e-ISSN: 0976-516(
p-ISSN: 2231-385(

- For the Facebook Social Network (Figure 6.a): In these datasets, the Connected Components algorithm with MGAOFSM performs well for high threshold value (25) whereas PageRank algorithm performs better with all four datasets. Triangle closing and Single Source Shortage Path algorithm performance same with 20 and 15 support value, but it drastically reduces with 25 threshold value due to optimization at this level. If we evaluate on the basis of execution time with different threshold value triangle closing algorithms for MGAOFSM with Facebook Social Network is better performance.
- Coronavirus (COVID-19) tweets (Figure 6.b): For this dataset, the PageRank, Connected Component, Single Source Shortest Path (SSSP) algorithm has the highest execution time with the lowest value for triangle closing algorithm for threshold 10. On an average Triangle Closing is the smallest execution time for all thresholds. The Connected Component has the highest average execution time in the dataset. For the PageRank, Connected Component, Single Source Shortest Path algorithms, the execution time with the decreasing threshold value.
- Google Web (Figure 6.c): For PageRank, Connected component, Single Source Shortest Path algorithms, the execution time increases with a decrease in threshold value. The average execution time for all threshold values is the smallest i.e. (56). Whereas connected components have maximum average execution time.
- Patent citation network (Figure 6.d): In these datasets, the SSSP has the lowest execution time with all values of thresholds and the highest with connected components. On average Triangle Closing was the lowest among all algorithms we tried to implement with our proposed methods.

Secondly, we consider our comparison with execution time with four algorithms with our proposed methodology.

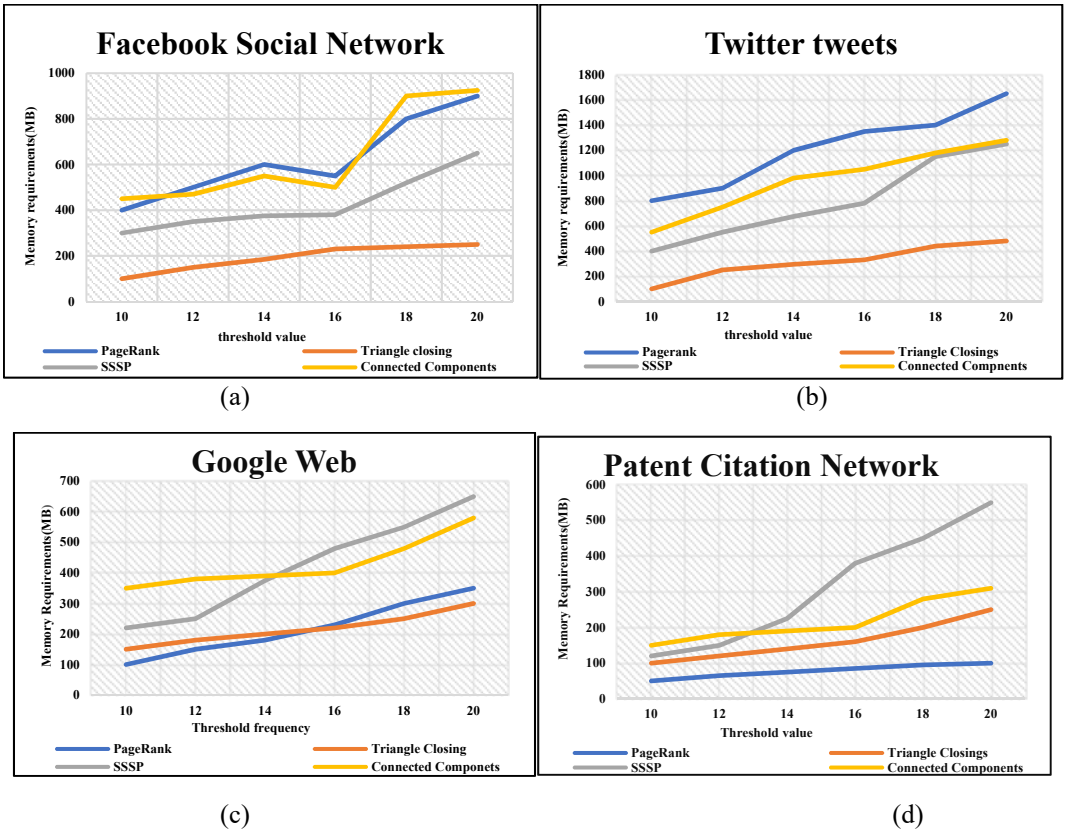

Figure 7. Memory requirement vs Threshold values

Except the Patent Citation Network, our proposed algorithm requires less memory with Triangle closing algorithm whereas in PageRank is lowest with all threshold values. The Single Source Shortest Path algorithm with GAOFSM has same memory requirement for 10, 12 percentage threshold value, then it increase with threshold value.

6. Conclusion and Future work

We investigated several graph mining algorithm for frequent subgraph with our proposed algorithm. The Giraph system has lowest execution time with Triangle Closing algorithm associated with MapReduce Geometric Multi-way Advanced Optimized Frequent Subgraph Mining. In this paper, we are able to merge both Giraph and

MapReduce frameworks to get better results as well as memory consumption less in a distributed system. PageRank methodology with our proposed algorithm required the lowest memory in all four graph datasets.

7. Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Anchuri, P.,Zaki, M. J.,Barkol,O.,Golan,S. and Shamy M.(2013), 'Approximate graph mining with label costs'. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2013).
[2] D. Dai, W. Zhang, and Y. Chen, "Iogp: An incremental online graph partitioning algorithm for distributed graph databases," in Proceedings of the 26th International Symposium on High- Performance Parallel and Distributed Computing. ACM, 2017, pp. 219–230.
[3] Mccune, R. R., Weninger, T., and Madey, G.(2015) 'Thinking like a vertex: A survey of vertex-centric frameworks for large-scale distributed graph processing', ArXiv: 1507.04405 (2015).
[4] Tomasz Kajdanowicz, Przemyslaw Kazienko, and Wojciech Indyk. 2014. Parallel Processing of Large Graphs. Future Generation. Comput. Syst. 32 (March 2014), 324–337. DOI:http://dx.doi.org/ 10.1016/j.future.2013.08.007.
[5] Yue Zhao, Kenji Yoshigoe, Mengjun Xie, Suijian Zhou, Remzi Seker, and Jiang Bian. 2014. LightGraph: Lighten Communication in Distributed Graph-Parallel Processing. In Proceedings of the 2014 IEEE International Congress on Big Data (BIGDATACONGRESS '14). IEEE Computer Society, Washington, DC, USA.
[6] Bruce Hendrickson a,*, Tamara G. Kolda," Graph partitioning models for parallel computing", Parallel Computing 26 (2000) , www.elsevier.com/locate/parc.
[7] Saeed Salem, Mohammed Alokshiya , Mohammad Al Hasan "RASMA: a reverse search algorithm for mining maximal frequent subgraphs", BioData Mining 14, 19 (2021). https://doi.org/10.1186/s13040-021-00250-1
[8] Lam B. Q. Nguyen, Loan T. T. Nguyen, Ivan Zelnika, Vaclav Snasel, Hung Son Nguyen ,Bay Vo," A Method for Closed Frequent Subgraph Mining in a Single Large Graph" Digital Object IEEE, Identifier 10.1109/ACCESS.2021.313366, VOLUME 9, 2021,December 23, 2021.
[9] Lam B. Q. Nguyen , Ivan Zelinka, Quoc Bao Diep" CCGraMi: An Effective Method for Mining Frequent Subgraphs in a Single Large Graph" MENDEL. 27, 2 (Dec. 2021), 90-99. DOI:https://doi.org/10.13164/mendel.2021.2.090.
[10] Saif Ur Rehman, Kexing Liu, Tariq Ali, Asif Nawaz, Simon James Fong "A Graph Mining Approach for Ranking and Discovering the Interesting Frequent Subgraph Patterns" International Journal of Computational Intelligence Systems volume 14, Article number: 152 (2021) .
[11] Giulia PretiMatteo Riondato "MaNIACS: Approximate Mining of Frequent Subgraph Patterns through Sampling, KDD '21: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data MiningAugust 2021 Pages 1348–1358https://doi.org/10.1145/3447548.3467344.
[12] Xiaohong Zhang, Zhiyong Zhong, Shengzhong Feng, Bibo Tu, Jianping Fan," Improving Data Locality of MapReduce by Scheduling in Homogeneous Computing Environments", Conference: IEEE International Symposium on Parallel and Distributed Processing with Applications, ISPA 2011, Busan, Korea, 26-28 May, 2011.
[13] Lam B. Q. Nguyen,Ivan Zelinka,Vaclav Snasel,Loan T. T. Nguyen,Bay Vo" Subgraph mining in a large graph: A review", Wires Data Mining and Knowledge Discovery.08 March 2022
[14] Zigang Cao,Gang Xiong,Yong Zhao,Zhenzhen Li,Li Guo," A Survey on Encrypted Traffic Classification", International Conference on Applications and Techniques in Information Security ATIS 2014: Applications and Techniques in Information Security pp 73-81.
[15] Yufei Gao,[1] Yanjie Zhou,[2] Bing Zhou,[3] Lei Shi,[4] and Jiacai Zhang "Handling Data Skew in MapReduce Cluster by Using Partition Tuning", Journal of Healthcare Engineering , Recent Advances and Developments in Mobile Health, Volume 2017 |Article ID 1425102 ,"
[16] Hill S, Srichandan B, and Sunder Raman R(2012). 'An iterative MapReduce approach to frequent subgraph mining in biological datasets'. In Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine (2012).
[17] Jaliya Ekanayake, Hui Li, Bingjing Zhang, Thilina Gunarathne, Seung-Hee Bae, Judy Qiu, and Geoffrey Fox. 2010. Twister: A Runtime for Iterative MapReduce. In Proceedings of the 19th ACM.International Symposium on High Performance Distributed Computing (HPDC '10). ACM, New York, NY, USA, 810–818.
[18] Jaliya Ekanayake, Hui Li, Bingjing Zhang, Thilina Gunarathne, Seung-Hee Bae, Judy Qiu, and Geoffrey Fox. 2010. Twister: A Runtime for Iterative MapReduce. In Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing (HPDC '10). ACM, New York, NY, USA, 810–818.
[19] Hamilton Wilfried Yves Adoni, Tarik Nahhal, Moez Krichen, A survey of current challenges in partitioning and processing of graph-structured data in parallel and distributed systems,Springer,june2020 , Distributed and Parallel Data Database.
[20] Charu C. Aggarwal, Haixun Wang," Managing and Mining Graph Data", IOP Conference Series Materials Science and Engineering 180(1):012065, March 2017.
[21] Minyang Han, Khuzaima Daudjee, Khaled Ammar, M Tamer Ozsu, Xingfang Wang, and Tianqi Jin. 2014. An Experimental Comparison of Pregellike Graph Processing Systems. Proceedings of the VLDB Endowment 7, 12 (2014), 1047–1058.
[22] Ribeiro, P., and Silva, F(2014),'G-Tries: A data structure for storing and finding subgraphs.'Data Mining and Knowledge Discovery 28, 2 (2014).
[23] Saeed Salem,Mohammed Alokshiya, Mohammad Al Hasan,RASMA: a reverse search algorithm for mining maximal frequent subgraphs,BMC Part of Springer nature,march 16,2021.
[24] C. Sakouhi, S. Aridhi, A. Guerrieri, S. Sassi, and A. Montresor, "Dynamicdfep: A distributed edge partitioning approach for large dynamic graphs," in Proceedings of the 20th International Database Engineering & Applications Symposium. ACM, 2016, pp. 142–147.
[25] D. Dai, W. Zhang, and Y. Chen, "Iogp: An incremental online graph partitioning algorithm for distributed graph databases," in Proceedings of the 26th International Symposium on High-Performance Parallel and Distributed Computing. ACM, 2017, pp. 219–230.
[26] Sergey Edunov,Dionysio Logothesis,Cheng Wang,Avery Ching and Maja Kabiljo,"Generating synthetic social graphs with Darwini" IEEE International Conference on Distributed Computing Systems, 2018.

Authors Profile

Ms Sadhana Priyadarshini is a Phd scholar in Department of Computer Science and Engineering at GITAM (Deemed to be University), Vishakhapatnam, India. She completed MTech (CSE) from SQA University in 2010.Her research interests in field of Data Mining.

Dr. Sireesha Rodda is a Head of Department of Computer Science & Engineering, GITAM (Deemed to be University). She has 20 years of research experience in the fields of Artificial Intelligence, Data Mining and Machine Learning. She has more sthan 45 papers published in referred journal.