GRAPHCHI: LARGE-SCALE GRAPH COMPUTATION ON JUST A PC
Aapo Kyrölä, Guy Blelloch (Carnegie Mellon University), Carlos Guestrin (University of Washington)

Contact author: akyrola@cs.cmu.edu

OSDI’12

DISK-BASED GRAPH COMPUTATION ON A BASIC PC / LAPTOP (OSDI’12)

• Developing distributed algorithms remains hard.
• Distributing graph computation is particularly challenging.
• Could we compute Big Graphs on just a single machine?

Natural graphs have random structure → How to eliminate random disk seeks?

Our Solution: Parallel Sliding Windows

Results: GRAPHCHI can solve as big problems as distributed systems, almost as fast

Future Work: GRAPHCHI on a big machine

Research question: How to compute multiple tasks on the same graph efficiently?

Background

- Computing recommendations for users requires executing graph algorithms with different settings for different groups → hundreds of tasks:
  - Languages
  - Countries
  - Customer segments
  - Interest groups, etc.

Idea: Use RAM for the algorithm state and let Parallel Sliding Windows process the graph from disk.

Early experiments: On a machine with 144 GB of RAM, compute 20 parallel personalized PageRanks on the current Twitter graph → 20x throughput of Hadoop running > 100 workers.