ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

Survey on Large Graph Analysis and Visualization

Rabina P

Assistant Professor, Dept. of CSE, Malabar Institute of Technology, Kannur, Kerala, India

ABSTRACT: Graphs are abstract representations that can describe a large set of real world phenomena having hundreds of thousands of nodes and millions of edges. Graphs are commonly used to solve many problems by representing it in the form of graphs. Large graphs are commonly used in social network, computer networks etc. Large graph is analysed to find the patterns of interest and it can be performed by means of visual interaction. There are some challenges in analysing large graph that is its processing requirements and visualization.

KEYWORDS: Graph visualization, Large graph, Sub graph, Graph representation, Graph analysis, Data structures, Graph mining.

I. INTRODUCTION

A graph is a representation of a set of objects where some pairs of objects are connected by links. The interconnected objects are represented by mathematical abstractions called vertices, and the links that connect some pairs of vertices are called edges. Graphs can be used for ranking hyperlinks, social network analysis etc.

Large graph consists of hundreds to thousands of nodes and millions of edges. Web graphs, social networks, computer network, Transportation network are some examples of large graph. As it is a complex data structure such graphs require excessive processing, more memory for storage and knowledge of a pattern of the graph. It is very difficult to predict the exact size and pattern of a large graph as it changes with time. Large graph analysis starts with division of the input graph into number of small parts called as sub graph because whole graph cannot fit into memory for processing at given time. Graph summarization finds the strong connected component that is a node which is connected to maximum nodes in the sub graph. All such components are then used to maintain connection between different sub graphs by using hierarchical representation.

The large graph [1] problem has been treated through graph hierarchies, according to which a graph is recursively broken to define a tree of sets of partitions. However, the efforts on this matter fail on the task of integrating the information from multiple partitions, disregarding mining techniques to fine inspect each sub graph. Conversely, for understanding a graph hierarchy, it is worthwhile to have systems that provide aids for answering the following questions.

- Hierarchical navigation: What is the relation between arbitrary groups (partitions)of nodes.
- Representation and processing: What are the adjacencies of a given graph node considering the entire graph, and not only its particular partition?
- Mining: Given a subset of nodes in the graph, what is the induced sub graph that best summarizes the relationships of this subset?
- Visualization: How to see through the levels of the graph hierarchy?
- Interaction: How to perform all these tasks efficiently and intuitively?

Visualization approach can be top down or bottom up. Commonly bottom up approach is preferred because top down has the drawback of lack of entire overview.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

II. APPLICATIONS OF LARGE GRAPH ANALYSIS

Some of the applications [3] of large graph analysis is listed below:

Social network: A social network is a social structure made up of a set of social actors (such as individuals or organizations) and a set of the dyadic ties between these actors. The social network perspective provides a set of methods for analysing the structure of whole social entities as well as a variety of theories explaining the patterns observed in these structures. The study of these structures uses social network analysis to identify local and global patterns, locate influential entities, and examine network dynamics. Here the challenge is to understanding communities, intentions, population dynamics, pandemic spread, transportation and evacuation. Face book adds more than one lakh users, 55M status updates, 80M photos daily; 1B active users with an average of 130 friend connections each. All data is rich, irregularly connected to other data. All is a mix of good and bad data. And much real data may be missing or inconsistent .So from all these analysing and summarization is difficult.

Chemical and Biological Data: Graphs can be used to represent chemical data efficiently. Graph nodes are used to show atoms and graph edges or links shows bonds between these atoms. Detailed structure of chemicals can be obtained by representing sub-units of the molecules using data and their bonds using links. For example Protein-Protein interaction shows Nodes corresponding to Proteins and edges corresponding to interaction between these proteins. This type of data becomes very important when analysed to seek out any new drug. Repeating representation of same node allows one to find certain patterns from them. Pattern finding results into sub graphs which in turn leads to give detailed results about specific pattern (chemical or compound). The same method is used to represent biological data. Nodes as amino acids and edges as a link between them are represented using graphs.

Computer Networks: A computer network or data network is a telecommunications network which allows computers to exchange data. To show computer network using a graph is the best way to represent it. Computers are connected to each other in a network to form interconnection. Computers are nodes and their interconnections are edges in this scenario. This type of structure is used to traverse the graph to find out various route, shortest route, etc.The connections between nodes are established using either cable media or wireless media. The best-known computer network is the Internet.

Web Data: World Wide Web is a vast collection of web pages connected together. Web data can be exhibited using graphs. Web pages can be shown using graph nodes and edges exhibits links or interconnection between these web pages.In this way, web is generally structured in the form a graph. This graph data is mined to search particular web page, to find patterns in web pages, to personalize the web, etc. Social web site is the best example of web data represented as a graph.

XML Data: XML data is naturally represented in a structured format. Queries are nested into one another; therefore this structure can be viewed as a labeled graph. This structured data becomes more complicated when attributes are added to the nodes. Since mining these queries to get required data has become essential, graph mining is widely used in this field.

III. LITERATURE SURVEY

The survey explores different approaches for analysis and visualizing large graph.

A. *MULTISCALE VISUALIZATION OF SMALL WORLD NETWORKS.*

Small world networks gather highly clustered subsets of nodes that are a few steps away from each other. This work [7] focuses on the use of the small world properties of networks to support the visualization process. This technique allows to compute the decomposition of a small world network into its highly connected components prior to the visualization, and to offer the user an abstract view of the network. D.Auber,Y.Chiricota, F. Jourdan, and G. Melanc Describes a metric that has been designed in order to identify the weakest edges in a small world network leading to an easy and low cost filtering procedure that breaks up a graph into smaller and highly connected components.

Advantage:

- User can easily find groups and subgroups of actors and understand their dynamics.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

B. ***TOPOLOGICAL FISH EYE VIEWS FOR VISUALISING LARGE GRAPHS.***

In this paper [2] D.Emden R. Gansner, Yehuda Koren, and Stephen C. proposes a topological zooming method. It pre computes the hierarchy of graph and the level of detail dependent on distance from one or more foci. This technique combines multi scale display which conveys the global structure of graphs with fish eye display for exploration of small regions. The display shows a detailed view of a region around a focus that the user chooses which provides fewer details as the distance from the focus increases. In one click user can move the focus to expand other areas of the graph and obtain new displays.

Advantages:

- Efficient use of available display and can examine local areas in detail.

Disadvantage:

- Very hard to relate two visualization of the same graph with different foci.

C. ***ASK GRAPH VIEW: A LARGE SCALE GRAPH VISUALIZATION SYSTEM.***

For interaction and navigation,[4] systems uses a node link layout of a clustered graph. It allows the user to collapse and expand the cluster to show the sub graph that is the user is allowed to navigate the hierarchy in a top down manner by interactively expanding individual cluster. Here the data structure used is hierarchy tree and clustering is done by separating articulation point.

Disadvantage :

- This architecture requires preprocessing the entire dataset being able to visualise it.
- ASK-Graph View was not designed to answer a particular visualization question.
-

D. ***GROUSE FLOCKS: STEERABLE EXPLORATION OF GRAPH HIERARCHY SPACE.***

Method in [5] creates the hierarchy using clustering and user can investigate all possible hierarchy instead of one. Here set of operations are provided, using these user can create and modify their hierarchy on their selection. Interaction specifies a cut ,and system draws the parts of the graph on demand as requested by the user that is leaves and met anodes below the cut are hidden from the user. The system can Interactively explore the graph by expansion/contraction of clusters without losing their mental map.

Advantage:

- It eliminates the clutter of node by using this many hierarchies can be investigated within minutes.

Disadvantage:

- The layout of the graph is constantly changing as the user explores the cut of the hierarchy.

E. VISUAL ANALYSIS OF LARGE GRAPHS USING (X,Y) CLUSTERING AND HYBRID VISUALIZATIONS.

In this paper[5]V.Batagelj,W.Didimo,G.Liotta,P.Palladino, and M.Patrignani proposes a new clustering technique where goal is to produce both inter cluster and intra cluster with desired topological properties. (X,Y) cluster means X class defines topological properties of inter cluster and Y defines properties of inter cluster. This allows the users to interactively explore the graph by expansion or contraction of clusters without losing their mental map. One of the challenges is effective use of automatic graph drawing algorithm. This approach guarantees both inter cluster and intra cluster properties.

Advantage:

- Guarantee highly connected graphs inside the cluster and few edge crossings for the graph clusters

F. TUGGING GRAPHS FASTER: EFFICIENTLY MODIFYING PATH PRESERVING HIERARCHIES FOR BROWSING PATHS.

In this paper[8]D. Archambault,T.Munzner,and D.Auber, proposed a system for exploring paths and proximity around nodes and sub graphs in a graph. Tug Graph creates met anodes that contain elements of the underlying graph that are directly connected to the sub graph or node of interest by an edge and call these sub graphs as proximal components. In Tug Graph technique algorithms are implemented for exploring a region of the graph located near a feature. Tug Graph approach by modifying, rather than completely deleting and reconstructing, the hierarchy each time a node or met anode is tugged. Both edge and connectivity conservation are required in order to visualize paths and proximity information. These approaches provide algorithms to find the nodes which exist between or around a node or sub graph of interest in a large graph. When a node in the graph is selected, the layout is adjusted so that the nodes directly adjacent to the focus node are brought in close to it Sub sequently, an adjacent node can be selected and the view is smoothly animated with zoom-in and out for context. It is suitable for displaying the structure near a small set of nodes in a larger graph.

Advantage:

- Used to explore structure near a feature.

Disadvantage:

- Can't handle weighted graph.
- Large number of proximal components can't be easily summarized in path preserving way.

G. LARGE GRAPH ANALYSIS IN THE GMINE SYSTEM.

In this paper[1] Jose F. Rodrigues Jr,Hanghang Tong, Agma J.M. Traina, Caetano Traina Jr, and Christos Faloutsos concentrates on issues of organising,visualysing and summarising large graph. GMine introduceses the concept of super graph ,Graph tree and CEPS summarisation. It allows to grasp the neighbourhood or group of node in a single click. CEPS is used for local analyzation of sub graph ,it contains collection of paths connecting sub graph of interest. Using the CEPS method, a user can specify a set of query graph nodes and GMine will summarize and present their internal relationship through a small (say, with tens of nodes), yet representative connection sub graph. Here user can specify the size of the sub graph. All nodes have the score corresponding to each query node, and find the total score of each node by adding all its score. After finding the scores selects the node having good scores . CEPS uses Random walk with restart technique to calculate the important score between nodes. Normally adjacency matrix ,adjacency list and binary decision diagrams are used for graph representation. These three are limited to main memory. Here Graph tree structure is used for graph representation. It stores the sub graph on disk and only reference is stored on main memory. CEPS can discover a collection of path rather than a single and is preferable to other methods on describing the multifaceted relationship between entities in social network. EXTRACT algorithm is used for CEPS summarisation. Effective visualization is possible with this technique and it can retain the relations between groups after partitions.

Advantage:

- Effective Visualization
- Memory consumption less.
- Retain the relations between groups after partitions.
- Can answer a particular visualization.

IV. CONCLUSION

Large graph analysing techniques has been discussed. Among the techniques analysed, the large graph analysis using GMine is found out to be better. Lot of research works are going in this area.

REFERENCES

1. Jose F. Rodrigues Jr,Hanghang Tong, Agma J.M. Traina, Caetano Traina Jr, and Christos Faloutsos. "Large Graph Analysis in the GMine System", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING,VOL. 25,NO. 1,JANUARY 2013 .
2. D. Emden R. Gansner, Yehuda Koren, and Stephen C., "Topological Fisheye Views for Visualizing Large Graphs" IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 11, NO. 4, JULY/AUGUST 2005

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 11, November 2016

3. Christos Faloutsos, ”Graph Mining: Laws”,Generators and Tools, CMU
4. J. Abello, F. van Ham, and N. Krishnan, ”Ask-Graphview: A Large Scale Graph Visualization System”IEEE Trans. Visualization and Computer Graphics, vol. 12, no. 5, pp.669-676, Sept./Oct. 2006.
5. D. Archambault, T. Munzner, and D. Auber , ”Grouseflocks: Steerable Exploration of Graph Hierarchy Space”,IEEE Tranthis papers. Visualization and Computer Graphics,vol. 14, no. 4, pp. 900-913, July/Aug. 2008.
6. V. Batagelj, W. Didimo, G. Liotta, P. Palladino, and M. Patrignani, ”Visual Analysis of Large Graphs Using (x,y)-Clustering and Hybrid Visualizations,”, Proc. IEEE Pacific Visualization Symp. (PacificVis), pp. 209-216, 2010.
7. D. Auber, Y. Chiricota, F. Jourdan, and G. Melanc on”Multiscale Visualization of Small World Networks” Proc. IEEE Ninth Conf. Information Visualization (InfoVis), pp. 75-81,2003.
8. D. Archambault, T. Munzner, and D. Auber, ”Tugging Graphs Faster: Efficiently Modifying Path-Preserving Hierarchies for Browsing Paths” ,EEE Trans. Visualization and Computer Graphics, vol. 17, no. 3, pp. 276-289, Mar. 2011.
9. Hanghang Tong, Christos Faloutsos ”Center-Piece Subgraphs: Problem Definition and Fast Solutions”.

BIOGRAPHY

Rabina P is an Assistant Professor in Computer Science Department, Malabar Institute of Technology. She received Master of Technology (M.TECH) degree in 2015 from Govt. College of Engineering Cherthala, Alappuzha. Her research interests are Graph Theory, Theory of Computation etc.