COL761 Assignment 1 Q2

Kartik Arora (2021CS50124) Piyush Chauhan(2021CS11010)

Ayushman Kumar Singh (2021CS51004)

February 10, 2025

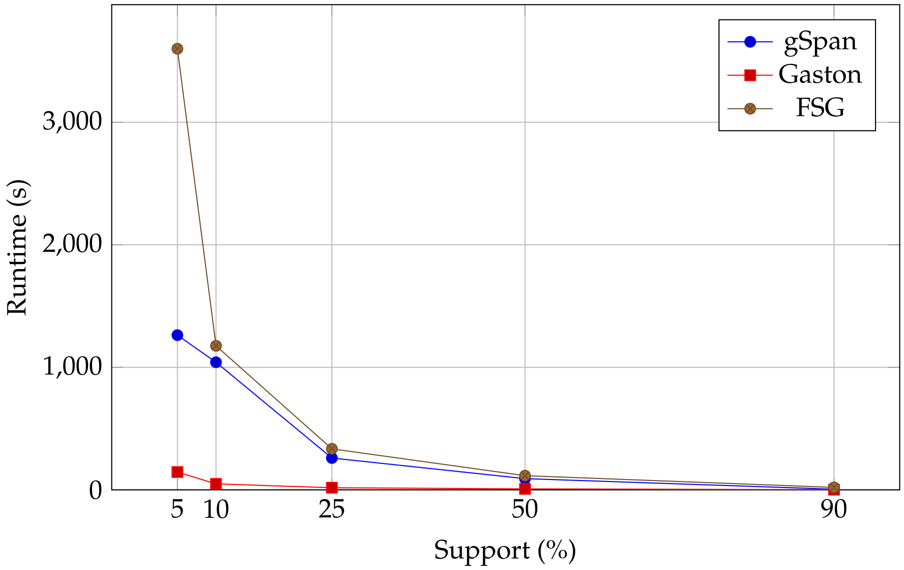
§1. Introduction

We perform an empirical analysis comparing the efficiency of Gaston, gSpan, and FSG for frequent subgraph mining. Using existing libraries, we evaluate their performance on the given dataset.

§2. Results

Support	gSpan Time (s)	Gaston Time (s)	FSG Time (s)
5%	1262.5666(MLE)	145.2926	3600(TLE)
10%	1041.3976	48.5983	1175.5403
25%	259.6201	17.1260	334.7641
50%	91.4732	7.9460	115.5521
90%	5.1786	1.0229	19.8195

Table 1: Comparison of gSpan, Gaston, and FSG Execution Times



§3. Analysis

- Gaston is the fastest across all methos because it uses an optimized pattern growth approach. It separately process paths trees and graphs, expanding one another in the aforementioned order.
- gSpan performs better that FSG for very high support threshold because for DFS approach. For 25% and 50% thresholds, gSpan slows down a little because of memory requirements.
- FSG is generally the slowes as it uses Breadth first search kind of approach , generating candidates and taking multiple passes over dataset.
- At low threshold , FSG gives TLE as the number of candidates generated becomes very large.
- Similarly, gSpan fails at lower threshold because of Memory overflow, because it need to store and process many DFS embeddings.

§4. Conclusion

- Gaston is the fastest approach for frequent subgroup mining because of its optimized search strategy.