Graph Twiddling in a MapReduce World
Jonathan Cohen

Adv. MapReduce Algorithms winter term 09/10
HPI
Winter presentation II – implementation
Arvid Heise, Michael Leben
Use Case: Find Domains in Wikipedia

Graph Twiddling | Michael Leben, Arvid Heise | 12/15/2009
Process Overview

- Original Graph → Simplify → Simplified Graph
- Edges with Degrees → Augmentation → Edges that form Triangles
- Find Trusses → Edges that are within trusses
- Find Components
Process: Finding Triangles

- Insolvency
- Arcandor
- Karstadt
- Quelle
- GM
- Car
- Opel
- Audi
- Short-time working
A K-Truss is a subgraph, in that each edge is part of k-2 triangles within the truss.
A k-truss is a subgraph, such that each edge is reinforced by at least \( k - 2 \) pairs of edges making a triangle with that edge.

A K-Truss is a subgraph, in that each edge is part of \( k - 2 \) triangles within the truss.
Figure 8. Trusses of a graph. Each truss has a randomly assigned color: (a) 3-trusses, (b) 4-trusses, and (c) 5-trusses. Vertices and edges not in trusses are black; such vertices are also hollow.
Our practical evaluation

- Cohen used Social Networks
  - More or less equal degrees

- Dbpedia data: links between articles are edges
  - Few nodes are extremely central
  - Most are very isolated

- Examples from our sample data:
  - USA (Degree of 88,000)
  - France (Degree of 33,000)
  - 2008 (Degree of 20,000)
High vertex degree leads to huge number of triads

Combination of any pair of neighbors
Solution of Triad Problem

- Before: each potential triangle part (triad) traverses the cluster

- Solution with "distributed cache":
  - each reducer accesses the complete edge file
First results

- Sample data contains x% of the vertices of the complete dataset (900 Mbyte)
  - “40%” (150 Mbyte) 5617 vertices in 41 9-Trusses
  - 4869 vertices in one garbage cluster
  - “30%” (83 Mbyte) 1389 vertices in 26 9-Trusses
  - but one garbage cluster contains 1016 vertices

- the bigger K is, the smaller and fewer clusters become
- What is the best cluster size?