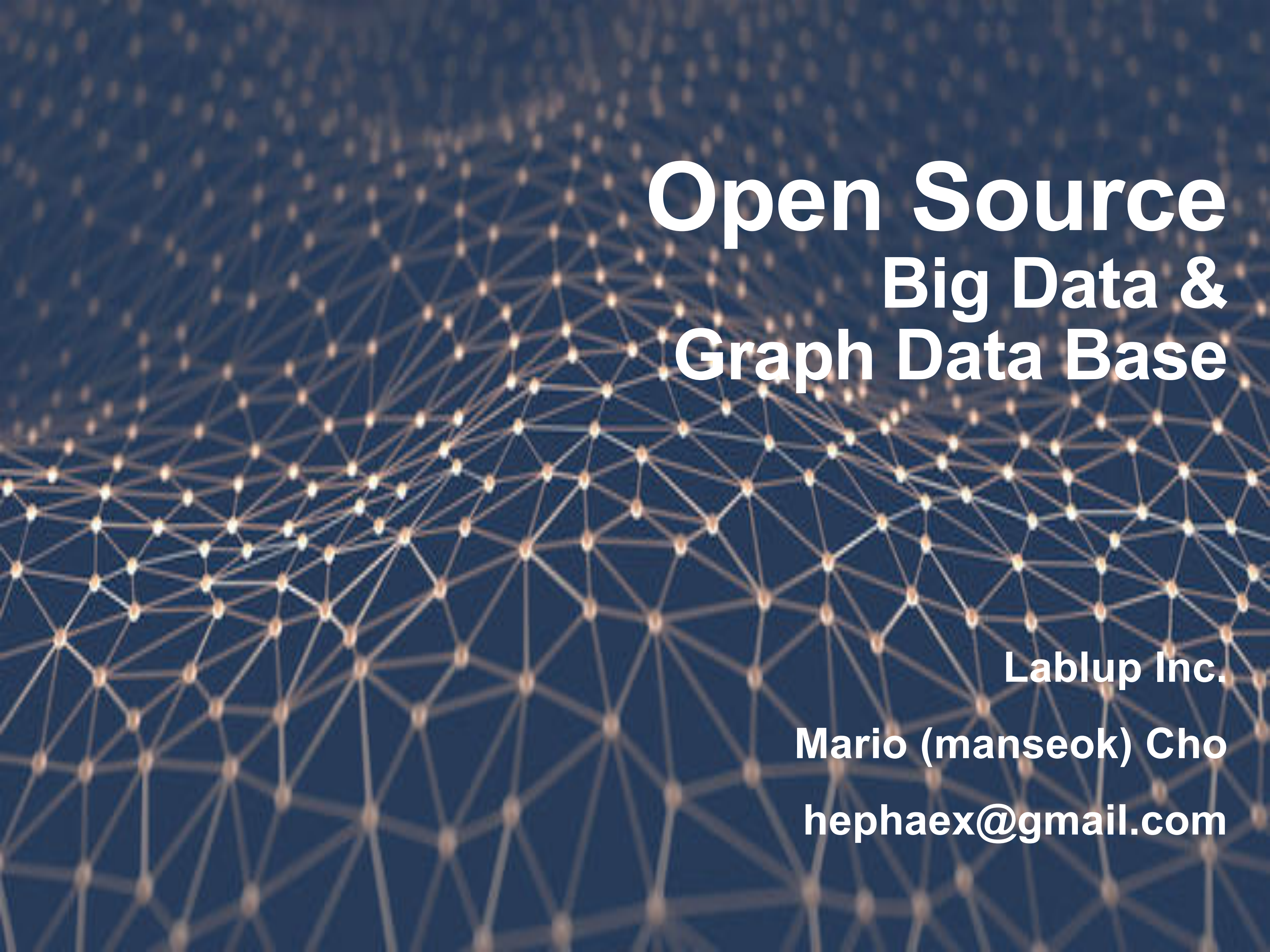


Open Source

Big Data & Graph Data Base

Lablup Inc.

Mario (manseok) Cho

hephaex@gmail.com

Mario (manseok) Cho

Development Experience

- Image Recognition using Neural Network
- Bio-Medical Data Processing
- Human Brain Mapping on High Performance Computing
- Medical Image Reconstruction(Computer Tomography)
- Enterprise System Architect
- Artificial Intelligence for medicine decision support

Open Source Software Developer

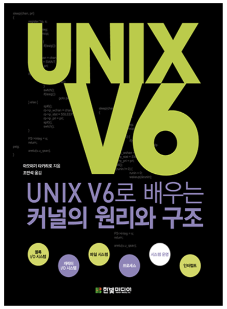
- Committer: (Cloud NFV/SDN)
  OPNFV
- Contribute:
  - TensorFlow (Deep Learning)
  - OpenStack (Cloud compute)
  - LLVM (compiler)
  - Kernel (Linux)

TensorFlow

openstack™ CLOUD SOFTWARE

Book

- Unix V6 Kernel



UNIX V6

UNIX V6로 배우는 커널의 원리와 구조

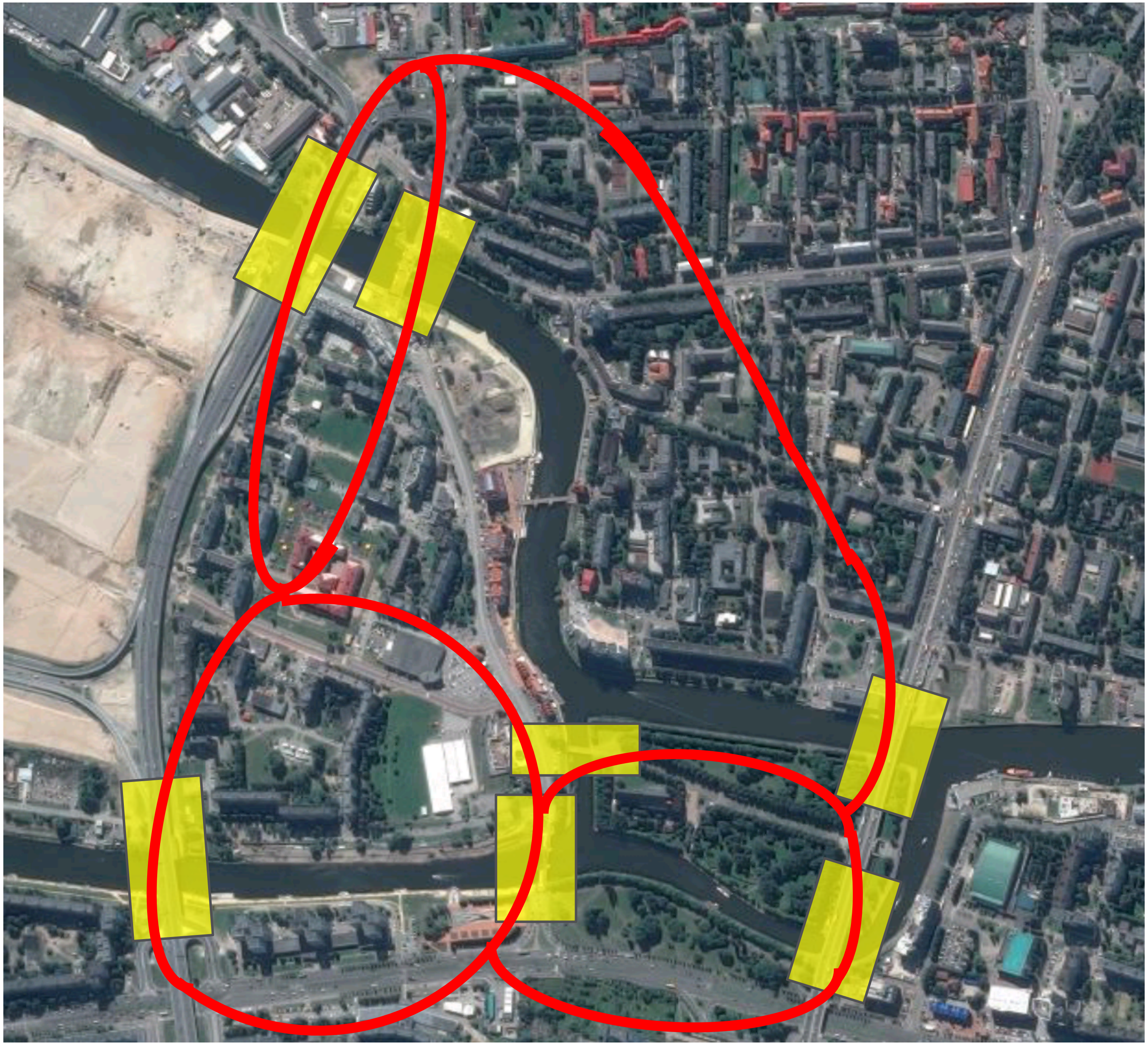
한빛미디어

Lablup Inc.

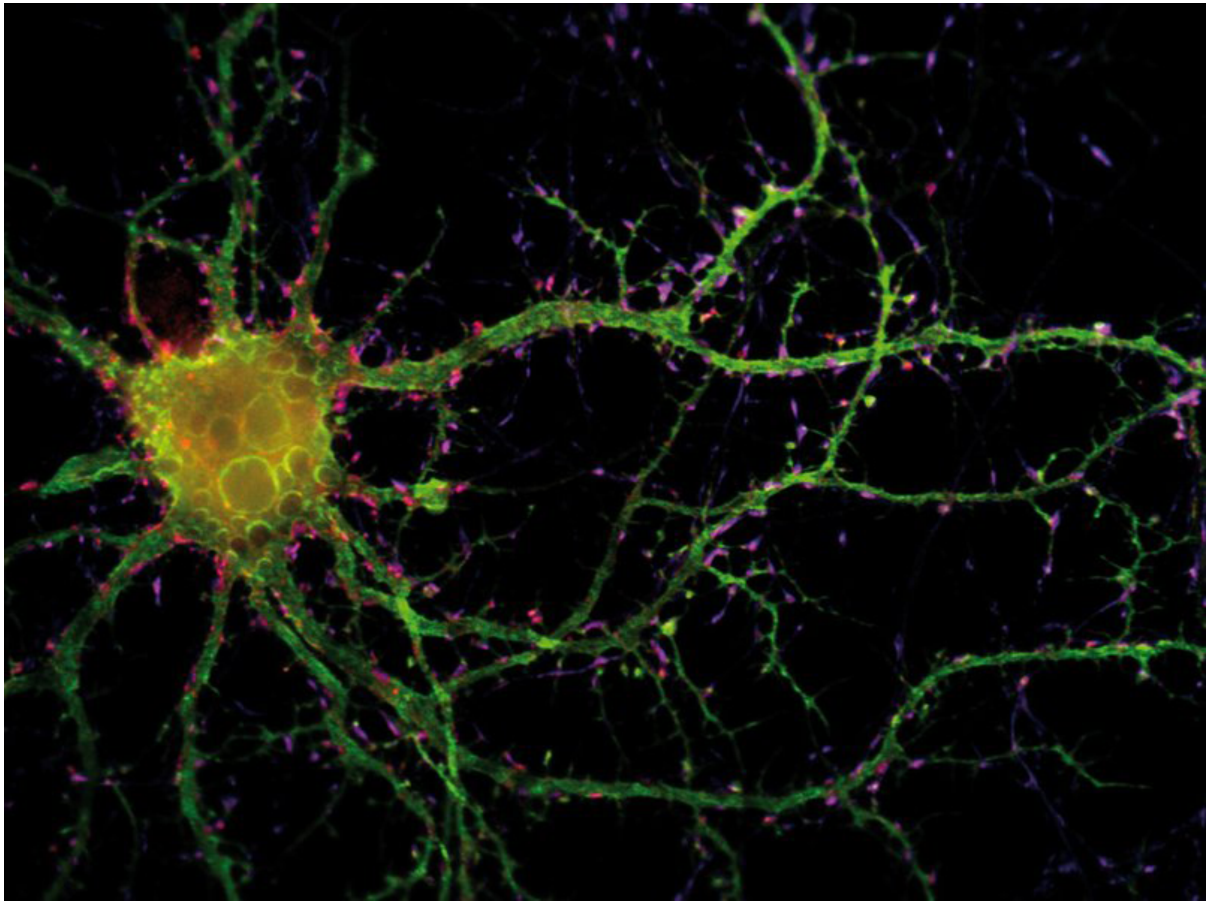
Mario Cho

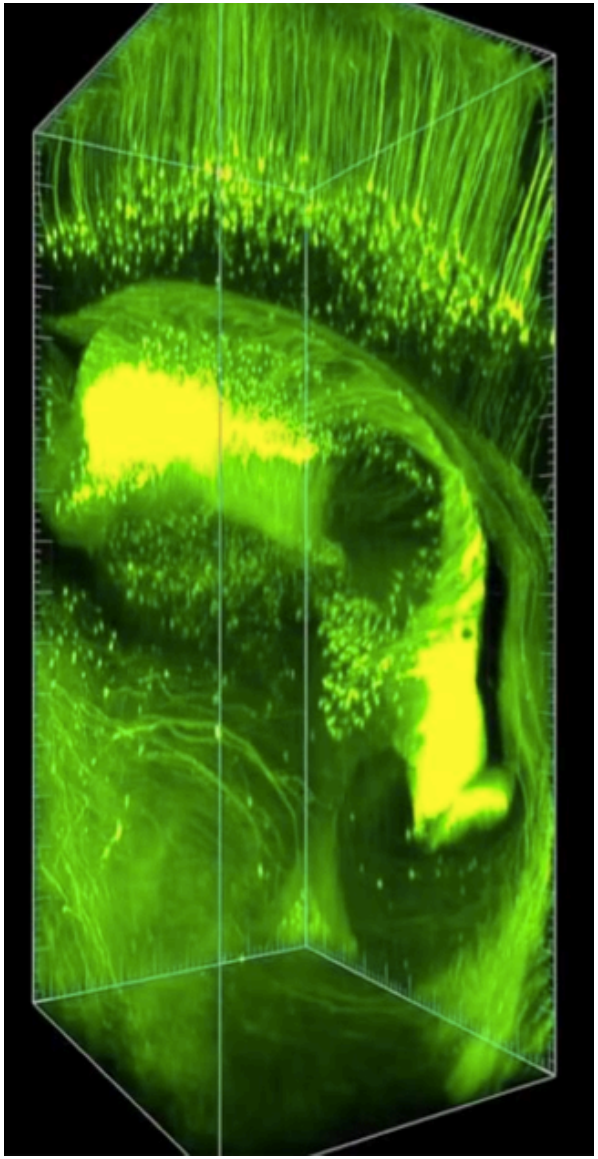
hephaex@gmail.com

7 Bridges of Königsberg *since 1735.*

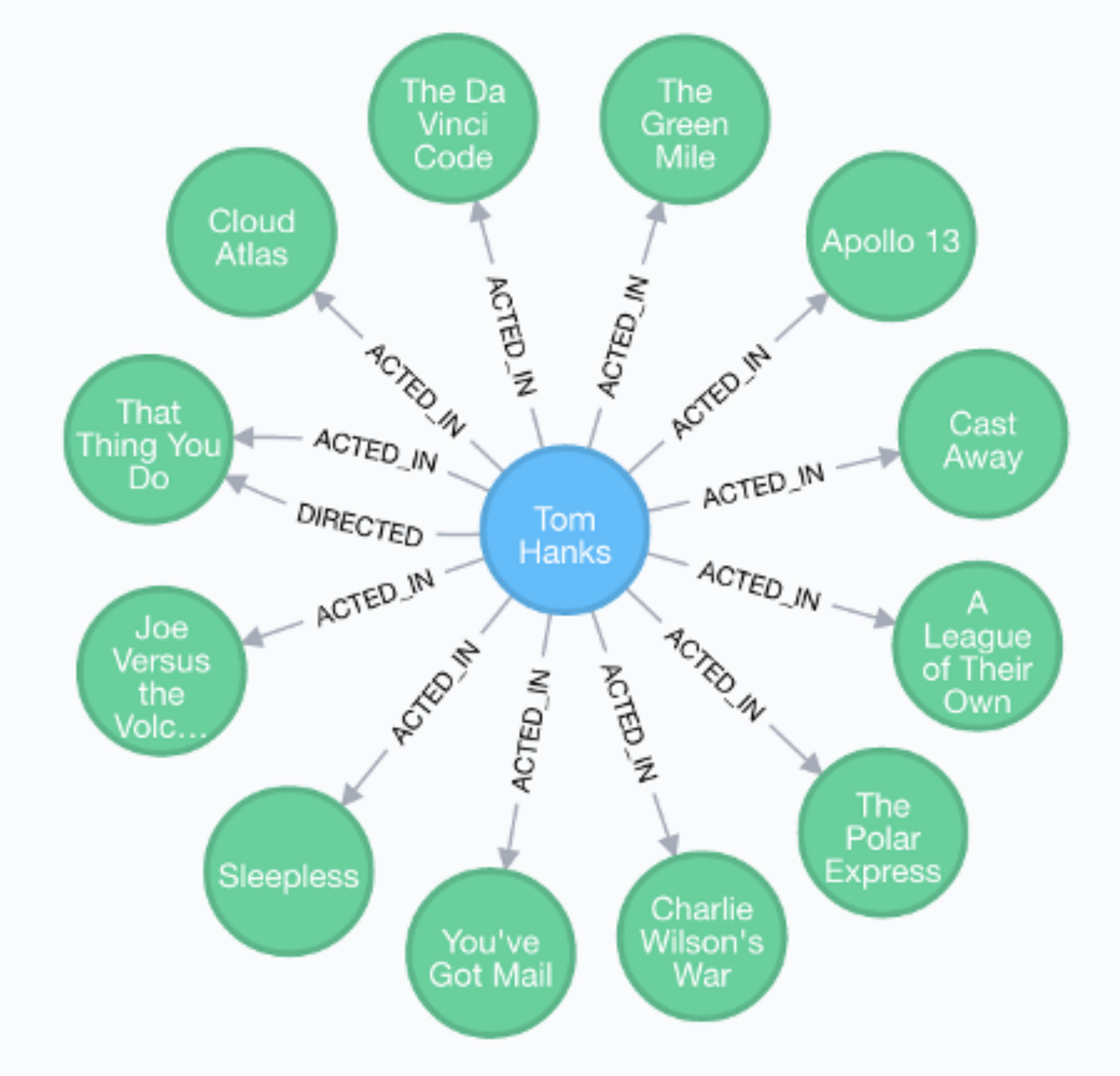


Neural network, Hippocampus



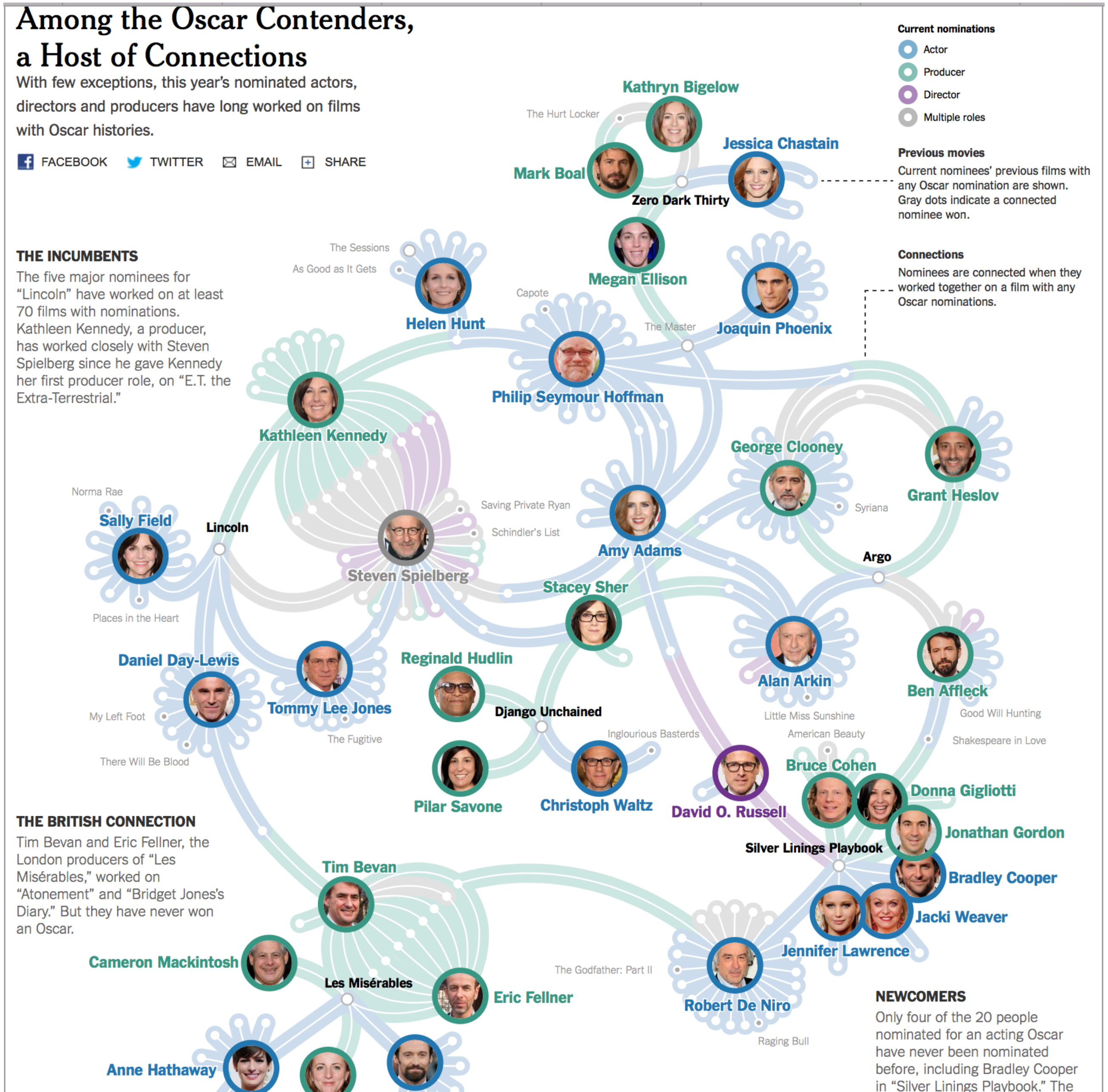


Movie star Social Graph



- Ref: neo4j.com

the Oscar Contenders, a Host of Connections



Among the Oscar Contenders, a Host of Connections

With few exceptions, this year's nominated actors, directors and producers have long worked on films with Oscar histories.

FACEBOOK TWITTER EMAIL SHARE

Current nominations

- ○ Actor
- ○ Producer
- ○ Director
- ○ Multiple roles

Previous movies

Current nominees' previous films with any Oscar nomination are shown. Gray dots indicate a connected nominee won.

Connections

Nominees are connected when they worked together on a film with any Oscar nominations.

THE INCUMBENTS

The five major nominees for "Lincoln" have worked on at least 70 films with nominations. Kathleen Kennedy, a producer, has worked closely with Steven Spielberg since he gave Kennedy her first producer role, on "E.T. the Extra-Terrestrial."

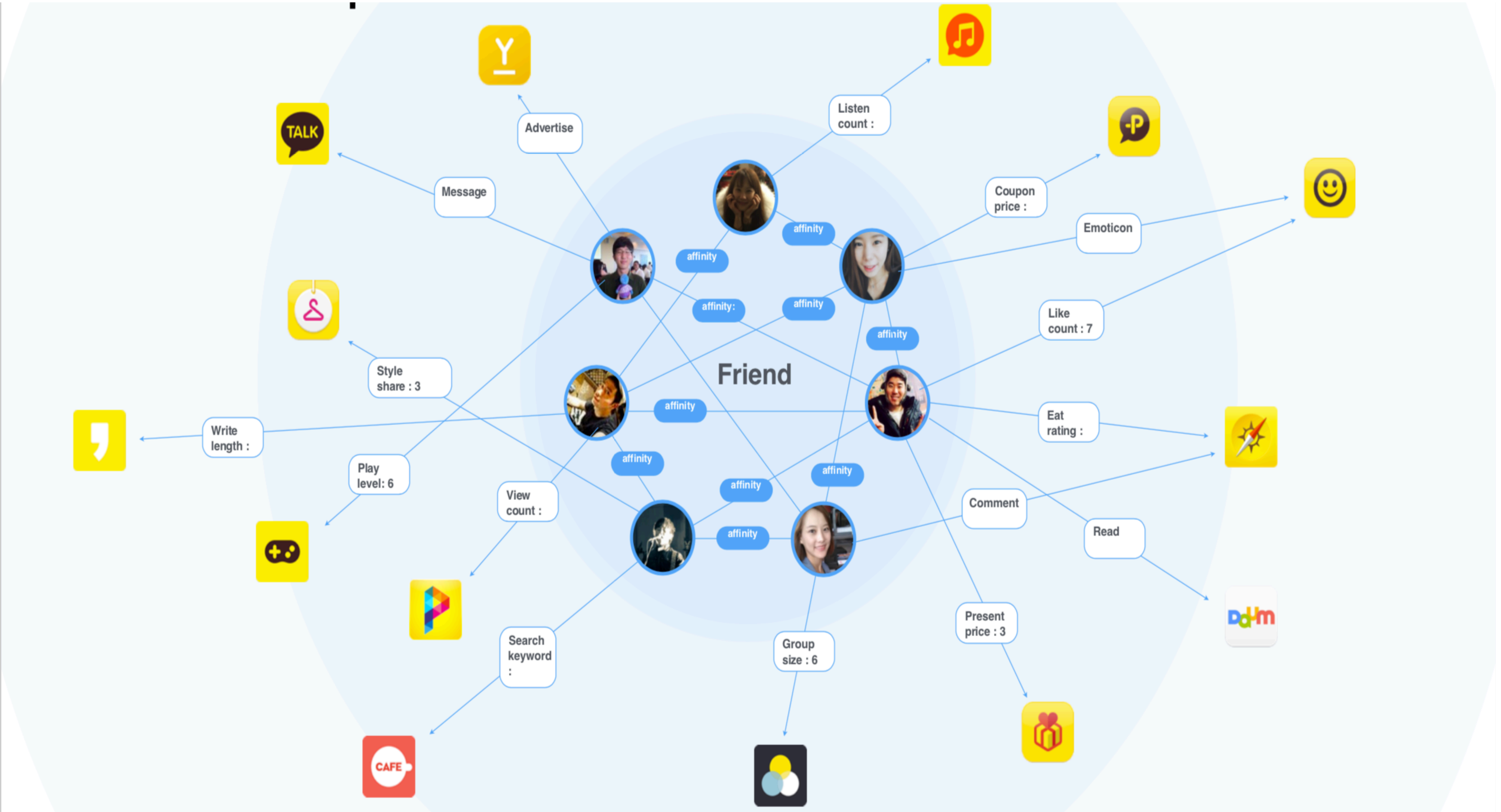
THE BRITISH CONNECTION

Tim Bevan and Eric Fellner, the London producers of "Les Misérables," worked on "Atonement" and "Bridget Jones's Diary." But they have never won an Oscar.

NEWCOMERS

Only four of the 20 people nominated for an acting Oscar have never been nominated before, including Bradley Cooper in "Silver Linings Playbook." The

KakaoTalk: Korean Most Popular SNS

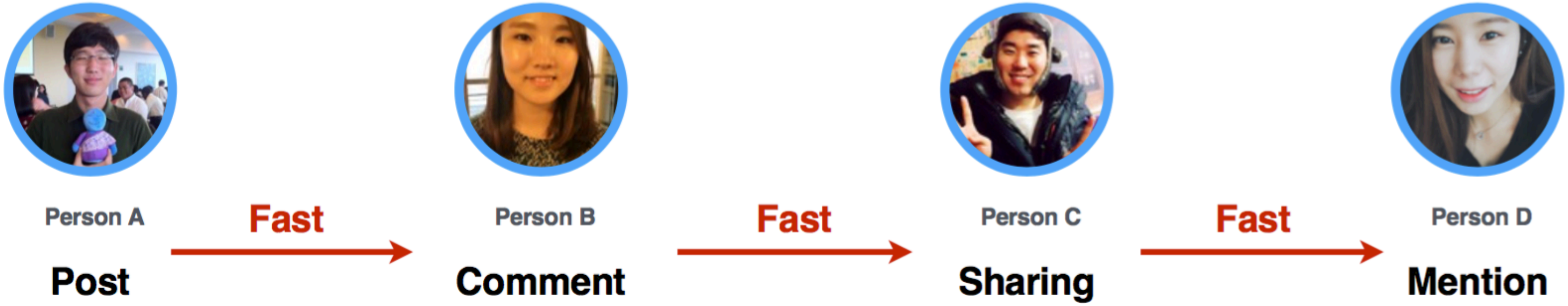


SNS connectivity Technical Challenge

- Scale: Large, constantly changing social graph
    - 200 million users (vertices)
    - It has 10 billion relationships (edge; link; relation)
    - More than 50 million relationships change every day
    - More than 3 billion activities are added daily
- Performance: Width-first search for linked data in real time
    - 65,000 queries per second at peak time
    - Maximum response time 50ms
- Dynamic ranking logic support
    - Push strategy: Difficult to change the ranking logic dynamically.
    - Pull strategy: Various ranking logic can be applied.

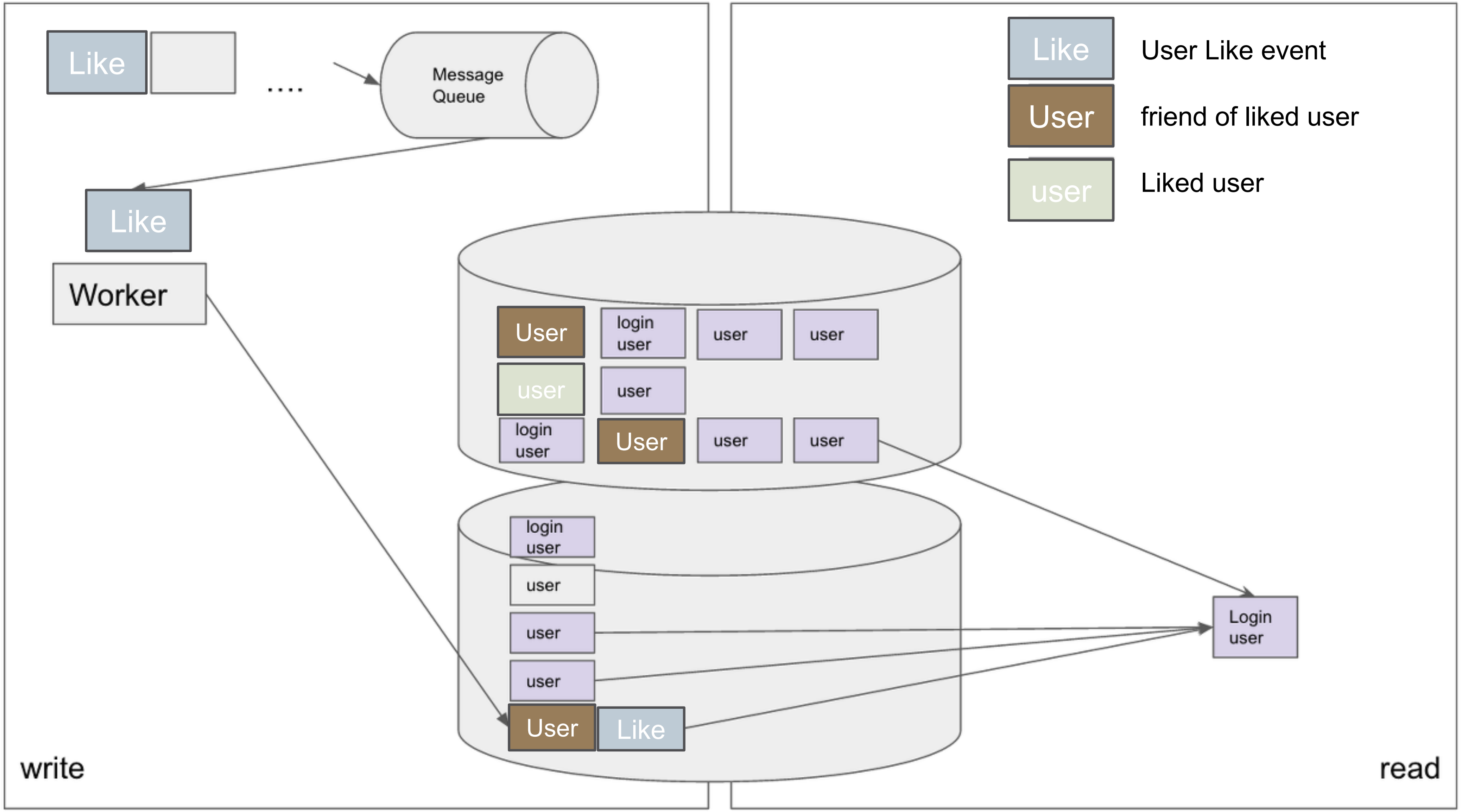
SNS connectivity Technical Challenge

- Real-time updates for viral effects
  - The biggest change brought about by the combination of social and mobile is "real”
  - Data can be analyzed on a daily or hourly basis
  - Recommendations based on the analyzed information can not be adjusted to the speed of users' consumption.
  - If you recommend the news that people have seen many times this time tomorrow,
  - it will become "the last hot news event".
  - If you share the news with your friends
  - Even if it is not hard real-time, it will be possible to handle soft real-time to prevent bloodshed.

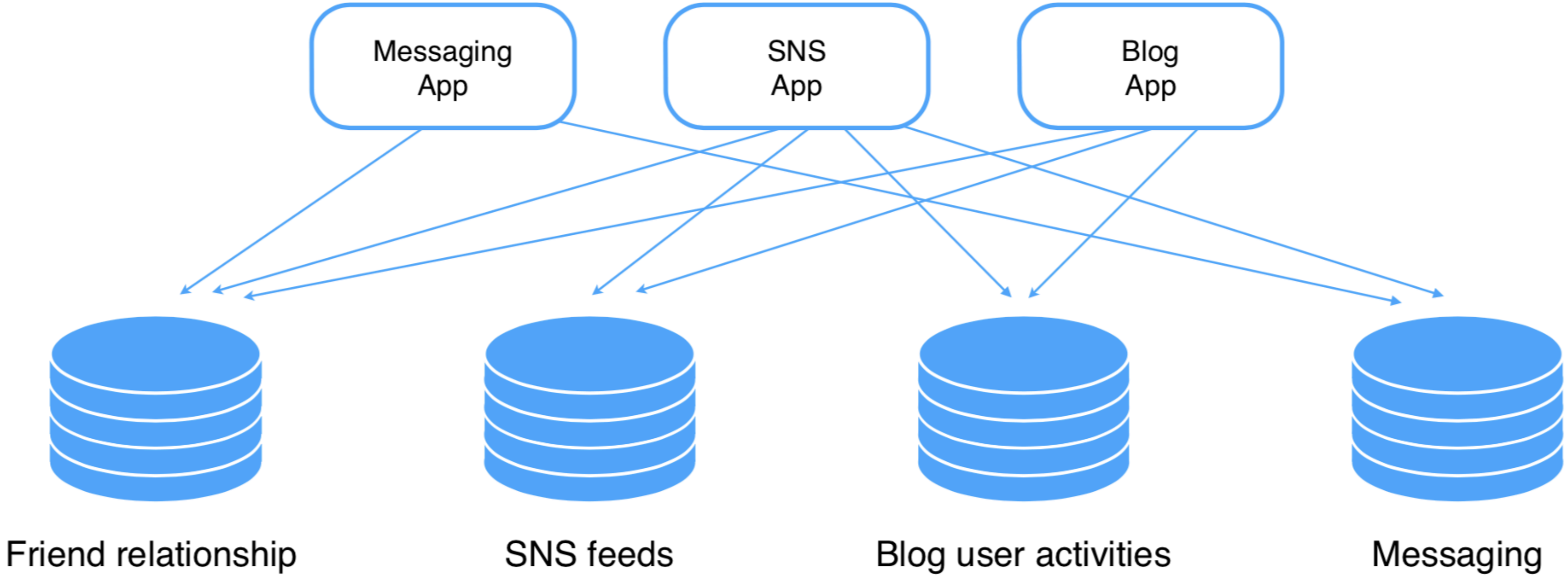


Person A Post → Fast → Person B Comment → Fast → Person C Sharing → Fast → Person D Mention

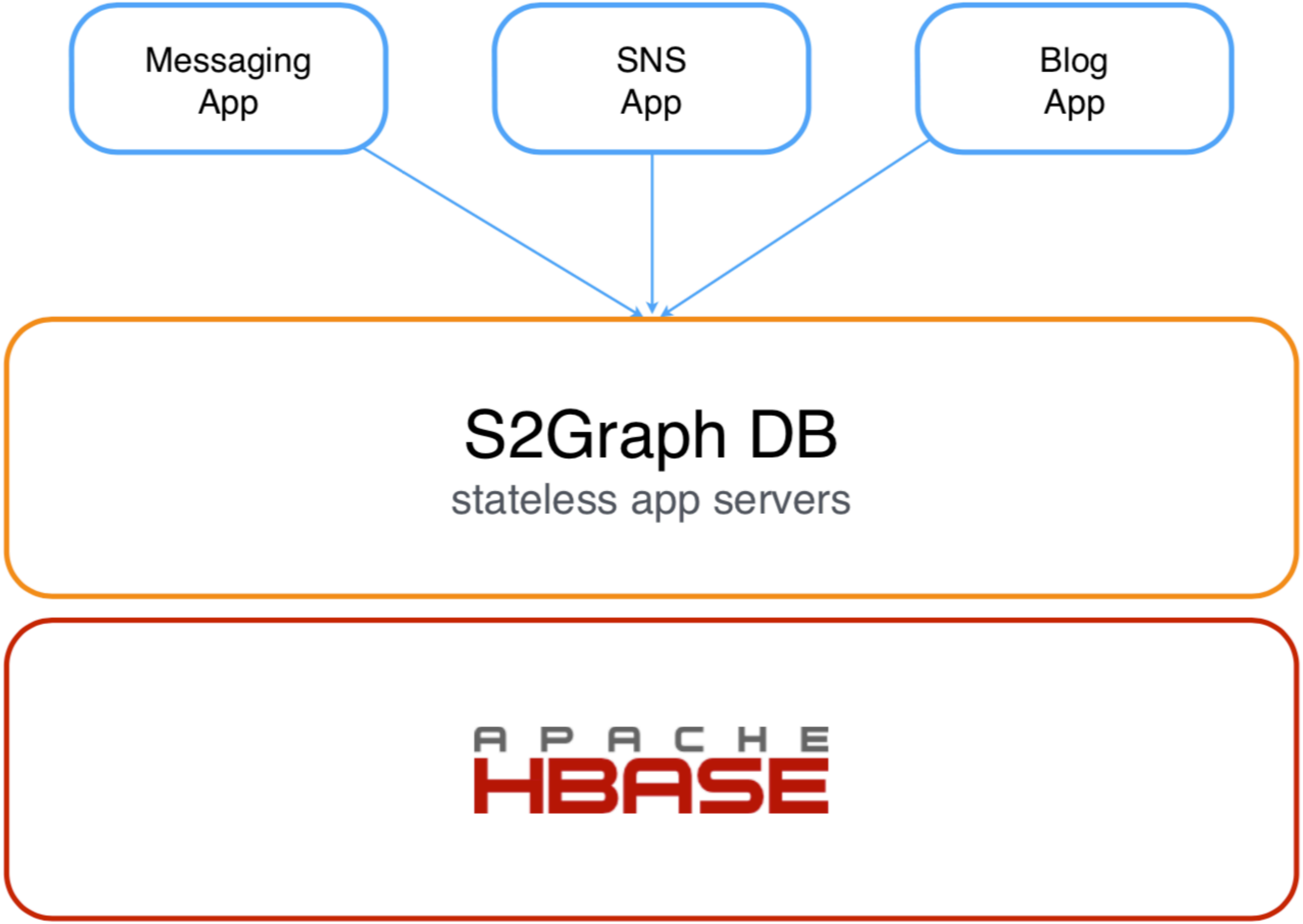
read fanout method



Before S2Graph



After S2Graph



Messaging App

SNS App

Blog App

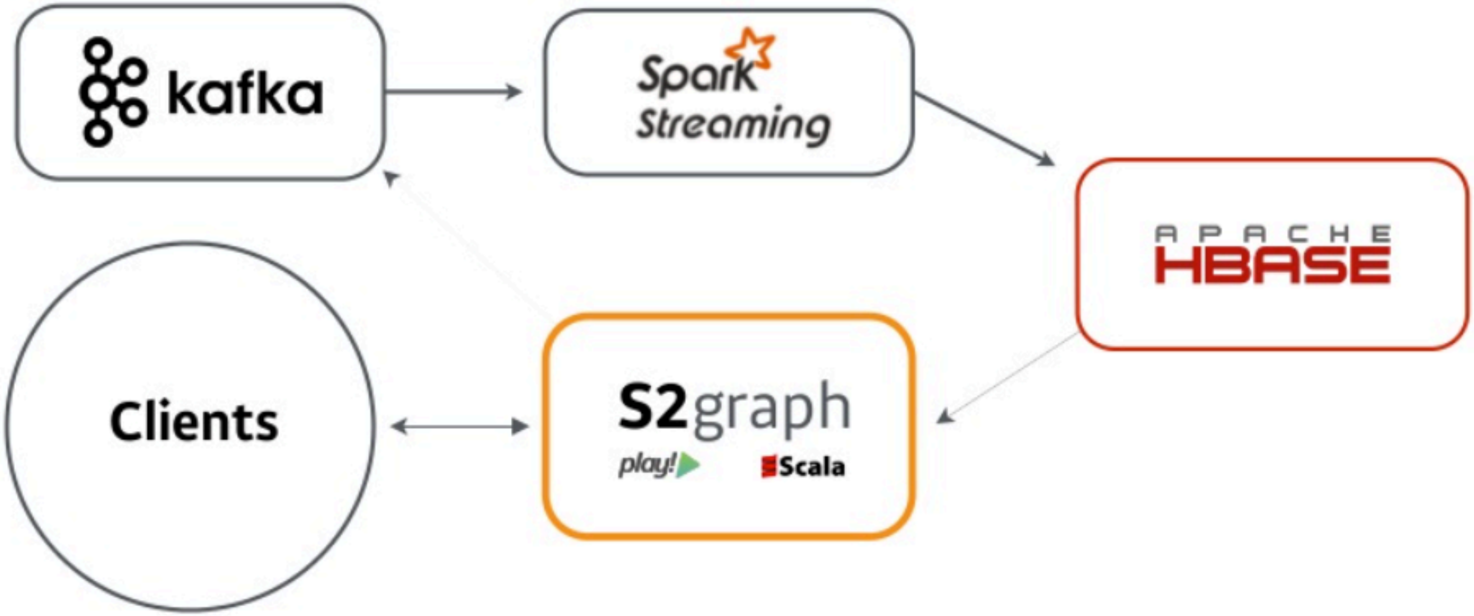
S2Graph DB

stateless app servers

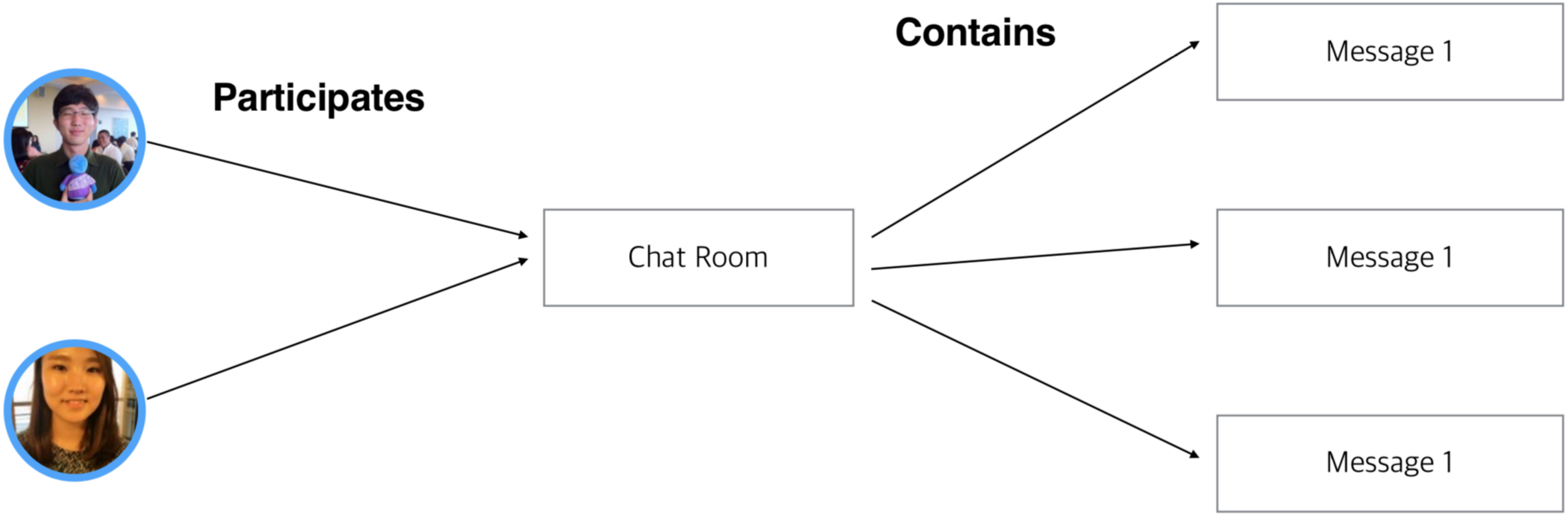
What is S2Graph?

Storage-as-a-Service + Graph API = Real time Breadth First Search

S2Graph is Not
Not support global computation(not like Apache Giraph, graphX).
Not support graph algorithm like page rank, shortest path.

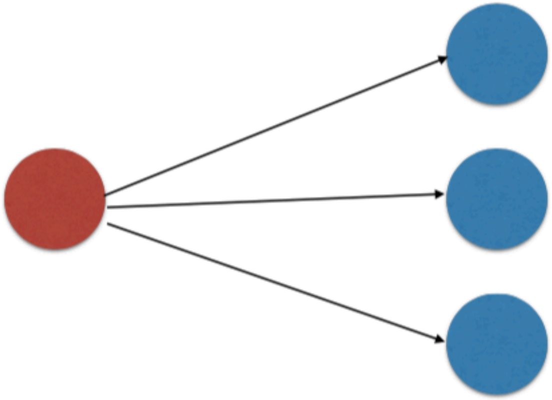


Messenger Data Model



S2Graph API: Vertex

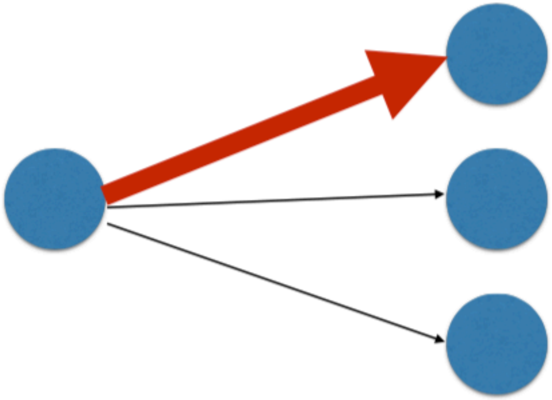
- **Vertex:**
  1. insert, delete, getVertex
- **2. vertex id: what user**
- **provided(string/int/long)**



ID	1231-123
Prop1	Val1
Prop2	Val2
…	…

S2Graph API: Edge

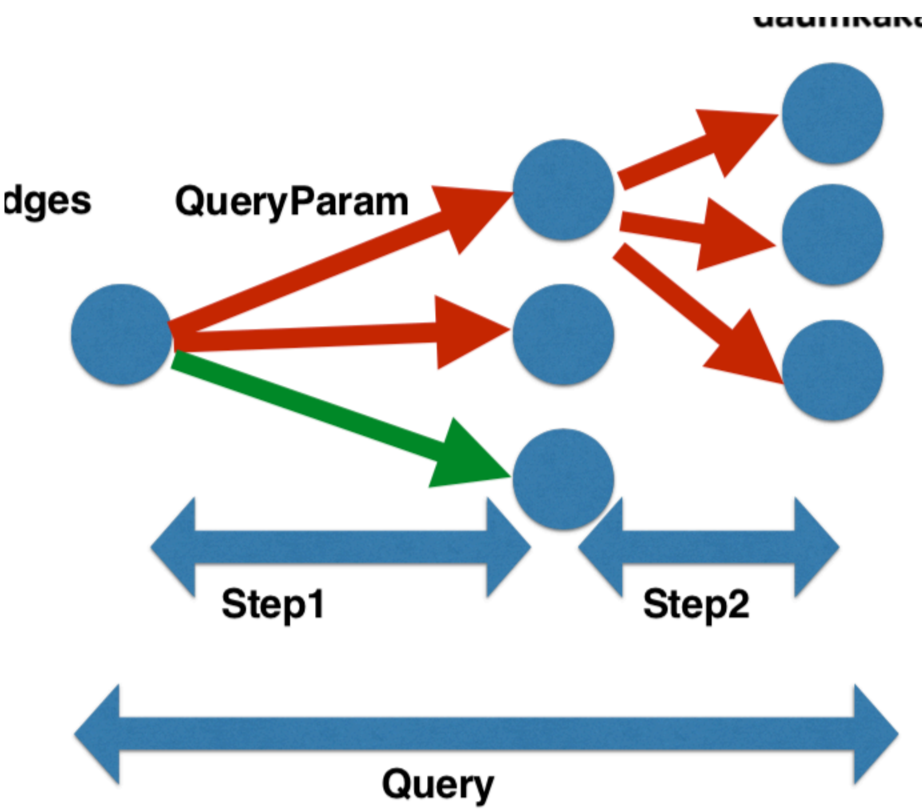
- Edges:
- 1. Insert,delete,update,getEdge(like CRUD in RDBMS)
- 2. Edgereference:(from,to,label, direction)
- 3. Multiplepropsonedge.
  4. Everyedgesareordered(details
- follow).



Edge Reference	1,101,”friend”,”out”
Prop1	Val1
Prop2	Val2
…	…

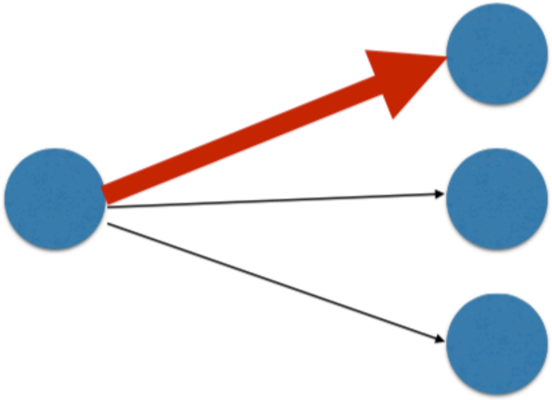
S2Graph API: Query

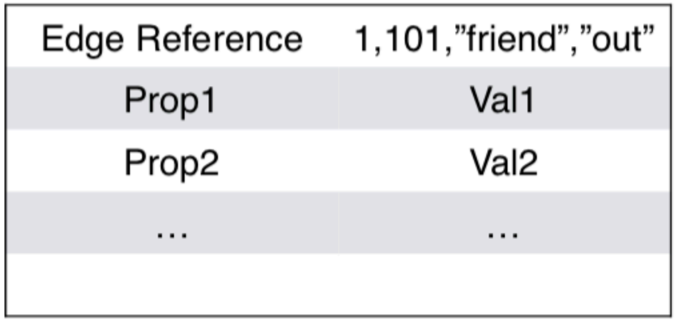
- Query: getEdges, countEdges, removeEdges
- Class Query {
  // Define breadth first search
  List[VertexId] startVertices; List[Step] steps;
- }
  Class Step {
- // Define one breadth
- List[QueryParam] queryParams; }
- Class QueryParam {
  // Define each edges to traverse for current
- breadth
  String label; String direction; Map options;
- }



S2Graph API: indices

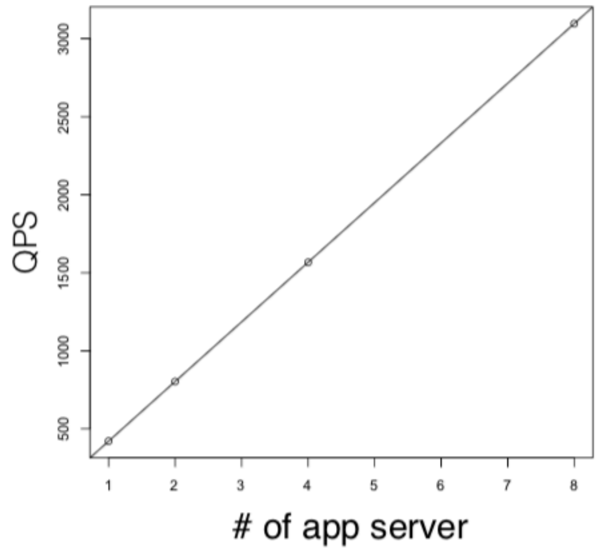
- Indices
- **1. addIndex, createIndex**
- 2.**Automatically keep edges ordered for multiple indices.**
- **3. Support int/long/float/string data types.**
- **class Index {**
  // define how to order edges. String indexName; List[Prop] indexProps;
- **}**

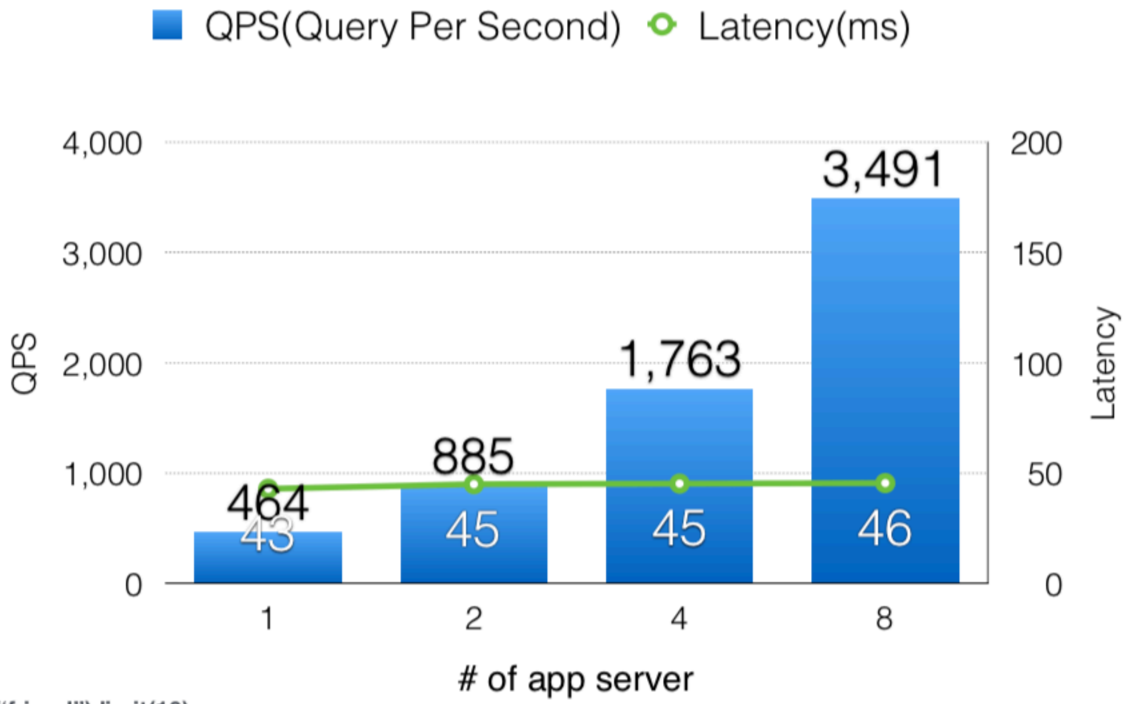




Edge Reference	1,101,"friend","out"
Prop1	Val1
Prop2	Val2
…	…

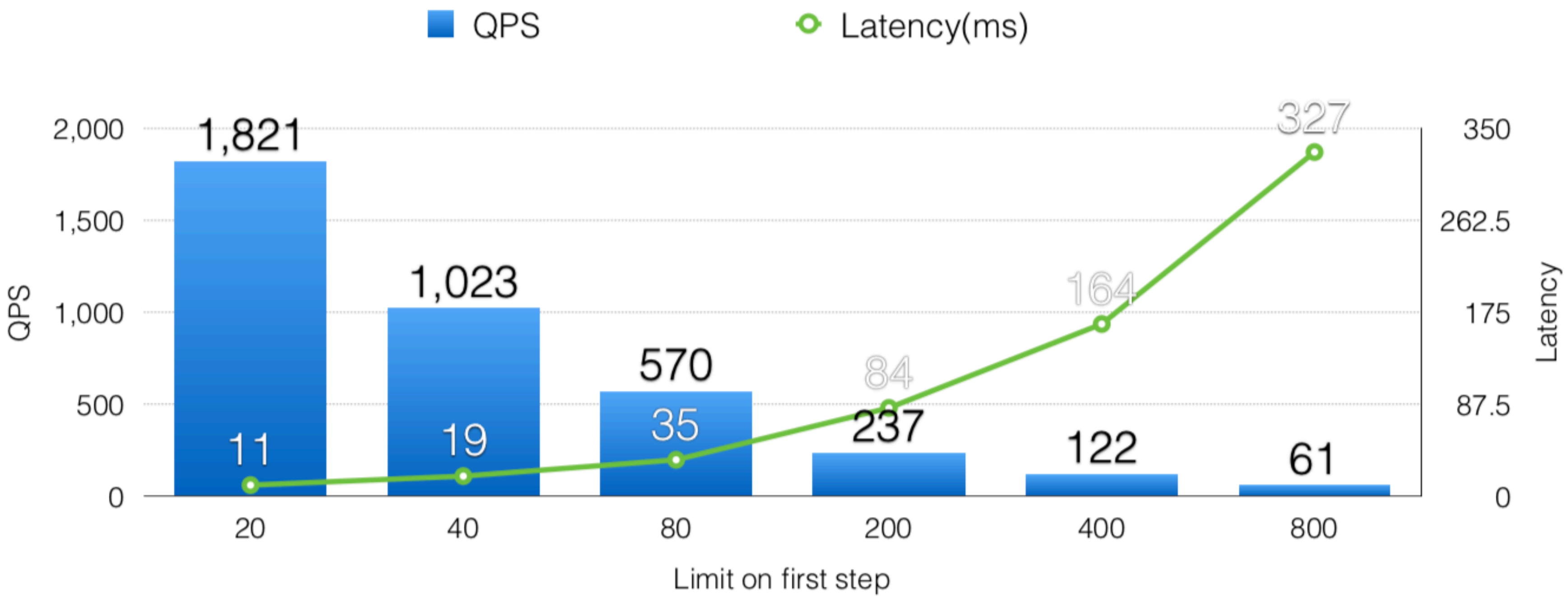
1. Linear scalability





- Benchmark Query : src.out("friend").limit(100).out("friend").limit(10)
- Total concurrency: 20 * # of app server

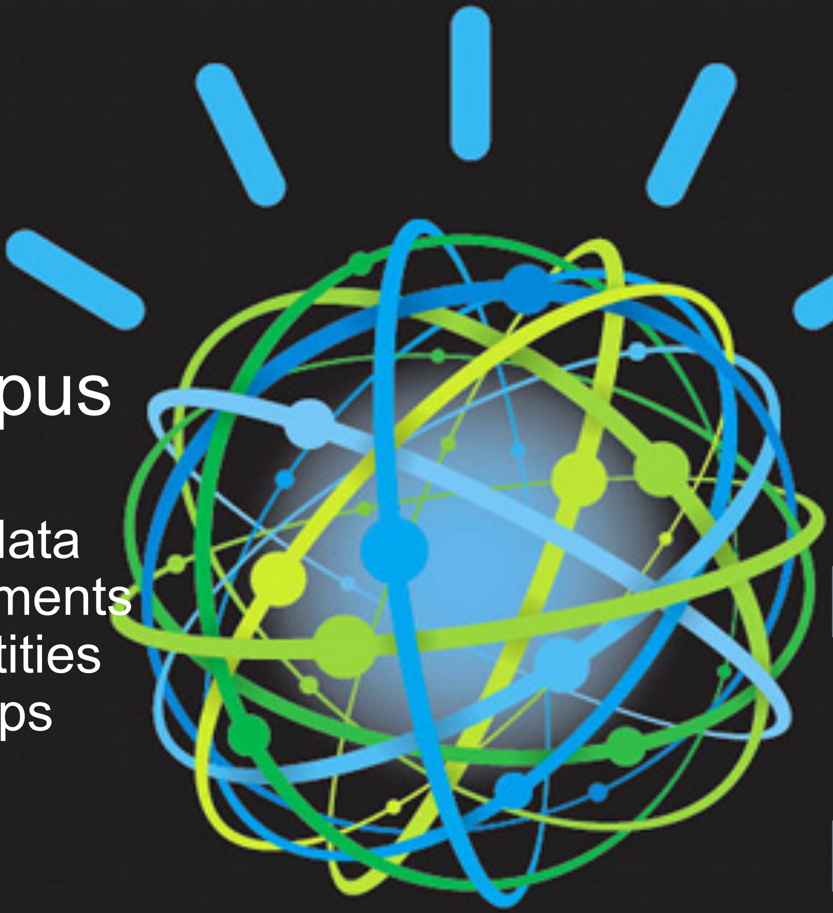
2. Varying width of traverse (tested with a single server)



- Benchmark Query : src.out(“friend”).limit(x).out(“friend”).limit(10)
- Total concurrency = 20 * 1(# of app server)

Watson Discovery Advisor

- Researches can’t innovate fast enough to create truly breakthrough therapies
- To anticipate the safety profile of new treatments

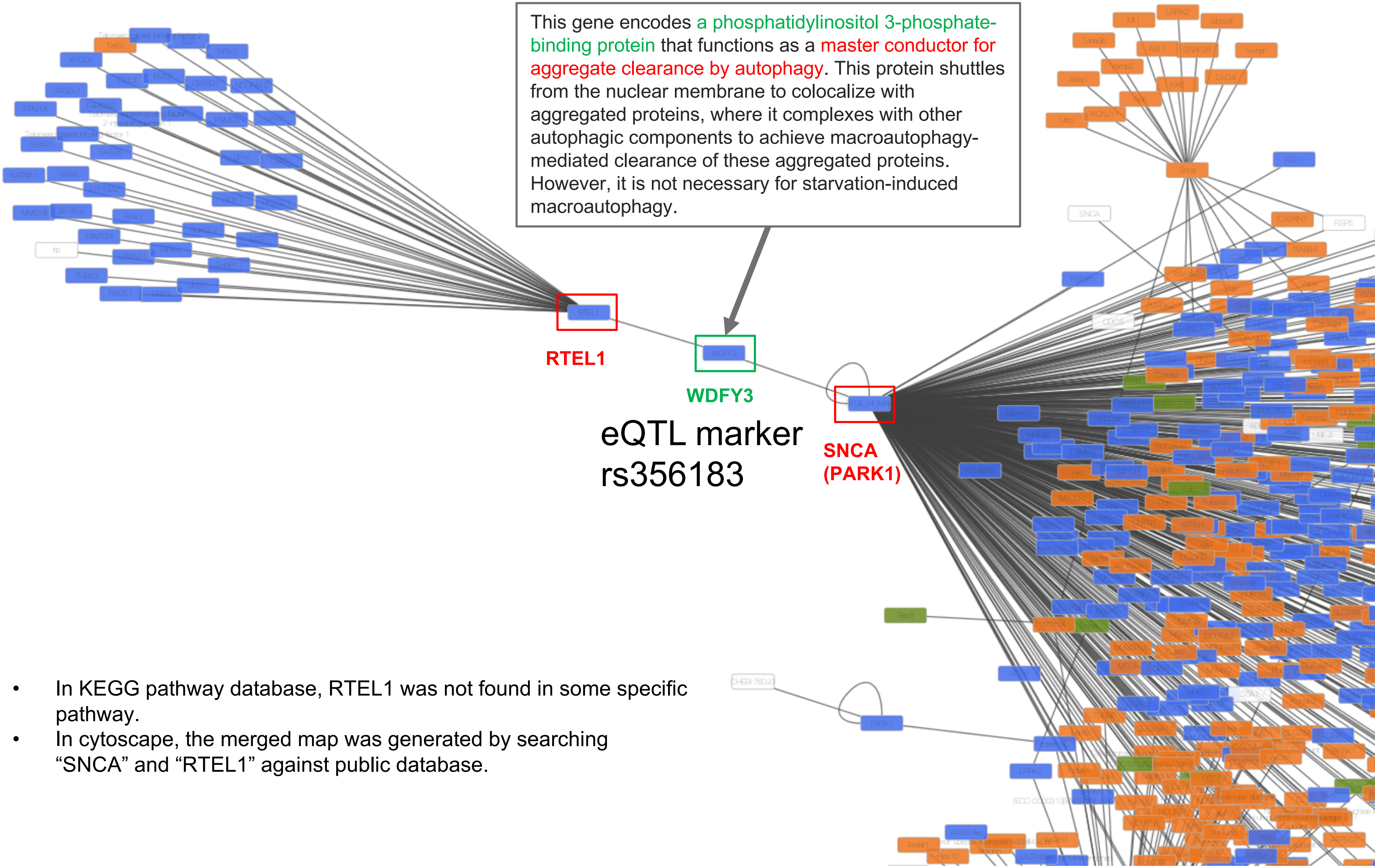


Watson Corpus

Over 1TB of data
Over 40m documents
Over 100m entities
& relationships

Chemical	12M+ Chemical Structures
Genomics	20,000+ genes
MD Text	50+ books
Medline	23M+ abstracts
Journals	100+ journals
FDA drugs	11,000+ drugs
Patents	16M+ patents

Bio Connectivity using S2Graph



This gene encodes a phosphatidylinositol 3-phosphate-binding protein that functions as a master conductor for aggregate clearance by autophagy. This protein shuttles from the nuclear membrane to colocalize with aggregated proteins, where it complexes with other autophagic components to achieve macroautophagy-mediated clearance of these aggregated proteins. However, it is not necessary for starvation-induced macroautophagy.

RTEL1

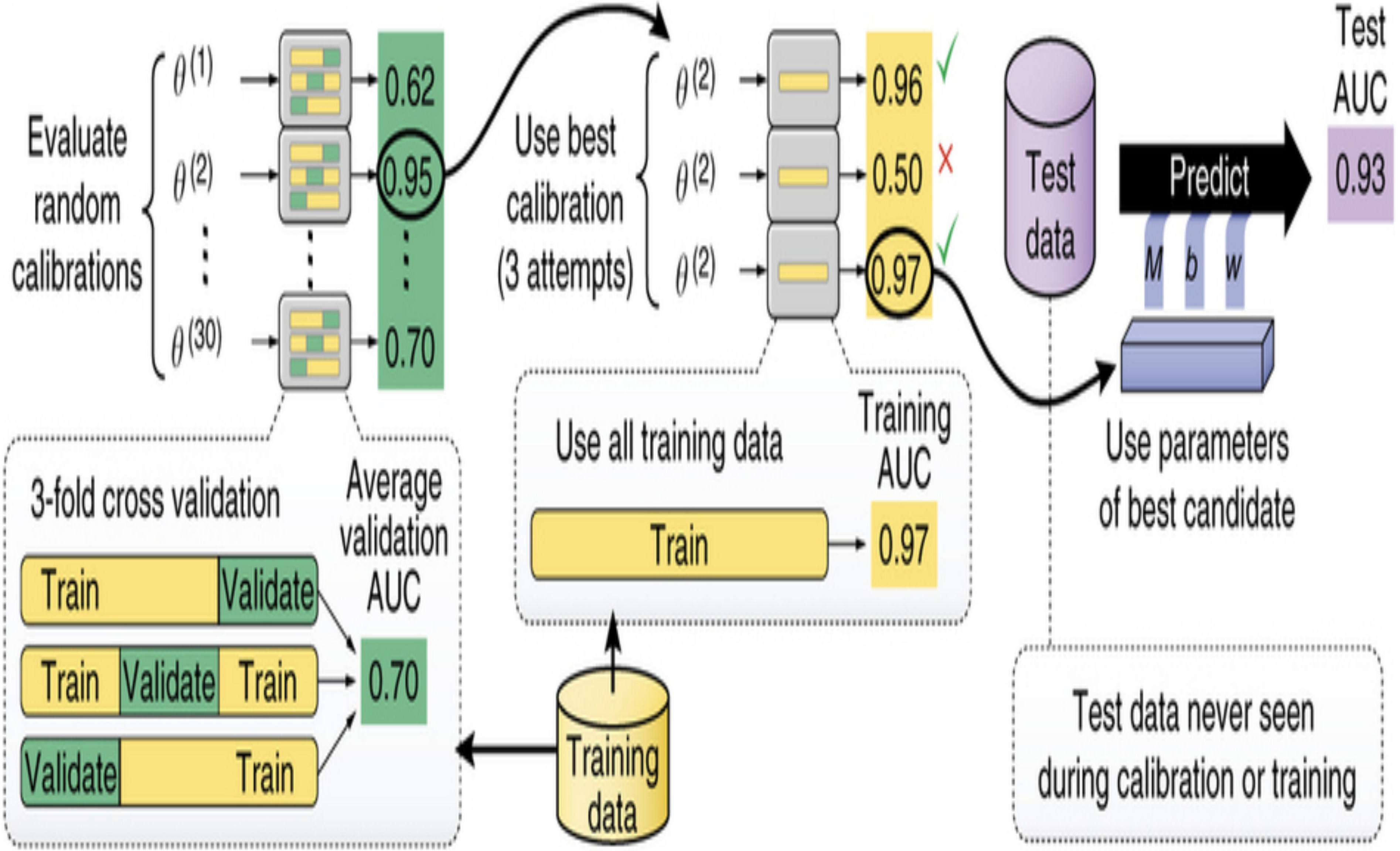
WDFY3

eQTL marker rs356183

SNCA (PARK1)

- In KEGG pathway database, RTEL1 was not found in some specific pathway.
- In cytoscape, the merged map was generated by searching "SNCA" and "RTEL1" against public database.

Genomics A.I.



Evaluate random calibrations
$\theta^{(1)}$
$\theta^{(2)}$
$\theta^{(30)}$
0.62
0.95
0.70
Use best calibration (3 attempts)
$\theta^{(2)}$
$\theta^{(2)}$
$\theta^{(2)}$
0.96 ✓
0.50 ✗
0.97 ✓
Test data
Predict
M
b
w
Test AUC
0.93
Use parameters of best candidate
3-fold cross validation
Average validation AUC
Train
Validate
Train
Validate
Train
Validate
Train
0.70
Use all training data
Training AUC
Train
0.97
Training data
Test data never seen during calibration or training

Thanks you!

Q&A