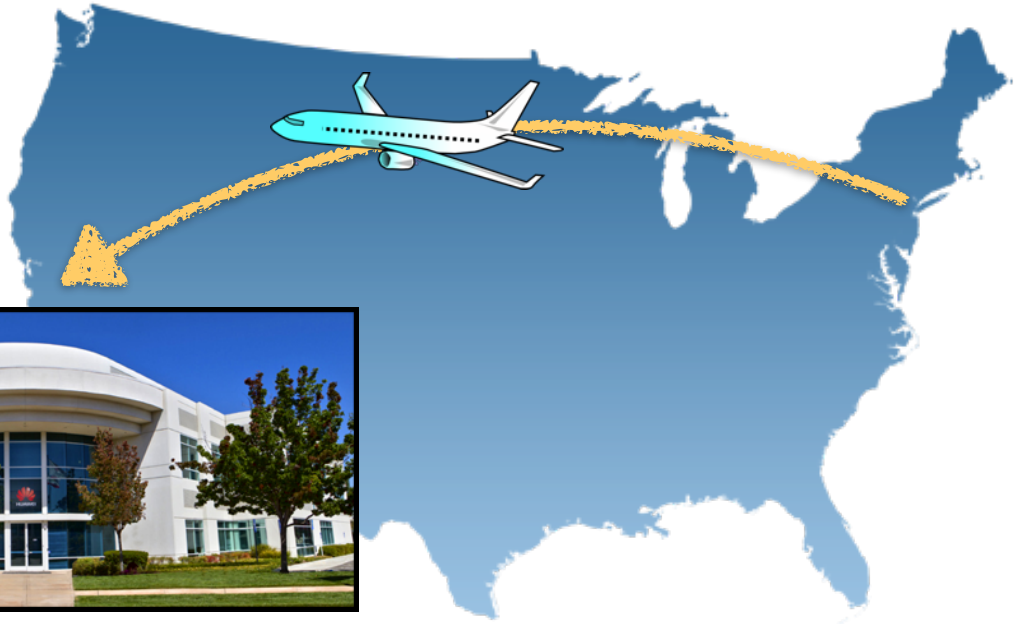
Big Graph Analytics Engine

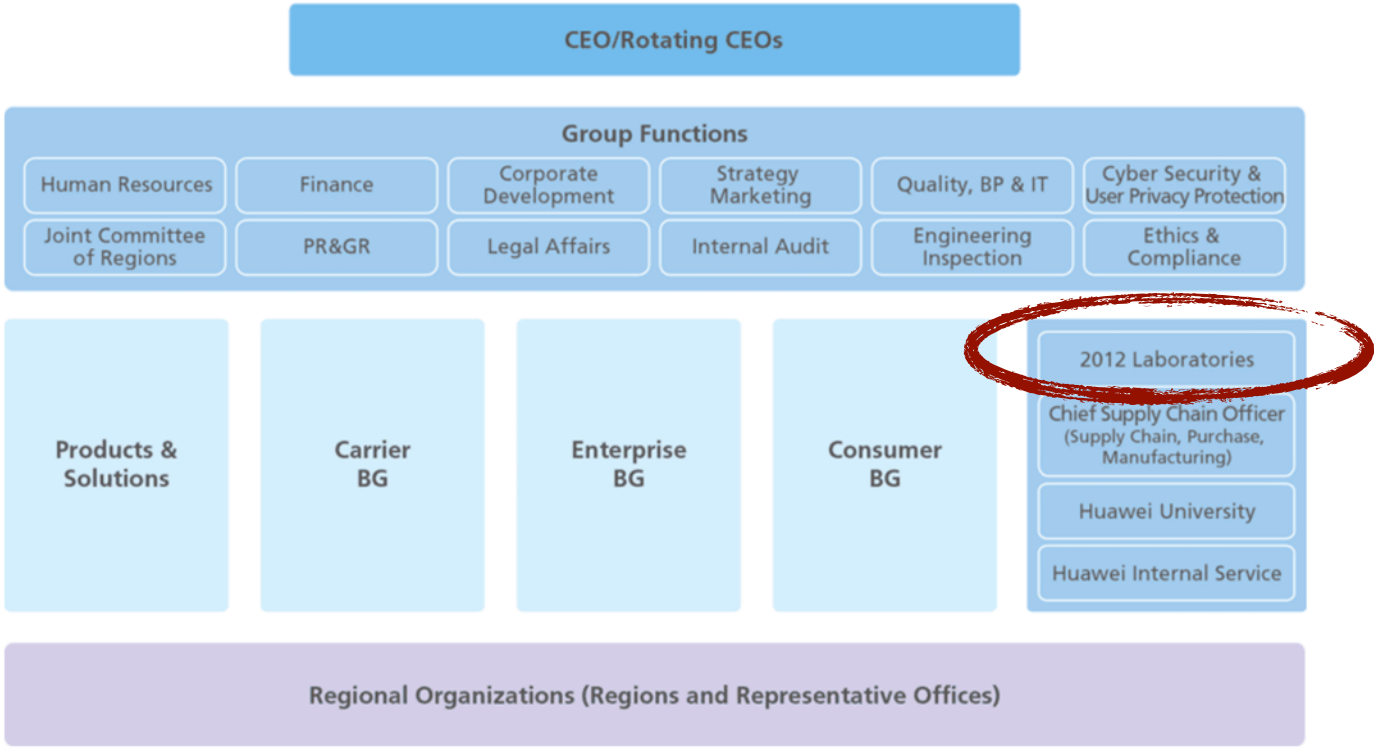
Yinglong Xia
6/23/2016

8th Linked Data Benchmark Council TUC Meeting@Oracle Conference Center

Introduction



Introduction



http://www.huawei.com/en/about-huawei/corporate-governance/corporate-governance

HUAWEI

Huawei headquarters in Shenzhen, Guangdong, China

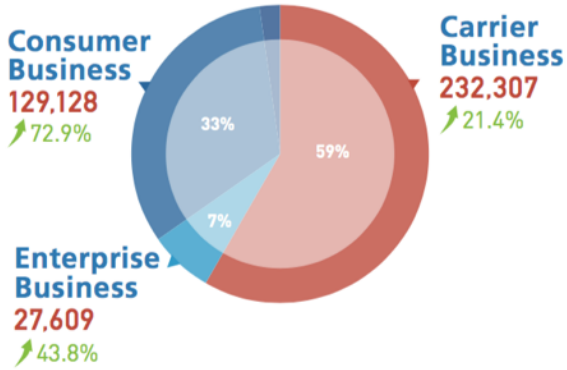
Native name	华为技术有限公司
Type	Private
Industry	Telecommunications equipment Networking equipment
Founded	1987; 29 years ago
Founder	Ren Zhengfei
Headquarters	Shenzhen, Guangdong, China
Area served	Worldwide
Products	Mobile and fixed broadband networks, consultancy and managed services, multimedia technology, smartphones, tablet computers, dongles

Recent Growth

CNY 395,009 million

Revenue

CNY Million	2015	2014	YoY
Carrier Business	232,307	191,381	21.4%
Enterprise Business	27,609	19,201	43.8%
Consumer Business	129,128	74,688	72.9%
Others	5,965	2,927	103.8%
Total	395,009	288,197	37.1%



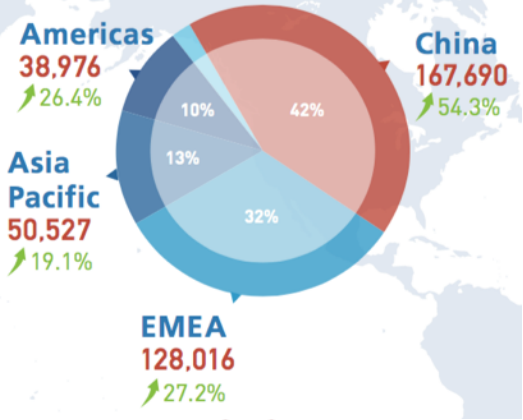
CNY 36,910 million

Net Profits

CNY 49,315 million

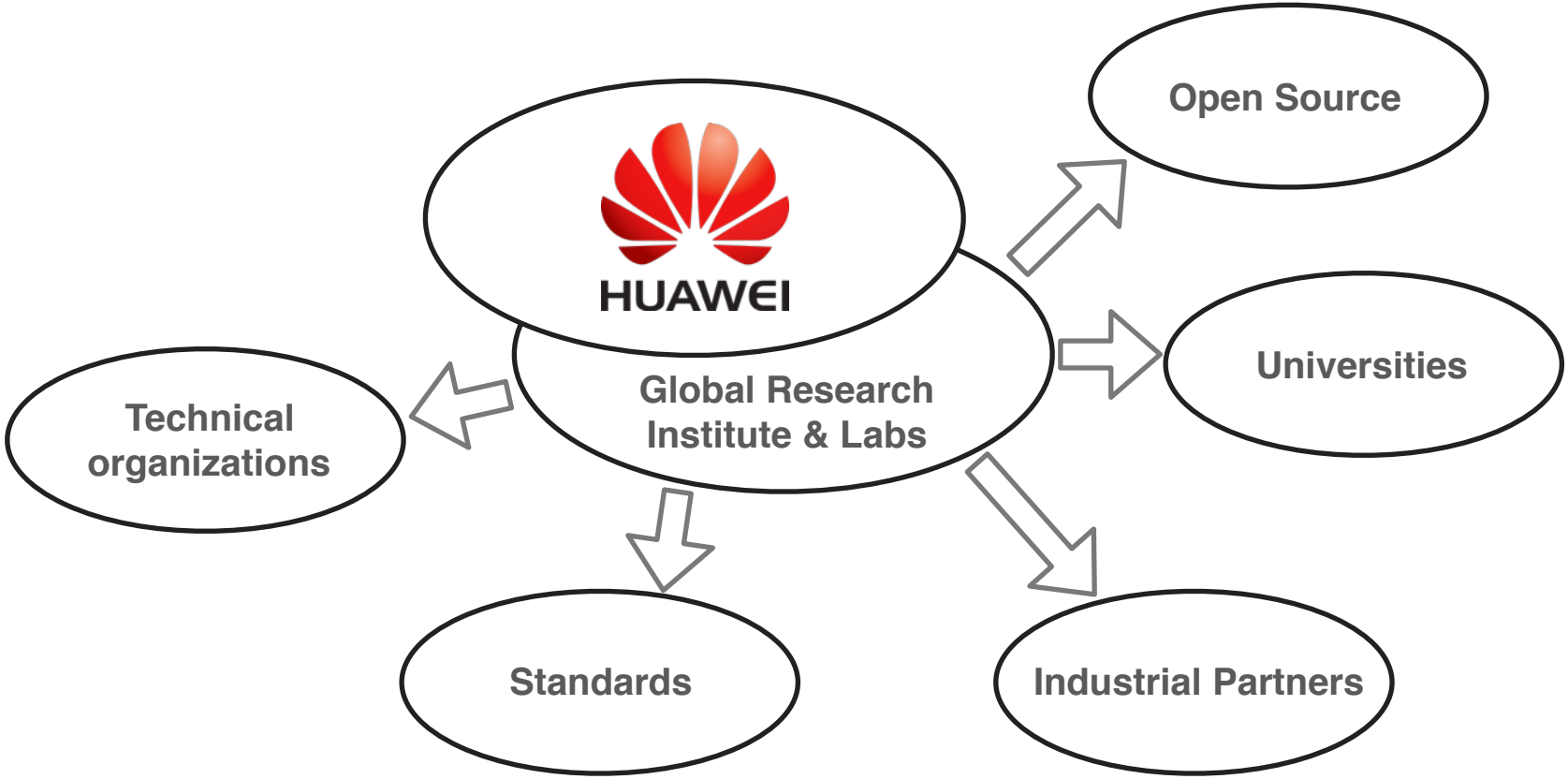
Cash flow from operating activities

CNY Million	2015	2014	YoY
China	167,690	108,674	54.3%
EMEA	128,016	100,674	27.2%
Asia Pacific	50,527	42,409	19.1%
Americas	38,976	30,844	26.4%
Others	9,800	5,596	75.1%
Total	395,009	288,197	37.1%



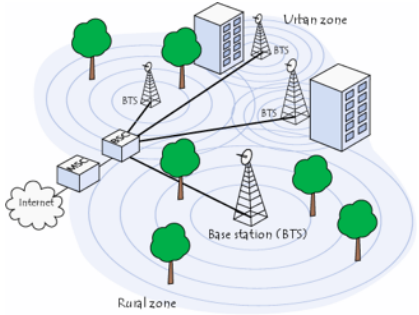
http://www.huawei.com/en/about-huawei

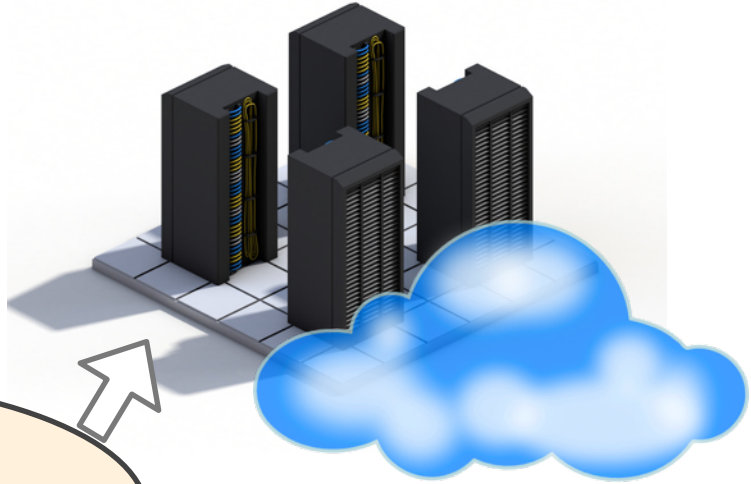
Collaboration



HUAWEI

Graph Analytics for Smart Big Data





Big Data Analytics & Management

NLP

Graph

Machine Learning

Deep Learning

HUAWEI

Graph in ONOS

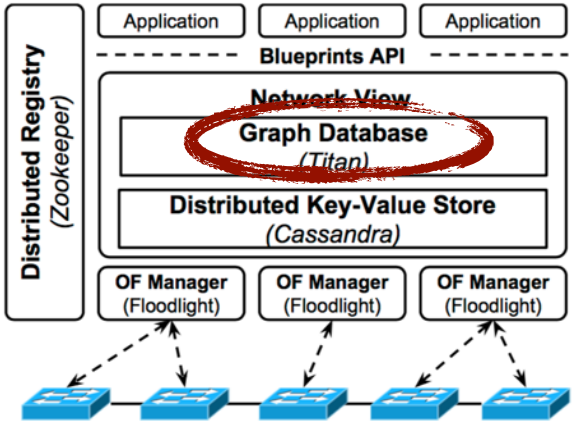


Figure 2: Prototype 1 Architecture

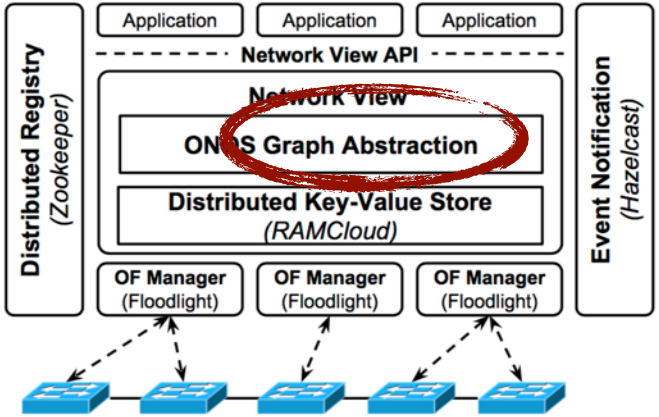
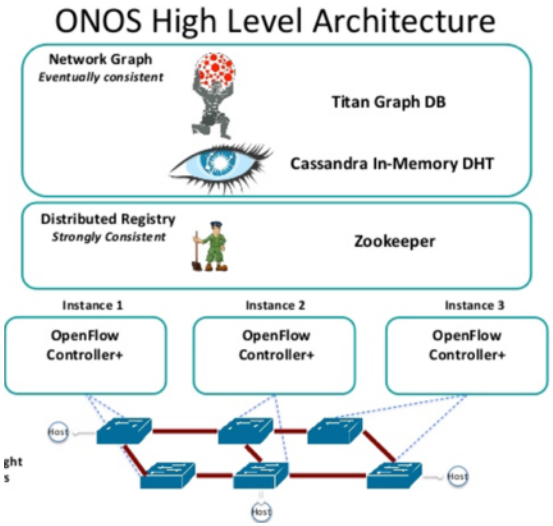


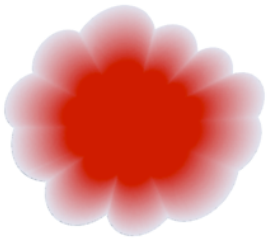
Figure 3: Prototype 2 Architecture

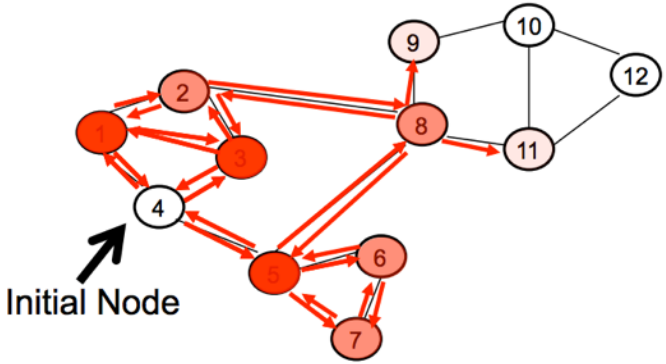


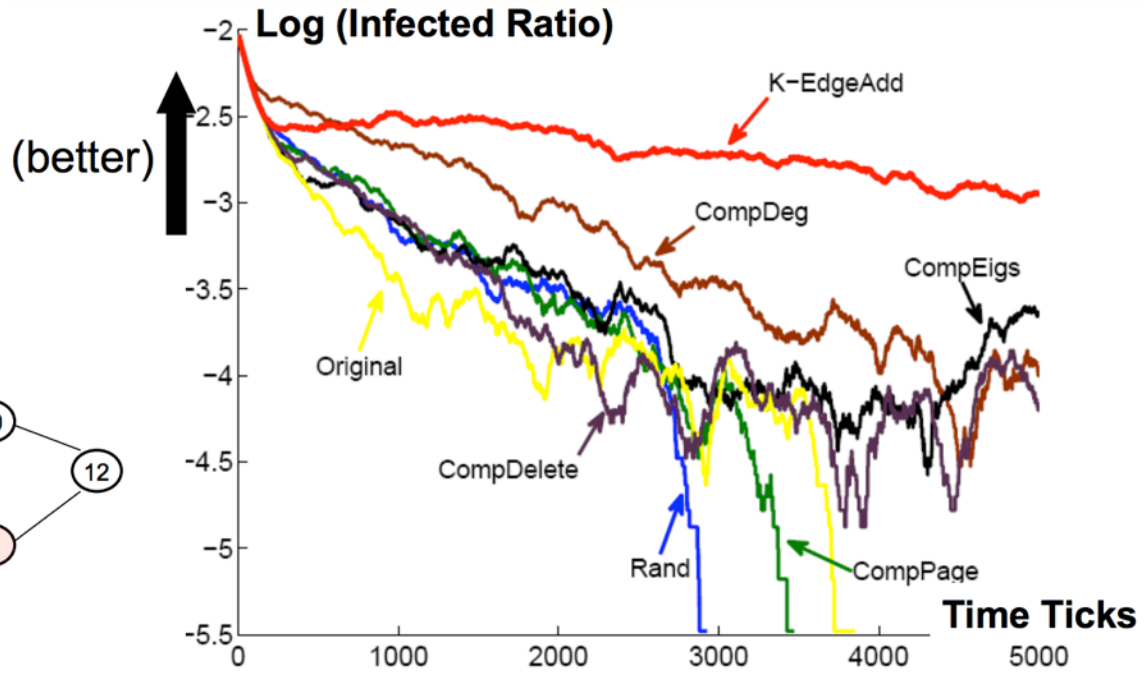
HotSDN'2014

HUAWEI

Topology Impact on Information Propagation

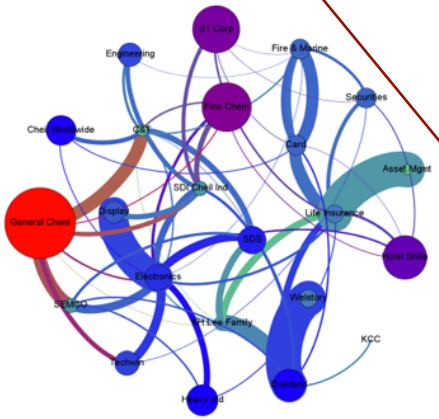


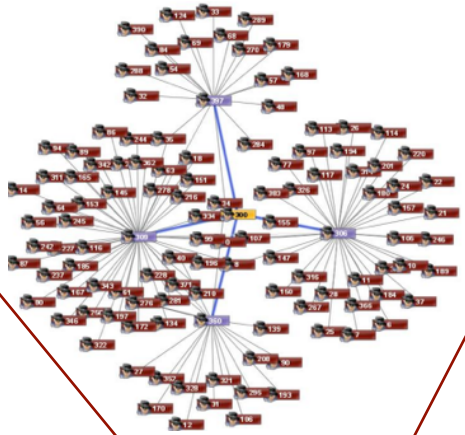


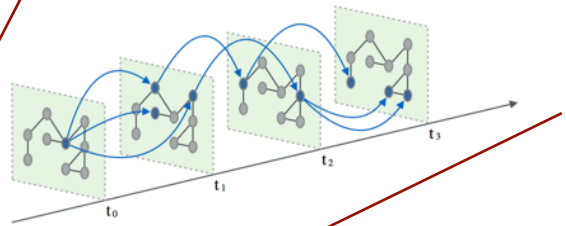


HUAWEI

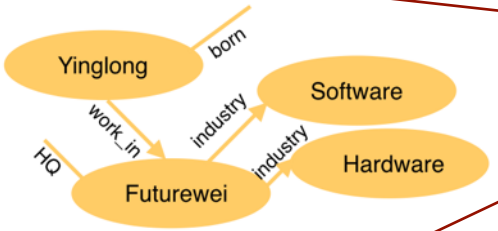
Explore the Variety in Graph Analytics

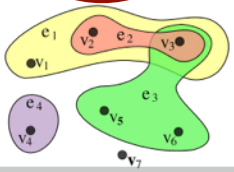


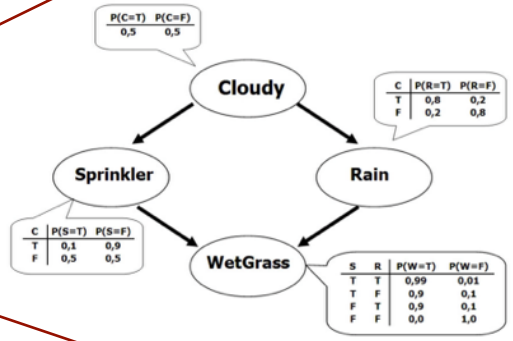




Graph





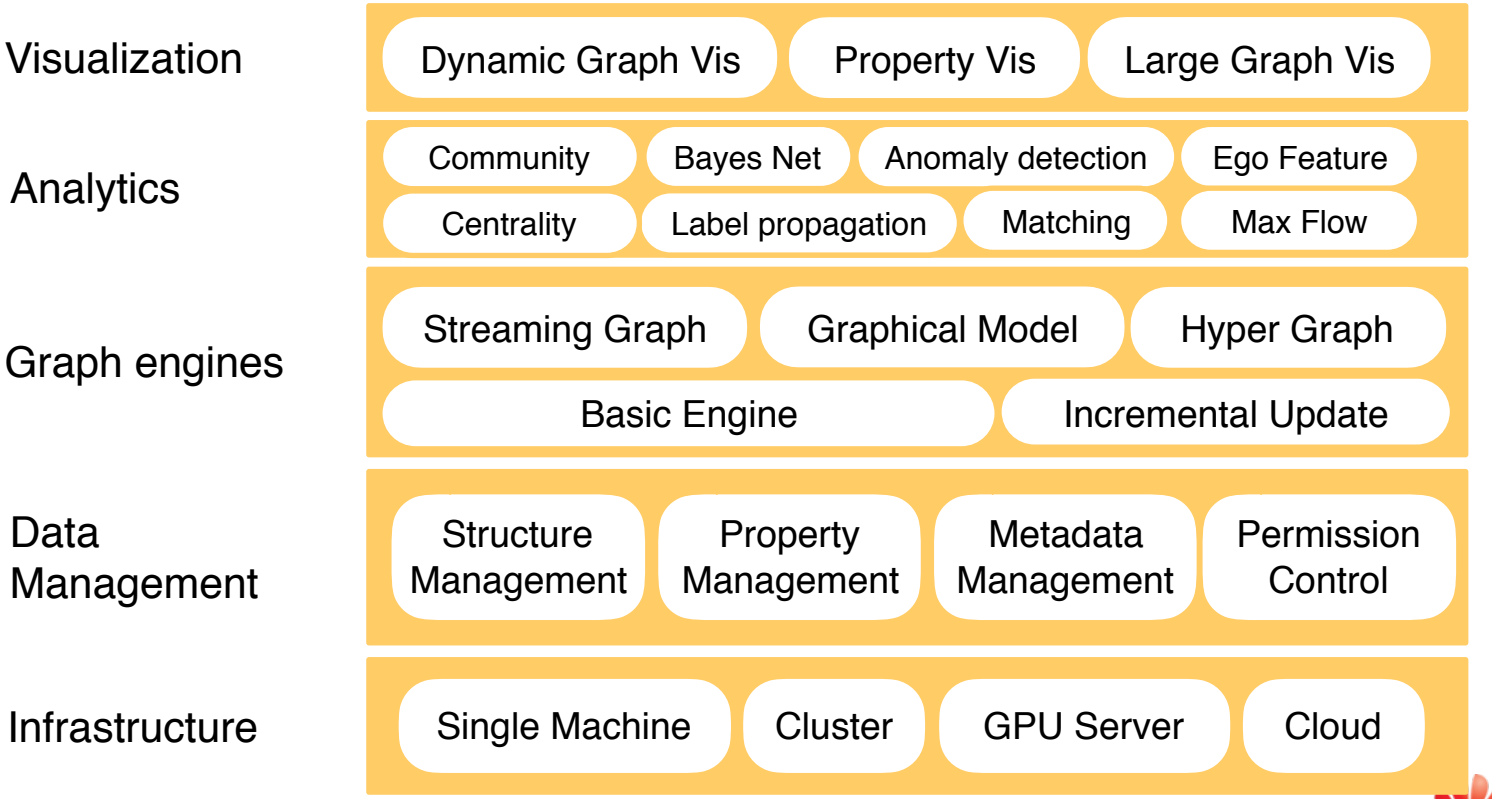


HUAWEI

Challenges

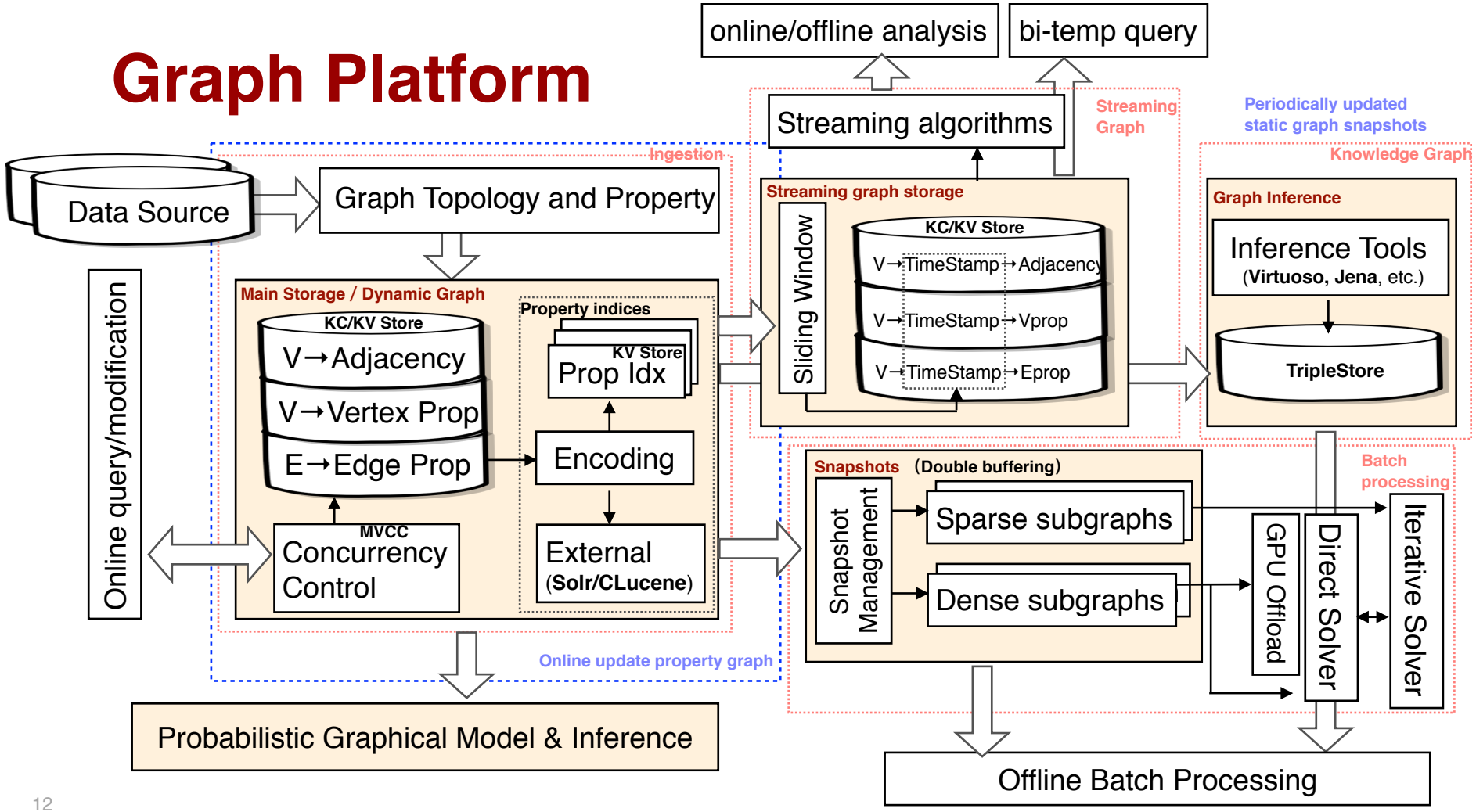
- Very large scale graphs for analysis
  - 10B~1000B in terms of the number of vertices
  - a few hundreds of properties, static and dynamic
  - distributed communication introduces additional overhead
- Irregularity in graph data access
  - Low data locality results in high disk/communication IO overhead
  - Data access patterns are diverse among graph analysis algorithms
- Near real-time requirement
  - Incorporate with incremental graph updates
  - Approximate query & analysis should be considered
- Efficiency and productivity to balance

Graph Platform for Smart Big Data

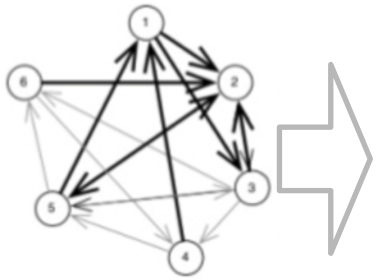


HUAWEI

Graph Platform



Unified Graph Data Access Patterns



Iteration *i*

step 1

shard 1 (1, 2)

src	dst	value
1	2	0.3
3	2	0.2
4	1	1.4
5	1	0.5
	2	0.6
6	2	0.8

shard 2 (3,4)

src	dst	value
1	3	0.4
2	3	0.3
3	4	0.8
5	3	0.2
6	4	1.9

shard 3 (5,6)

src	dst	value
2	5	0.6
3	5	0.9
	6	1.2
4	5	0.3
5	6	1.1

step 2

src	dst	value
1	2	0.3
3	2	0.2
4	1	1.4
5	1	0.5
	2	0.6
6	2	0.8

src	dst	value
1	3	0.4
2	3	0.3
3	4	0.8
5	3	0.2
6	4	1.9

src	dst	value
2	5	0.6
3	5	0.9
	6	1.2
4	5	0.3
5	6	1.1

step 3

src	dst	value
1	2	0.3
3	2	0.2
4	1	1.4
5	1	0.5
	2	0.6
6	2	0.8

src	dst	value
1	3	0.4
2	3	0.3
3	4	0.8
5	3	0.2
6	4	1.9

src	dst	value
2	5	0.6
3	5	0.9
	6	1.2
4	5	0.3
5	6	1.1

equivalent

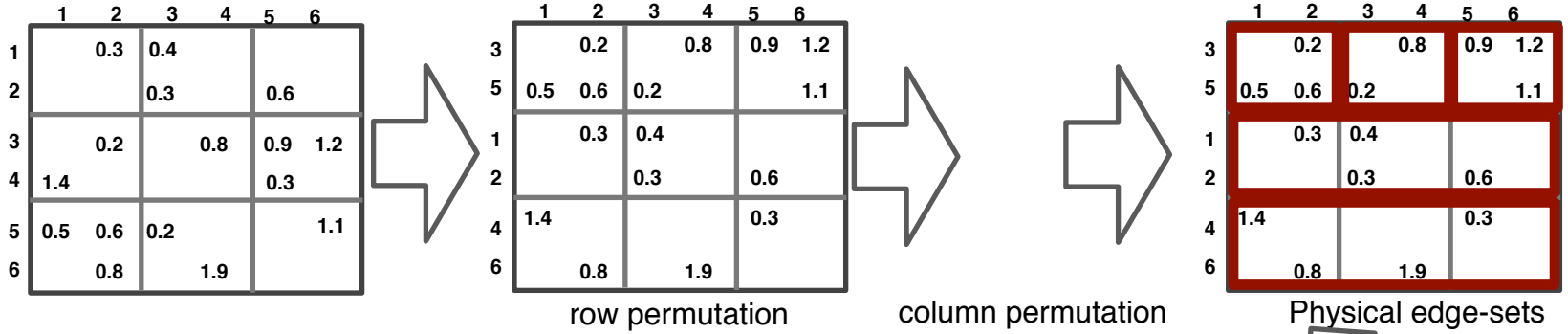
	1	2	3	4	5	6
1		0.3	0.4			
2			0.3		0.6	
3		0.2		0.8	0.9	1.2
4	1.4				0.3	
5	0.5	0.6	0.2			1.1
6		0.8		1.9		

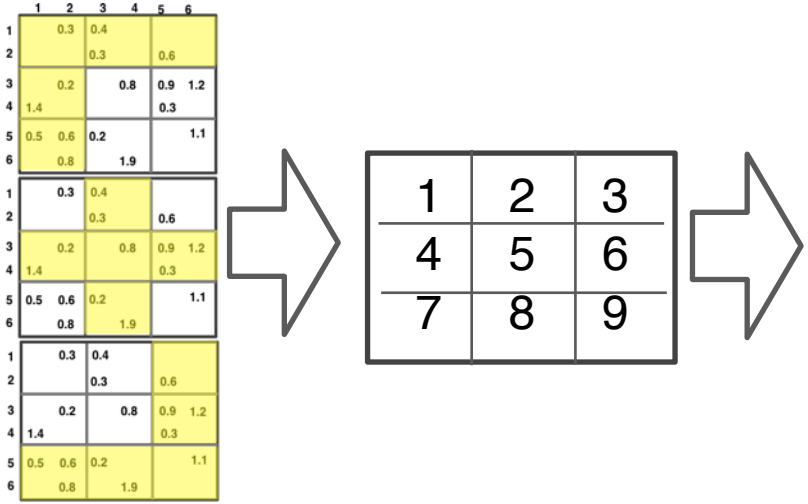
	1	2	3	4	5	6
1		0.3	0.4			
2			0.3		0.6	
3		0.2		0.8	0.9	1.2
4	1.4				0.3	
5	0.5	0.6	0.2			1.1
6		0.8		1.9		

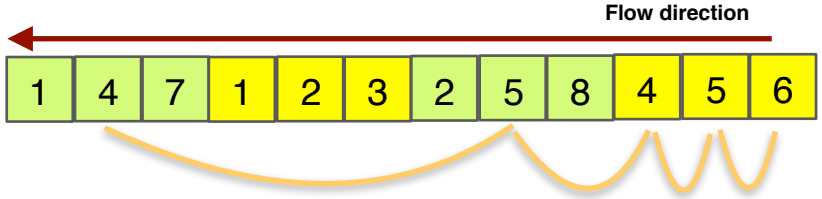
	1	2	3	4	5	6
1		0.3	0.4			
2			0.3		0.6	
3		0.2		0.8	0.9	1.2
4	1.4				0.3	
5	0.5	0.6	0.2			1.1
6		0.8		1.9		

observation on PSW data access patterns inspires highly efficient sharding representation

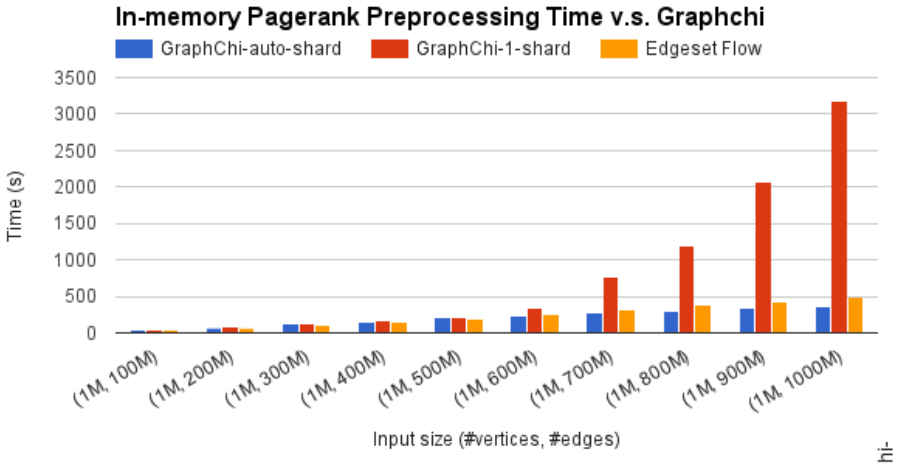
Construct Edge-set Flows





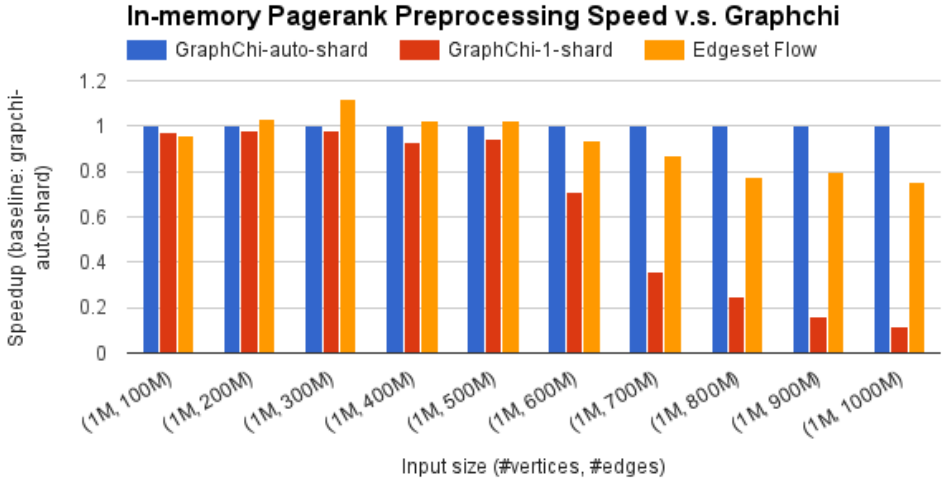


Preliminary Experiments - Preproc.

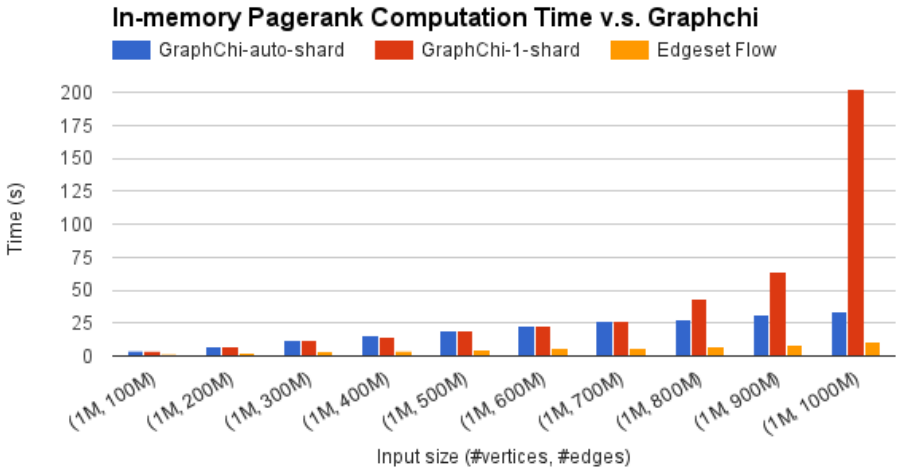


Create the data in our format

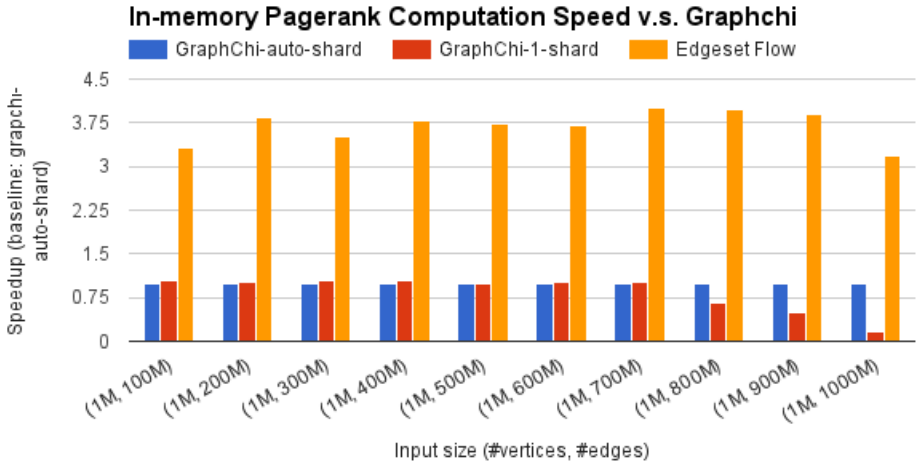
Graph Ingestion/Preprocessing Time



Preliminary Experiments - Comp.

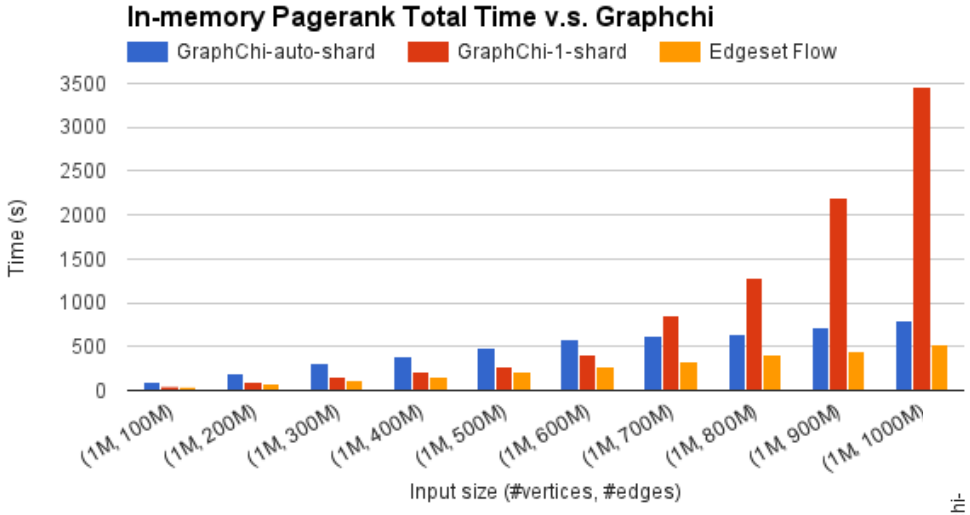


PageRank w/o Loading Time

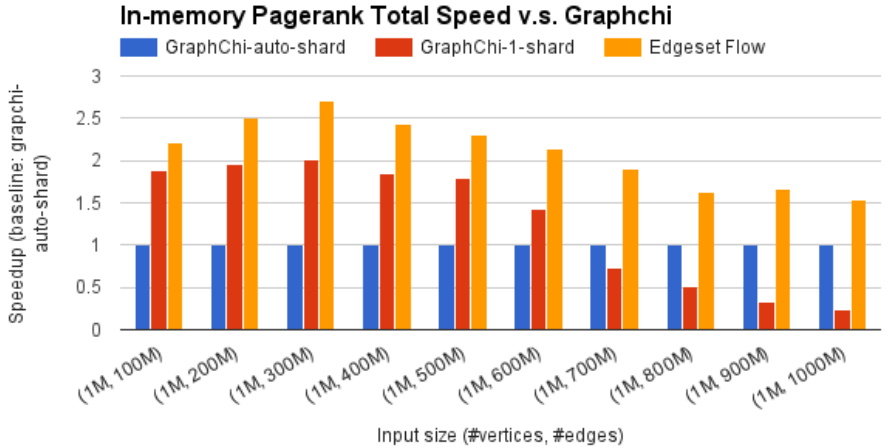


Decent speedup achieved w/ or w/o loading time

Preliminary Experiments



PageRank Total Time



Conclusion

- Many big data problems involve links among a lot of entities, naturally represented as a graph
- Property graph is highly expressive
- Industry is looking for graph/graphical model engines for complex network analysis, streaming graph, probabilistic graphical models, and RDF graph computing
- Efficiency is the key in many industry graph analysis systems, especially when the data volume is big
- Eventually, the graph engine should serve for AI Business systems

Thanks

Yinglong Xia
yinglong.xia.2010@ieee.org