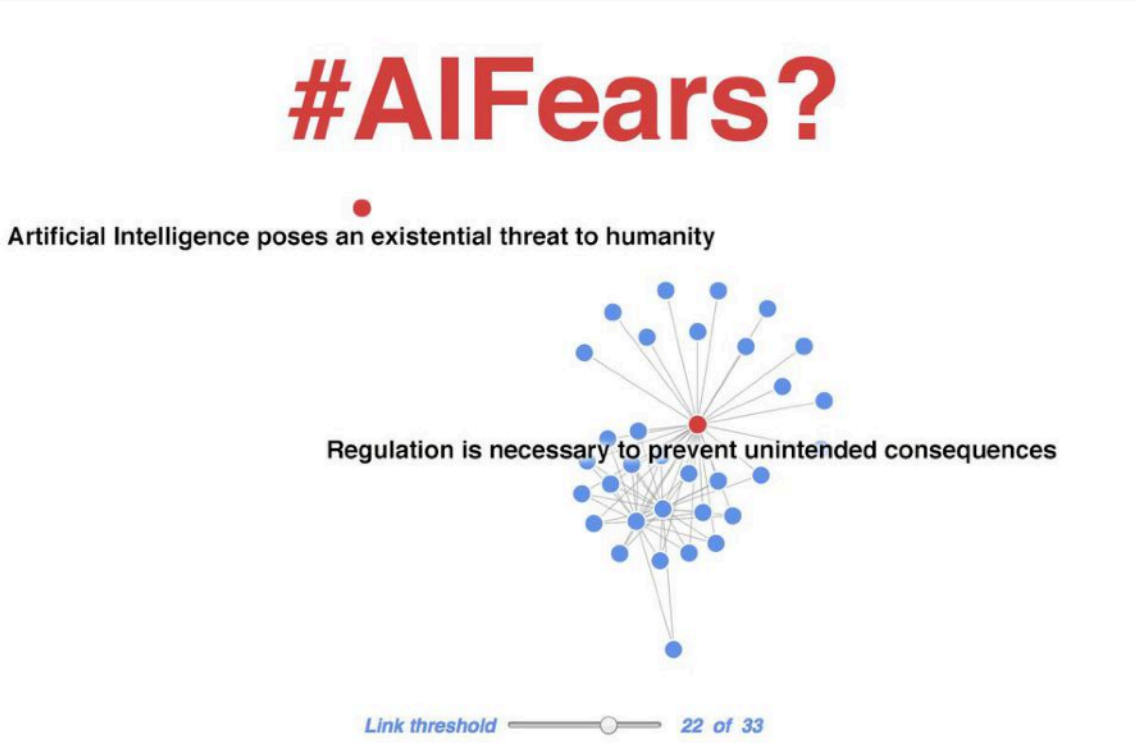
A Graph Engine Service for Cloud AI Platforms

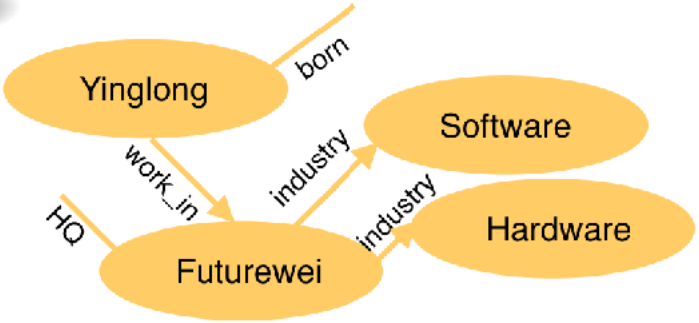
Yinglong Xia
Huawei Enterprise Intelligence
06/08/2018

Trend of AI


#AIFears?
Artificial Intelligence poses an existential threat to humanity
Regulation is necessary to prevent unintended consequences
Link threshold
22 of 33

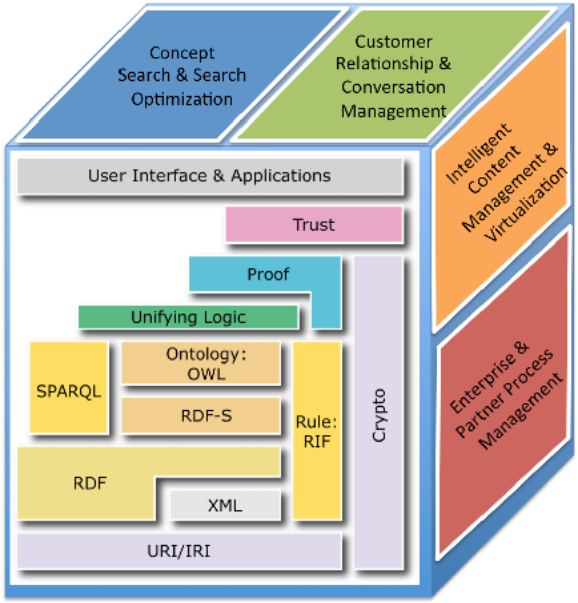
Reinvented Wheels in Parallel Universe?

Semantic Web



subject	Predicate	Object
Yinglong	work_in	Futurewei
Yinglong	born	1980
Futurewei	has_HQ	Shenzhen

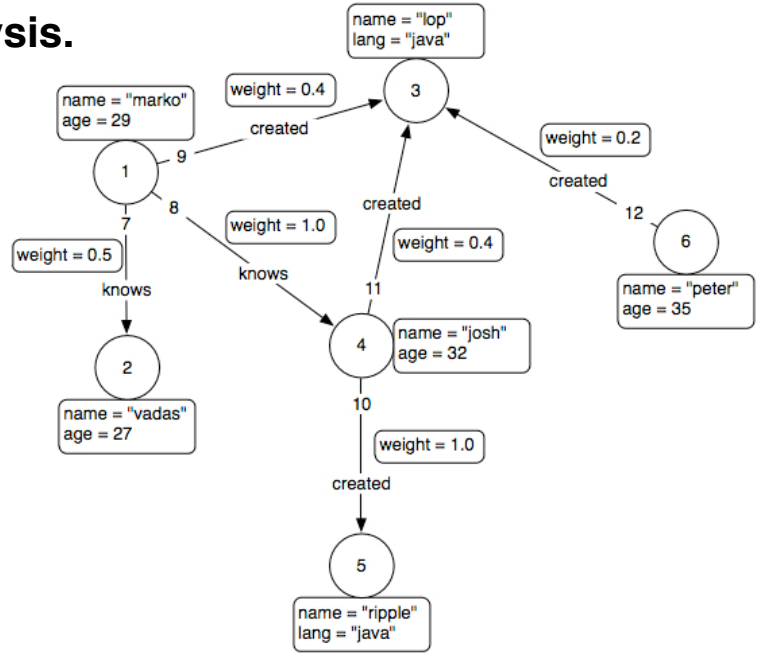
Sir Berners-Lee believes the core value of RDF sementic web is to transofrm WWW into media for information exchange



RDF Graph = A collection of triples, linking the description of resources

Labeled Property Graph

- **Property graph is a data representation model with strong expressiveness**
- **Property graph is supported by most graph databases (NoSQL) and also forms the foundation of graph analysis.**
  - **Vertices**
    - **Unique ID for each**
    - **A set of (directed) edges**
    - **Property: a set of key-value pairs**
  - **Edges**
    - **Unique ID for each**
    - **Two end vertices**
    - **With at least a label**
    - **Property: a set of key-value pairs**



Heterogeneous Information Network

CONTEXT-RICH RECOMMENDATION: INTEGRATING LINKS, TEXT, AND SPATIO-TEMPORAL DIMENSIONS

August 11, 2017

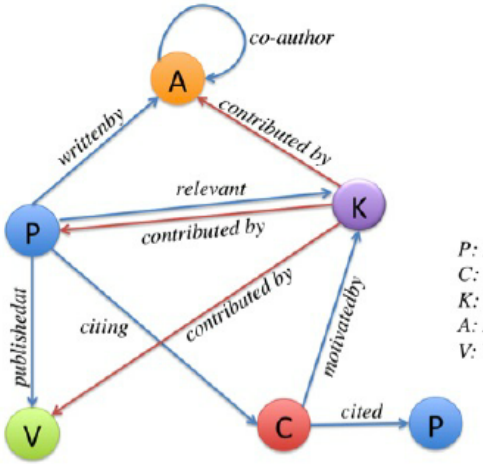
A Survey of Heterogeneous Information Network Analysis

Authors:
Chuan Shi — Beijing Key Lab of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing, China
Yitong Li — Beijing Key Lab of Intelligent Telecommunications Software and Multimedia, Beijing University of Posts and Telecommunications, Beijing, China
Jiawei Zhang — University of Illinois at Chicago, Chicago, IL
Yizhou Sun — University of California, Los Angeles, CA
Philip S. Yu — Computer Science Department, University of Illinois at Chicago, Chicago, IL

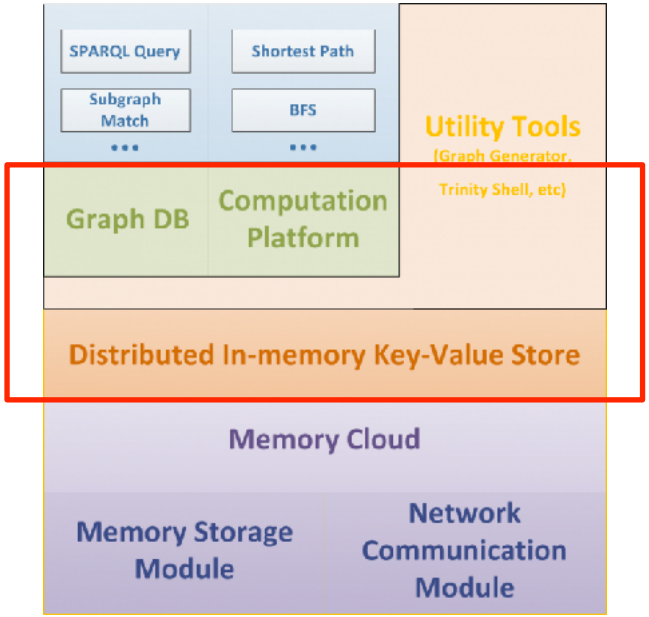
2017 Article
orig-research

Bibliometrics
- Citation Count: 3
- Downloacs (cumulative): n/a
- Downloacs (12 Months): n/a
- Downloacs (6 Weeks): n/a

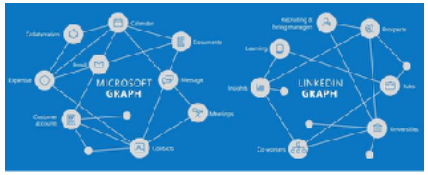
Published in:
- Journal
IEEE Transactions on Knowledge and Data Engineering archive
Volume 29 Issue 1, January 2017
Page 17-37
IEEE Educational Activities Department Piscataway, NJ, USA
table of contents doi>10.1109/TKDE.2016.2598561

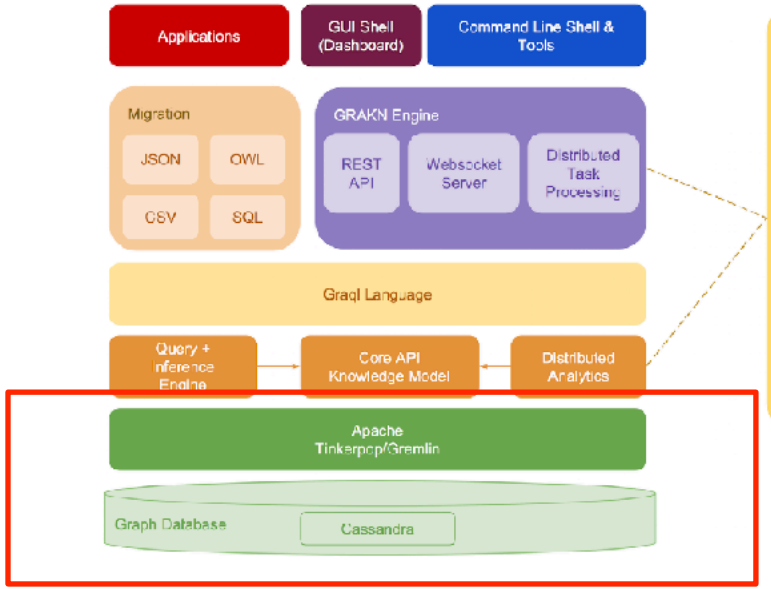


Find Answer?



Microsoft Satori knowledge graph utilizes **Trinity** and GraphDB

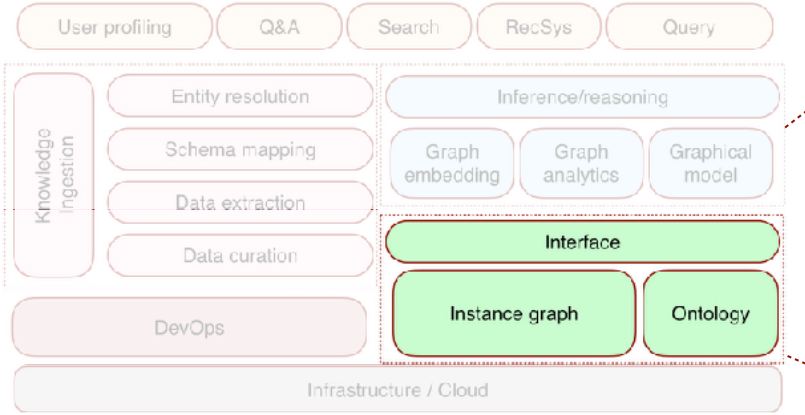




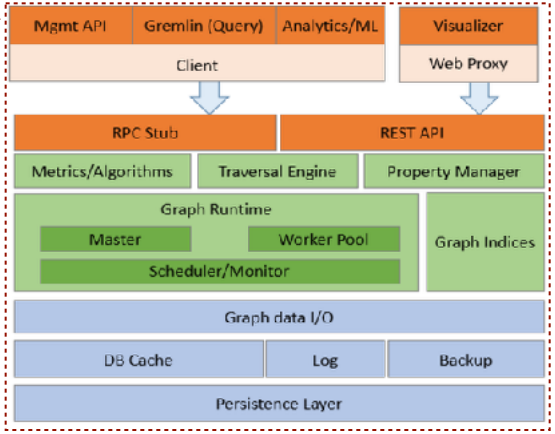
GraKn.ai utilizes a property graph as its data layer

- Y. Xia, et. al, *C-Graph: A Highly Efficient Concurrent Graph Processing Framework*, ICS, 2018
- Y. Xia, et. al, *An Edge-Set Based Large Scale Graph Processing System*, IEEE Big Data 2016

Property Graph Storage

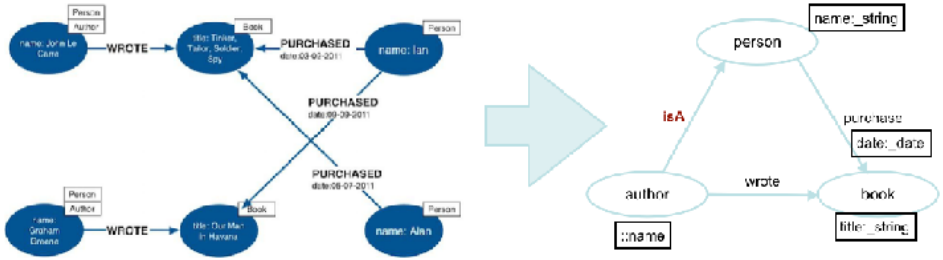


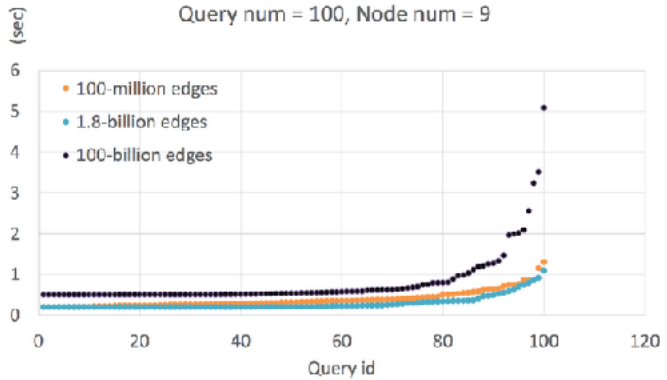
Eywa: Unified property graph analysis and query system



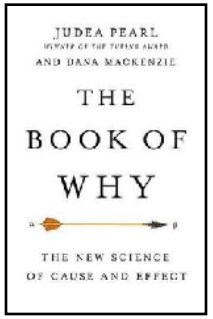
- 9 VMs from data center
- Each with 22-core CPU , 125 GB memory
- Orkut: |V|=3.07M, |E|=117M
Friendster: |V|=65M, |E|=1.8B
Kronecker: |V|=984M, |E| =106.5B

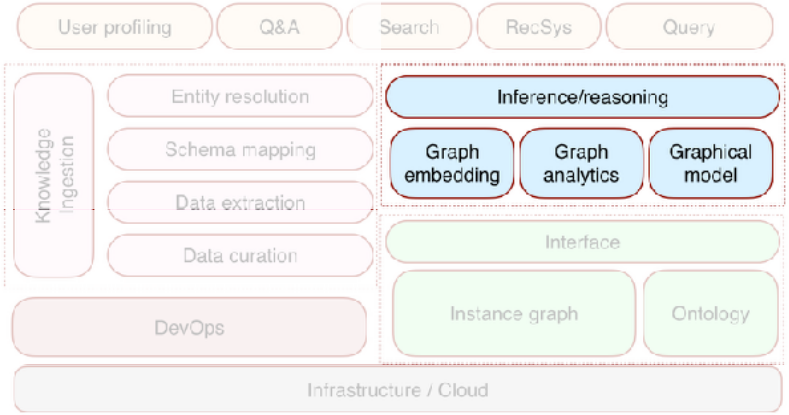
Labeled property graph model and its schema description

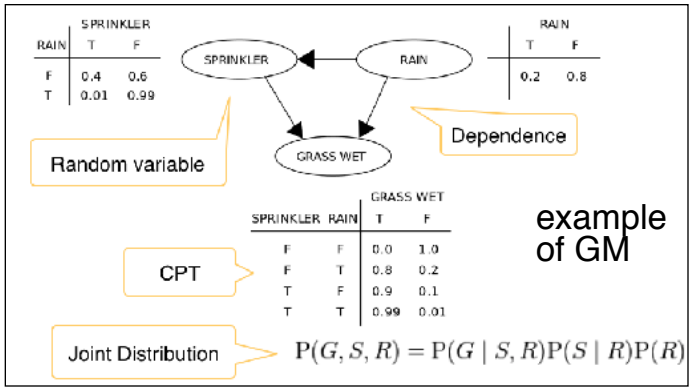




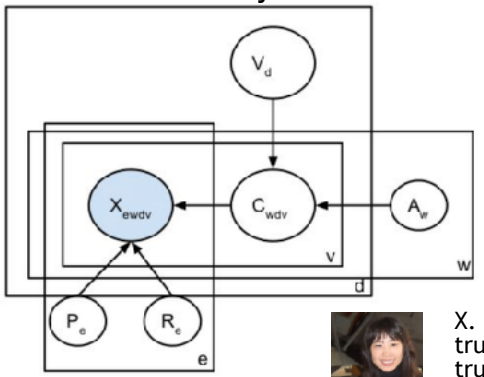
Graphical Models







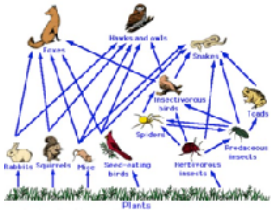
Predict the correctness of extraction and accuracy

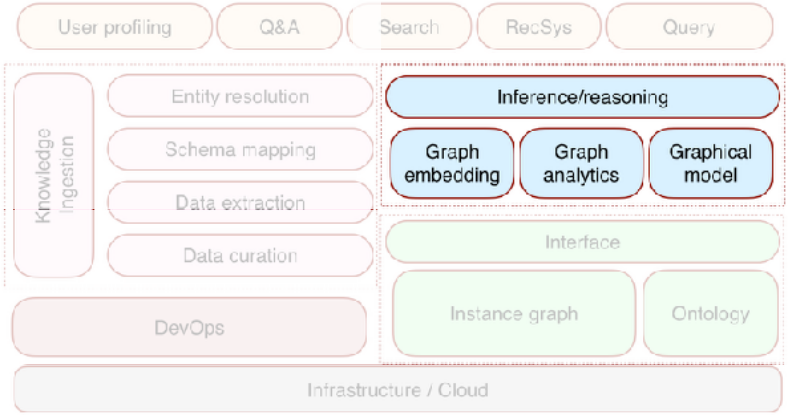


X. Dong et. al., Knowledge-based trust: estimating the trustworthiness of web sources. In *VLDB*, 2015

- Observations
  - X_{ewdv}: whether extractor e extracts from source w the (d,v) item-value pair
- Latent variables
  - C_{wdv}: whether source w indeed provides (d,v) pair
  - V_d: the correct value(s) for d
- Parameters
  - A_w: Accuracy of source w
  - P_e: Precision of extractor e
  - R_e: Recall of extractor e

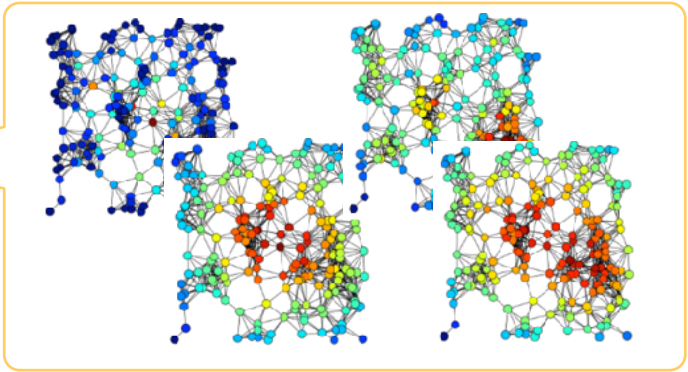
Complex Network Analysis





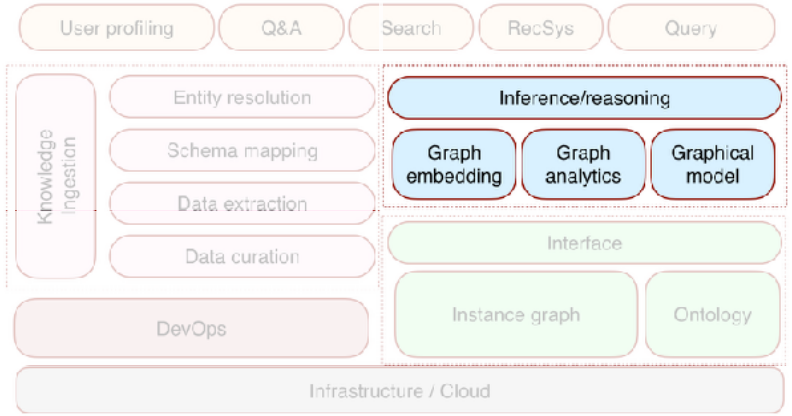
- Effeciveness of knowledge inference is upon the completeness and **quality** of the linked data
- **Transitivity** between two vertices may reveal redudant links or missing connections
- **Clusters** on instance graph helps manage knowledges effeciently
- **Inconsistency** can be identified through such analysis

Import properties/metrics:
- Small-world effect
- Betweenness
- Eccentricity/**Centrality**
- Transitivity
- Resilience
- Community structure
- Clustering coefficient
- Matching index

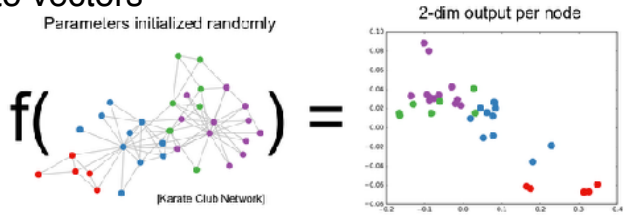


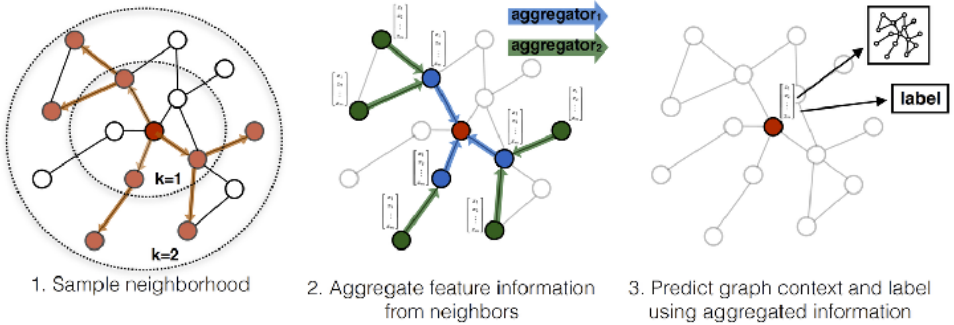
Complex network models:
- Poisson random graph
  - degree~Poisson
  - Small world effect
- Watts and Strogatz graph
  - Transitivity
  - Small world effect
- Barabasi and Albert graph
  - Small world
  - Power law

Knowledge Graph Embedding



Transform local topology and properties into vectors





1. Sample neighborhood

2. Aggregate feature information from neighbors

3. Predict graph context and label using aggregated information

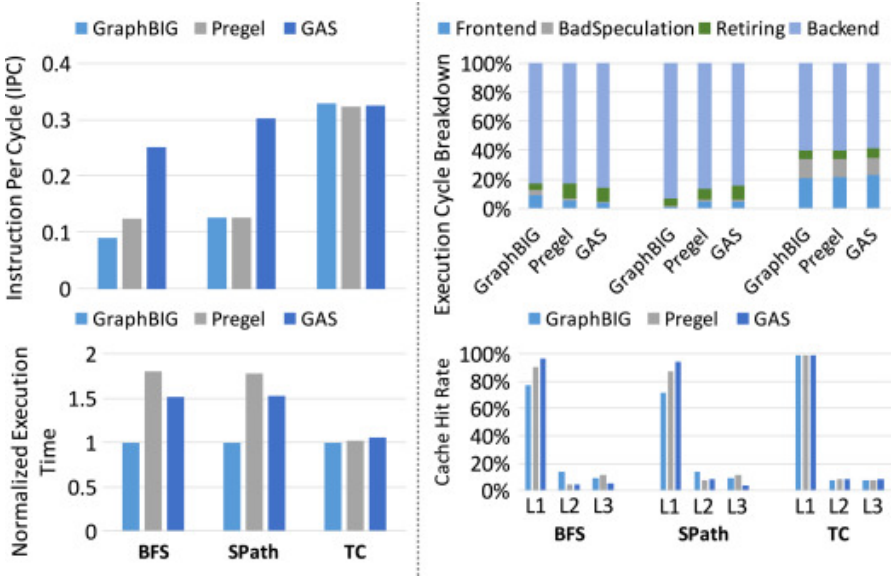
- Most embedding frameworks are inherently **transductive** and can only generate embeddings for a single fixed graph.
- These transductive approaches do not efficiently generalize to unseen nodes (e.g., in **evolving** graphs or **hidden** vertices)
- In contrast, GraphSAGE is an inductive framework that leverages node attribute information to efficiently generate representations

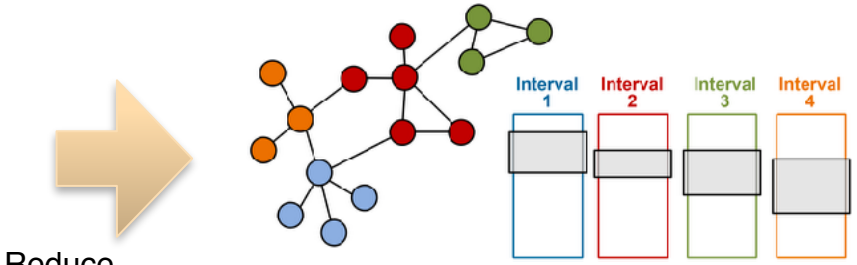
Eywa - Integrated Graph Prcessing Platform

- Y. Xia, et. al, *C-Graph: A Highly Efficient Concurrent Graph Processing Framework*, ICS, 2018
- Y. Xia, et. al, *An Edge-Set Based Large Scale Graph Processing System*, IEEE Big Data 2016

Improved Graph Data Oraganization

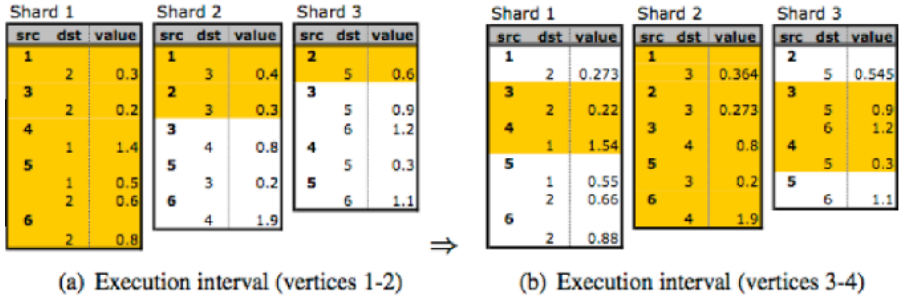
Observations on graph computing





Reduce randomness in data access

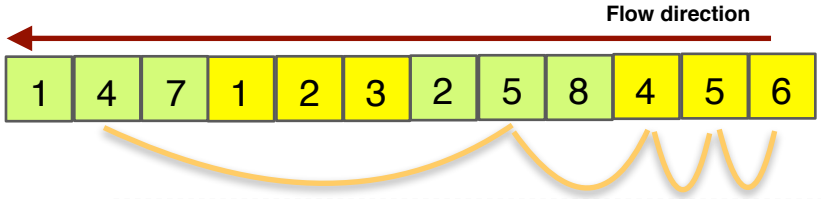
Operation



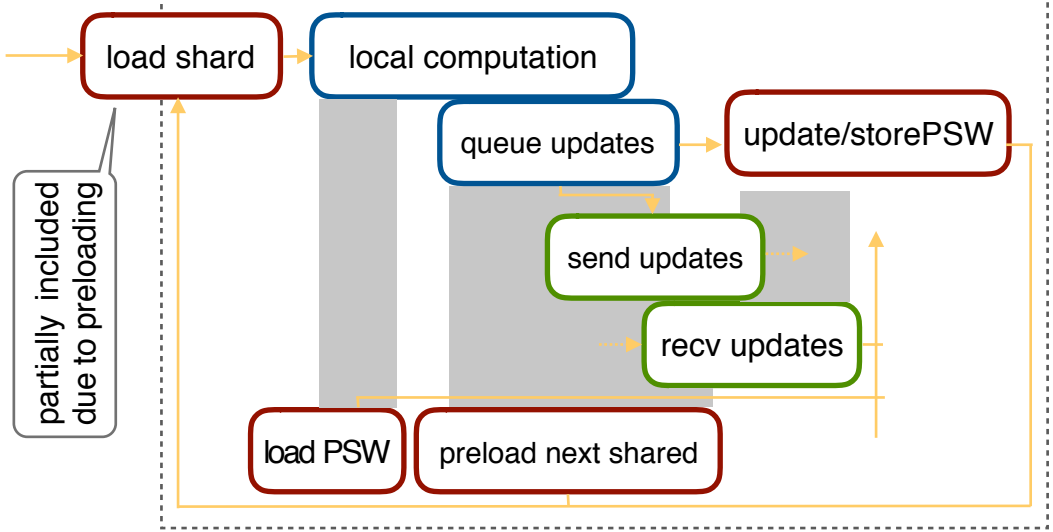
(a) Execution interval (vertices 1-2)

(b) Execution interval (vertices 3-4)

Scheduler/Prefetch of Relaxed BSP



load shard | load PSW | local computation | queue updates | send updates | recv updates | update/storePSW



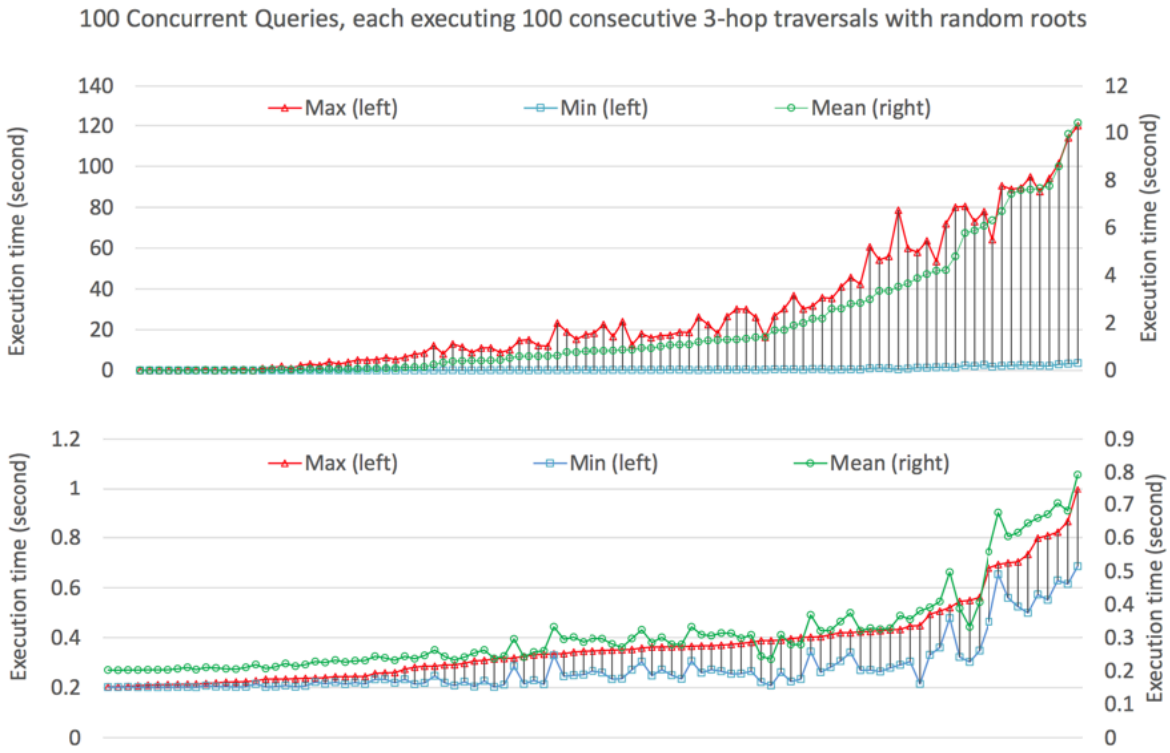
speedup due to overlapping

Dash box includes all tasks within an iteration

task type: disk IO | CPU | network IO

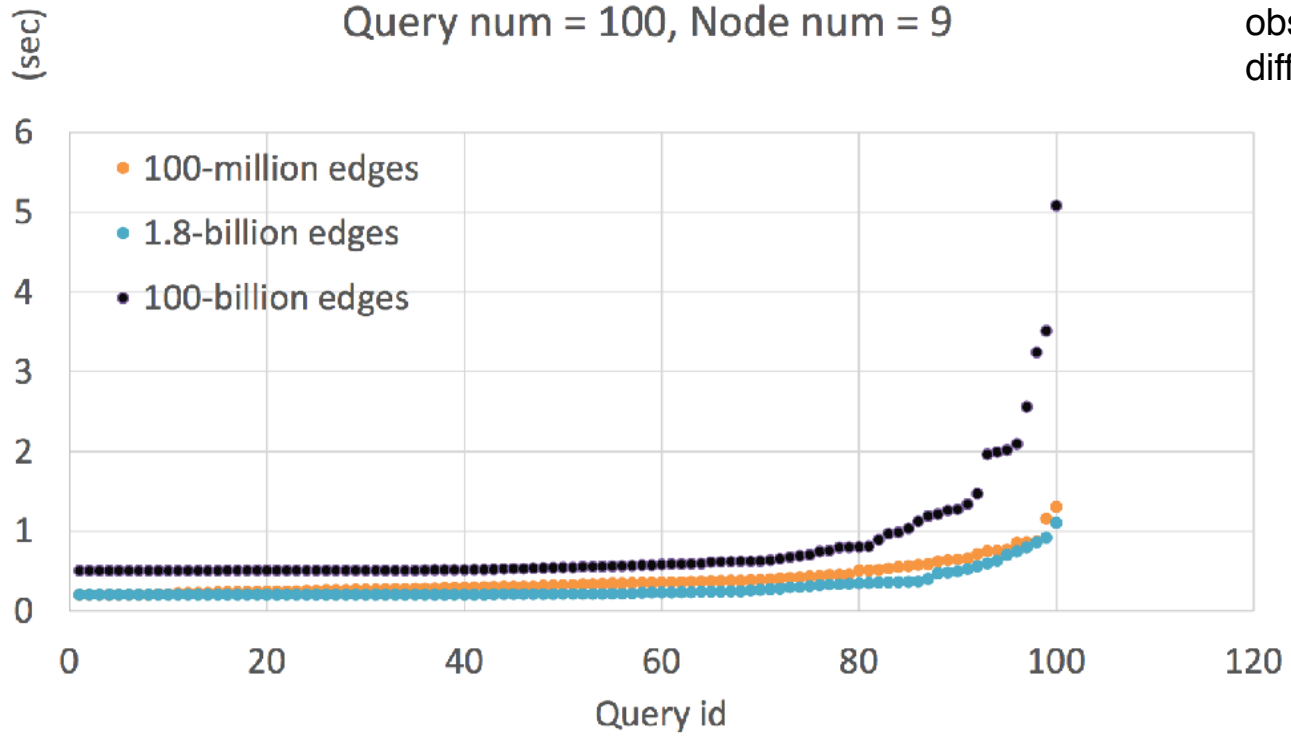
Experiments

Eywa VS. Titan (|V|: 3072441, |E|:117185083)



- Eywa outperformed the baseline method
- Eywa shows consistent running time

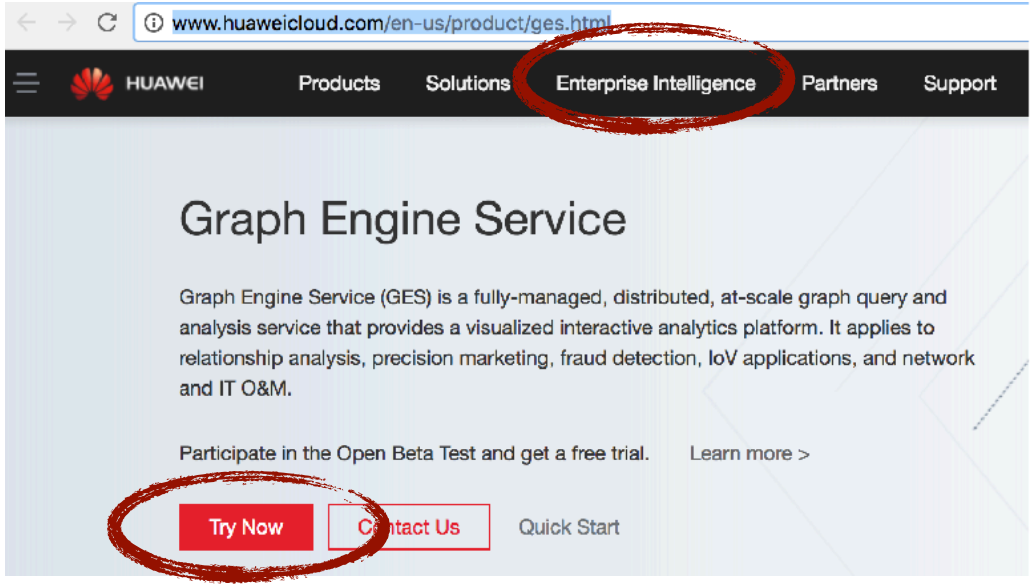
Experiments - 3



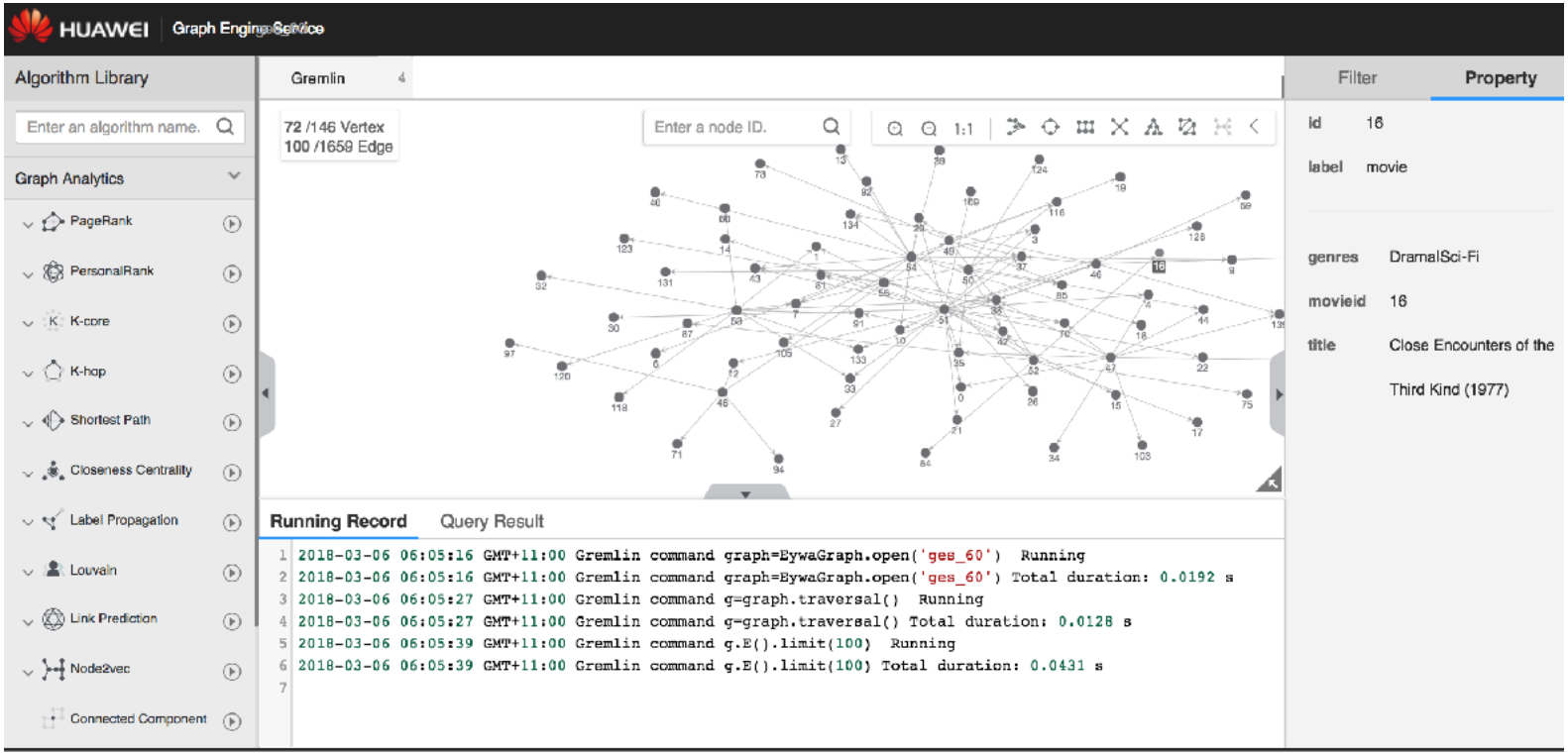
- ***Consistent*** performance observed across graphs of different scales

Demo

http://www.huaweicloud.com/en-us/product/ges.html



Demo - 2


HUAWEI
Graph Engine Service
Algorithm Library
Enter an algorithm name.
Graph Analytics
PageRank
PersonalRank
K-core
K-hop
Shortest Path
Closeness Centrality
Label Propagation
Louvain
Link Prediction
Node2vec
Connected Component
Gremlin
72 /146 Vertex
100 /1658 Edge
Enter a node ID.
Filter
Property
id 16
label movie
genres Drama|Sci-Fi
movieId 16
title Close Encounters of the Third Kind (1977)
Running Record
Query Result
2018-03-06 06:05:16 GMT+11:00 Gremlin command graph=EywaGraph.open('ges_60') Running
2018-03-06 06:05:16 GMT+11:00 Gremlin command graph=EywaGraph.open('ges_60') Total duration: 0.0192 s
2018-03-06 06:05:27 GMT+11:00 Gremlin command g=graph.traversal() Running
2018-03-06 06:05:27 GMT+11:00 Gremlin command g=graph.traversal() Total duration: 0.0128 s
2018-03-06 06:05:39 GMT+11:00 Gremlin command g.E().limit(100) Running
2018-03-06 06:05:39 GMT+11:00 Gremlin command g.E().limit(100) Total duration: 0.0431 s

Thanks

Yinglong Xia
yinglong.xia@huawei.com
yinglong.xia.2010@ieee.org