

- The main issue with lower support threshold is that they end up with an exponentially expanded search space which would affect the performance a lot
- The most basic approach to fight this problem should be to take into account the different possibilities of pruning the search space. This kind of pruning can either be such that the information that would be lost won't be useful at all for the given query. Or the pruning might be more time efficient but end up with some crucial information loss, thus leading to approximate outputs.
- In such cases, it all leads to developing different pruning methods such that any theoretical bounds on their approximate outputs can be guaranteed. Thus making such methods practically useful when needed
- Frequent Subgraph mining techniques have been classified into two major categories: (1) apriori-based approaches and (2) pattern growth-based approach. ^{[6],[7],[8]}
- There is a large overhead in the case of Apriori-based (BFS-based) methods, and curb those issues FP-growth based (DFS-based) methods were introduced.
- Now, in the case of FP-grow techniques, we can think of different techniques which would prune the paths according to certain conditions so that the method would not need to go through each of these paths.
- Now we can think of using a similar concept of pruning the frequent subgraphs as used in the case of pruning the number of possible frequent itemsets. That is the notion of a "closed subgraph", just like "closed itemsets"
- Now just as a closed itemset is defined as an itemset in which no superset has the same support, we can define a closed subgraph fragment as a fragment in which no superstructure of it has the same support.
- As in the case of item sets, restricting the search to the closed itemset fragments does not lose any information. Similar thing would happen in the case of closed subgraph fragments.
- The reason behind this will be that all the frequent fragments can be constructed from the closed fragments just by forming all the substructures for the closed fragments that are not closed fragments themselves.
- Thus we can simply restrict the search to these closed fragments and end up pruning the search space without losing any information. A method similar to this one (they have some modifications over it and use the term of *Equivalent Occurrence Pruning*) can be found in the paper of Closegraph^[2]. It is similar to the canonical representation used in the case of gSpan^[3].
- Different aspects regarding this method are also discussed in detail in the paper^[1].
- Similar technique to prune the items that can not generate any frequent patterns based on their Minimum Support values are also discussed and used in these research works as CFP growth algorithm. ^{[4],[5]}

References

- [1] Borgelt, Christian, and Thorsten Meinl. "Full Perfect Extension Pruning for Frequent Subgraph Mining." Mining Complex Data. Springer, Berlin, Heidelberg, 2009. 189-205.
- [2] X. Yan and J. Han. Closegraph: Mining Closed Frequent Graph Patterns. Proc. 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD2003, Washington, DC), 286–295. ACM Press, New York, NY, USA 2003
- [3] X. Yan and J. Han. gSpan: Graph-Based Substructure Pattern Mining. Proc. 2nd IEEE Int. Conf. on Data Mining (ICDM 2003, Maebashi, Japan), 721–724. IEEE Press, Piscataway, NJ, USA 2002
- [4] Kiran, R. Uday, and P. Krishna Reddy. "Novel techniques to reduce search space in multiple minimum supports-based frequent pattern mining algorithms." Proceedings of the 14th international conference on extending database technology. 2011.

- [5] Hu, Ya-Han, and Yen-Liang Chen. "Mining association rules with multiple minimum supports: a new mining algorithm and a support tuning mechanism." *Decision Support Systems* 42.1 (2006): 1-24.
- [6] Lakshmi, K., and T. Meyyappan. "A comparative study of frequent subgraph mining algorithms." *International Journal of Information Technology Convergence and Services* 2.2 (2012): 23.
- [7] K. Lakshmi and T. Meyyappan, "Frequent subgraph mining algorithms—a survey and framework for classification," in *Proceedings of the Conference on Innovations in Theoretical Computer Science (ITCS '12)*, pp. 189–202, 2012.
- [8] Rehman, Saif Ur, et al. "Performance evaluation of frequent subgraph discovery techniques." *Mathematical Problems in Engineering* 2014 (2014).