A STUDY ON NETWORK VISUALIZATION TOOLS

Harish Babu.Kalidasu[1], Battina Pullarao[2], Venkateswarao Molli[3]

[1,2,3] Assistant Professor, St.Jospeh University in Tanzania, Dar-es-salaam, Arusha campus

ABSTRACT

Visual datamining supports interactive and scalable networks visualization and analysis which allows effective exploration and statement of ideas inside the various number of biological and biomedical fields. Huge networks like protein interactome or transcriptional regulatory networks contains tons and tons of objects and millions of relationships. As these networks are constantly developing a new knowledge which becomes available. Then their content is abundantly noted and it can be presented in several different methods. However, by trying to discover the knowledge and new models of theories within particular complex datasets which can involve various workflows represents many designs of source data distributes data sources and merging heterogeneous database searching and integrating results from several computational and mathematical analyses which effectively analyzing their properties and results. As the researchers required the needs of the address and in the design and development of scalable and interactive network visualization and analyses.

I. INTRODUCTION

The exploration and understanding of relationships between biological molecules, networks and ideas become a major key of blockage in systems biology. The pure amount of data and their heterogeneity delivers a test for the visualization of data. Here, we are having different graph representations which are available. Furthermore, it often maps the data on 2D graphs to visualize biological interactions. These procedures are acceptable to a wide range of problems. Though many of them to reach their boundaries in terms of sociability, approachability, responsiveness, friendliness and pleasantness when millions of nodes and connections have to be analyzed and visualized. In this study, we reviewing on some of the visualization tools that are presently available in visualization and biological networks generally developed in the earlier years. We mention some of the tools functionalities, limitations and explicitly developed in future in a way of data integration and information sharing.

Everyday bio-informatics has been changed, expanded, extended, developed continuously and its grown is important connection and linking with life science research. The qualities and quantities extends data obtained by a new high-throughput technologies which includes micro-chips, large-scale OMICS data such as genomics, proteomics and transcriptomics remains huge biological data repositories which are growing exponentially in size. Moreover, every minute in scientific knowledge increases thousands of thousand pages to read the new scientific data or information produced in 24x7 hours while a scholar / researcher takes several years. To monitor the scientific output produced from a single disease for example breast cancer, a researcher have to scan read, search more than a hundred various journal papers per day. The basic datasets shows increasing complexity and dynamics that are produced by various application areas. The combination of various types of data is important. Presently various biological types of data such as protein structures and families, sequences,

ontologies, proteomics data, gene expression and other experimental data which are stored in dissimilar databases. At present databases or data collection will be focused and frequently store the information using detailed data formats. Many researchers or scientists overlapping but not exactly match the information with other databases which introduces another difficulty to combine the information. Now, the understanding into complexities and dynamics of biological systems, the data or information stored in the various data repositories needs to be combined and linked in well-organized ways. To report this issues, the data integration[1] is a key subject in the previous years.

In modern days different data sources are available. For example DNA micro array tests produce millions of gene expression measurements which provides modest way of collecting massive amounts of data or information within a short time period. It is used to collect the data from tissues and cell samples about the differences in gene expression while comparing the traditional diagnostic methods which are based on mainly the morphological appearance of growth, the methods depends on gene expression profiles are more precise, natural, objective, dependable, consistent and reliable[2]. In this paper we reviewed about some tools from this we try to simplify the analysis and understanding of biological data by transforming the fresh data into logically structured and visually concrete representations. The objective is to find the structures and patterns which remains unknown in the raw unstructured data sets.

II. ANALYSIS OF NETWORK VISUALIZATION TOOLS

In this section we discuss about commonly used network visualization tools that are developed in earlier days. These tools are selected to cover the series of different functionalities and structures important for data analysis and visualization. While these tools are mostly applicable. In this, we point-out some strengths and weakness and mention their specific features. While these tools are analyzing for protein-protein interactions, pathways, heterogeneous networks, gene networks or tools for reviewing evolutionary relations between proteins and requires distinct and much more brief analysis on this review. The criteria for the assessment of visualization tools which includes efficiency, power and quality of network visualizations formed, the compatibility of all other tools and data sources, the analytical functionalities accessed limitations in relation of broad applicability, data quantity and user-friendliness. Following the detailed assessment of the particular strength of an each tool.

III. CYTOSCAPE

Cytoscape[3] is a individual java application. It is an open source project which is under LGPL licnese. The Cytoscape generally delivers 2D representations which are suitable for large-scale network analysis with millions of nodes and edges. Cytoscape supports directed, undirected and weighted graphs and derives with powerful visual styles which allows the user to modify the properties of nodes or edges. This tools provides multiple layout algorithms which includes spring-embedded and cyclic layouts. Moreover, expression data can be plotted as label, border thickness, color or border color. The compatability of Cytoscape derives with other tools. Cytoscape allows the user to import the mRNA expression profiles, gene functional annotations from gene ontology and KEGG.

The functionalities of Cytoscape is highly integrative and user can zoom in or zoom out and look into the network. The position of the network as well as the edge properties or node properties can be reloaded and saved. Cytoscape approaches with network manager to simply organize various networks.

The strengths of Cytoscape is the visualization of molecular interaction network and their integration with gene expression profiles and other related data. It also allows the user to work and manipulate various multiple network.

IV. BIOLAYOUT EXPRESS 3D

BIOLAYOUT EXPRESS 3D[7] is written in java and it is released under GNU public license. To run this software we need standard graphics card and high range graphics cards. It is a tool for layout visualization and clustering of large scale network in 3D and 2D representations. It supports weighted graphs and un-weighted graphs together with the edge interactions. It supports the input of data in a minimum number of standard graph formats, while suitable data from any sources can be visualized by using this tool. A multiple colour schemes extract the networks more than the informative and clusters can be visualized easily. It uses a graphics types of approaches and it is limited in size of network in the process.

The compatibility comes with the simple input file format which requires the user to provide only the list of connections. This tool is compatible with Cytoscape and also supports expressions, layout and sif file formats. Presently it is not connected with data sources only.

The functionalities of highly interactive and the user can control both 2D and 3D representations, users move around the view, move or rotate, zoom in or zoom out the current network. The strength of the BIOLAYOUT EXPRESS 3D offers various analytical methodologies to microarray data analysis.

V. PATIKA

Pathway Analysis Tools for Integration and Knowledge Acquisition (PATIKA)[4] is a web-based non-open source application and it is openly used for non-commercial use. It provides 2D representations of directed graphs or single graphs. There is no limitations for size of the graphs. PATIKA offers a very natural and broadly accepted representation for cellular processes by using directed graphs where nodes resembles to molecules and edges corresponding to interactions between them. However, the implementation of variety of layout algorithms is slightly limited, PATIKA is capable to support the shared graph representation of states and transitions. It signifies various types of edges: product edges-which provides the source and target nodes of a product edge that defines the change and a product of this transition, Activator edges which defines the initiating state and the transition is activated by this state.

The functionalities of PATIKA describes that a user can connect to the server and query the databases which consists the preferred pathway. While the pathways are created and drawn automatically. The user can handle pathway through operations such as adding a new state or remove an current transition, edit their contents such as explanation of a state or transition or change the graphical representation of a particular pathway components.

The strengths of this PATIKA tool is for data integration and analysis of pathway. It is an combined software environment designed which provides researchers a complete solution for analyzing and modelling a cellular processes. This PATIKA is one of the tools that allows to visualize the transitions capably.

VI. PIVOT

PIVOT[5] is a Java application, it is available free for academicians and it has its own license agreement. It plans all in 2D representation and it uses single non directed lines to display the relationships between bio-entities. PIVOT is not a partial regarding to the size of data it can present. Generally the range of combined layout algorithms is inadequate, but the PIVOT works on detailed layout algorithms for visualizing relations.

The compatibility of PIVOT configured their work with proteins from different species like human, yeast, drosophila and mouse, it present an functional annotations and identification of homologs from the four species, and the protein data or files are stored in an MS-Access file, it is easily modifiable by the users to enter their particular data or information.

The functionalities of a PIVOT can develops the network to show all the proteins to a particular stated distance, to detect the shortest path of a communications or unfold the relationships between "distant" proteins, which gives reply similarly to a conditions occurring in resulting experiments. It focusses on dense areas of the map and procedure to visualize a window and subarea of a huge networks.

The strengths of a PIVOT is finest way of well-matched visualizing protein-protein interactions and also identifies the relationships between them.

VII. SUMMARY AND DISCUSSION

The area of data visualization presently looks on few major challenges: constantly increasing the quantity of data which is to be visualized and analyzed, the combination of various dissimilar data and the representation of several links between the nodes with heterogeneous biological values. As the analysis shows that every visualization tools has detailed features and then these tools differ in how they address the bounded challenges. Once the heterogeneity of data is the key challenge, integrate tools like PIVOT offered possible results or explanations. Whenever the pure form, but fewer the heterogeneity carries the bottleneck, tools with high determination and descent scaling functionalities like Cytoscape are well organized to help over their boundaries.

Now-a-days biological projects and tests becomes more complex and larger in space and the data produced greater amounts than earlier. The cumulative increasing use of high-throughput technologies enlarges the amount of data produced per each experiments and quickly increases the sizes of the databases. In the analyses stage we can classify the visualization of data is already major problem. While the pure quantity of data and their heterogeneity carries a challenge for well-organized visualization tools. The main objective of the visualization tools must be natural representation of data which provides an effective understanding and to permits a theory focused on planning of the succeeding experiments. The tools which are represented in the review are related to the wide range of problems and their different structures or features makes suitable for a widespread range of applications. Though, constant improvement of visualization algorithms and best surviving

visualization methods spread the limitations in terms of user friendliness whenever millions of nodes have to be visualized and analyzed. In order to expand the big datasets having maximum of layout algorithms. While the layout problems are quiet remains and one of the crucial problem in network visualization. As the quicker, faster and more understandable algorithms are necessary to carry the large scale networks into a form and it can be easily understood by the human brain.

In-addition to the extension of layout algorithms to comprise a third aspect or dimension and it would be unique central step concerning a new generation of visualization tools. It becomes more significant like pathway or heterogeneous data sets visualization. Furthermore, the aspects allows a clear structure and less tangled views and it could be strongly ease an enhanced navigation within the network. The rapid development of data demands the combination of controlling filters into visualization tools. In-order to reduce the noise in a particular dataset restrict their users attention to an essential set of nodes of a particular interest which improves visualization. Likewise, more efficient and cooperation of Graphical User Interfaces (GUIs) which allows the user to visualize and search the relevant subnetworks or inadequate ranges of a whole network without consuming to mesh their vast data masses. While increasing the performance of a visualization tools handling very efficiently and allocating the memory is essential. It is succeeded by loading the necessary parts of the graphs in memory. In this method, the quantity of data and the classifications that it can be visualized. While the Graphical User Interfaces (GUIs) performance increases over the time and allows the visualization tools to more sources on demanding algorithms which are advanced graphics designs.

The aim of the future generation visualization tools should reduce the gap between analysis and visualization. Almost all the existing visualization tools integrate a restricted number of data analysis functionalities which produces necessarily and continuously difference between the dissimilar applications. The users have to be aware of the multiple tools which are suitable to analyze his data and must change between them. Due to the standard file formats the sharing between the information and data should be simple task, this should be supported by newly developed visualization tools. Unlike standard formats are applicable for various different data types, this will be features for producing the integrated heterogeneous data into a network.

REFERENCES

[1]. LENZERINI- Data integration: A theoretical perspective. PODS 2002, pg.no: 243-246

[2]. B.HANCZAR, M.COURTINE, A.BENIS, C.HENNEGER, K.CLEMENT, J.D.ZUCKER- Improving classification of microarray data using prototype based feature selection, ACM SIGKDD exploration, 5(2), Pg.no: 23-30, 2003

[3]. SHANNON P, MARKIEL A, OZIER O, BALIGA NS, WANG JT, RAMAGE D, AMIN N, SCHWIKOWSKI B, IDEKER T: Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 2003, 13(11):2498-2504

[4]. DEMIR OB E, DOGRUSOZ U, GURSOY A, NISANCI G, CETIN-ATALAY R, OZTURK M: PATIKA: An Integrated Visual Environment for Collaborative Construction and Analysis of Cellular Pathways. Bioinformatics 2002, 18:996-1003

[5]. NIR ORLEV RS, YOSEF SHILOH: PIVOT: protein interactions visualization tool. Bioinformatics 2003.

[6]. KANEHISA M, GOTO S, KAWASHIMA S, OKUNO Y, HATTORI M: The KEGG resource for deciphering the genome. Nucleic Acids Res 2004, (32 Database):D277-280

[7]. FREEMAN TC, GOLDOVSKY L, BROSCH M, VAN DONGEN S, MAZIERE P, GROCOCK RJ, FREILICH S, THORNTON J, ENRIGHT AJ: Construction, visualisation, and clustering of transcription networks from microarray expression data. PLoS computational biology 2007, 3(10):2032-2042

[8]. Hooper SD, Bork P: Medusa: a simple tool for interaction graph analysis. Bioinformatics 2005, 21(24):4432-4433