GraphTwist: Fast Iterative Graph Computation with Two-tier Optimizations

Yang Zhou, Ling Liu, Kisung Lee, Qi Zhang, Calton Pu, Karsten Schwan
College of Computing, Georgia Institute of Technology

Big Graphs are Everywhere

Popular graph datasets in current literature

	n (vertices in millions)	m (edges in millions)	size
AS-Skitter	1.7	11	142 MB
LJ	4.8	69	337.2 MB
USRD	24	58	586.7 MB
BTC	165	773	5.3 GB
WebUK	106	1877	8.6 GB
Twitter	42	1470	24 GB
YahooWeb	1413	6636	120 GB

[Paul Burkhardt, Chris Waring 2013]

Challenges of Big Graphs

- Graph size v.s. limited resource
- High-degree vertices
- Skewed vertex degree distribution
- Skewed edge weight distribution

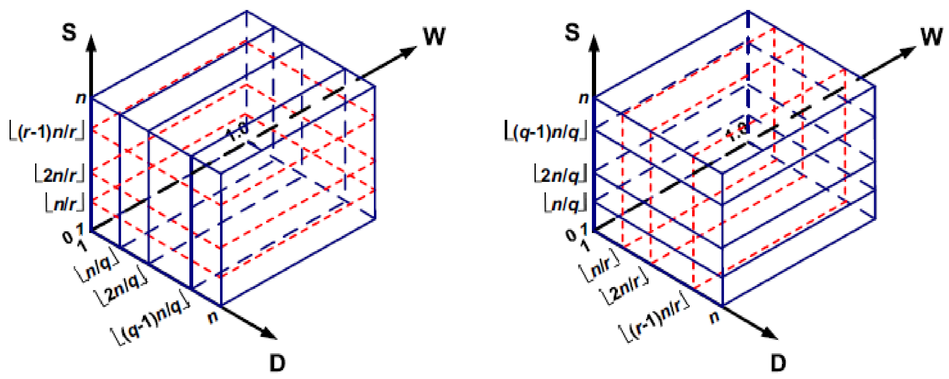
Real-world Big Graphs

Graph	Type	#Vertices	#Edges	AvgDeg	MaxIn	MaxOut
Yahoo	directed	1.4B	6.6B	4.7	7.6M	2.5K
uk-union	directed	133.6M	5.5B	41.22	6.4M	22.4K
uk-2007-05	directed	105.9M	3.7B	35.31	975.4K	15.4K
Twitter	directed	41.7M	1.5B	35.25	770.1K	3.0M
Facebook	undirected	5.2M	47.2M	18.04	1.1K	1.1K
DBLPS	undirected	1.3M	32.0M	40.67	1.7K	1.7K
DBLPM	undirected	0.96M	10.1M	21.12	1.0K	1.0K
Last.fm	undirected	2.5M	42.8M	34.23	33.2K	33.2K

Challenges of Graph Processing Systems

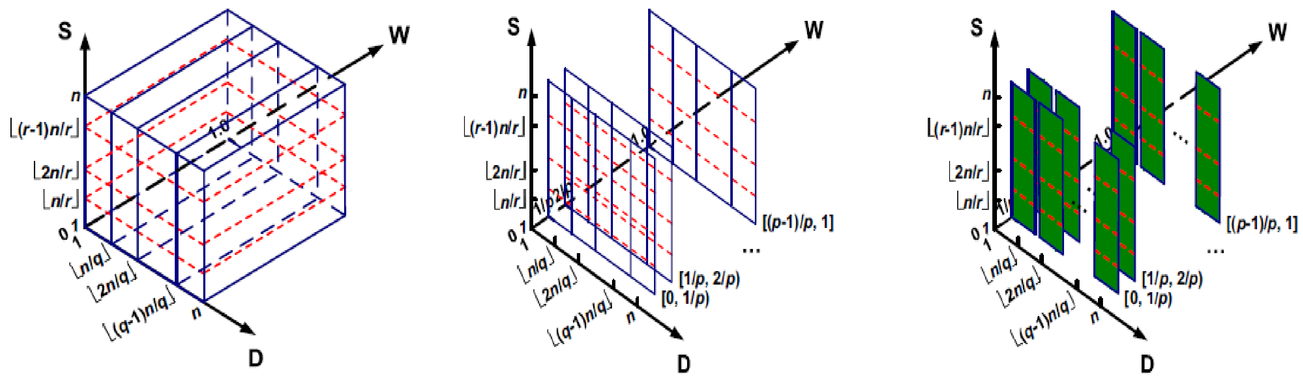
- Diverse types of processed graphs
- Different kinds of graph applications
- Random access
- Workload imbalance
- Exploring graph utility-aware pruning

Modeling a Graph as a 3D Cube



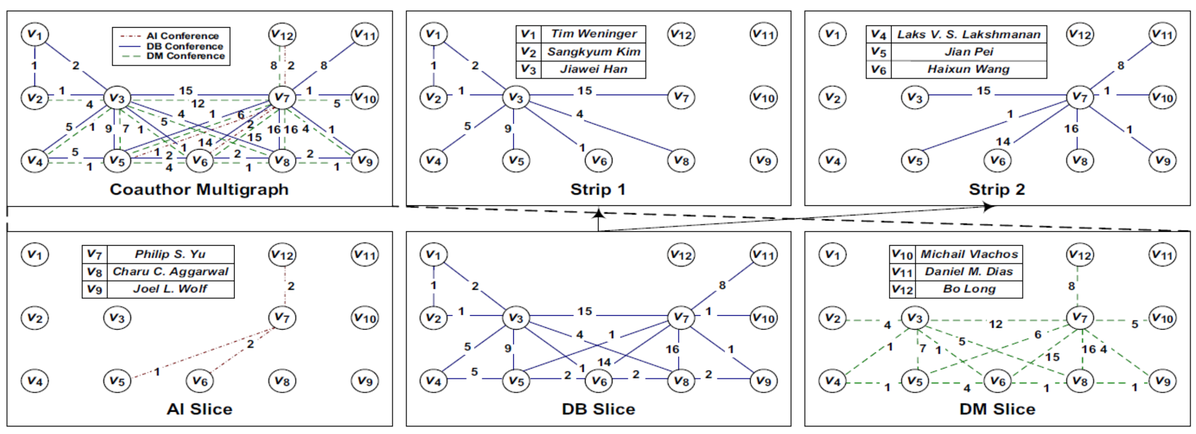
(a) In-edge Cube (b) Out-edge Cube

Hierarchical Graph Parallel Abstractions

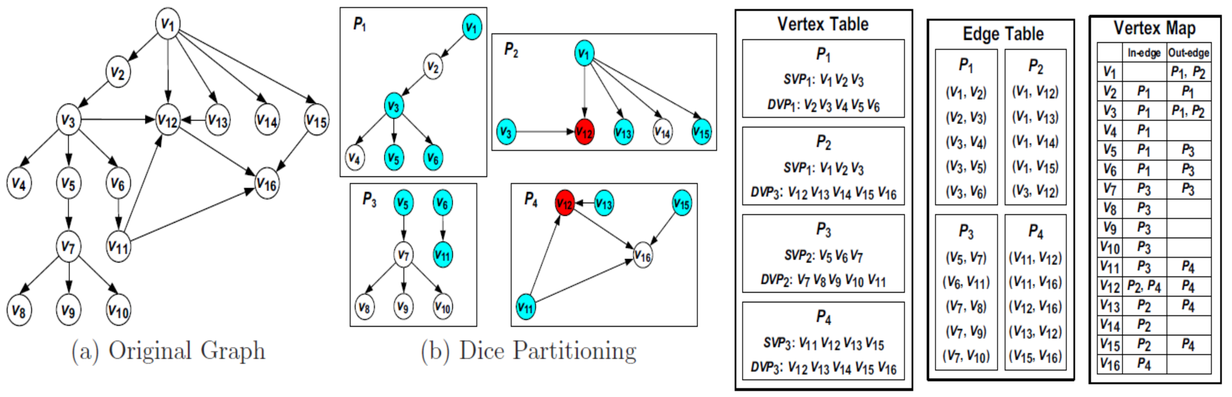


(a) In-edge Cube (b) In-edge Slice (c) In-edge Strip

Slice Partitioning and Strip Partitioning

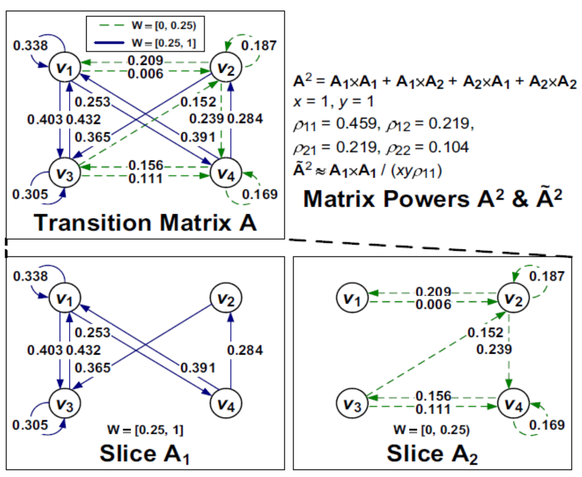


Dice Partitioning

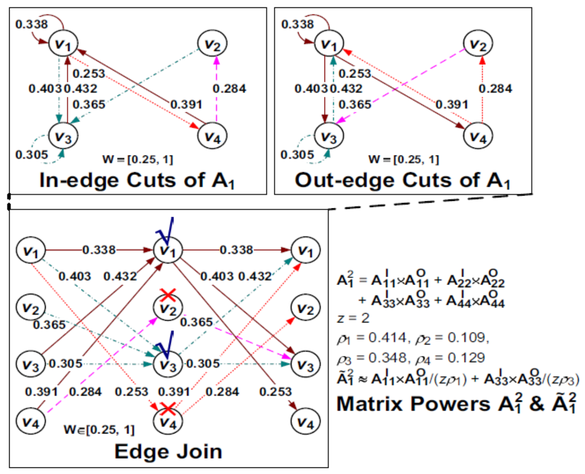


(a) Original Graph (b) Dice Partitioning

Slice Pruning and Cut Pruning

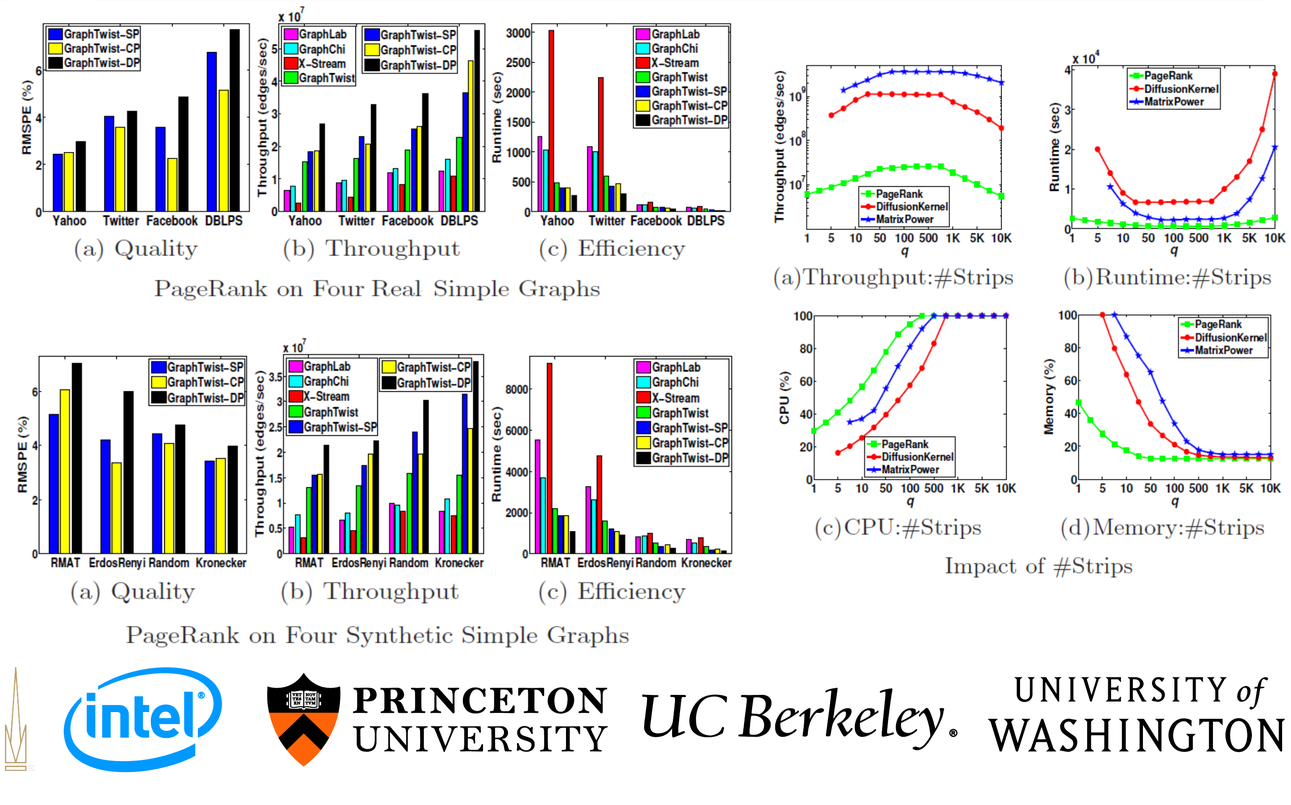


Matrix Power with Slice Pruning



Matrix Power with Cut Pruning

Experimental Evaluation



(a) Quality (b) Throughput (c) Efficiency

PageRank on Four Real Simple Graphs

(a)Throughput:#Strips (b)Runtime:#Strips (c)CPU:#Strips (d)Memory:#Strips

Impact of #Strips

(a) Quality (b) Throughput (c) Efficiency

PageRank on Four Synthetic Simple Graphs

Intel Science & Technology Center for Cloud Computing

Carnegie Mellon University · Georgia Tech · intel · Princeton University · UC Berkeley · University of Washington