

Algorithms for Dynamic Geometric Data Streams

Dissertation of Gereon Frahling

Oktober 2006

The increasing interconnectivity of modern computer systems generates huge amounts of data, which cannot be stored. In this thesis we develop data streaming algorithms, which process the data sequentially without storing. We only store a small sketch or summary of the data, and use it to approximately answer queries about the content of the data.

In particular we look at the model of dynamic geometric data streams. Here we are given a data stream consisting of m insert and delete operations of points from a d -dimensional space into a set M . We develop data stream algorithms, which use at most $\text{poly}(\log m)$ bits of memory:

First we show how to efficiently maintain a sample of M in dynamic data streams. Then we develop a coresets technique, which can be used to maintain $(1+\epsilon)$ -approximate k -Median, k -Means, and MaxCut clusterings of the dynamic point set M . We furthermore show many more applications of the coresets technique and use it to accelerate iterative clustering algorithms in practice.

In the last chapter we introduce data stream algorithms to approximately count minors like triangles or bipartite cliques in huge graphs given as data streams. They can be used to obtain structural information about social graphs and the webgraph.