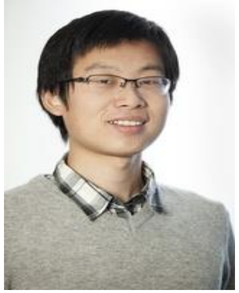
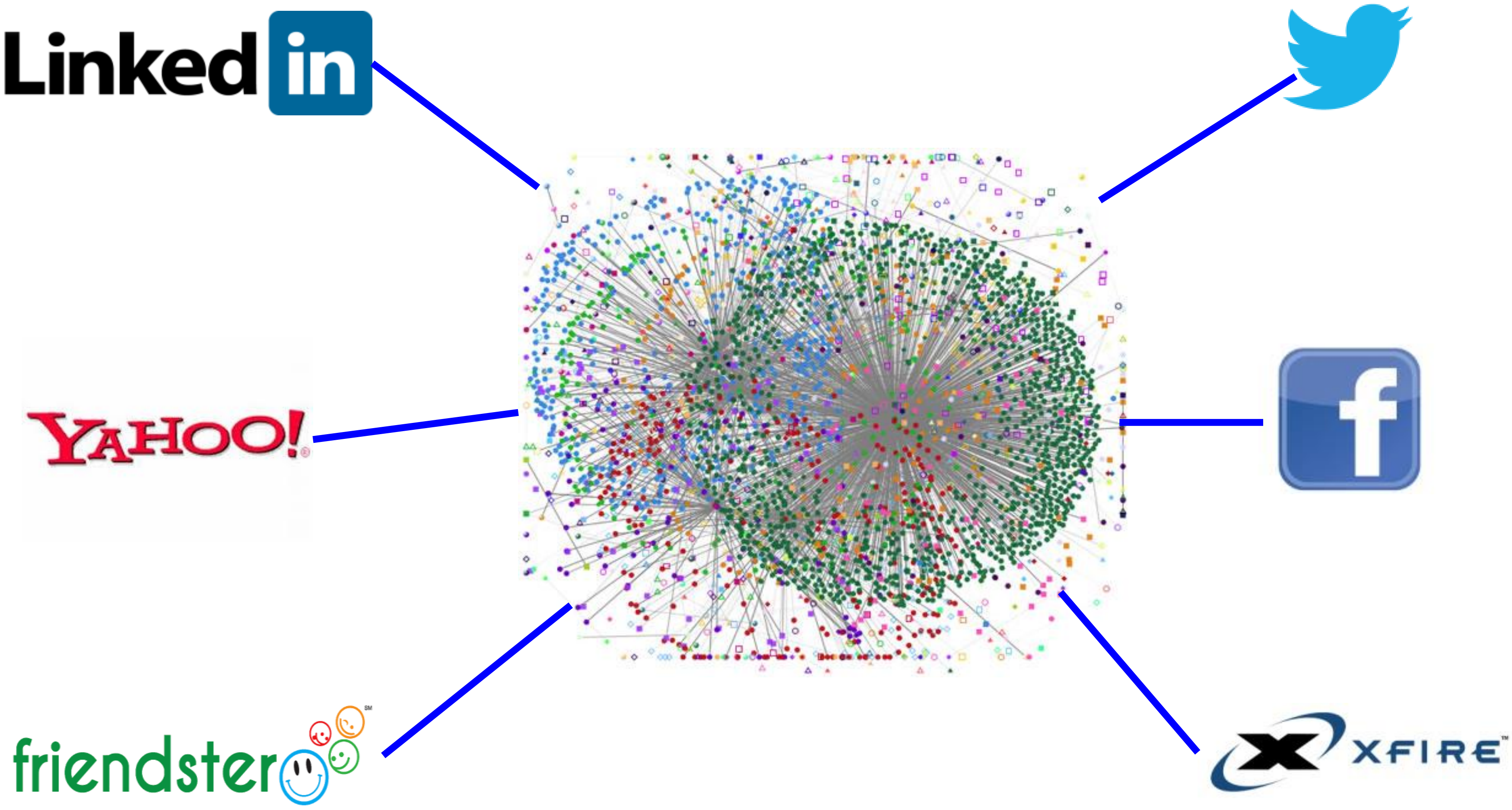
Benchmarking Graph-Processing Platforms: A Vision (A SPEC Research Group Process)



Yong Guo, Ana Lucia Varbanescu, Alexandru Iosup, Claudio Martella, Theodore L. Willke

Delft University of Technology, University of Amsterdam,
VU University Amsterdam, Intel Corporation

The data deluge: large-scale graphs



Platform diversity

- Platform: the combined hardware, software, and programming system that is being used to complete a graph processing task.

Neo4j
the graph database
hadoop
APACHE GIRAPH
HyperGraphDB
YARN
GraphLab
Trinity
StratoSphere
Above the Clouds
PROJECT PEGASUS

What is the performance of these platforms?

Performance Metrics

Graph Diversity

Algorithm Diversity

- Graph500
  - Single application (BFS), Single class of synthetic datasets
- Few existing platform-centric comparative studies
  - Prove the superiority of a given system, limited set of metrics

Our vision: a benchmarking suite for graph processing across all platforms

A Call to Arms

- Defining workloads
- Understanding the metrics, datasets, and algorithms used in practice: fill in our survey http://goo.gl/TJwkTg
- Evaluating and reporting on various platforms

Join us within the SPEC RG Cloud Working Group

http://research.spec.org/working-groups/rg-cloud-working-group.html

Our Vision for Benchmarking Graph-Processing Platforms

- Methodological challenges
  1. Evaluation process
  2. Selection and design of performance metrics
  3. Dataset selection
  4. Algorithm coverage

- Practical challenges
  5. Scalability of evaluation, selection processes
  6. Portability
  7. Result reporting

Y. Guo, A. L. Varbanescu, A. Iosup, C. Martella, T. L. Willke: Benchmarking graph-processing platforms: a vision. ICPE 2014: 289-292

Selection and Design of Performance Metrics for Graph Processing

- Raw processing power
  - Execution time
  - Actual computation time
  - Edges/Vertices per second
- Resource utilization
  - CPU, memory, network

- Scalability
  - Horizontal vs. vertical
  - Strong vs. weak
- Overhead
  - Data ingestion time
  - Overhead time

- Elasticity (?)

Challenge 2. Metric selection

Dataset Selection: Application Domains

- Number of vertices, edges, link density, size, directivity, etc.

	Graphs	#V	#E	d	$\bar{D}$	Directivity
🟨 G1	Amazon	262,111	1,234,877	1.8	4.7	directed
🟨 G2	WikiTalk	2,388,953	5,018,445	0.1	2.1	directed
🟦 G3	KGS	293,290	16,558,839	38.5	112.9	undirected
🟨 G4	Citation	3,764,117	16,511,742	0.1	4.4	directed
🟦 G5	DotaLeague	61,171	50,870,316	2,719.0	1,663.2	undirected
🟩 G6	Synth	2,394,536	64,152,015	2.2	53.6	undirected
🟨 G7	Friendster	65,608,366	1,806,067,135	0.1	55.1	undirected

SNAP 🟨

GRAPH 500 🟩

The Game Trace Archive 🟦

https://snap.stanford.edu/ http://www.graph500.org/ http://gta.st.ewi.tudelft.nl/

Challenge 3. Dataset selection

Graph-Processing Algorithms

- Literature survey of of metrics, datasets, and algorithms
  - 10 top research conferences: SIGMOD, VLDB, HPDC …
  - Key word: graph processing, social network
  - 2009–2013, 124 articles

Class	Examples	%
Graph Statistics	Diameter, PageRank	16.1
Graph Traversal	BFS, SSSP, DFS	46.3
Connected Component	Reachability, BiCC	13.4
Community Detection	Clustering, Nearest Neighbor	5.4
Graph Evolution	Forest Fire Model, PAM	4.0
Other	Sampling, Partitioning	14.8

Y. Guo, M. Biczak, A. L. Varbanescu, A. Iosup, C. Martella, and T. L. Willke. How Well do Graph-Processing Platforms Perform? An Empirical Performance Evaluation and Analysis,IPDPS'14.

Platforms we have evaluated

Challenge 6. Portability

- Distributed or non-distributed
- Graph-specific or generic

hadoop

YARN

StratoSphere
Above the Clouds

Distributed (Generic)

APACHE GIRAPH

GraphLab

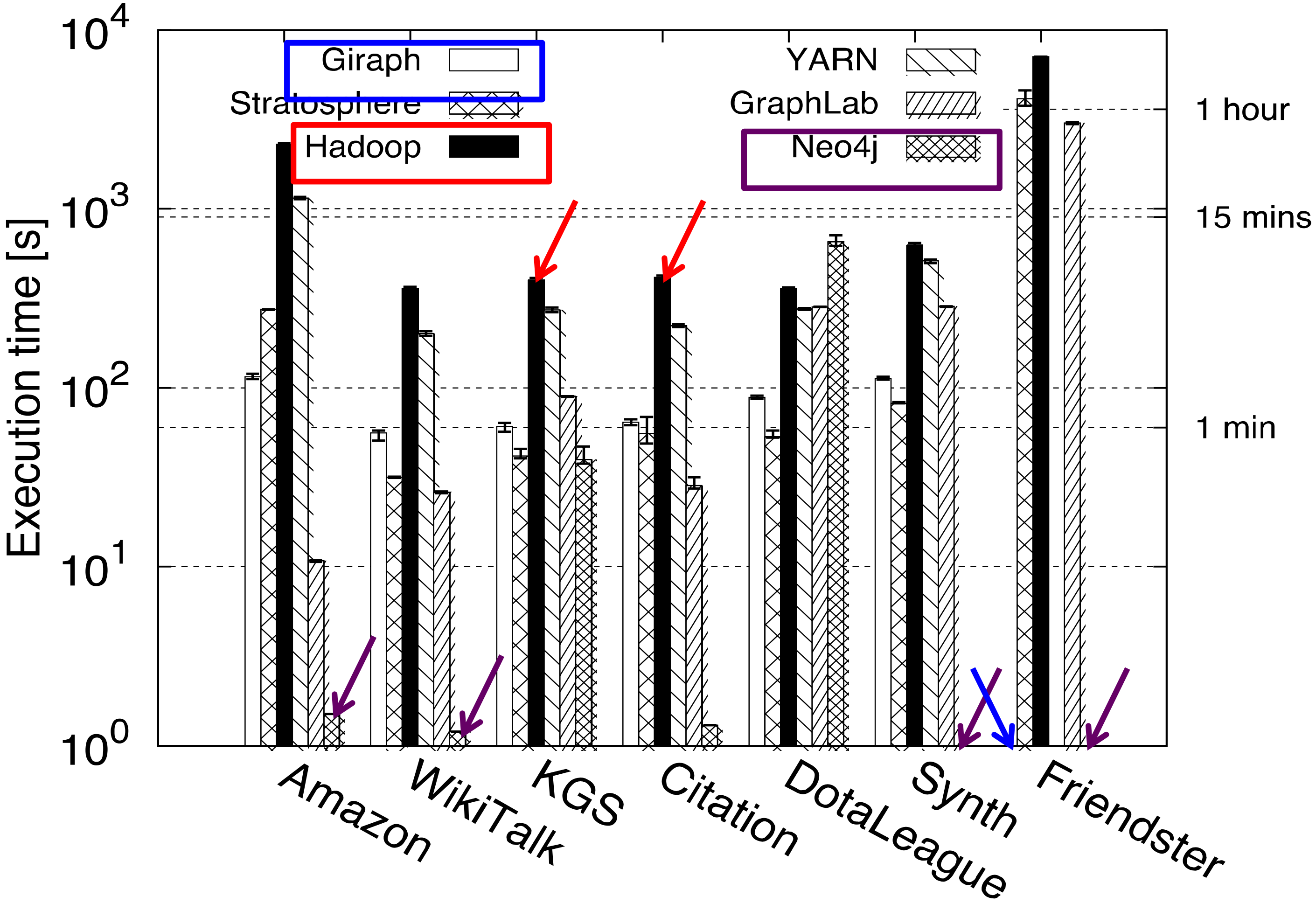
Distributed (Graph-specific)

Neo4j
the graph database

Non-distributed (Graph-specific)

Y. Guo, M. Biczak, A. L. Varbanescu, A. Iosup, C. Martella, and T. L. Willke. How Well do Graph-Processing Platforms Perform? An Empirical Performance Evaluation and Analysis,IPDPS'14.

BFS: results for all platforms, all graphs



Challenge 7. Result reporting

Y. Guo, M. Biczak, A. L. Varbanescu, A. Iosup, C. Martella, and T. L. Willke. How Well do Graph-Processing Platforms Perform? An Empirical Performance Evaluation and Analysis,IPDPS'14.

Key Findings From the Study of 6 Platforms

- Performance is function of (Dataset, Algorithm, Platform, Deployment)
  - Previous performance studies may lead to tunnel vision

- Platforms have their own drawbacks (crashes, long execution time, tuning, etc.)
  - Best-performing is not only low response time
  - Ease-of-use of a platform is very important

- Some platforms can scale up reasonably with cluster size (horizontally) or number of cores (vertically)
  - Strong vs weak scaling still a challenge—workload scaling tricky

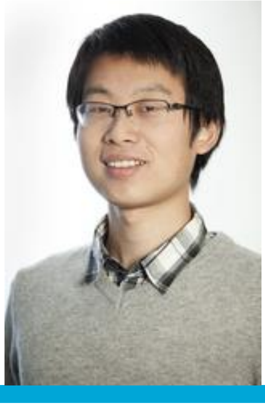
Y. Guo, M. Biczak, A. L. Varbanescu, A. Iosup, C. Martella, and T. L. Willke. How Well do Graph-Processing Platforms Perform? An Empirical Performance Evaluation and Analysis,IPDPS'14.

Thank you for your attention! Comments? Questions? Suggestions?

- Join us, join the SPEC Research Group
- Fill in our survey on Big Data Use Cases http://goo.gl/TJwkTg

- Ask about other results

Yong Guo
Yong.Guo@tudelft.nl
http://www.pds.ewi.tudelft.nl/yong/
Parallel and Distributed Systems Group
Delft University of Technology



Experimental Setup

- DAS4: a multi-cluster Dutch grid/cloud
  - Intel Xeon 2.4 GHz CPU (dual quad-core, 12 MB cache)
  - Memory 24 GB
  - 1 Gbit/s Ethernet network

- Size
  - Most experiments take 20 working machines
  - Up to 50 working machines

- HDFS used here as distributed file systems

Y. Guo, M. Biczak, A. L. Varbanescu, A. Iosup, C. Martella, T. L. Willke, *How Well do Graph-Processing Platforms Perform? An Empirical Performance Evaluation and Analysis*, IPDPS 2014

Design of a Survey

- Survey on Big Data Use Cases

  http://goo.gl/TJwkTg

- In total, 33 questions
- 15-20 minutes avg to finish the survey, at most 40 minutes

- Includes 4 major parts
  - Institution profile—Who processes Big Data?
  - System infrastructure—On what infrastructure?
  - Data collection, storage, and process—Which workloads, metrics, etc.? What are the challenges?
  - Framework stacks—Who do they rely on?

Our method: A benchmark suite

- Identifying the performance aspects and metrics of interest
- Defining and selecting representative datasets and algorithms
- Implementing, configuring, and executing the tests
- Analyzing the results

Selection of algorithms

A1: General Statistics (STATS: # vertices and edges, LCC)

- Single step, low processing, decision-making

A2: Breadth First Search (BFS)

- Iterative, low processing, building block

A3: Connected Component (CONN)

- Iterative, medium processing, building block

A4: Community Detection (CD)

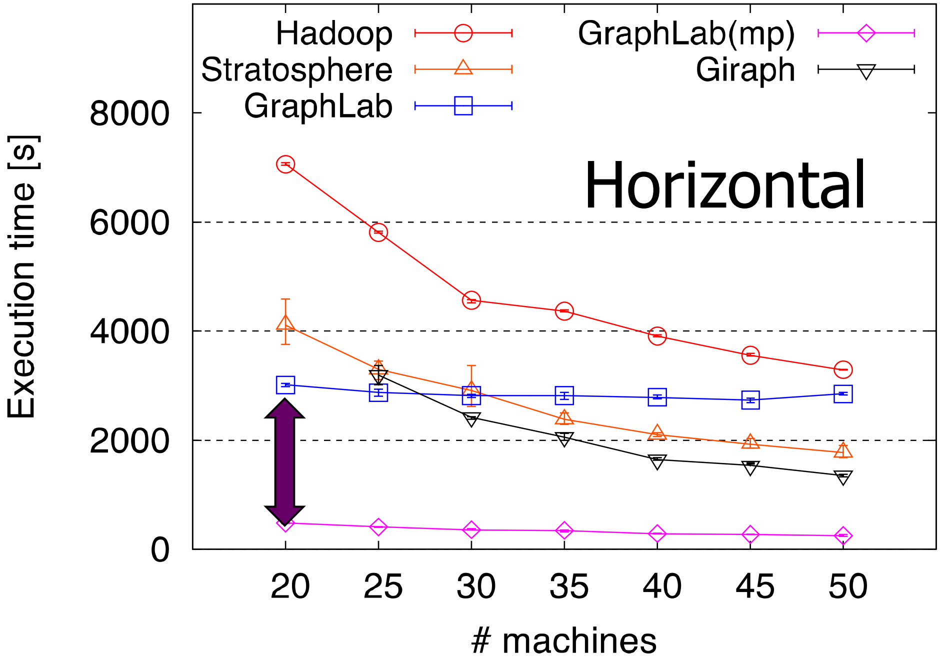
- Iterative, medium or high processing, social network

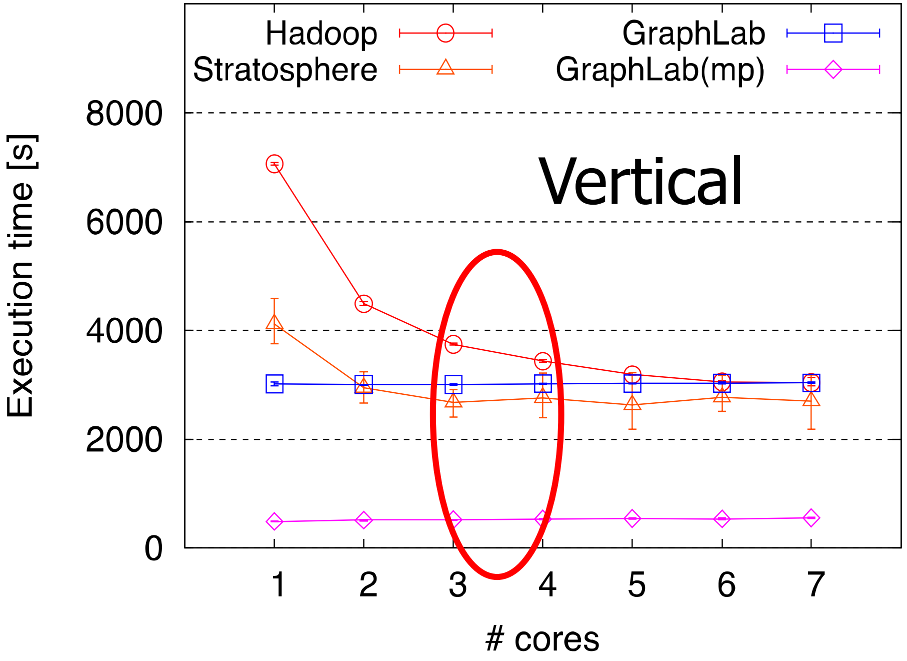
A5: Graph Evolution (EVO)

- Iterative (multi-level), high processing, prediction

Challenge 4. Algorithm selection

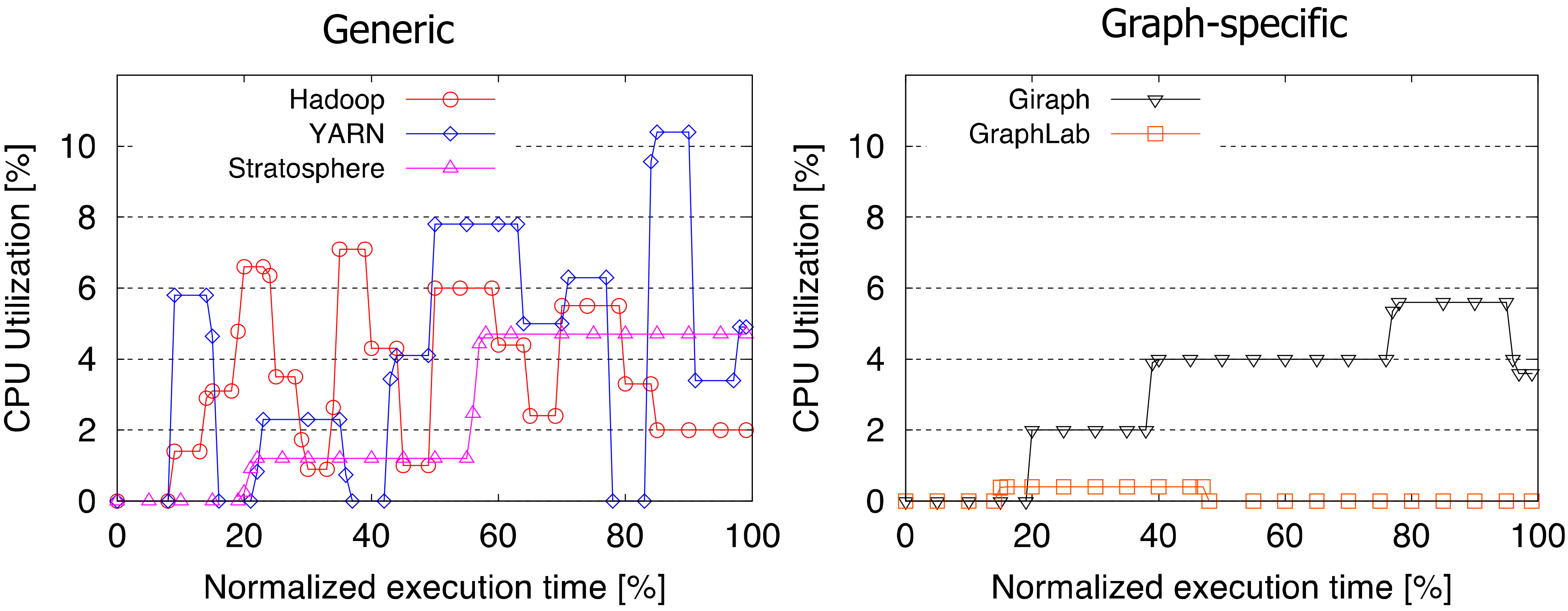
Scalability: BFS on Friendster





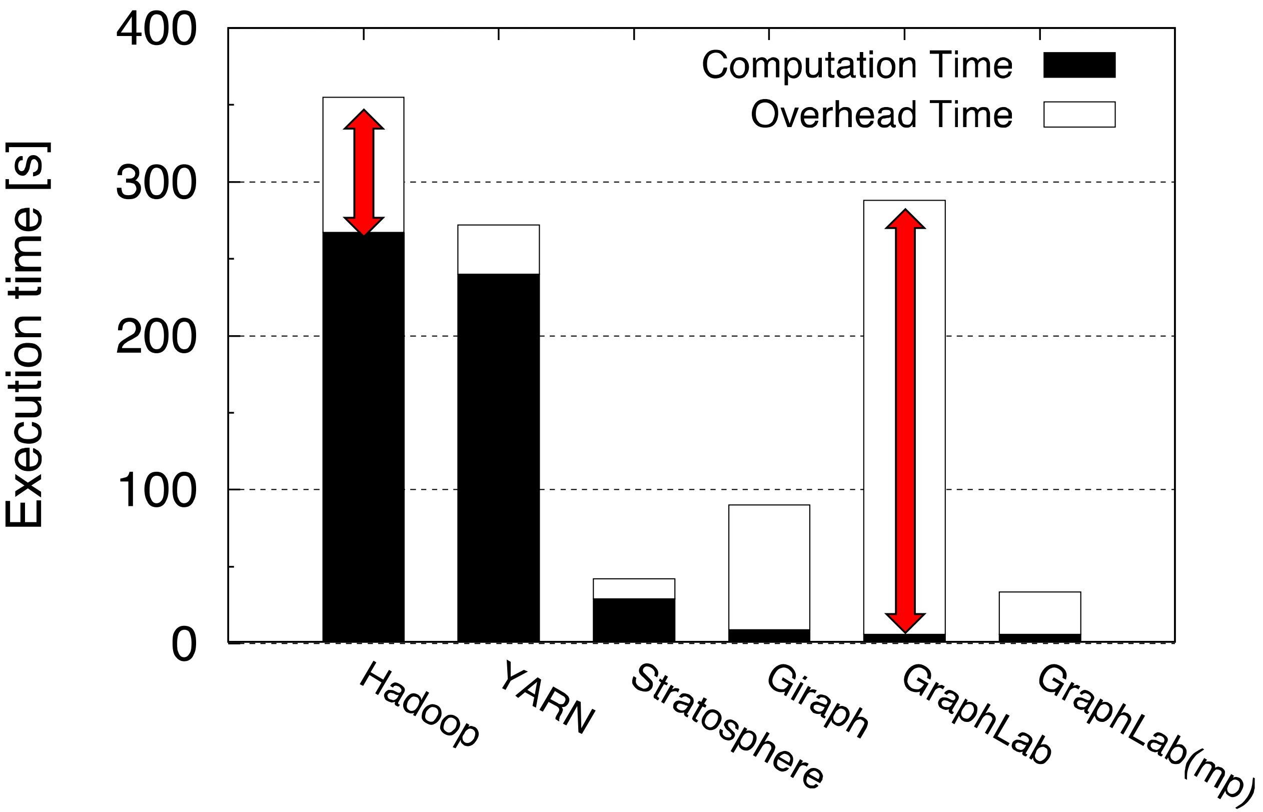
- Using more computing machines/cores can reduce execution time
- Tuning needed, e.g., for GraphLab, split large input file into number of chunks equal to the number of machines

The CPU utilization: computing node



- YARN and Hadoop exhibit obvious volatility
- The CPU utilization of graph-specific platforms is lower

Overhead: BFS on DotaLeague



- The percentage of overhead time of generic platforms is smaller
- The percentage of overhead time is diverse across the platforms, algorithms, and graphs

Additional Overheads
Data ingestion time

- Data ingestion
  - Batch system: one ingestion, multiple processing
  - Transactional system: one ingestion, one processing

- Data ingestion matters even for batch systems

	Amazon	**DotaLeague**	**Friendster**
HDFS	1 second	7 seconds	5 minutes
Neo4J	4 hours	6 **days**	**n/a**

Guo, Biczak, Varbanescu, Iosup, Martella, Willke.
How Well do Graph-Processing Platforms Perform?
An Empirical Performance Evaluation and Analysis