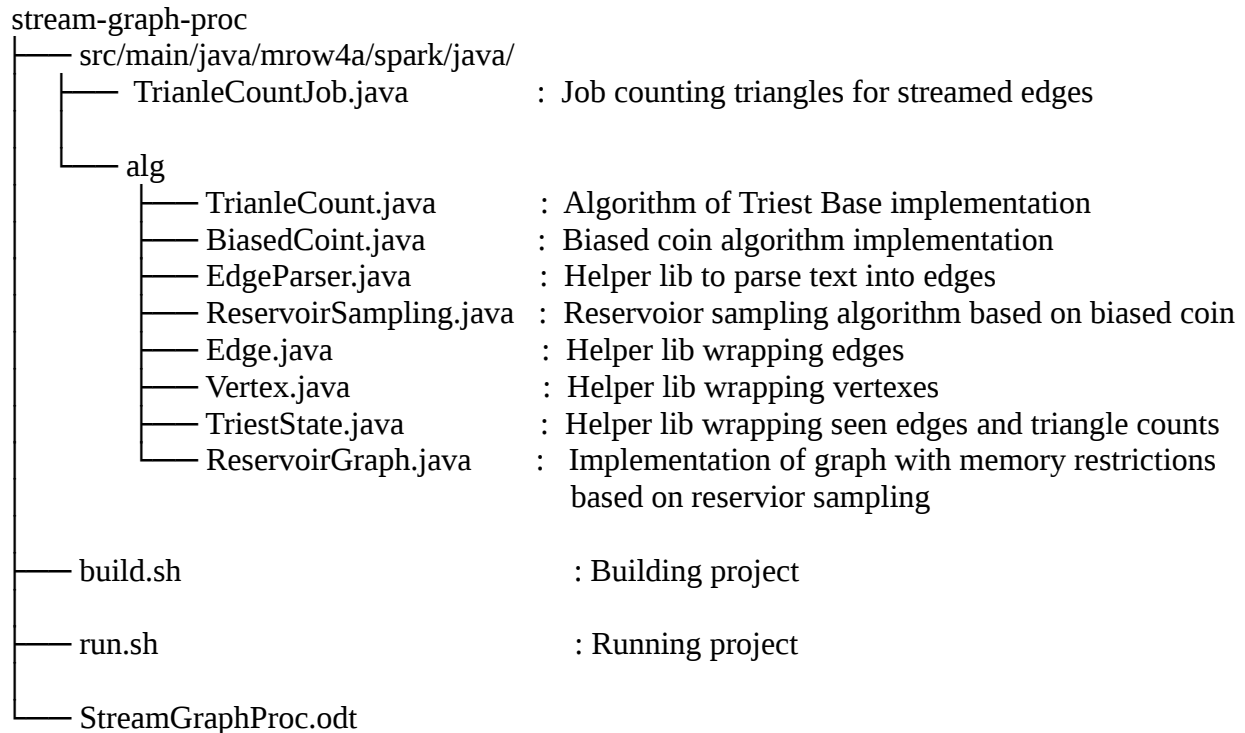


Data Mining 26.11.2017 – Piotr Mrowczynski – Assignment 3

The objective of this assignment was to allow triangle counting of huge graphs, where edges are being streamed continuously. The algorithm also cares about memory requirement, and uses probabilistic methods to store only M edges in memory and still allow accurate triangle count predictions.

Project structure



Building (Docker required for maven single assembly)

```
chmod +x stream-graph-proc/build.sh
chmod +x stream-graph-proc/run.sh
stream-graph-proc/build.sh
```

Running

```
stream-graph-proc/run.sh
```

Dataset: 16,000,000 edges

Triangle count exact: 7,515,023

Results:

Memory limit: 2,000,000 edges

Current triangle count estimate: 7,594,782 at t=16,518,947

Memory limit: 1,000,000 edges

Current triangle count estimate: 7,626,910 at t=16,518,947

Memory limit: 100,000 edges

Current triangle count estimate: 9,015,504 at t=16,518,947