

J. Skála: Clustering Geometric Data Streams

Data stream algorithms have been extensively studied in connection with databases and network statistics. However, there is not much research dedicated to geometric data streams. Since geometric models are growing larger and larger like those from Stanford's Michelangelo project (David – 28 mil. points, St. Matthew – 187 mil. points), data stream approach is becoming essential to process such models.

In my upcoming talk I would like to present a method for clustering geometric data in a streaming fashion. I will describe the so called facility location algorithm which is used when we don't know the exact number of clusters in data. I will then explain how the data stream clustering works, how it can be improved and how to set various parameters to get a nice clustering in reasonable time.