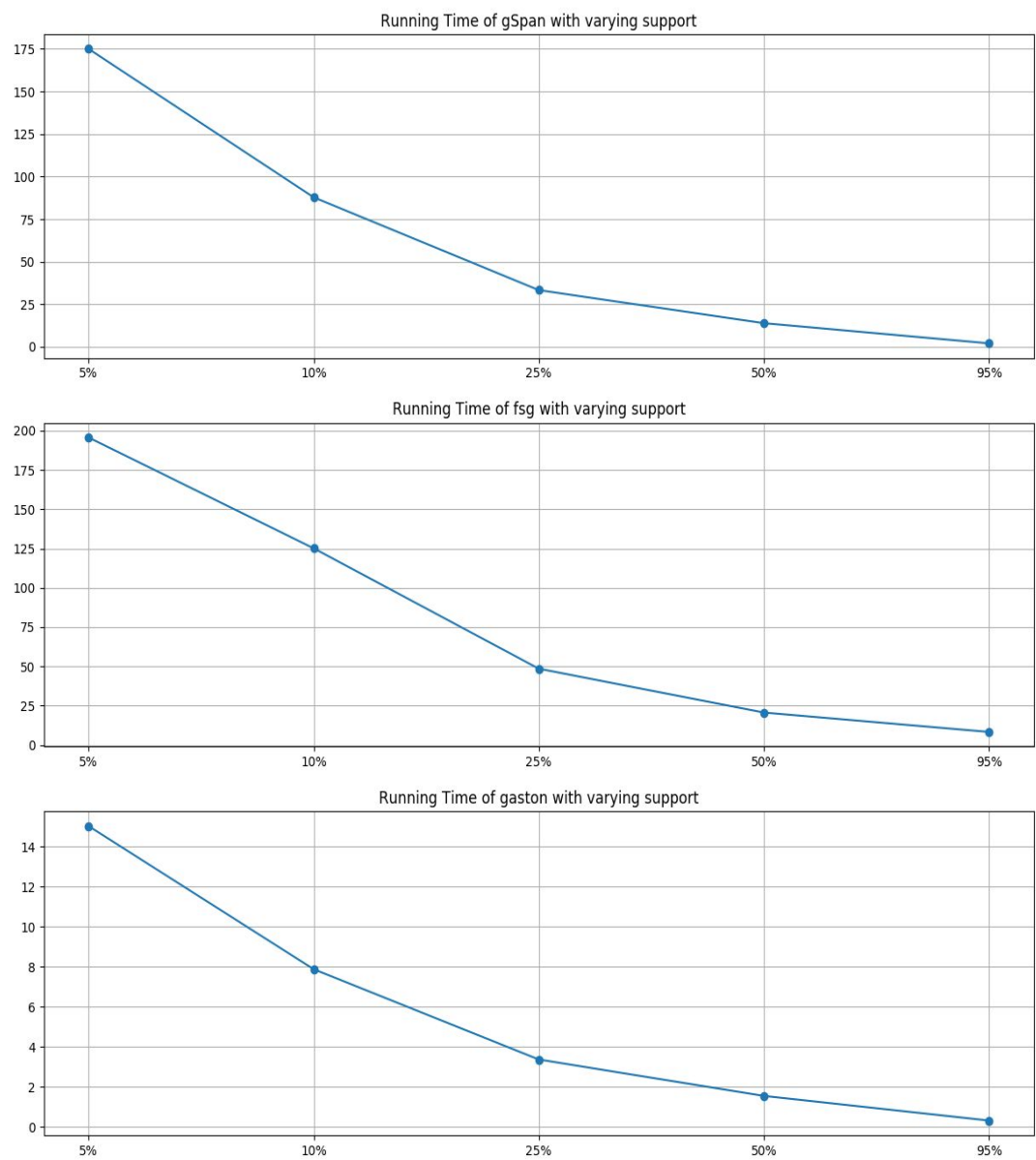


# Comparison between gSpan, FSG and Gaston



### **Observations :-**

- Running time in decreasing order is  $FSG > gSpan > Gaston$ .
- Running time decreases as support values are increased which is in alignment to the expected result. This is because the number of frequent subgraphs decreases when the support values are increased.
- Difference in the running times among the 3 algorithms decreases with increase in support value.
- Growth Rate - As the support values are reduced, the running time increases by most rate in case of FSG followed by gSpan and then Gaston. This means that the slopes of the running time plot for the 3 algorithms is  $FSG > gSpan > Gaston$ .
- Due to the above mentioned points, we can call Gaston as more scalable and faster which is then followed by gSpan and then by Gaston

### **Reasons :-**

#### **FSG vs gSpan**

- FSG uses BFS(Breadth First Search) strategy. So, by nature of BFS strategy it explores all the nodes level by level. It is thus slow because it processes nodes which are not even relevant.
- gSpan uses DFS(Depth First Search) strategy, hence making gSpan an order faster than FSG because it removes all infrequent edges at every step and thus avoids generating candidates and pruning steps like FSG. It thus eliminates repeated database scans required in FSG Algorithm.

#### **Why is Gaston fastest?**

- Gaston uses the fact that most frequent substructures in practical graph databases are actually free trees.
- The algorithm thus "quickstarts" its search by using a highly efficient enumeration strategy for enumerating the frequent free trees first.
- It follows a global order on cycle closing edges and generates those cycles which are bigger than the previous ones.
- In addition, duplicate detection is done efficiently in two phases which include hashing for pre sorting and isomorphism testing for final duplicate detection.
- All the above points make Gaston work the fastest.

## References -

- gSpan: Frequent Graph Mining Package  
<https://www.cs.ucsb.edu/~xyan/software/gSpan.htm>
- Gaston open source library  
<http://liacs.leidenuniv.nl/~nijssensgr/gaston/download.html>
- PAFI Software Package for Finding Frequent Patterns in Diverse Datasets  
<http://glaros.dtc.umn.edu/gkhome/pafi/overview/>