, Genome Biology 4(3):R22, 2003.
8. Royer, L., “Unraveling Protein Networks with Power Graph Analysis “, PLoS Comput Biol 4(7): e1000108, 2008.
9. Pavlopoulos, G.A. & O'Donoghue, S.I., “Arena3D: Visualization of Biological Networks in 3D”, BMC Systems Biology, 2008, pp 102-104.
10. Pavlopoulos, G.A., “A Survey of Visualization Tools for Biological Network Analysis”, Bio Data Mining, 2008, pp 1-12.
11. Pavlopoulos, G.A., “jClust: A Clustering And Visualization Toolbox”, Bioinformatics 25(15): 2009, pp 1994-1996.
12. Freeman, T.C., Goldovsky, L. & Brosch, M., “Construction, Visualization, and Clustering of Transcription Networks From Microarray Expression Data”, PLoS Computational Biology, 3(10), 2007, pp 2032-2042.
13. Lima-Mendez, G .& van Helden, J., “The Powerful Law of The Power Law and other Myths In Network Biology”, MolBiosyst 5(12), 2009, pp. 1482-1493.
14. Almende B.V., “vis library Copyright (C) 2010
15. Javier De Las Rivas, Celia Fontanillo, “Protein – Protein Interactions essentials: Key Concepts to Building And Analyzing Iteractome Networks”, Bioinformatics & Functional Research Group, Cancer Research Center,Vol. 6, issue 6, 2010, pp 1-8
16. William Andreopoulos, Dirk Labudde, “Protein-Protein Interaction Networks”, pp 1-24.
17. James Abello, Frank Van Ham, Neeraj Krishan, “ASK GraphView: A Large ScaleGraph Visualization System”, IEEE Transactions on Visualization and Computer Graphics”, Vol. 12, No.5, October 2006, pp. 1587-1598.
18. J. Zhang., “The interaction of internal and external representations in a problem solving task”, Proc. Thirteenth Annual Conference of Cognitive Science Society, 1991.
19. Weiwei Cui “A Survey on Graph Visualization”, Hong Kong University of Science and Technology Clear Water Bay, Kowloon, Hong Kong, pp. 1-52

AUTHOR PROFILE

Professor Dr. Varsha H. Patil is Doctorate in Computer Engineering. She is serving as reviewer for national and international journals and conferences of repute. Her areas of interest include Image Processing, Data Mining and Graph Analysis. She has 50+ papers to her credit. She is life member of various professional bodies. She is recipient of many prestigious honors and awards. Currently she is chairman of board of studies (Computer Engineering, Savitribai Phule Pune University, Pune)

Ms. Swati Bhvsar is assistant professor at Savitribai Phule Pune University. Her topics of research include image processing, data mining and graph analysis. She is life member of various professional bodies.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)