

Load balancing techniques for CHAOS

May 18, 2016

Author:
Vlad Ioan HAPRIAN

Supervisors:
Laurent BINDSCHAEDLER
Prof. Willy ZWAENEPOEL

1 Abstract

Chaos is a bulk synchronous processing system for graph analytics which scales from secondary storage to multiple machines in a cluster. It stripes graph data across all nodes in the cluster in order to achieve I/O balance. Instead of relying on expensive preprocessing of the input graph to achieve locality and load balance, Chaos opts instead for minimal partitioning and dynamic load balancing. The system splits the graph in equal vertex slices and their associated edges and assigns the resulting partitions to machines. As a result, power-law graphs partitioned in this manner suffer from significant imbalance between the edge sets of different partitions. Chaos sidesteps this issue by relying on work stealing to achieve computation balance at run-time but doing so comes at the cost of significant overhead.

This work explores alternate partitioning schemes that balance the computation and I/O between machines without incurring the full overhead of work stealing. In particular, we present a new partitioning scheme based on identical size edge partitions as well as different approaches inspired from other state-of-the-art graph processing systems.