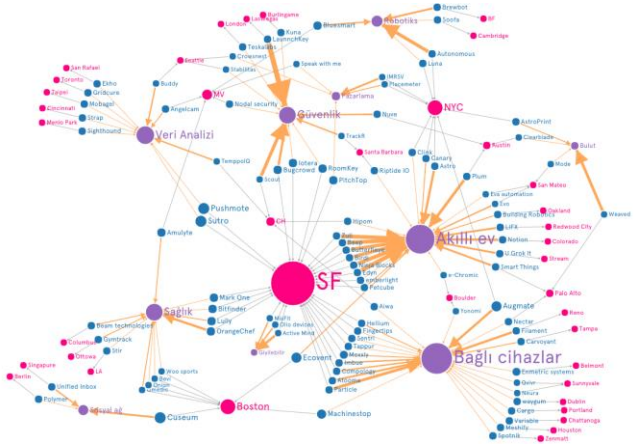
CRIS Center for Research in Intelligent Storage

Large-scale graph processing with emerging storage devices

Huibing Dong

GRAPHS ARE EVERYWHERE



ALGORITHMS

Center for Research in Intelligent Storage

Algorithm	Description
PageRank	Ranking nodes based on their incoming/outgoing edges
Breath-First Search	Graph traversal based on breath
Shortest Path	Finding the path between 2 nodes where the sum of weights is minimal
Pattern Matching	Finding certain structures (e.g. path, star)
Triangle Count	Counting the number of triangles
Connected Component	Finding the subgraphs in which any two vertices are connected

ALGORITHMS

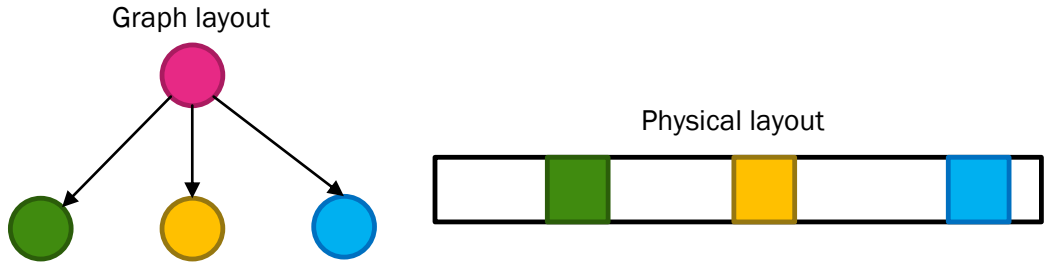
Algorithm	Description
PageRank	Ranking nodes based on their incoming/outgoing edges
Breath-First Search	Graph traversal based on breath
Shortest Path	Finding the path between 2 nodes where the sum of weights is minimal
Pattern Matching	Finding certain structures (e.g. path, star)
Triangle Count	Counting the number of triangles
Connected Component	Finding the subgraphs in which any two vertices are connected

$$\mathrm{PR}(v^t) = 1 - d + d \times \sum_{\mathrm{inedges}(v)} \mathrm{PR}(u^{t-1}) \;/\; |\mathrm{outedges}(u)|$$

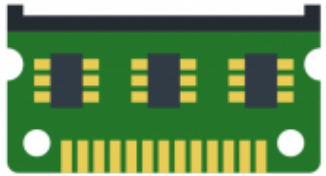
Iterative

Neighbors matter

CHALLENGES



Irregular access



Expensive DRAM

GRAPH PROCESSING PLATFORMS

Center for Research in Intelligent Storage

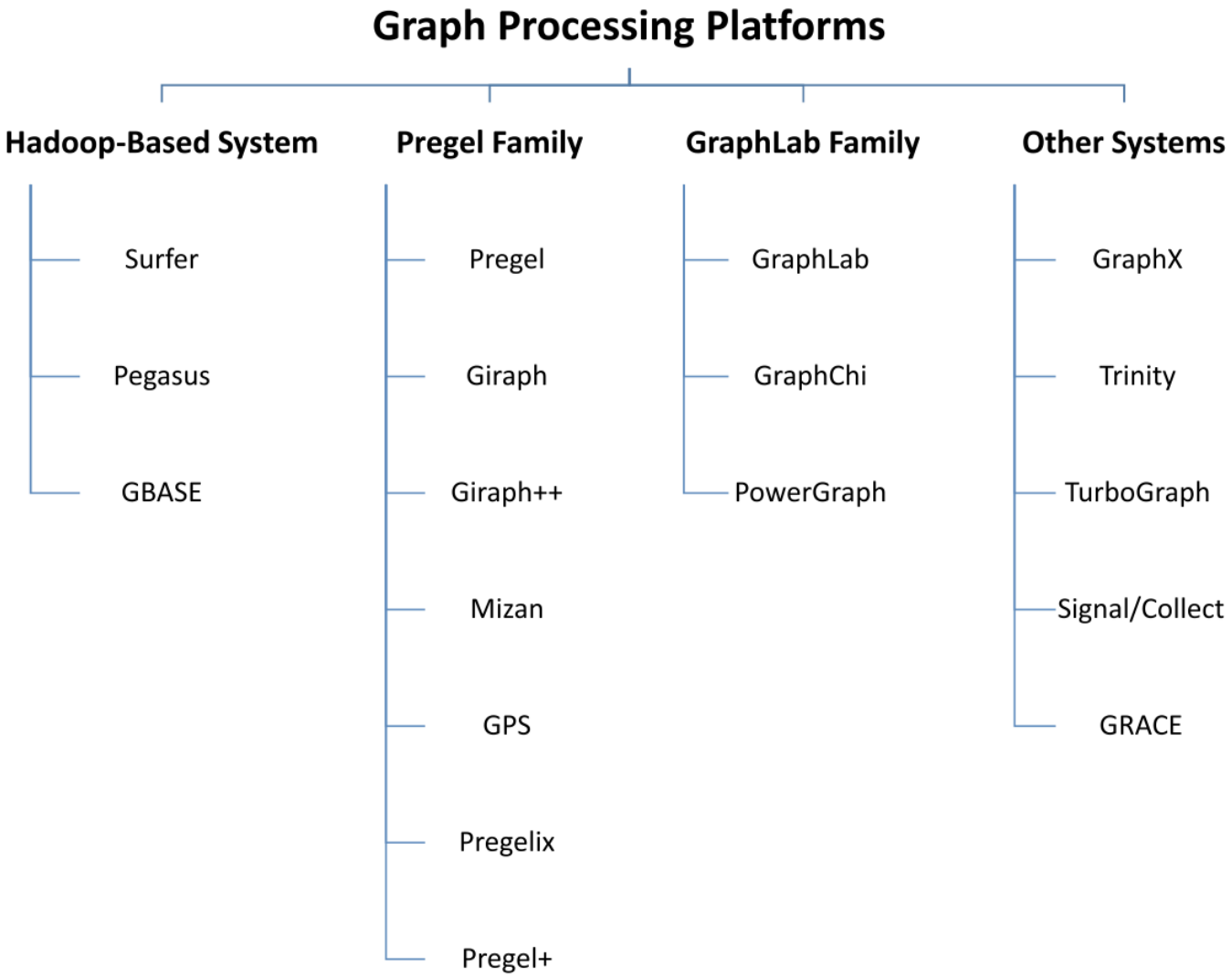
Shared-memory	Distributed	External-memory
Memory	Memory	Memory + Storage
Single machine	Clusters	Single machine
Limited graph size	Costly	Larger size + cost efficient
2013-Ligra, 2014-GraphX, 2015-Chaos	2010-Pregel, 2010-Graphlab, 2012-PowerGraph, 2014-GraphX, 2014-Blogel, 2015-Chaos, 2016-Gemini	2012-GraphChi, 2013-X-stream, 2013-TurboGraph, 2015-GridGraph, 2015-FlashGraph, 2017-Graphene, 2017-Mosaic, 2018-GraFBoost, 2019-GraphOne, 2019-Lumos

**Blue: Semi-external systems*

Others architecture

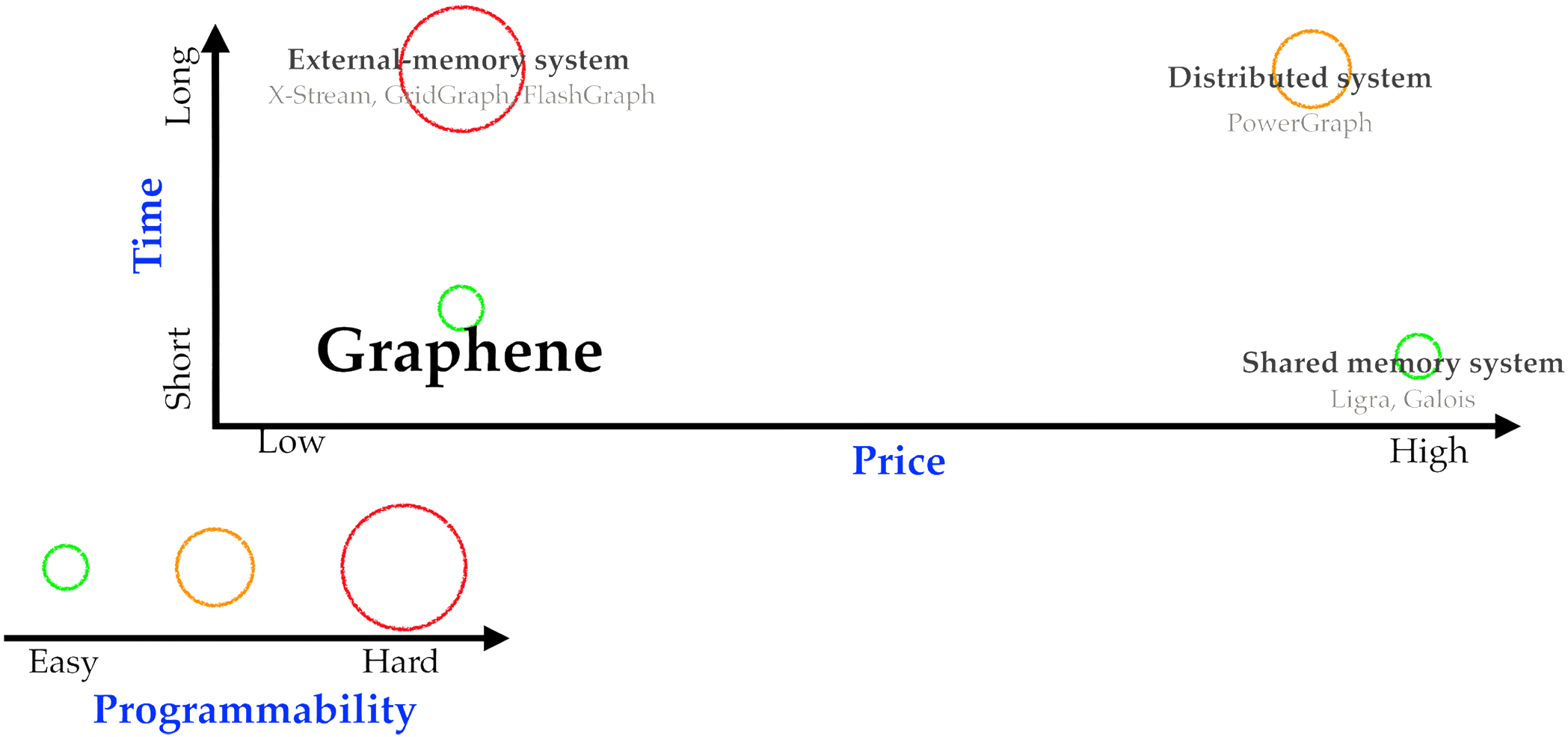
- Graph-optimized database: Neo4j

PLATFORM FAMILY



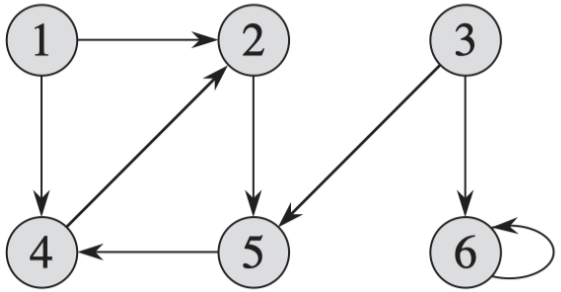
Large scale graph processing systems: survey and an experimental evaluation, Omar Batarfi et. al, Cluster Computing'15

EXTERNAL/OUT-OF-CORE SYSTEM

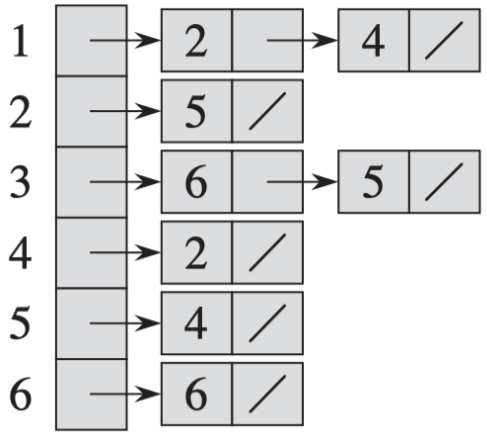


Graphene: Fine-Grained IO Management for Graph Computing, Hang Liu and H. Howie Huang, The George Washington University, FAST'17

GRAPH REPRESENTATION



Graph



Adjacency list

	1	2	3	4	5	6
1	0	1	0	1	0	0
2	0	0	0	0	1	0
3	0	0	0	0	1	1
4	0	1	0	0	0	0
5	0	0	0	1	0	0
6	0	0	0	0	0	1

Adjacency matrix

- ✓ Sparse matrix
- ✓ Much less storage space needed

Storage format: Compressed Sparse Column (CSC) & Compressed Sparse Row (CSR) files

Introduction to Algorithm, 3rd edition, Page 590, Thomas H. Cormen Charles E. Leiserson Ronald L. Rivest Clifford Stein

EMERGING STORAGE DEVICES

Center for Research in Intelligent Storage

Semi-external systems

- Vertex data
  - In the main memory
  - Fine-grained accesses, byte-addressable
- Edge data
  - On the secondary storage
  - Coarser accesses

External systems

- Even the vertex data itself is too large
- Both Vertex & Edge data on the secondary storage

HDD 

Slow
Random access is a disaster

SSD 

Continue to scale
Large capacity with lower latency

MODEL

Programming Model

Vertex-centric

Edge-centric

*IO-centric

Execution Model

Bulk synchronous

Asynchronous

PAPER LIST

Center for Research in Intelligent Storage

External systems

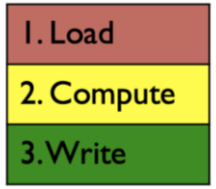
OSDI'12	GraphChi	Parallel Sliding Window	8GB DRAM + 256GB SSD + 750GB HD
KDD'13	Turbograph		12GB DRAM + 2x 512GB SSD
SOAP'13	X-Stream	Edge-centric	64GB DRAM + 2x 200GB
ICDE'15	Venus		
ATC'15	GridGraph	2D partition	30.5GB DRAM + 2x 2TB HDD/1800 GB SSD
FAST'15	FlashGraph	semi-external; merges I/O	512GB DRAM + 15x OCZ Vertex 4 SSD
SC'16	G-Store		
FAST'17	Graphene	Fine-grained IO	128GB DRAM + 16x 500GB SSD
EuroSys'17	Mosaic		
ISCA'18	GraFBoost		48GB DRAM + Xilinx VC707 FPGA + 2x 512GB SSD
FAST'19	GraphOne		512GB DRAM + 512GB SSD
ATC' 19	Lumos	Dependency-Driven	64GB DRAM + 4x 1.9TB SSD

PAPER LIST

Center for Research in Intelligent Storage

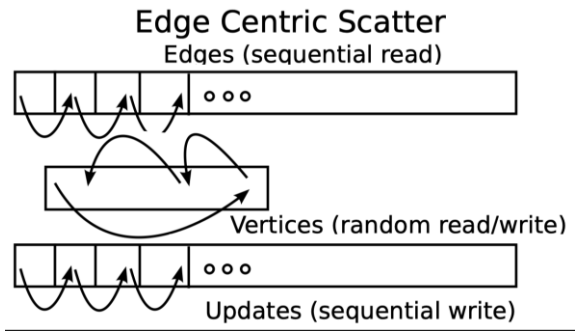
GraphChi

Parallel Sliding Window
- one sub-graph at a time
- 3 phases



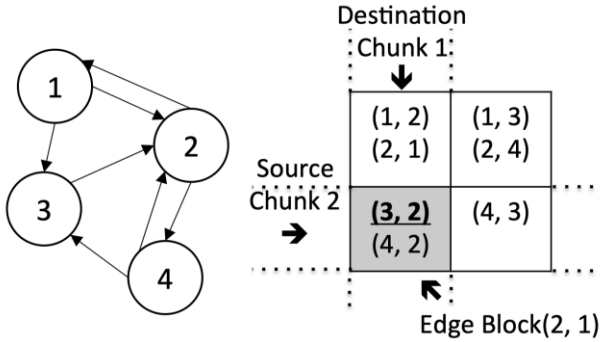
X-stream

Edge-centric



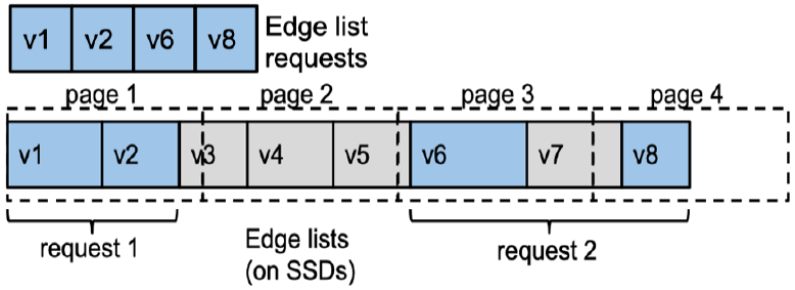
GridGraph

2D partitioning

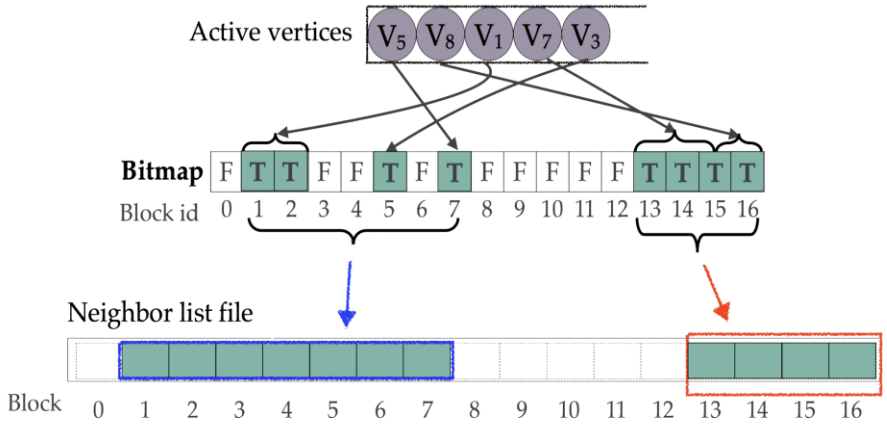


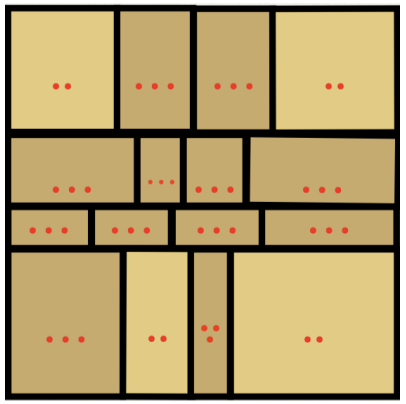
FlashGraph

- Semi-external
- Merge edges I/O requests



Graphene





POTENTIAL RESEARCH PROBLEMS

CRIS Center for Research in Intelligent Storage

- Programming model optimization
- Execution model support
- Partitioning
- Serializing
- Emerging Storage devices selection:
  - Zone-named space SSD
    - ZAC/ZBC
  - Open-channel SSD
  - Hybrid devices
  - Storage tiers