ISSN No: - 2456- 2165

Improve performance of Interactive Visualization for Large Scale Big Data Networks

Neelam Matva
M.Tech Scholar , Computer Science
Jaipur National University , Jaipur
matwa.neelam@gmail.com

Manish Suroliya
Assistant Professor
Jaipur National University , Jaipur
ms786.suroliya@gmail.com

Abstract :- In this particular thesis, we are providing statistical modelling of a big data network. We are exploring the properties of the network like node degree distribution and coefficient of cluster for the purpose of analyzing various network, such as Directed Weighted, Directed Underweight, Undirected Weighted as well as Undirected Unweighted. There are thousands of different data points, which are normally statistical graphics which can come together to provide a cluttered view. If we consider this need to better the interactive visualization for a much bigger database then we may be developing a Pixel Based Overlay Network algorithm inside the MATLAB for the purpose of statistical modelling and visualization of the big picture data network. We are also testing the Pixel Based Overlay Network for the purpose of real world and big scale networks.

Keywords: *Statistical modeling, Large Scale Network, Big Data,Degree Distribution, Cluster Coefficient.*

I. INTRODUCTION

Visualization for the big data is very important part for any kind of analysis data. Most importantly in big data visualization of any particular network which can help the analyst too know and visualize how exactly the big data is being spread over the network. Currently the computing advancements may lead to many big processes towards graphical capacities and their many different possible options for the purpose of displaying data and visualizing it in this present research. Knowing the complicated network on the metric is an example of the degree of cluster coefficient. On the other hand if we look at the network as a smaller stature then it might not be a big problem it is only when the network is huge this kind of problems might take longer time to compute as it can transform with the help of device memory and many different resources for the purpose of computing such required factors. One more very important factor that has to be considered is the algorithm which are going to be used in the system as it can harm the tuning time similarly as in the sub linear complexities. (Sub linear space or sub linear time)[1].

Some of the challenges that are being faced by the graph of the problem have many working similarities such as weight of the minimum tree and the overall dynamic algorithm in which the graph is not static and can evolve with the time. It is mainly that all these changes are not extreme and if the change could be in few nodes [1]. One more factor in this particular challenge is faced for the purpose of implementation of any perspective in which most of the work is on the basis of the theoretical computers. There are some of the alternative approaches which can come from the reach of the theoretical computers. [1] There are also some different approaches which can come from the reach of the analyzing bigger network graph and distributed computing.

Most of the big data chances for the purpose is for the purpose of statistical inference, but may be a lot bigger challenges are forward if we compare the overall analysis of this more carefully and collectively. Most of the times a small set of data is used. From January to June of 2015, Canadian Statistical Science Institute has organized many thematic programs on the Statistical Inference and Learning and Models in the big Data. It has become a lot apparent with the first of the two weeks of this program which consists of many common issues which can rise indifferent practical settings. This particular paper has risen from the attempt to dissolve all these common problems from which the presentation and any kind of discussion has taken place during this particular thematic program. Technically, this particular program will put a lot of emphasis on the rolled of the statistics and computer science along with mathematics in the process of obtaining the insights into the big data. The two complementary stands are introduces in cross cutting of foundation, most of the research which underpins the analysis and a particular domain specific research which is focused on a particular application in this area. Most of the former categories may include a machine learning, network analysis, statistical inference, optimization, as well as visualization. These particular division are not very rigid, indeed for the purpose of application and foundation of the application of this area the feedback cycle is very important in which each of this inspires the development in others. Some of the most essential application areas are where the big data is very important and may not be able to focus on the workshops, most importantly all these application will not feature any individual presentation. Some of the programs which started with the starting conference which provides a complete overview of all the topics in the past 6 months of program 2. All these talks are mainly put forward in the field of institutes which is available online with the help of fields live 3. Technically more advances can let us collect more

ISSN No: - 2456- 2165

data. In the year 2012 itself, it was calculated that data collection has grown at a rate of 50% every year. As there is a lot of important research that is underway for the purpose of advancing the overall processing, rapid accessing of records and storing them (Chen et al, 2014; Schadt et al., 2010), this particular program will put a lot of emphasis on the overall challenged for the purpose of modelling, analysis and interference.

II. ANALYST-CENTERED LINEARLY SCALABLE ALGORITHM

Requirement to enlarge this kind of big data networks imagination is getting acute over the last few decades just as the overall size of the big scale network has increased. Currently most of the study signifies that with the help of an effective imagination technique the overall efficiency of the operator increases and can solve most of the big problems. [7]

Another very effective approach can put the focus more towards the developing algorithm which can linearly be scalable. This is why helping in the process of effective imagination for this kind of network with big data points with the use of network adjacency matrices, clustering coefficient and degree of distribution with standard related relationship. Also the developed algorithm is the element for the purpose of modelling out imagination for this king of network. There are many statistical measures which have already been proposed by the Chopade et al [8][9].

A. Degree distribution

The number of edges connected to a vertex is called its degree. The degree distribution $P(k)$ of many empirical networks has a power law, $P(k) \sim k^{-\alpha}$, where α is typically between 1 and 3 [10].

B. Clustering coefficient

The clustering coefficient, CC_i , of a vertex i is the ratio between the actual number of edges that exist between the vertex and its neighbors and the maximum number of possible edges between these neighbors. The CC of the network is defined as [11] [12]:

$$CC = \frac{1}{n}\sum_{i \epsilon V} CC_i = \frac{1}{n}\sum_{i \in V} \frac{M_i}{k_i(k_i - 1)/2} \quad \ldots\ldots (1)$$

where CC_i is the local clustering coefficient, M_i is the number of edges that exist between the neighbors of vertex i, and k_i is the number of neighbors for vertex i. The denominator k_i (k_i - 1) /2 is the maximum possible number of edges that can exist between the neighbors of vertex.

Creating an adjacency matrix from the dataset is an appropriate way to represent the dataset and make it much more uncomplicated and simpler to analyze the dataset and to visualize it. Splaying the dataset onto an adjacency matrix helps to display the graph and gather all the information that the dataset would have and thus can be represented [13].

The network adjacency matrix A of size $N * N$ be written as given by equation 2 [14],

$$A_{ij} = \begin{cases} 1 \;\; if\ there\ is\ an\ edge\ from\ vertix\ i\ vertix\ j \\ 0 \;\; if\ there\ is\ no\ edge\ from\ vertix\ i\ vertix\ j \end{cases} \quad \ldots (2)$$

III. PROBLEM STATEMENT

Getting the interactive visualization of the big data or the big scale connection is not a very easy task and can get very difficult to read from this kind of database as you are trying to imagine this kind of network which will be containing billions of data point which cannot be very easy like the normal statistics graphics. This can normally get very chaotic and can get ever difficult to read [2]. This is why the more effective way is to sub sab sample all the data for the purpose of providing us a less chaotic and a lot better imagination for the purpose of handling and revealing the complete heap of data which was not capable for reading with the help of chaotic view. This was made available with all the common standard statistical and graphical imagination. Just like the need of a good network imagination of this kind of big data all the research can prove that complete size of this data normally getting bigger and all the need for this kind of enlarging network imagination is a lot important and can be the biggest effect on the overall efficient as it can be able to provide a better imagination of this king of big picture networks.

Most of the big networks are not just restricted to any single area but can in reality the big network can be very easily found in many different area in which they can produce this directly without any need of machine or computer. Different data sources like this have data in the readable format, most of these big networks for example cannot be effective for the purpose of transportation network of which the example contains data of Airlines and the data from various different airports. It can be belonging to different things like weather station in which all the machines and different panels are used for the purpose of monitoring all the changes that take place in the weather condition and can be easily recorded. Most often this kind of record and set of data which is required to be recorded is very big because the data has been collected on daily basis. This is why all these big networks cannot analyses or even are able to imaging for the purpose of providing the required information with the use of these normal network monitoring tools.

IV. PROPOSED METHODOLOGY

A. Overlay Network

There is an overlay network of the computer which has been constructed over one more network. There are many different nodes that are present in the overlay network which can teach of getting connected with the logical or even the virtual link. Each of these can have their different ways, which are through

ISSN No: - 2456- 2165

different links that have underlying network. Example, it can be distributed system for example the peer to peer network and the client server network application that are overlay network as all the nodes that run over the internet. The reason behind launching of internet was to first buildan overlay on the telephone network but today with the help of VoIP, telephone network is getting built as an overlay over internet.

B. Use Of Overlay Network

Use of overlay network in the telecommunication has been because it is available in the digital circuit of switching equipment's and the optical fiber. The overall telecommunication that moves the network and the IP are together overlaid optical fiber layer, IP circuit and the transport layer.

There are many organization specific networks which were before overlaid on the telecommunication network like in the frame relay and the Asynchronous Transfer Mode which is switching the infrastructure for the purpose of migration from all these the infrastructure to the IP which is based on the MPLS networks and the overall private networks that have started in the year 2001 -02.

There is also a physical standpoint to this overlay network as they are very complex and as they are made up of different logical layers which are operated and constructed with the help of different things like universities, businesses, government, etc. All of these let the separation concern over the time that is permitted for the purpose of buildup of the broader set of all these services. This cannot have been proposed by only one telecommunication operator which is ranging from a broadband internet access or for voice over internet protocol or for IPTV, communicative telecom operator and so on.

C. Over The Internet

As of today net has become an important factor for more overlaying the networks which can be built so that permit routing of this message to the destination which is not said by the IP address. This is the address which is distributed for the tables and can also be used for the route message to the node for getting a specific logical address of who's the IP address is not there in the advance.

There are networks which have been proposed for the way of improving the Internet routing. This is with the help of quality of the services which can guarantee and provide you a better quality streaming media.

Overlay networks have also been proposed as a way to improve Internet routing, such as through quality of service guarantees to achieve higher-quality streaming media. Previous proposals such as Internet Server, Differs, and IP multicast have not seen wide acceptance largely because they require modification of all routers in the network. On the other hand, an overlay network can be incrementally deployed on end-hosts running the overlay protocol software, without cooperation from ISPs. The overlay has no control over how packets are routed in the underlying network between two overlay nodes, but it can control, for example, the sequence of overlay nodes a message traverses before reaching its destination.

Akamai technology is the prime example for this which is an overlay network that can provide you a very trustworthy and effective content delivery. The academic research will also include the End System Multicast and the Overcast for this multicast, Over Qi's for the purpose of quality of the service that is guarantees and the Resilient Overlay Network which is for the purpose of resilient routing.

D. Proposed Algorithm Steps

Step 1:- Let the input data be a matrix for the Pixel Based Overlay Network which is algorithm.

Step 2:- This is the methodology which we are going to consider for data of this image matrix.

Step 3:-Use the mining, zooming and panning method with the help of a paranoma for the purpose of creation of the two images.

Step 4:-Panaroma is very specific method which can be used for the purpose of combining the two images and generating a bigger size image.

Step 5:-After that we should be applying the median, intersection and the regression methodology for the degree of distribution with which we are going to get the low gone time, lesser errors, and high performing processor.

V. IMPLEMENTATION

In the implementation , we show the Network construction and GUI for the proposed network. Network graphs, as representations of network systems of elements and their interactions. But often it is some quantity associated with the elements that is of most interest, rather than the network. Nevertheless, such quantities may be influenced by the interactions among elements, here are few examples,

- Behaviors and beliefs influenced by social interactions.
- Functional role of proteins influenced by their sequence similarity.
- Computer infections by viruses may be affected by 'proximity' to infected computers.

Figure 1 is showing the design the GUI for the Erdos Collaboration Network, Connectivity of internet routers and Pixel Based Overlay Network designs . When we click at Statistical Modeling , then we will get the results of Erdos collaboration network , Connectivity of internet routers network . It will show the Data Visualization for the Erdos collaboration network and Connectivity of internet routers .

ISSN No: - 2456- 2165

Fig 1 :- GUI for Data Visualization for Erdos collaboration network , Connectivity of internet routers and Pixel Based Overlay network

As we click at Paranoma Image The input image folder will get open . It will how the input data for the proposed network. We give the second input image to the Pixel Based Overlay Network. It is creating the Data Sets for the Pixel based Overlay network. As we click at PAN / ZOOM the Image get Zoom. For compare the all networks , we have to click at comparison analysis in between three networks .

VI. RESULTS

Figures 2 and 3 are presenting you the overall degree of distribution for the purpose of Erdos collaboration network and the overall connectivity of all the internet routers. All the weights that are characterizing different connections in the exhibit are very complex and have statistical features which can be very highly different from the distribution and the power law behavior. This analysis can also be used for weighing the quantities and the overlay study of correlation in between the topology and the weights which can provide a complementary perspective on the complete structural institution of a particular network which can be undetected by al the quantities that are based on this kind of topological information.

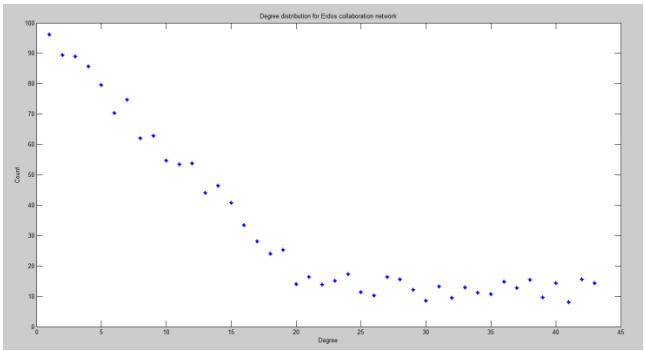

Fig 2:- Degree distribution for Erdos collaboration network

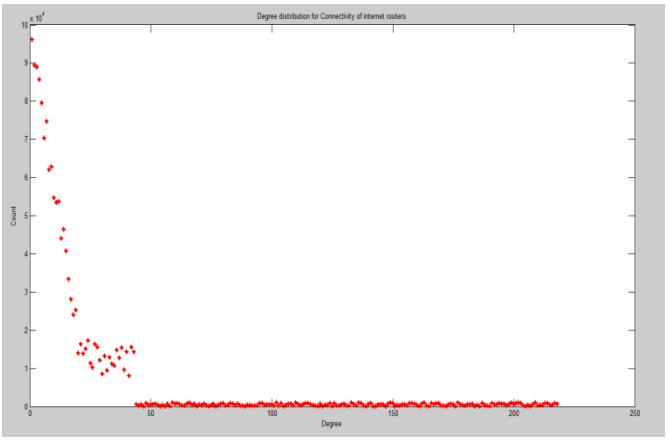

Fig 3:- Degree distribution for Connectivity of internet routers

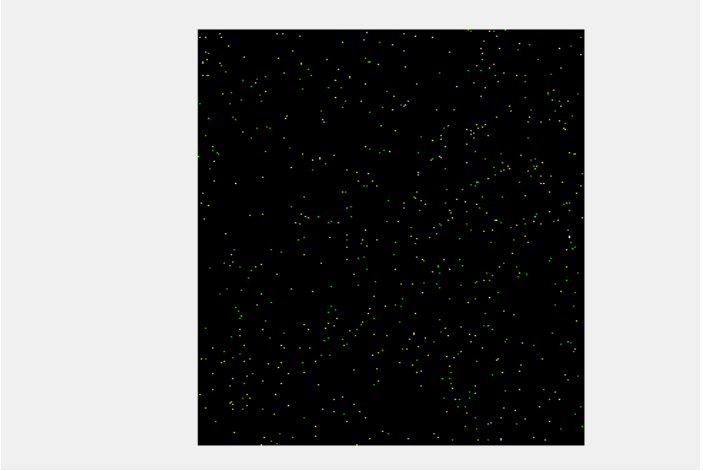

Fig 4:- Fetching Paranoma network Data Image

As we mention in the proposed methodology , We are designing a paranoma network which is showing the Big data Visualization in the form of image . So in the GUI as we click at the Fetching then the data image will fetch for complete the network visualization process.

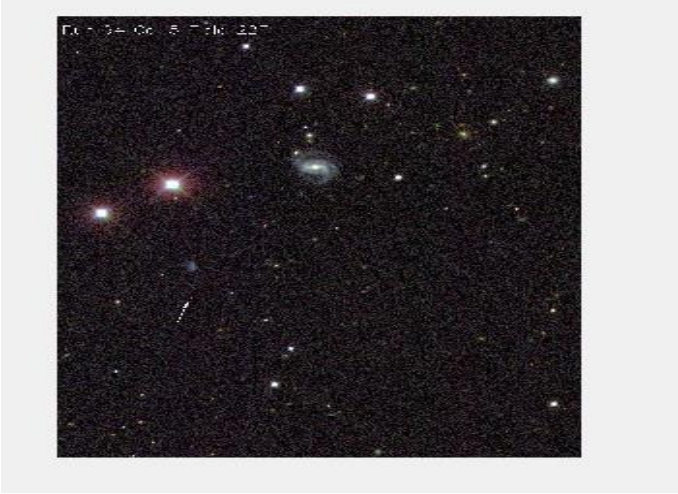

Fig 5:- Zoomed Image Data for Paranoma network

ISSN No: - 2456- 2165

Figure 5 is showing the Zoomed Paranoma Network image . As the image fetch for data visualization calculation . Figure 6 is showing the comparison of Data Visualization for Erdos collaboration network , Degree distribution for Connectivity of internet routers and Paranoma Network .

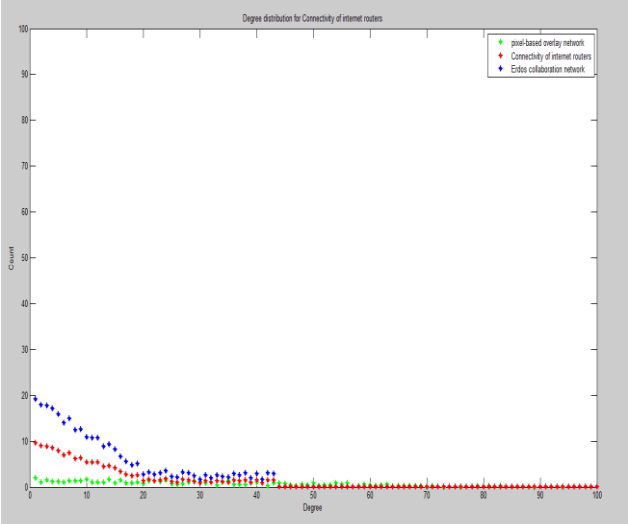

Fig 6:- Comparison of Erdos collaboration network , Degree distribution for Connectivity of internet routers and Paranoma Network Data Visualization

VII. CONCLUSION

In this study, they are comparing 2 types of networks which are divided in terms of the data of visualization. The research that is conducted in this particular field of imagination and the overall network representation is a very big topic and normally this topic consists of always changing over the last few months in which there have been many approached and the representation which has to be presented in this always improving as the technology is getting enhanced in the processor once it's getting fast which is helping in the visualization get a lot faster them ever before.

Overall study of the network is very essential as it can help in knowing what all is actually happening inside the network and by imagining it can help in understand what all are the valuables data which was not made clear to us before because of the big size of the data set which was being used.

Analysis of the Pixel Based Overlay Network is an algorithm which can be very effective for the purpose of analyzing a real big scale data network. This is the most important discovery in which the distribution of all the vertex degrees of all the big scale data network are very heterogeneous. There are many different vertices which have only some neighbors that are coexisting with the help of some of the vertices along with the neighbors. In many cases the tail of such kind of distribution can also be described as the power law which has a very effective approximation and thus all the expression of the scale free network.

In the future for further improvement in the research, we can improve the performance of the Erdos collaboration network , Degree distribution for Connectivity of internet routers and Paranoma Networks by reduce packet loss Ratio , Energy consumption , Data Transmission Speed . These parameters will be able to improve the performance of the Erdos collaboration network , Degree distribution for Connectivity of internet routers and Paranoma Networks.

References

[1] Graph-theoretic challenges in network science- Network of Excellence in Internet Science-January-2014.
[2] http://strata.oreilly.com/2013/12/interactive-visualization-of-bigdata.html.
[3] E. Kolaczyk, "Statistical Analysis of Network Data", SAMSI Program on Complex Networks, Opening Workshop, Dept of Mathematics and Statistics, Boston University.
[4] T. Eliassi-Rad,T. Critchlow,G. Abdulla, "Statistical Modeling of Large- Scale Simulation Data", SIGKDD '02, July 23-26, 2002, Edmonton, Alberta, Canada.
[5] A. Freno, M. Keller and M. Tommasi, "Fiedler Random Fields: A Large-Scale Spectral Approach to Statistical NetworkModeling", Advances in Neural Information Processing Systems 25 (NIPS 2012 Proceedings).
[6] D. Nga, O. See, D. Quang, C. Xuen, L. Chee, "Visualization Techniques in Smart Grid", *Scientific Research, Journal*,Smart Grid and Renewable Energy, 2012, 3, pp.175-185.
[7] Chopade, P.; Flurchick, K. M.; Bikdash, M.; Kateeb, I., "Modeling and visualization of Smart Power Grid: Real time contingency and security aspects", Southeastcon, 2012 Proceedings of IEEE , vol., no., pp.1,6, 15-18 March 2012.
[8] Chopade, P.; Bikdash, M., "Critical infrastructure interdependency modeling: Using graph models to assess the vulnerability of smart power grid and SCADA networks,' Emerging Technologies for a Smarter World (CEWIT), 2011 8th International Conference & Expo on , vol., no., pp.1,6, 2-3 Nov. 2011.
[9] Chopade, P.; Bikdash, M., "Modeling for survivability of Smart Power Grid when subject to severe emergencies and vulnerability", Southeastcon, 2012 Proceedings of IEEE , vol., no., pp.1,6, 15-18 March 2012.
[10] K. Sun, "Complex Networks Theory: A New Method of Research in Power Grid", 2005 IEEE/PES Transmission and Distribution Conference & Exhibition: Asia and Pacific Dalian, China, pp. 1-6.
[11] J. Watts and H. Strogatz, "Collective dynamics of small-world networks", *Nature*, Vol. 393, No. 6, June 1998, pp. 440-442.
[12] N. Eggemann, S. Noble, "The clustering coefficient of a scale-free random graph", ScienceDirect, *Discrete Applied Mathematics*, 159 (2011).
[13] Riaz, F.; Ali, K.M., "Applications of Graph Theory in Computer Science", Computational Intelligence, Communication Systems and Networks (CICSyN), 2011 Third International Conference on , vol., no., pp.142,145, 26-28 July 2011.

ISSN No: - 2456- 2165

[14] Chopade, P.; Bikdash, M., "Structural and functional vulnerability analysis for survivability of Smart Grid and SCADA network under severe emergencies and WMD attacks", Technologies for Homeland Security (HST), 2013 IEEE International Conference on , vol., no., pp.99,105, 12-14 Nov. 2013
[15] http://www.mathworks.com/help/bioinfo/ref/biograph.html
[16] http://mathinsight.org/degree_distribution.