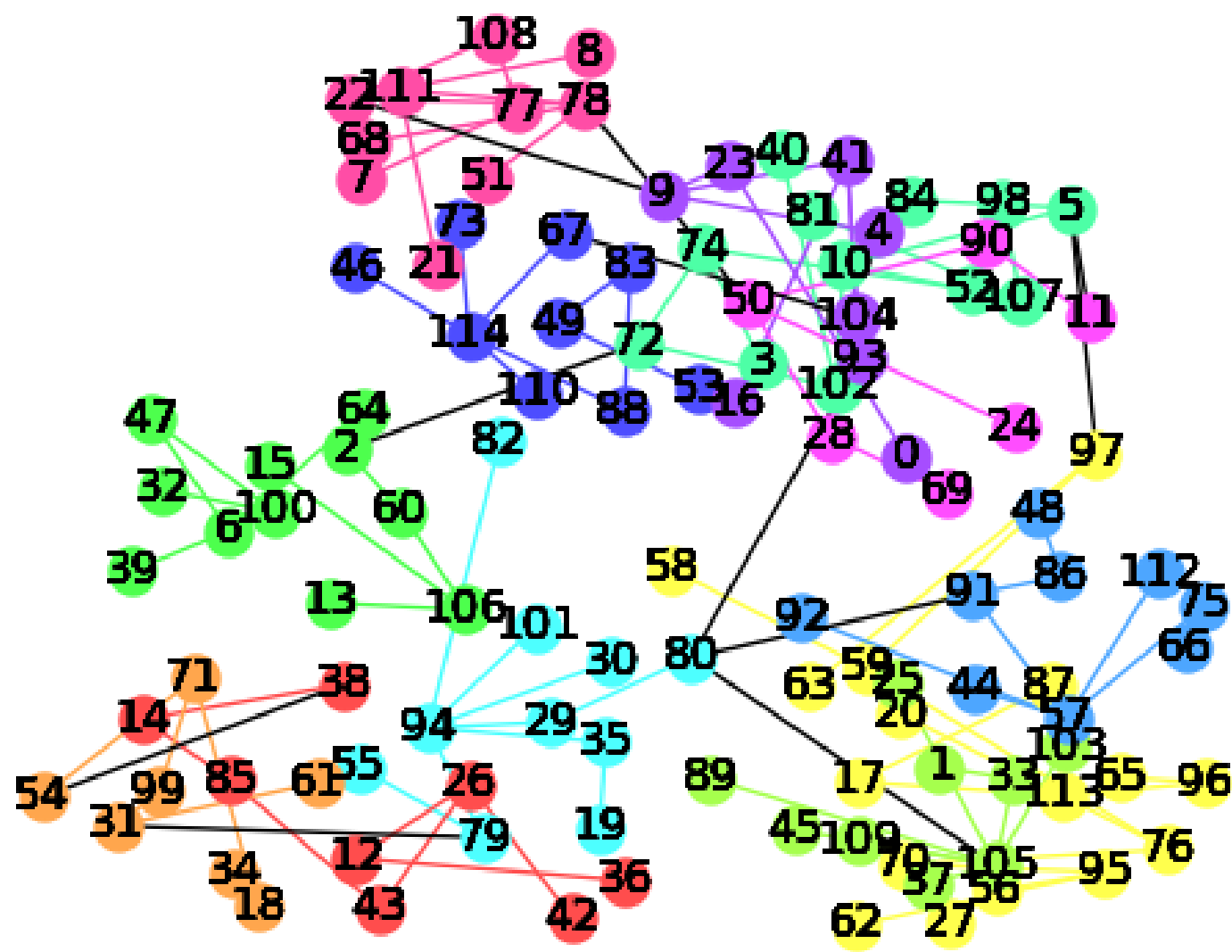


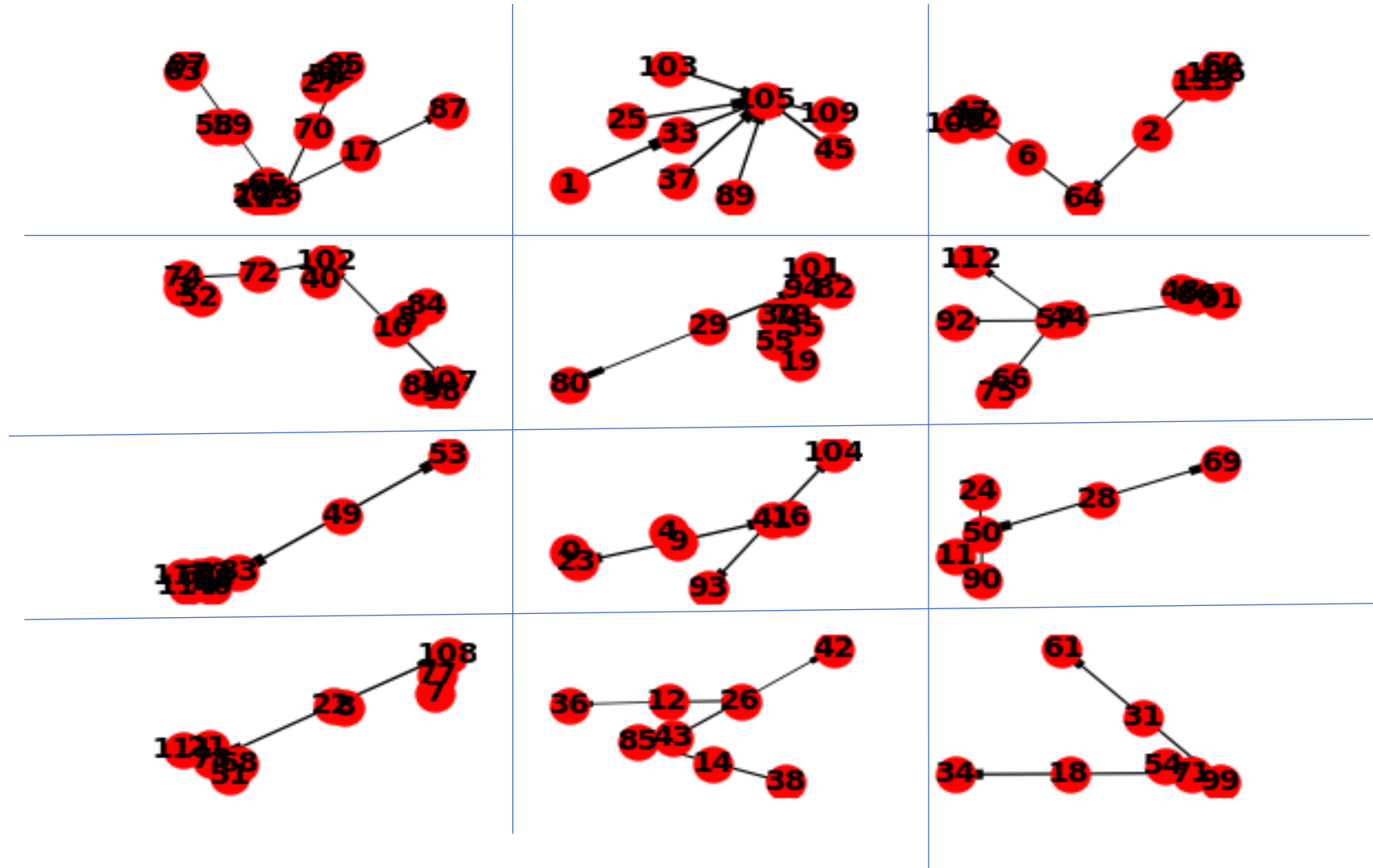
Graph mining

Results of dcut

- Create density tree
- Cut density tree
- Find all neighbors of the two cutted nodes
- Create two new trees with this nodes
- Repeat until n-trees were created
- Find nodes in original graph
- Color them in clusters



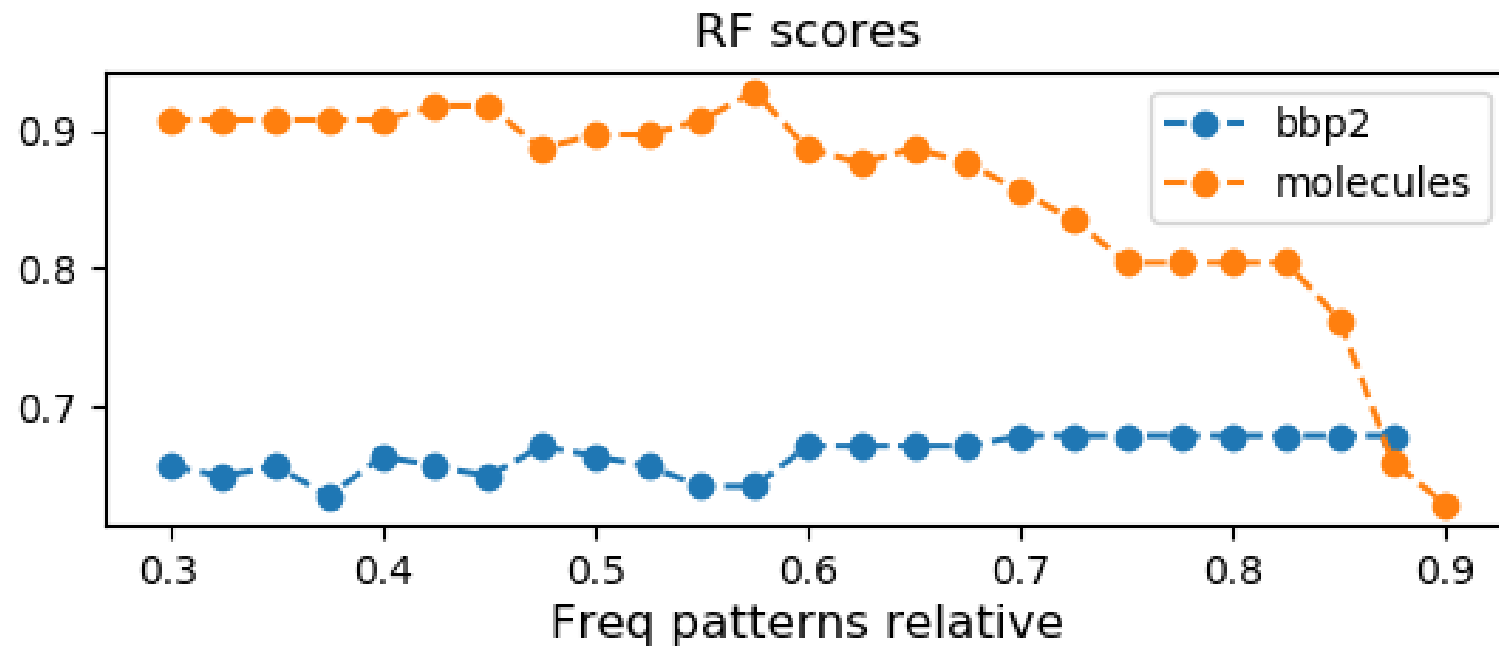
Density trees after 12 cuts



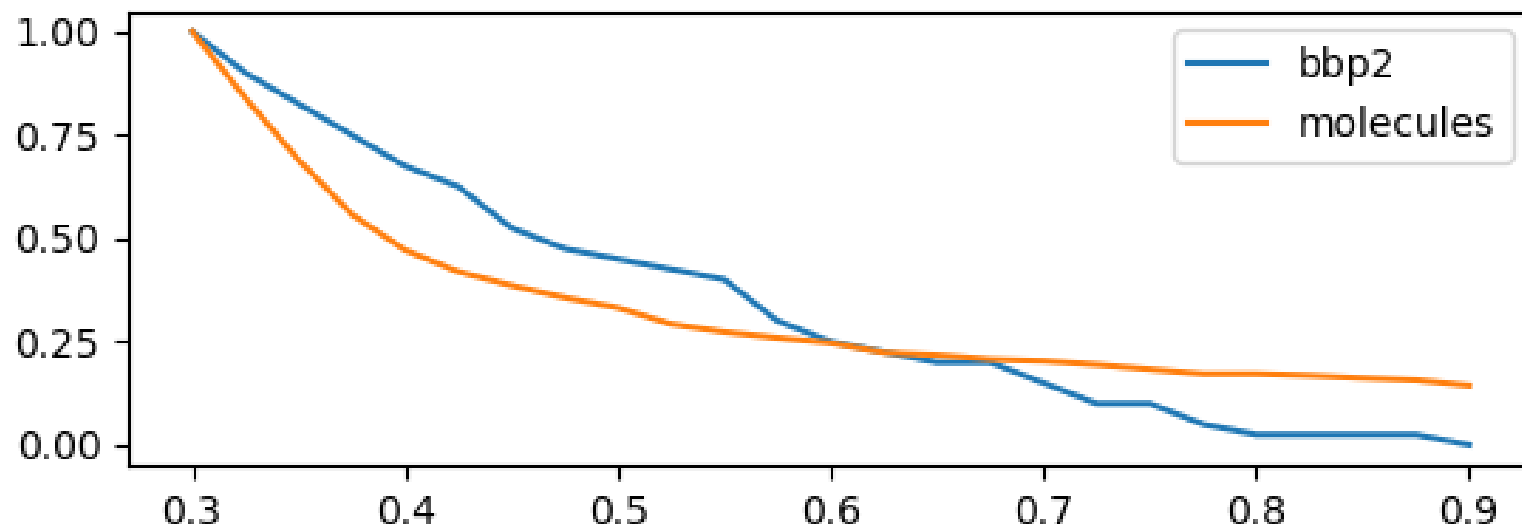
gspan

- Load graphs
- Split data in training and test data
- Create matrices with results of gspan algorithm
 - Column should be freq patterns
 - Rows should be the input graphs
 - 1 if pattern in graph else 0
- Hand train matrix to an RandomForest classifier
- Check performance with the test matrix

Different rf scores depending on Threshold of the gspan



Development of the amount of found patterns



SLR-kit

- Basic run + -showAUC -pruneSingletons
-pruneZeroKnowledge
- Statistics used:
 - Average graph degree
 - Density
 - Subgraphs

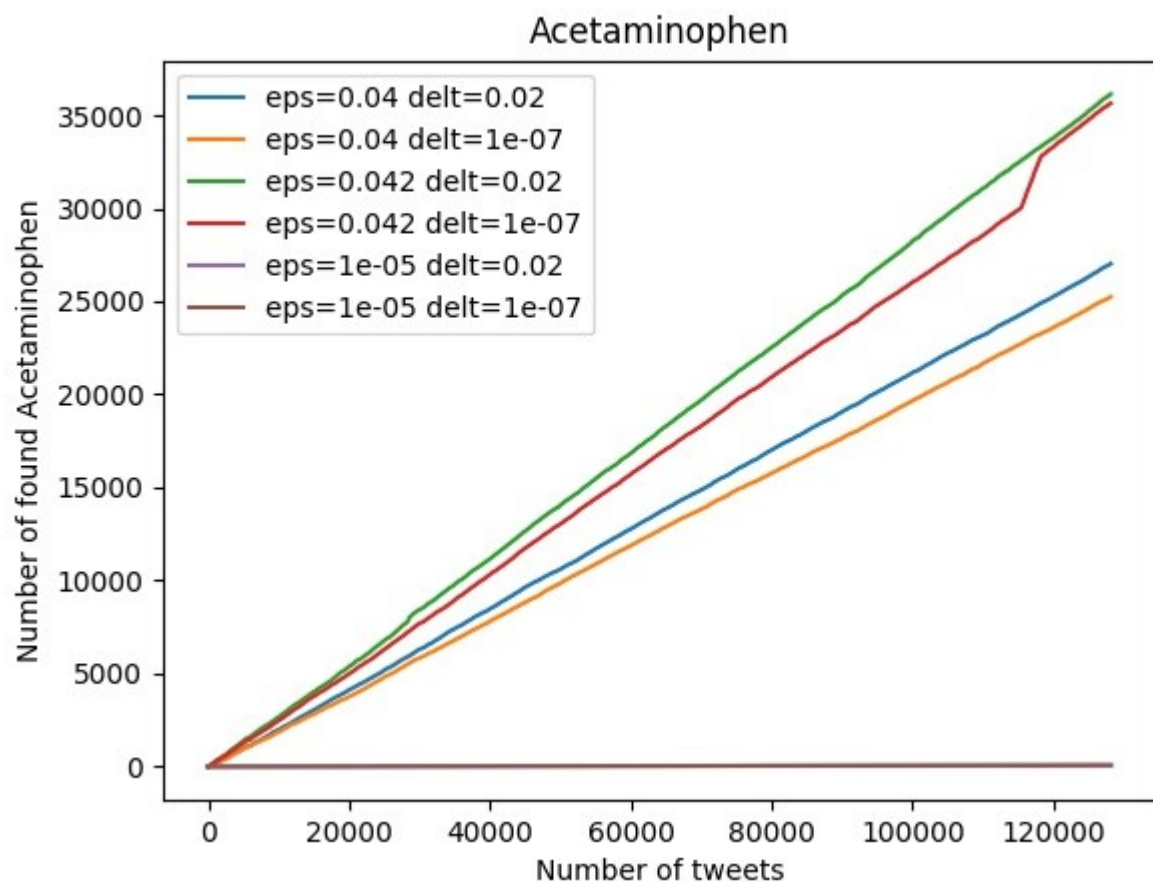
cora_cite

- Size of graph 4240
- Average degree 5.31
- Density 0.0025
 - Nodes with degree 0 , ausgangsgrad, eingsangsgrad (657, 657)
- Full connected graph, size of the subgraphs (False, [3385, 34, 11, 9, 6, 5x2, 4x6, 3x2, 2x46

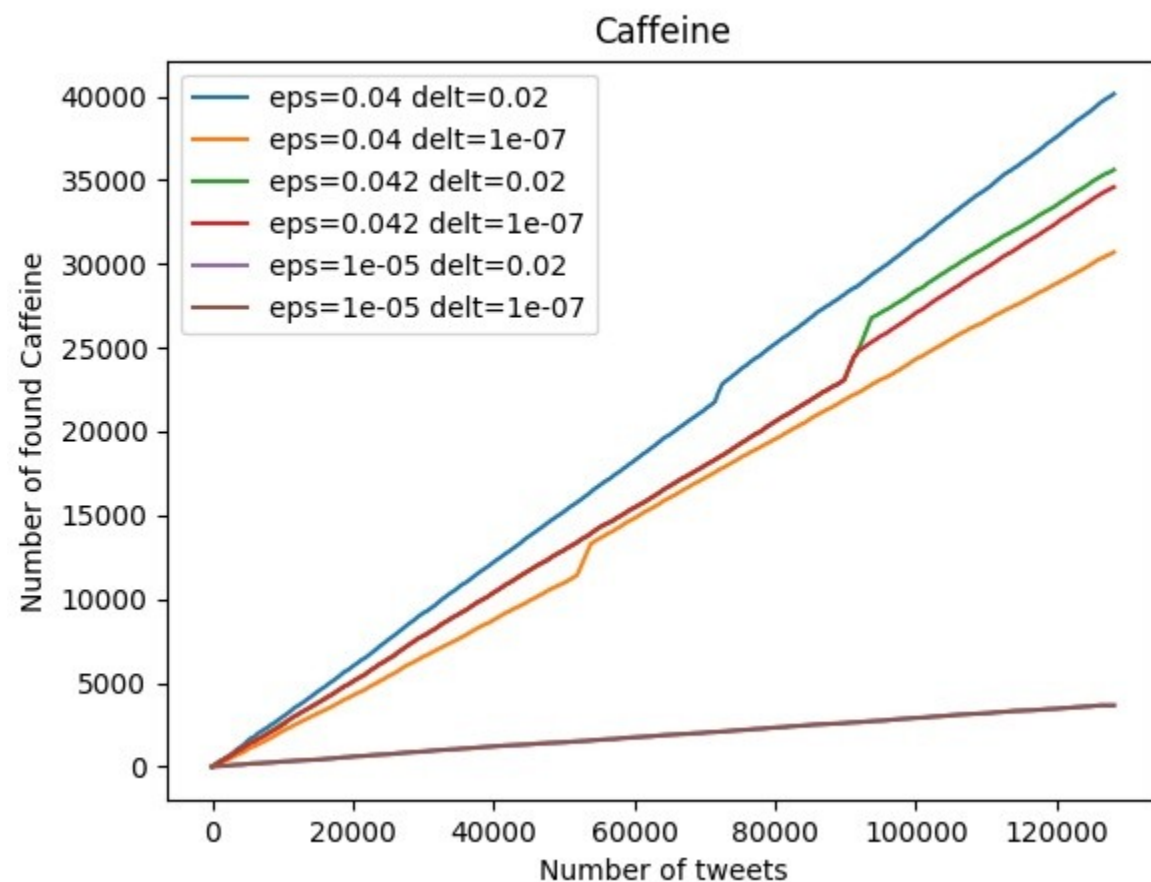
Count Min Sketch

- CMS table with murmurhash
- Read tweets & perform estimate over the drugs
-
- Epsilon --> how much error is added to our counts with each item we add to the cm sketch
- Delta --> with what probability do we want to allow the count estimate to be outside of our epsilon error rate

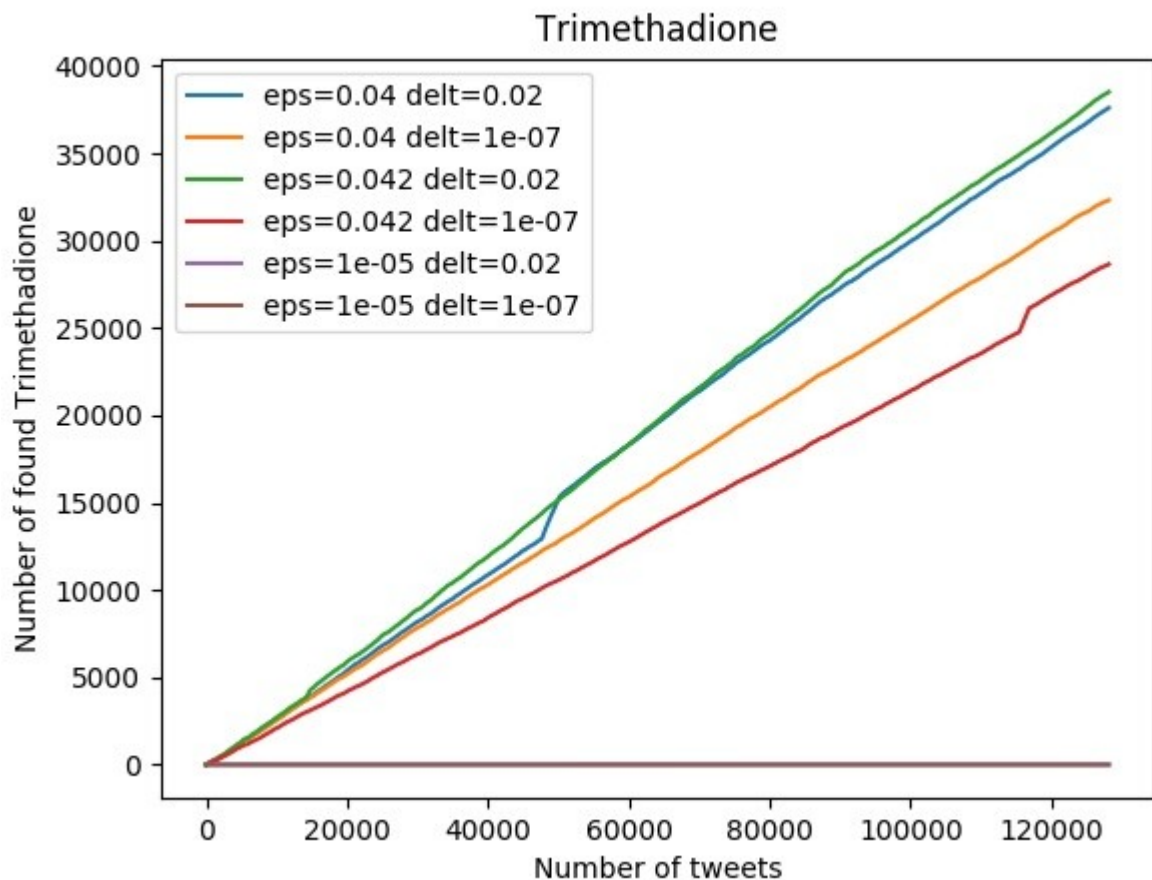
Acetaminophen



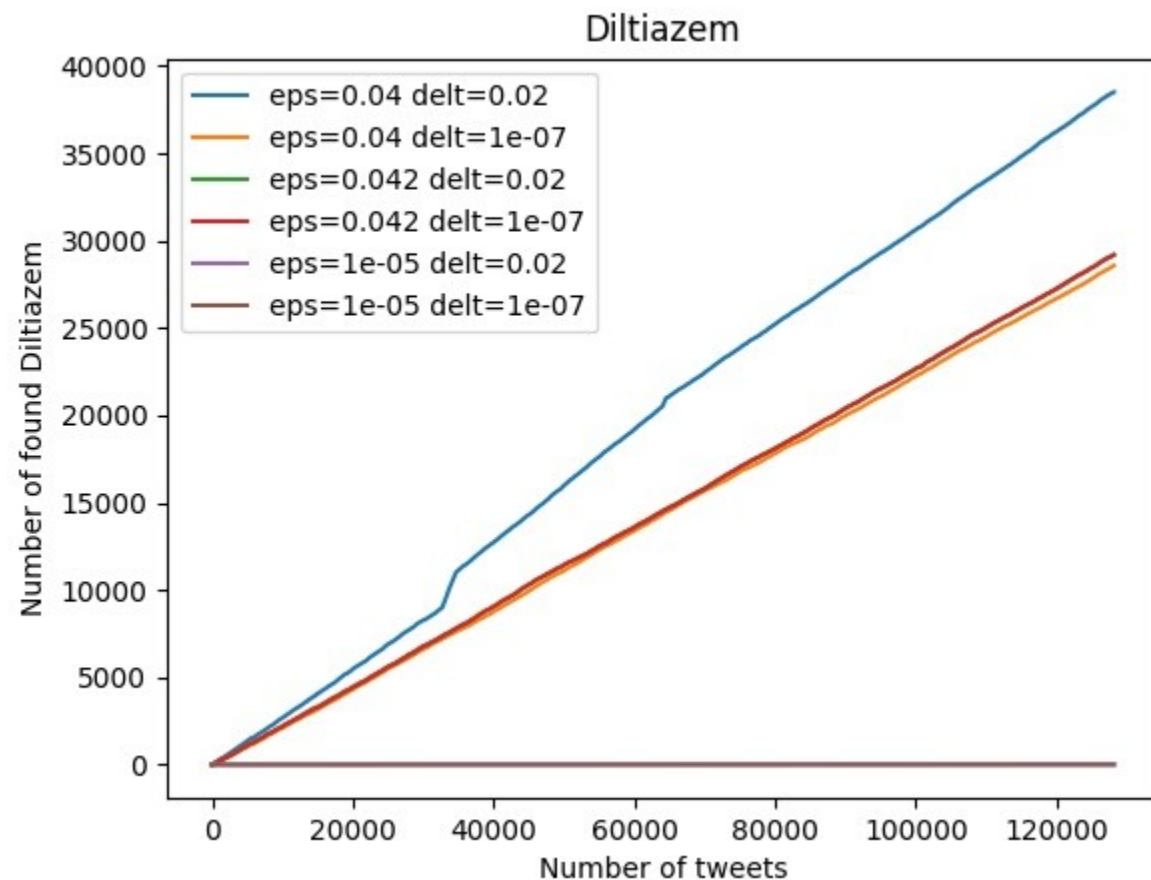
Caffeine



Trimethadione

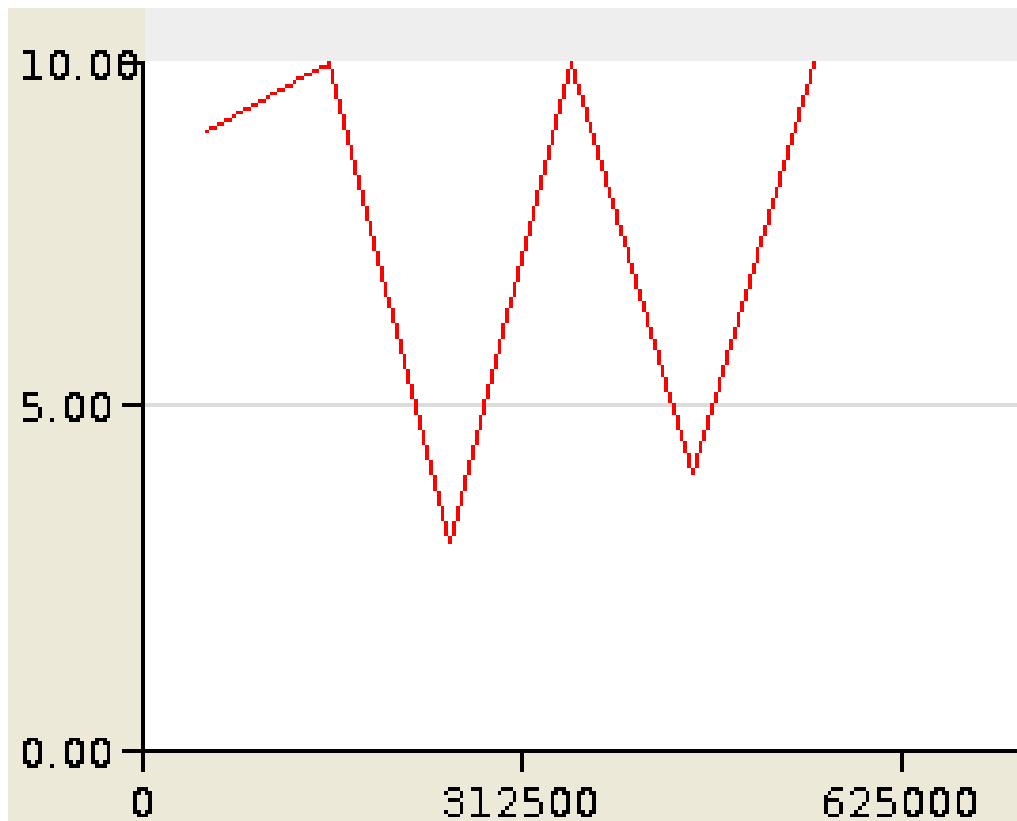


Diltiazem

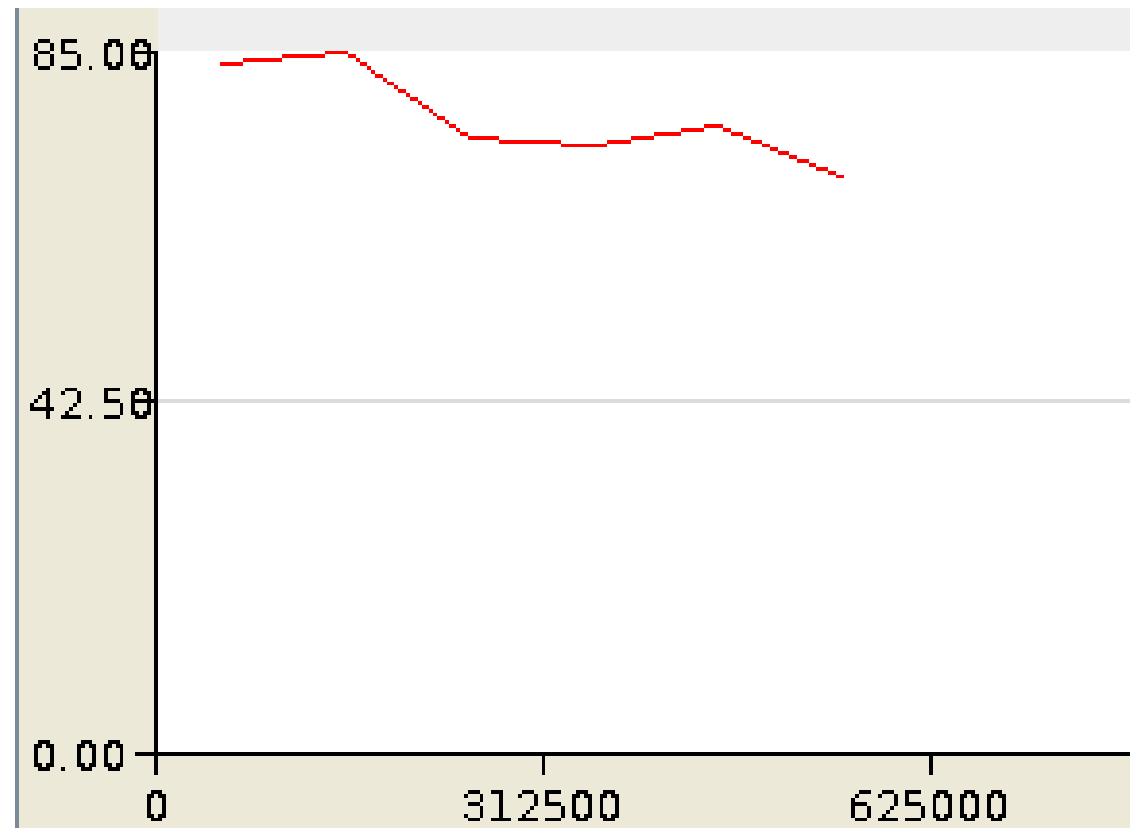


Stream Classification

CovtypeNorm Hoeffding Tree



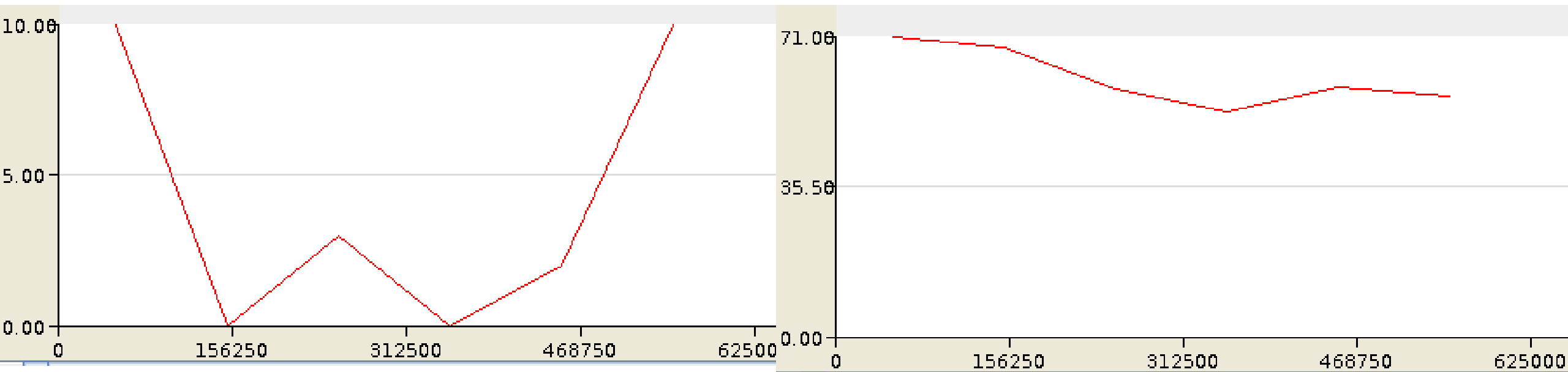
Window 10



Window 100000

Stream Classification

CovtypeNorm Naive Bayes



Window 10

Window 100000