



Survey on a Graph Partition Method for Large Graph

Charushila N.Gajare

M.E. Computer Engineering Department, Matoshree College of Engineering and Research Center, Maharashtra, India

Abstract— Large graph is one complex data structure. It is used to store and represent information. One must understand its structure and able to decompose it properly without any loss of data. The partition quality consequentially affects the overall computing performance. Balance factor and edge cut ratio is used to measured quality of graph partition. Partitioning or clustering methods are used to decompose a large graph. This paper work will use PAGE algorithm along with different partition technique to remove hardware dependency. Graph partitioning method decomposes a large graph into sub graphs. It finds most connected components of every sub graph which are used to form hierarchical representation of sub graph.

Keywords— Graph partition, graph computation, large graph, sub graph.

I. INTRODUCTION

Nowadays, Large graph consists of billions of nodes and trillions of edges. Web graph, social networks are some examples of large graph. However, This Large graph have more memory to store them. Hence to decompose a large graph into sub graph uses various graph partitioning method without any loss of data. The performance of graph computation system is depending on quality of graph partition. The quality of a graph partition is measured by the balance factor and edge cut ratio. A pregel, Giraph, GPS and Graph Lab [1]. [3], [4], [5] are the example of graph computation system. In graph partition original graph splits into number several sub graphs. The small edge cut ratio indicates that high quality partition and it improve the performance of the system. Because of the communication cost is reduces. The ratio of the edges crossing different sub graphs to the total edges is called edge cut ratio. In the existing graph system cannot obtained the effectively high quality graph partition. In the PAGE computation engine two tasks, first is PAGE worker this module is work with dual concurrent message processor.

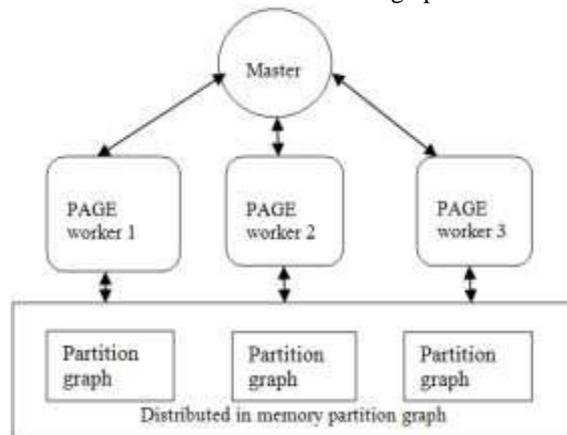


Fig.1 Framework of PAGE

But the PAGE algorithm can utilized normal system resources. Because of multiprocessor communication it required additional overhead time as communication cost. The drawback of PAGE algorithm is developed for dual concurrent message processor so it's hardware dependency in PAGE algorithm by using different technique and also used pattern growth approach with PAGE for sub graph mining. The Proposed system is also work in the huge whole lab of the system.

1.1 Graph computation system

Pregel and Giraph are the optimized only the simple random partition and cannot used for well partitioned graph. Based on pregel and Giraph, GPS [4] uses several other optimizations for the performance improvement. Trinity [5] uses the global message distribution with bipartite graph partition techniques to reduce the memory usage, but it does not use in single computing node. Center Piece Sub graph [11] is work with using random walk with restart(RWR) to optimize all possible graph. Another method is MING approach [10] which is extension of CEPS, it uses the Entity Relationship database to measure the related nodes. CEPS are also the well concept of connection of sub graph. But, faloutsos et al.

[12] states that both CEPS and MING approach are fail to measurement of some characteristics for social networks. Hence Faloutsos[13] has described another familiarity function but it does not used in comprehensive relationship because of it is essential of social network.

Table - I Techniques for Graph computation system

Paper	Year	Technique	Remark
G. Malewicz, M. H. Austern, A. J. Bik, J. C. Dehnert, I. Horn, N. Leiser, and G. Czajkowski. "Pregel: A system for large-scale graph processing."	2010	Simple optimization technique	Optimized only for the simple random partition and cannot efficiently use the well partitioned graph.
S. Sathoglu and J. Widom. "GPS: A graph processing system," in Proc. 25th Int. Conf. Sci. Statist. Database Manage.	2013	Centralized message buffer	It is very preliminary and cannot extend to a verity of of graph computation system.
B. Shao, H. Wang, and Y. Li. "Trinity: A distributed graph engine on a memory cloud."	2013	Bipartiate graph partition technique	It does not discuss the message processing of single computing node.

1.2 Graph partitioning method

A Good partition is depend on less number of edge cut ratio and the consistent graph partition is a divides equal size graph. Spectral Bisection partitioning [13] method is a matrix based approach ,the drawback of this it does not obtained acceptable output to the separation of large number of sub graph is achieved by frequent bisection and also the partition size is unidentified .Multilevel partitioning method is solved for that statistical problem using multigrain method .Karypis and Kumar proposed K-way partitioning called as METIS [9] which is based on multilevel partitioning which reduces the size of graph vertices and edges. The drawback is the for storing graph partition in adjacency matrix, and for run time it is not possible to addition and deletion in sub graph. For the parallel graph partitioning one of the method is Parallel incremental graph partitioning [7] for that method drawback is initial partition is estimate using linear programming based bisection method. After studying literature and taking various drawbacks into consideration, the proposed system aims to develop a "A partitioning method for large graph" which construct high quality graph partitioning using small edge cut ratio and balance factor. And also reduce the computational cost of graph partitioning.

Table - I Techniques for Graph computation system

Paper	Year	Technique used	Remark
Stephen T. Barnard and Horst D. Simon. A fast multilevel implementation of recursive spectral bisection for partitioning unstructured problems. In Proceedings of the sixth SIAM conference on Parallel Processing for Scientific Computing	1993	Matrix based approach	It does not always give satisfactory result and also minimum cut partitioning fails when the number of sub graph to be formed.
G. Karypis and V. Kumar. "Multilevel Graph Partitioning Schemes," Proc. IEEE/ACM Conf. Parallel Processing, pp. 113-122	1995	Multilevel partitioning problems	It uses static data structure to store partitions node or edges addition or deletion in subgraph at runtime is not passible.
Chao-Wei Ou and Sanjay Ranka "Parallel Incremental Graph Partitioning" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS	1997	Recursive spectral bisection based method	The drawback of the method is initial partition is to be calculated using linear programming based bisection method.

II. ARCHITECTURE OF PROPOSED SYSTEM

The figure 2 shows the architecture of the proposed system. There is huge graph database of vertices and edges from that database one pre-processing state are created in that how many vertices and edges are there and which edge is connected to vertices and also find weight of the edges. The system architecture shows following detail explanation:

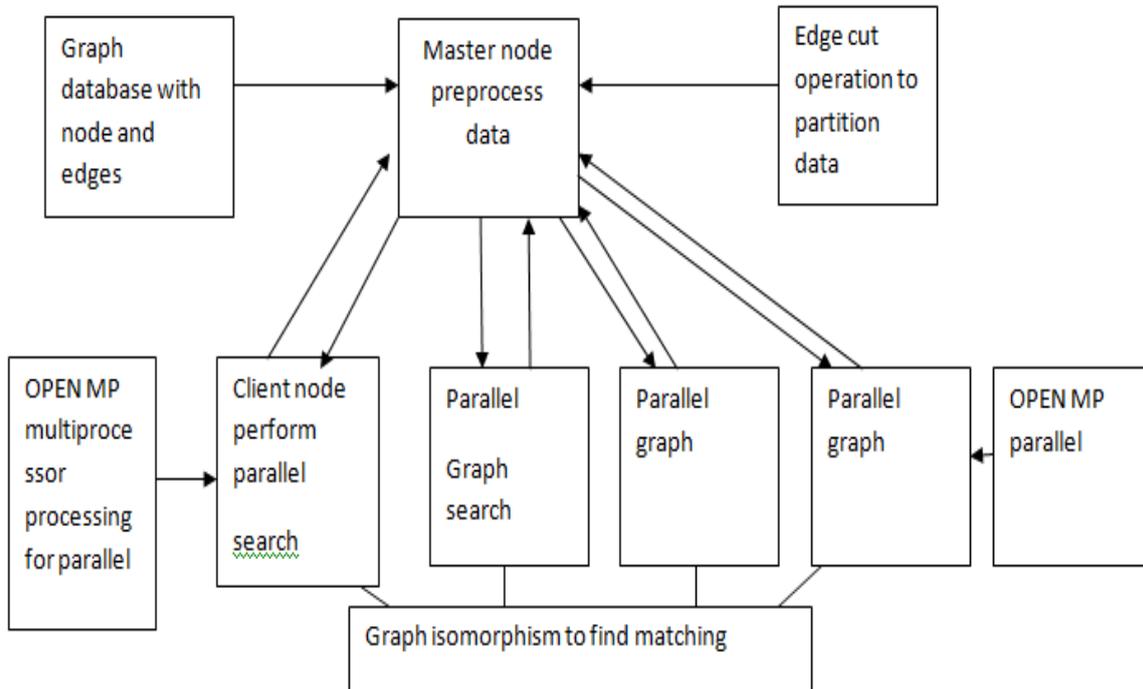


Fig. 2. Architecture of Proposed System

2.1 Master node

Master node is special thread it remove null value from graph database. It detects number of slave node connected and partition data equally on each slave node. Master node can give the input file and then partition creates using graph cut method. This method can create the multiple play graph into sub graph partition and these sub graph partition are send to slave node.

2.2 Slave/parallel node

Slave node get query and data from master and use graph isomorphism using OPEN-MP technique to detect sub graph present in give data D. slave master plays the role like if graph found slave node return data to master or find frequent match graph.

2.3 Graph isomorphism

Graph isomorphism used to match the sub graph. From that above discussion the given one big tree which are divided sub graph this subgraph are match to query graph to check those are same or not. This operation is locally performed using isomorphism algorithm. This local node result is return to send the master node. The advantage of chord core system or hexachord system performs the parallel using OPENMP multiprocessor. Hence the operation is performed by parallel and also the distributed. Therefore the overall performance of time is reduced.

2.4 Isomorphism technique

The duplicate copies of a candidate patterns are isomorphic to each other. For identifying graph isomorphism canonical coding scheme is used. The fig.3 shows the example of isomorphism technique this technique is used to in the proposed system to find the graph matching.

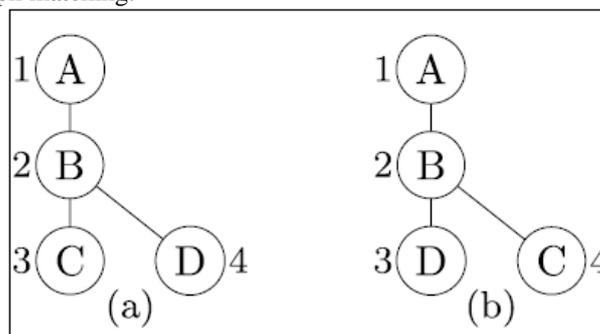


Fig. 3. Example of isomorphism technique

III. CONCLUSION

In this paper, the proposed system described Balance factor and edge cut ratio is used to measured high quality of graph partition. The existing system PAGE algorithm is developed for dual concurrent message processors so its hardware dependent. And this system cannot utilized in normal system resources because of multiprocessor communication also the algorithm is used to additional communication cost and synchronization barrier of the master node. The portioning method for PAGE algorithm is simple i.e without any optimization. Hence the proposed system remove hardware dependency in page algorithm by using different partition technique and achieve same performance on graph data set and also use pattern growth approach with PAGE for Sub graph mining.

ACKNOWLEDGMENT

At the outset, I thank to Prof. J.F. Rodrigues for providing the DBLP data set.

REFERENCES

- [1] GIRAPH. (2013). [ONLINE]. AVAILABLE:HTTPS://GITHUB.COM/APACHE/GIRAPH.
- [2] Y. Low, D. Bickson, J. Gonzalez, C. Guestrin, A. Kyrola, and J. M. Hellerstein, "distributed graphlab: A framework for machine learning and data mining in the cloud," Proc. VLDB Endowment. vol. 5, pp.716- 727, 2012 .
- [3] G. malewicz, M. H. Austem, A. J. Bik, J. C. Dehnert, I. Hom, N. Leiser, and G. Czajkowski, "Pregel: A system for large-scale graph processing," in proc. ACM SIGMOD Int. Conf. Manage. Data, 2010, pp. 135-146.
- [4] S. Salihoglu and J. Widom, "GPS: A graph processing system," in Proc. 25th Int. Conf. Sci. Statist. Database Managee, in Proc. 21st Int. Conf. World Wide Web, 2012, pp. 949-958.
- [5] J. C. Dehnert, I. Hom, N. Leiser, and G. Czajkowski, "Pregel: A system for large-scale graph processing," in proc. ACM SIGMOD Int. Conf. Manage. 2013, pp. 22:1-22:12.
- [6] C. Faloutsos, K. S. McCurley, and A. Tomkins, *Fast Discovery of Connection Subgraphs*, Proc. ACM 10th Intl-Conf. Knowledge Discovery and Data Mining (SIGKDD), pp. 118-127, 2004.
- [7] Chao-Wei Ou and Sanjay Ranka *Parallel Incremental Graph Partitioning* IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 8, NO. 8, AUGUST 1999.
- [8] C.R. Palmer and C. Faloutsos, *Electricity Based External Similarity of Categorical Attributes*, Proc Seventh Pacific- Asia Conf. Advances in Knowledge Discovery and Data Mining (PAKDD), pp. 486-500, 2003.
- [9] G. Karypis and V. Kumar, *Multilevel Graph Partitioning Schemes*, Proc. IEEE/ACM Conf. Parallel Processing, pp 113-122, 1995.
- [10] G. Kasneci, S. Elbassuoni, and G. Weikum, *Ming: Mining Informative Entity Relationship Subgraphs*, Proc. 18th ACM Conf. Information and Knowledge Management (IKM), pp. 1653- 1656, 2009
- [11] H. Tong and C. Faloutsos, *Center-Piece Subgraphs: Problem Definition and Fast Solutions*, Proc. 12th ACM SIGKDD Intl Conf. Knowledge Discovery and Data Mining (KDD), pp. 404-413, 2006.
- [12] J.F. Rodrigues Jr., H. Tong, A.J.M. Traina, C. Faloutsos, and J. Leskovec, *Large Graph Analysis in Gmine System* IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 1, JANUARY 2013.
- [13] Stephen T. Barnard and Horst D. Simon. *A fast multilevel implementation of recursive spectral bisection For partitioning unstructured problems*. In Proceedings of the sixth SIAM conference on Parallel Processing for Scientific Computing, pages 711-718, 1993.