International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2018 IJSRCSEIT | Volume 3 | Issue 1 | ISSN : 2456-3307

Analysis of Organized Data (web cloud data) Through Data Provenance Technique

Sabah Naseem Akhter*[1], Dr. Ashish B Sasankar[2]
*[1]Research Scholar, Department of Electronics & Computer Science, R. T. M Nagpur University, Nagpur Nagpur, Maharashtra, India
[2] Head of The Department, G. H. Raisoni College of Information Technology, Nagpur Hingna, Nagpur, Maharashtra, India

ABSTRACT

Data provenance signifies "the source" or "origin". It offers authenticity to the user. User incorporates data by demonstrating it as per ontology of their decision utilizing a graphical UI that computerizes a great part of the procedure. User at that point connects with the framework to modify the automatically created model. Amid this procedure, user can change the information as expected to standardize information communicated in various configurations and to rebuild it. Once the model is finished, user can distributed the incorporated data or store it in a database.

Keywords : Data Provenance , Cytoscape Tool , Cloud Data

I. INTRODUCTION

Cloud computing is dynamically accessible shared resources retrieved over a network. It is only pay for what you use, shared internally or with other customers. It is open for all user that's why it required authenticity. Provenance means lineage or ancestry which provides origin of data. The provenance of a data product contains information about how the product was derived, and is crucial for enabling scientists to easily understand, reproduce, and verify scientific results. To find the provenance graph we have to use karma tool. The cytoscape tool is an independent tool that can be added to existing digital framework for reasons for gathering and representation of provenance information. It uses a particular design that authorizations support for various instrumentation modules that make it usable in various structural settings [1].Used extensively in different domain, data provenance has already been used in many application such as physics ,biology , e-science etc. For data provenance , we need to understand the issues of data creation , alteration & copying. Provenance is metadata i.e. data about data, means how and when and by whom a particular set of data was collected, and how the data is formatted[2]. We make the case that provenance is crucial for data stored on the cloud and identify the properties of provenance that enable its utility. Using cytoscape tool we can find the origin of any type of data such as structured , unstructured , text ,image etc. Here we used structured data to find the origin & analyze the network.

II. CYTOSCAPE

Cytoscape is an open source software platform for complex network analysis and visualization[3]. Using cytoscape we can find the origin of any type of data such as structured , unstructured , text ,image etc .

1) Choose one file from cloud i.e structured or organized data. Create .sif file . .sif file is executable in cytoscape .Now , import file locally.

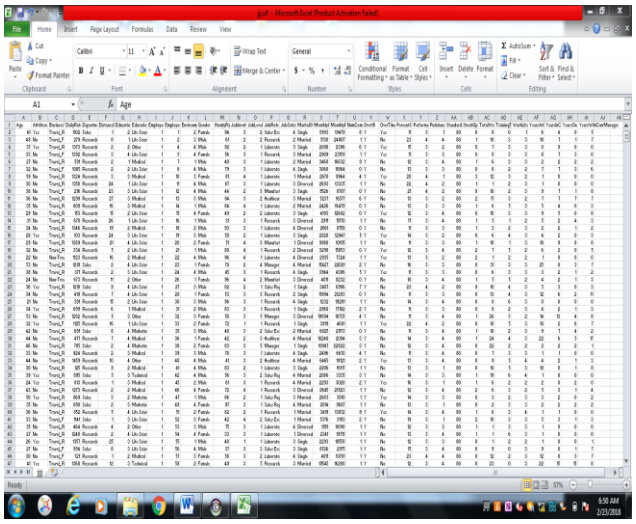

Fig-1_ online .sif file

2) Import file locally & display a network graph of that data , in the form of nodes & edges.

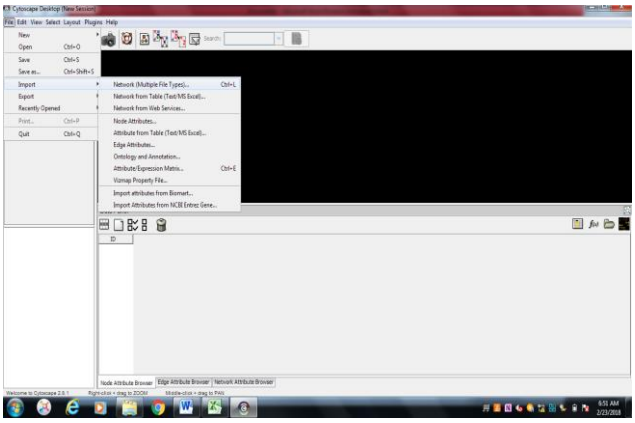

Fig-2_Import .sif file [source from cytoscape tool]

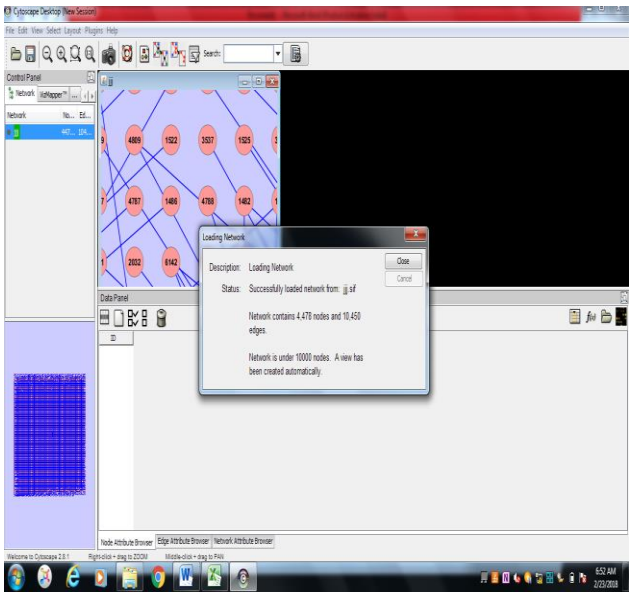

Fig-3_ Create network[source from cytoscape tool]

III. NETWORK ANALYSIS

The accompanying plugins settings can be designed .

A) Analyze Network:

a) Directed Network

The network contains solely directed edges. Here, NetworkAnalyzer provides three possible interpretations of the edge directions in the network. The user has to select one of the interpretations for further processing of the network.

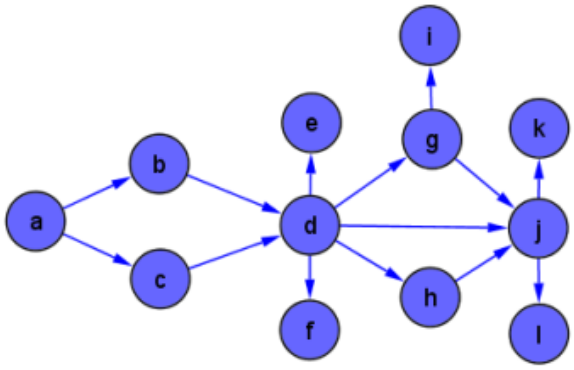

Fig-4_Directed graph [Source cytoscape tutorial]

b) Undirected Network

The network contains both undirected and directed edges. Note that undirected edges cannot be converted unambiguously to directed ones. Therefore, networks with mixed edges are handled as undirected ones[4].

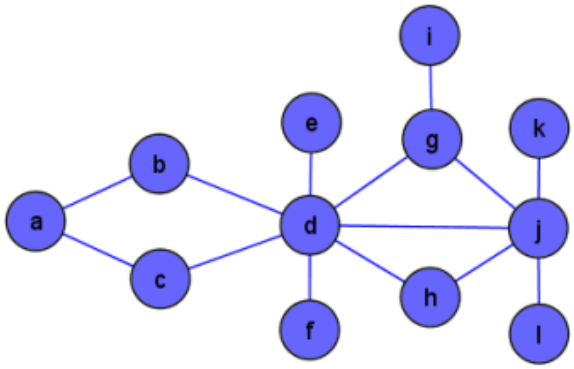

Fig-5_Undirected graph [source cytoscape tutorial]

c) Edge attributes

For every node in a network, NetworkAnalyzer computes its in degree and out degree for directed graph , its clustering coefficient, the number of self-loops, and a variety of other parameters. It also computes edge betweenness for each edge in the network. , NetworkAnalyzer stores the computed values as attributes of the corresponding nodes and edges. This enables the users to apply different visualizations or to filter nodes or edges based on the values of the computed attributes.

d) Use expandable interface for the dialog that displays analysis results

If this option is enabled, analysis results are presented in a window in which all charts are placed below each other in expandable boxes. If this option is disabled, analysis results are presented in a window that contains tabs for the group of simple parameters and for every complex parameter. Users who wish to view simultaneously two or more complex parameters of one network, should enable this option.

e) parameter visualization

Using parameter visualization we can change the default setting of parameter. Such as background color , brightness , darkness etc.

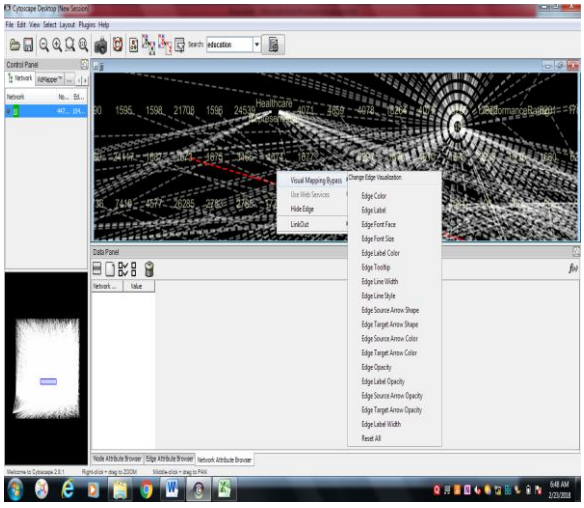

Fig-6_ Visualization of graph [source from cytoscape tool]

f) Plot Parameters

In undirected networks, the clustering coefficient C_n of a node n is defined as $C_n = 2e_n/(k_n(k_n-1))$, where k_n is the number of neighbors of n and e_n is the number of connected pairs between all neighbors of n [5,6]. In directed networks, the definition is slightly different: $C_n = e_n/(k_n(k_n-1))$.In both cases, the clustering coefficient is a ratio N/M, where N is the number of edges between the neighbors of n, and M is the maximum number of edges that could possibly exist between the neighbors of n. The clustering coefficient of a node is always a number between 0 and 1.

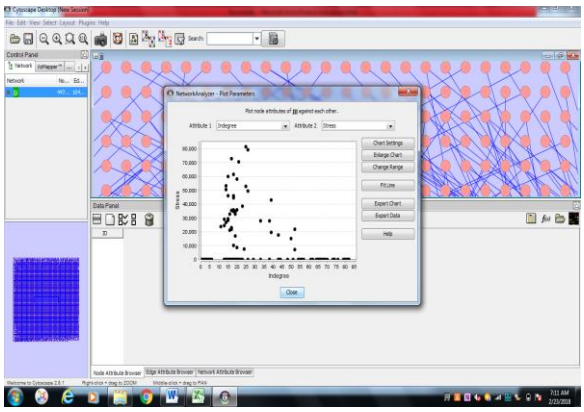

Fig-7_Plot parameters[source from cytoscape tool]

IV. CONCLUSION

This paper evaluates the performance of the structured data provenance framework in collecting and querying for provenance from workflow executions, and finds it to scale well with the size of the workflows and the number of concurrent clients. and is relevant to similar scientific projects. The workloads in themselves form a benchmark to compare and evaluate other Provenance systems and such a comparison is done with the network service. Cytoscape is currently deployed and being used in the biology test. Our future Work includes evaluating the performance of cloud data for real workflow runs and getting usable results for them by suppressing the I/O variations. In the data intensive applications – possibly by the use of local storage instead Of network file systems. In addition to visually browsing provenance graphs, we are also investigating other ways in which provenance can be put to use. Notable among these is on using data provenance as a factor in searching and ranking of

data products by applying quality metrics .

V. REFERENCES

[1]. Michael Goodman - PI, Marshall Space Flight Center Indiana University, 'Karma Provenance Collection Tool'

[2]. Research statement RagibHasan, Email: rhasan7@jhu.edu

[3]. Tutorial on using Karma Visualization Plug in for Cytoscape to visualize Instant Karma Dataset.

[4]. http://med.bioinf.mpiinf.mpg.de/netanalyzer/help/2.7/index.html#simple

[5]. Watts D.J., Strogatz, S.H.: Collective dynamics of 'small-world' networks. Nature 393 (1998) 440-442

[6]. Barabási, A.L., Oltvai, Z.N.: Network biology: understanding the cell's functional organization. Nat Rev Genet 5 (2004) 101-113

[7]. http://www.cytoscape.org/download.php