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.2, Special Issue 5, October 2014

A Parallelized Social Network Analysis Using Virtualization for Student's Academic Improvement

K.Geetha[1], Dr.A. Vijaya Kathiravan[2]

Ph.D. Research Scholar, Department of Computer Science, Periyar University, Salem, TamilNadu, India[1]

Assistant Professor, Department of Computer Science, Govt. Arts College, Salem,TamilNadu, India[2]

ABSTRACT: *"Big Data is the Next Frontier for Innovation, Competition, and Productivity"[1].*

Big data is massive and messy and it's generated at a very fast rate. These characteristics pose a problem for data storage and processing, but focusing on these factors has resulted in a lot navel-gazing. The data which are generated by social network includes structured (10%) and unstructured (90%), which are huge in volume and are becoming big challenge to process and analyze. Big data technology offers significant contributions over technology development. In addition to this, now a day it is also having higher impact on the academic performance of the students. The social networking sites like LinkedIn, Tweeter, and Face book are having such impact among the younger generation. This paper is concerned with student's involvement in social media and their interaction with others (student to student and student to tutor) through this network. Through this analysis of their interaction and communication over social network it will be helpful to find out their interest and requirements. This analysis also list out the usage of some web services and their impact over the linguistic and academic behavior of the young learners. In social network analysis the structured data are capable of storing and analyze using traditional analysis techniques. Here the unstructured data are from social media are considered as useless. But it is also needed to analyze and extract meaning full information from them. The unstructured huge volume of data could be processed under special environment with additional techniques. It is not feasible to process under traditional environment. Since the social network could be represented by socio graph, it is will be feasible by partitioning the big socio graph into computable size. Then it is taken as input into virtualized memory to analyze. Here the Mapper() and Reducer() programs are executed under virtual environment in distributed manner. The virtualization of process has been used to reduce the space complexity and time complexity. There are tools created to address the problem of big data. Hadoop is one of the best examples for dealing big data in a distributed environment. It involves breaking the huge socio graph into cliques of similar interest and analyzed to identify the students need for their development.

KEYWORDS: Big data, Social network, Virtualization, Pedagogical, Hadoop, Map Reduce, Cliques, Similarity, Academic performance.

I. INTRODUCTION

The research for data-driven decision-making is now being recognized broadly, and there is growing interest and involvement for the notion of ``Big Data [6]. It's Heterogeneity, scale, timeliness, complexity, and privacy problems with Big Data degrade the progress at all phases of the analysis that can create value from data. Where there are problems in acquiring the huge volume of data and also hard to collect them and to store. It is also hard to identify and classify which data items are to discard and to retain. Much data today is not natively in structured format. For example, tweets and blogs are weakly structured pieces of text, while images and video are structured for storage and display. Transforming such content into a required structured format for later analysis is a major challenge. The value of data explodes when it can be linked with other data, thus data integration is a major creator of value. Since most data is directly generated in digital format today, we have the opportunity and the challenge both to influence the creation to facilitate later linkage and to automatically link previously created data. Data analysis, organization, retrieval, and modeling are other foundational challenges.

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.2, Special Issue 5, October 2014

The survey regarding the student's involvement in social media represent both positive and negative impact on their academic as well in personal development. Since the 47% of American adults used social networking sites like Facebook, MySpace, Twitter, LinkedIn, and Classmates.com in 2011, up from 26% in 2008[1].

It shows higher recommendation on analyzing this to produce higher tutoring system, to develop their personal skill and to follow up their nature of communication. The conversation among the young groups through social media (age group of 12 to 24) has been taken for analysis. This will be helpful to classify them based on their area of interest. Here the challenges underlying this are the social media data are big data. It includes both structure, unstructured and Semi structured. The traditional techniques like RDBMS for storing and SQL for querying cannot be taken into account for analysis. The new technology is needed to proceed with big data. Even though there are challenges in handling and storing of Big Data, it is becoming important due to its value. Much data today is not natively in structured format. For example, tweets and blogs are weakly structured pieces of text, while images and video are structured for storage and display. Transforming such content into a required structured format for later analysis is a major challenge. These challenges could be dealt with a frame work called Hadoop.

The data generated by social network could be taken for the analysis of network entities. The social network could be represented by the graph using collection of nodes and vertices. Each node/ vertices of the graph represents the actor of the social network and the connection between the vertices is referred as the edges demonstrate the relationship among actors. Social network is termed as the interaction or relationship between any numbers of social groups with similar interest. They can be classified into two categories as link based and contend based[8].The many-to-many nature of social networking produces data that directly reflect what users actually do and what they're talking about.

Hadoop is one of the best examples for big data tool. Developed by Google and is adopted by various leading organizations to deal their big data problem. Hadoop is an open-source software framework for storing and processing big data in a distributed fashion on large clusters of commodity hardware. Essentially, it accomplishes two tasks: massive data storage and faster processing. It work different by break down the big data into smaller chunks and break down the computation into smaller pieces as well the piece of computation send to the piece of data and process under distributed environment by parallel.

Hadoop has three main components.

- **HDFS** – the Java-based distributed file system that can store all kinds of data in distributed manner.
- **MapReduce** – a software programming model for processing large sets of data in parallel.
- **YARN** – a resource management framework for scheduling and handling resource requests from distributed applications.
-

Hadoop act as the technological platform to execute Map Reduce programs in the distributed environment. This could speed up the data processing operations-namely managing and analysing the bigdata to get real value. Even though the Hadoop has some limitations in solving the big data challenges, it is able to enabling distributed MapReduce processing on arbitrary sized clusters of low-cost hardware. To enable this, the Hadoop contributors built a distributed file system – the Hadoop Distributed File System (HDFS) and a set of components to execute distributed Map Reduce java programs. In current era the use of Internet has developed with its development in all fields. The interaction nature of internet is developed by the social Medias. The social networking sites are used to build communities, Chatting forums and Blogs to transfer their experiences. It gives higher impact over student communities. It could be helpful for the academicians to share their knowledge and experience over internet. They could clarify their requirements over forums and blogs. By this they could interact with others who are all experienced in certain areas to gather knowledge and to utilize the social media in effective manner. This will also helpful for the researchers to get valuable information from social network. From this information we could classify and also cluster the participants based on their interest. Many countries initiate government funded learning systems like e-learning, e-governance, and e-health [1].

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.2, Special Issue 5, October 2014

In terms of analyzing the learner's interest we could provide online materials as per their requirement. Here the research undergone for academician's development. And also we could identify and improve those who are less interest in certain areas. For example if someone feel hard with studying statistic and to know more in that area, we could assist them by allocating tutor and by providing free resource materials. In this analysis the programming paradigm Map Reduce has been used to analyze the social media big data.

The data generated by social network could be taken for the analysis of network entities. The social network could be represented by the graph using collection of nodes and vertices. Each node (vertices) of the graph represents the actor of the social network and the connection between the vertices is referred as the edges demonstrate the relationship among actors. Social network is termed as the interaction or relationship between any numbers of social groups with similar interest. They can be classified into two categories as link based and content based [8]. In context with pedagogical the above mentioned two categories are taken for analysis. While the social network data are huge in volume, dynamic and heterogeneous in nature, it is considered as huge graph of interactions. Here the communication between two actors (nodes) is represented by links and weight gained between them is referred by the content shared among them. A computation on huge graph is become challenging task due to its increasing size. There is a need for a technique to partition it and distribute among several nodes.

II. LITERATURE REVIEW

Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process the data within a tolerable elapsed time [7]. In big data like social network graph, graph partitioning is a hectic issue during virtualization or parallel processing. In mathematics, the graph partition problem is defined on data represented in the form of a graph G = (V,E),with V vertices and E edges, such that it is possible to partition into smaller components with specific properties. A good partition is defined as one in which the number of edges running between separated components is small. For instance, a k-way partition divides the vertex set into k smaller components [14]. For a (k,v) balanced partition problem, the objective is to partition G into k components of at most size v (i.e., n/k partitions), while minimizing the capacity of the edges between separate components[15]. A multi-level graph partitioning algorithm works by applying one or more stages. Each stage reduces the size of the graph by collapsing vertices and edges, partitions the smaller graph, then maps back and refines this partition of the original graph. Uniform graph partition is a type of graph partitioning problem that consists of dividing a graph into components, such that the components are of about the same size and there are few connections between the components [15]. Recently, the uniform graph partition problem has gained importance due to its application for clustering and detection of cliques in social, pathological and biological networks. The Kernighan–Lin algorithm and Fiduccia-Mattheyses algorithm were the first effective 2-way cuts by local search strategies. Their major drawback is the arbitrary initial partitioning of the vertex set, which can affect the final solution quality. The most common example is spectral partitioning, where a partition is derived from the spectrum of the adjacency matrix. In a 2001 research report[8] and related lectures, META Group (now Gartner) analyst Doug Laney defined data growth challenges and opportunities as being three-dimensional, i.e. increasing volume (amount of data), velocity (speed of data in and out), and variety (range of data types and sources). Big data is high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization (Gartner) [6]. Many computation tasks are performed on large scale graph data which may amount to as many as billions or even trillions of vertices or edges (e.g. online social network graphs, knowledge graphs, and web graphs).

A standard way to scale up such computations is to partition an input graph into clusters of balanced sizes and small cut costs [6]. Sparse and unstructured graphs arise in many application domains including data mining, pattern recognition, social network analysis, scientific computing, parallel processing, etc,. As the capacity to collect information and to model complex systems has soared, so has the size of the graphs that are being generated and need to be processed and analyzed [6]. Partitioning the graph into disjoint parts results in many queries accessing multiple partitions, and replication approaches require a large amount of extra storage [4]. In 2004, Google published a paper on a process called Map Reduce that used such architecture. Map Reduce framework provides a parallel processing model and associated implementation to process huge amount of data. With Map Reduce, queries are split and distributed

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.2, Special Issue 5, October 2014

across parallel nodes and processed in parallel (the Map step). The results are then gathered and delivered (the Reduce step). The framework was incredibly successful. So the algorithm is replicated. Therefore, an implementation of Map Reduce framework was adopted by an Apache open source project named Hadoop [13].

Teenagers now use the Internet for the majority their daily activities and information gathering, as opposed to older generations who used resources like the television or newspaper [3].A recent survey showed that approximately ninety percent of teens in the United States have Internet access, and about seventy-five percent of these teens use the Internet more than once per day . This study also showed that approximately half of all teens who have Internet access are also members of social networking sites, and use the Internet to make plans and socialize with friends. In September, 2005, out of total adult internet users (18-29 years) 16% were those who were using any social networking site and this percentage increased to 86% on May, 2010[2].

III. THE PROPOSED WORK FOR STUDENT NETWORK ANALYSIS

The research work is mainly focused on Partitioning largest social network graph into smallest sub graphs to process in efficient manner. The proposed work is to identify the group with similar interest. Since the social network is a big data, it is hard to process in a single system. That is why we proposed Virtualization technique for graph partitioning based upon similarity in conversations. In this approach, we virtualizes the graph partitioning to multiple threads and allow it to execute in parallel using new parallel algorithm that executes within the Map Reduce framework. The Hadoop Distributed File System(HDFS) used to distribute the partitioned data in distributed environment to process in parallel using Map Reduce. In this contest the scalability is another challenge to handle. The NoSQL is the database designed to handle such issues. Usually social network data are stored in distributed and virtual (cloud) file system like GFS or Big table in the form of an adjacency matrix or adjacency list. The recursive back tracking algorithm have been used to find the cliques in a big social network graph after partitioning.

a. The Proposed Architecture

In the proposed architecture, the Exabyte of data is taken as a whole Big Graph then it is transformed into minimal size (Terabyte) graph. Where the unclear noisy data are transformed into required format for further analysis. This is really a complex graph containing several cliques. The cliques are identified based upon the centrality and thus the big graph is partitioned into several cliques called bi-partite graph. The following Fig.1 demonstrates the proposed architecture. Each clique analysis has been processed in parallel using virtualization based upon centrality.

The processing of big social network graph is initiated by giving adjacency matrix as input for graph partitioning. The complex graph is partitioned into smaller sub-graphs based upon the promising centrality nodes. From the promising centrality nodes, the cliques are identified and splitted that forms the partitions of our big graph, which is also represented as adjacency matrices. This clique formation is paralleled using virtualization technique.

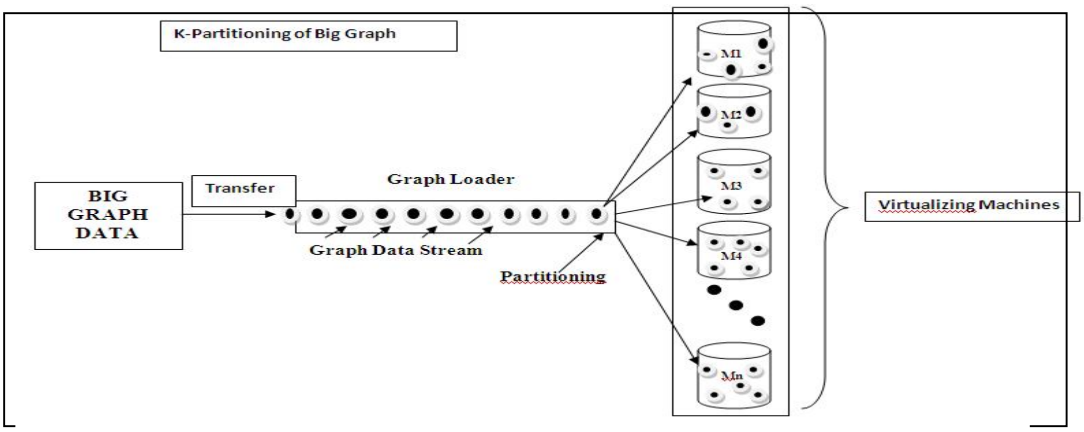

Fig1 Partitioning of Big graph data based on similarity measures

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.2, Special Issue 5, October 2014

b. Graph Partitioning

Formally, a graph can be written as G(V,E), where V is a set of vertices and E is a set of edges connecting vertices in V. The number of vertices is denoted n and the number of edges m. A second graph S (V',E') is a sub graph of G(V,E) if V' V and all edges in E' exists in E [15]. From this Graph representation we could analyze the strength of any actor (vertex) by measuring the strength of the relationship that each actor has with consequent actors. The graph partition problem is defined on data represented in the form of a graph G = (V,E), with V vertices and E edges, such that it is possible to partition G into smaller components called clique with specific properties. Graphs like social networks are notoriously difficult to partition (expander-like). Large data sets drastically reduce the amount of computation that is feasible – O(n) or less. The partitioning algorithms need to be parallel and distributed. While partition the larger graphs (G (V,E)) of social networks into sub graphs(S(V',E') the problem is to be balanced partitioning. Fig.2 shows the graph partitioning that creates sub graphs, which is of roughly in equal size. Here we could prefer graph partitioning algorithm that could run faster and could be parallelized[11]. We could partition the social network graph based on the following phenomena:

- **Clique Identification** – A complete bi-partite graph which is equivalent in centrality.
- **Homophily -** Persons being in similar nature.
- **Co-citation regularity -** Two persons have common interest.

In this approach, a k-way partition divides the vertex set into k smaller components called cliques using centrality based approach.

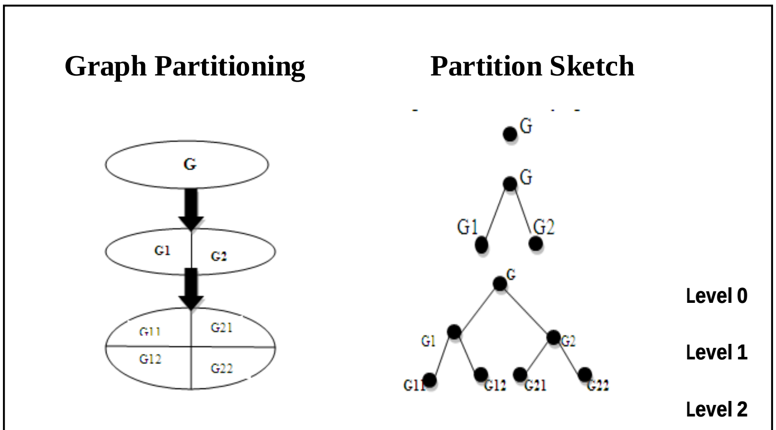

Fig2 Sample Graph Partitioning

Algorithm: Big Graph Partitioning

1. Start with weighted edges for graph
   - Give more penalties for cutting particular edges.
   - Redefine degree of nodes as sum of the adjacent edge weights instead of number of adjacent nodes.
2. Coarsen graph in steps, getting $G_1 \leq G_2 \leq G_3 \leq \ldots \leq G_m$.
3. At each coarsening step
   - Condense multiple nodes to form super nodes.
   - Increase edge weights to account for the condensation of the multiple edges into one.
4. After small enough graph is created apply spectral partitioning.
5. Uncoarsening to return to original graph.
   - Project coarse mesh G_{i+1} to next larger one G_i
   - Easily done because each (super) node in G_{i+1} consist of two nodes of Gi.
   - Since Gi is finer, more degrees of freedom become available to improve minimization.
   - Kernighan–Lin algorithm is used at each uncoarsening step to refine partitioning inherited from G_{i+1}.

6. k-way partitioning of graph G(V,E), where v∈V(vertex), s(v)-size of v, e∈E(edge), w(e)- weight.
7. Divide the V into k disjoint subsets $V_1,V_2\ldots V_{k..}$
8. Size of i^{th} subset is,

$$\sum_{v\in Vi} s(v) \le Bi$$

9. Minimum cut-cost between two subsets is, $D_a=E_a - I_a$, Where E_a- External cost and I_a- Internal cost.

c. Virtualization In Big Graph

Virtualization is an enabling technology which provides some or all of the following capabilities: Abstraction – Abstract the technical aspects of stored data, such as location, storage structure, API, access language, and storage technology. In graph partitioning, it abstracts the data of the graph to be partitioned. **Virtualized Data Access** – Connect to different data sources and make them accessible from a common logical data access point. The Fig.1 shows the virtualized data access. It creates the virtual environment for accessing and processing the sub graphs. **Transformation** – It transform the output to the consequent processing. **Data Federation** – Combine results sets from across multiple source systems to produce the final outcome of the work carried out. **Data Delivery** – It provide the facility to deliver the result sets as visual or data services when requested. Virtualization may include functions for development, operation, and/or management.

d. Clique Based Approach

Cliques are sub graphs in which every node is connected to every other node in the clique. As nodes cannot be more tightly connected than this, it is not surprising that there are many approaches to sub-graph detection in networks based on the detection of cliques in a graph and the analysis of how these overlap. The following Fig.3 display a graph with clique detection. One approach is to find the maximal clique, which is to find the cliques which are not the sub-graph of any other clique. The Fig.3 is a sample diagram that detects the graph of similar interest based upon the centrality based clique analysis. The classic algorithm to find these is the Bron-Kerbosch algorithm.

The overlap of these can be used to define cliques in several ways. The simplest is to consider only maximal cliques bigger than a minimum size (number of nodes). The union of these cliques defines a sub-graph, whose components (disconnected parts) define partitions.

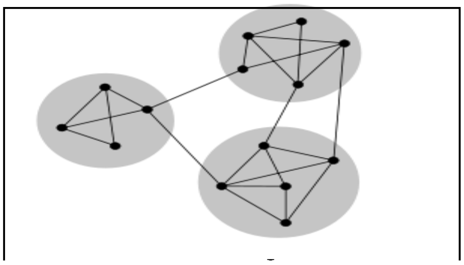

Fig3 Partition the Graph using clique

IV. METHODOLOGY USING VIRUALIZATION

The Social Network analysis technique used to analyze the social network communications and interactions between the actors, groups and communities. It helps to identify the interest among the individuals and groups. This techniques helps to guide with individual's interest among certain area. It helps to project the virtual information to the public. In this approach the data is collected from social network media like Twitter. It contains data which dropped from distinguished areas like e-learning, web blogs, and social networks etc. There we need a pre processing layer to filter irrelevant data.(i-e) The data from general interactions like personal tweets. We need the data of educational interaction between learners. When we get such data from Social networks, it will be huge and complex to store and

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.2, Special Issue 5, October 2014

difficult to process. So we go for the graph partitioning technique to reduce the complexity of the graph by minimizing its size.

This proposal is already handled by many researchers to solve such problems. But there are problem with handling memory and time complexity. Here the virtualization helps to handle these by executing Mapper () and Reducer () tasks in virtual machine where the partitioned graph data are stored. Each virtual machine is fault tolerance in nature. So that we could attain the data security by keeping the snapshot of each execution. The virtual memory allocation is only for time being till the execution complete its job. So the released memory could be availed by another virtual machine to run the job. The following Fig-4 shows the sequence of execution using virtualization techniques. It has four layers as follows,

- **Pre processing Layer**: To filter irrelevent data.
- **Graph partitioning Layer**: To reduce the Graph size.
- **Virtualization Layer**:To minimize the memory handling and improve the parallelism.
- **Visualization Layer : To display the groups with similar interest.**

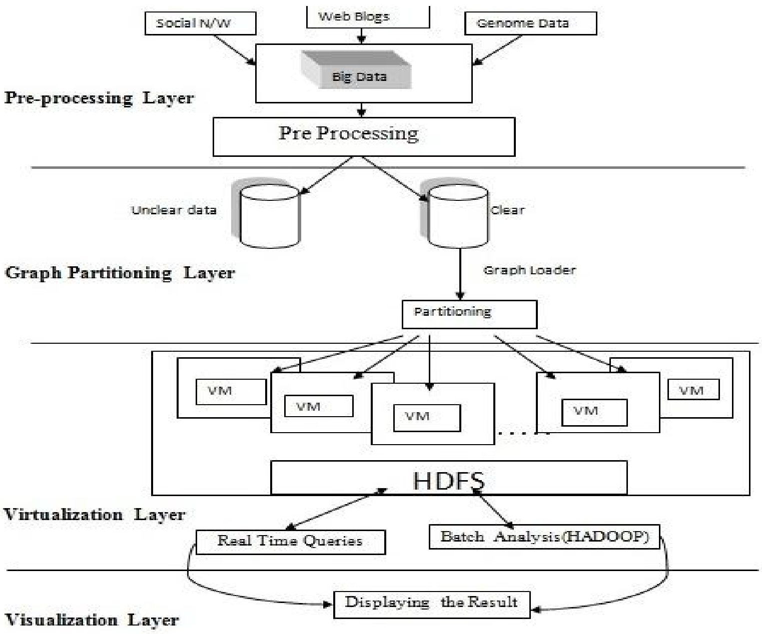

Fig4 Virtualization of Social graph analysis

Algorithm for graph processing: Graph Processing

- **Problem Statement:** There is a network of entities and relationships between them. It is required to calculate a state of each entity on the basis of properties of the other entities in its neighborhood. This state can represent a distance to other nodes, indication that there is a neighbor with the certain properties, characteristic of neighborhood density and so on.
-
- **Solution:** A network is stored as a set of nodes and each node contains a list of adjacent node IDs. Conceptually, MapReduce jobs are performed in iterative way and at each iteration each node sends messages to its neighbors. Each neighbor updates its state on the basis of the received messages. Iterations are terminated by some condition like fixed maximal number of iterations (say, network diameter) or negligible changes in states between two consecutive iterations. From the technical point of view, Mapper emits messages for each node using ID of the adjacent node as a key. As result, all messages are grouped by the incoming node and reducer is able to recompute state and rewrite node with the new state.

```
Algorithm: Graph Processing()class Mapper
 method Map(id n, object N)
       Emit(id n, object N)
       for all id m in N.OutgoingRelations do
         Emit(id m, message getMessage(N))

 class Reducer
    method Reduce(id m, [s1, s2,...])
       M = null
       messages = []
       for all s in [s1, s2,...] do
          if IsObject(s) then
             M = s
          else              // s is a message
             messages.add(s)
       M.State = calculateState(messages)
       Emit(id m, item M)
```

Fig5 Algorithm for graph processing

V. CONCLUSION

The social network analysis among student community will help to improve their educational standard and personal development. It will bring out the students interest apart from their academic participations. The partitioning of the social network graph is not an easy process due to its size as well as its complex structure. The graph partitioning of social will be network graph is already taken for research and it result with many difficulties in memory management as well as with time complexity. The usage of virtualization technique could manage the memory in efficient manner at the same time the time consumption will be reduced by the parallel processing of virtual systems. The goal of graph partitioning in big data is assigning equal amount of work on giving for each processor for load balancing, minimizing the total amount of communication, ignoring the overlapping of communication among nodes and representing the graph as matrix. This research will definitely satisfy all the above goals using virtualization. This work will be helpful in future for identifying the individual evaluation of an actor in social network. The resultant technique is also helpful for clustering the social network graph. Since the social network is dynamic in nature its graph representation will also be dynamic. So the virtualization techniques will be helpful in dealing with any dynamic big graph. This research could be good educational research to undersatand the student interaction and participation,and also used to improve the teaching techniques.

REFERENCES

[1] Mohammad Hassan, et. Al."The Impact of Social Performance Visualization on Students", 12th IEEE International Conference on Advanced Learning Technologies, 2012.
[2] Saba Mehmood, Tarang Taswir, "The Effects of Social Networking Sites on the Academic Performance of Students in College of Applied Sciences, Nizwa, Oma n."- International Journal of Arts and Commerce, January 2013
[3] Matthew J. Stollak, Getting Social: The Impact Of Social Networking Usage On Grades Among College Students, ASBBS Annual Conference.Las Vegas,2011.
[4] George Karypis, Big data analytics 2013- Partitioning & Clustering Big Graphs, University of Minnesota, 2013.
[5] Heer J. Boyd, D. Vizster: Visualizing Online Social Networks, IEEE Symposium on Information Visualization.InfoVis- 2005.
[6] Florian Bourse, Marc Lelarge, Milan Vojnovic, Big data analytics 2013- Algorithms for Large-Scale Graph Partitioning, Cambridge UK,2013.
[7] Snijders, C., Matzat, U., & Reips, U.-D. 'Big Data': Big gaps of knowledge in the field of Internet. International Journal of Internet Science, 2012.
[8] Laney, Douglas. "3D Data Management: Controlling Data Volume, Velocity and Variety", 6 February 2001.
[9] Laney, Douglas. "3D Data Management: Controlling Data Volume, Velocity and Variety", 6 February 2001.
[10] Webster, John. "MapReduce: Simplified Data Processing on Large Clusters", "Search Storage", 2013.

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.2, Special Issue 5, October 2014

[11] Hendrickson, B. and Leland, R. "A multilevel algorithm for partitioning graphs". Proceedings of the 1995 ACM/IEEE conference on Supercomputing, 1995.

[12]Karen Stepanyan, Kerstin Borau, Carsten Ullrich ,"A Social Network Analysis Perspective on Student Interaction Within the Twitter Microblogging Environment", 10th IEEE International Conference on Advanced Learning Technologies, 2010.

[13]Philip Howard ,"Big Data Analytics with Hadoop and Sybase IQA", White Paper by Bloor Research, April 2012.

[14]Sanders, Peter and Schulz, Christian, "Think Locally, Act Globally: Highly Balanced Graph Partitioning", Proceedings of the 12th International Symposium on Experimental Algorithms, 2013.

[15] Andreev, Konstantin and Räcke, Harald, "Balanced Graph Partitioning". Proceedings of the Sixteenth annual ACM symposium on Parallelism in algorithms and architectures (Barcelona, Spain), 2004.

BIOGRAPHY

K.GEETHA, is currently working as an Assistant Professor in Computer Science Department of Salem Sowdeswari College, Salem-10. She is handling both UG and PG classes. She received her M.Phil. in Computer Science from PRIST University, Thanjavur. She is guiding PG students for the academic project completion. She is also pursuing Ph.D., in Computer Science under Periyar University, Salem-11. She presented two papers in International and State level conferences. Her area of interest is DataMining.

Dr.A.VIJAYA KATHIRAVAN is working as an Assistant Professor in Computer Applications in PG and Research Department of Computer Science, Govt. Arts College (Autonomous), Salem-07, TamilNadu, INDIA. She received her M.Phil. in Computer Science from Bharathiar University, Coimbatore and she awarded her doctoral degree in Computer Applications from University of Madras, Chennai. She has published 6 Books, 3 papers in National Journal, 30 papers in International Journal, 35 Papers in National Conference Proceedings, 38 Papers in International Conference Proceedings and a total of 112 publications. Her research interests include data structures and algorithms, data/text/web mining, search engines, web communities, social network mining, machine learning, Natural Language Processing, Organizational leadership and human resource management.