G* The Dynamic Graph Database

Introduction: Why Graphs?

Alan G. Labouseur, Ph.D.

Marist College — alan.labouseur@Marist.edu
3NF Consulting, Inc. — alan@3NFconsulting.com

Daily Deluge of Data

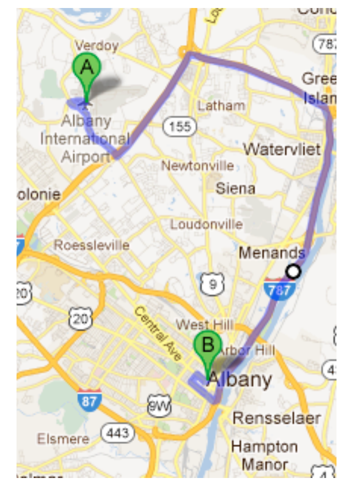
Today, we are awash in a daily deluge of data.

- It's constantly growing.
- It's constantly changing.
- Some is structured.
- Some is unstructured.
- Some is semi-structured.

Piling up data is easy.

Gaining insight from the data pile is hard.

This is one of many challenges of Big Data.



Structured Data

Some data is structured.

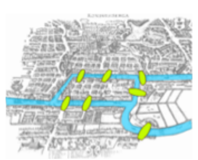
	doctor unknown	did integer	is_treating unknown	pid integer
1	doctor	1	is treating	0
2	doctor	1	is treating	2
3	doctor	1	is treating	3
4	doctor	1	is treating	4
5	doctor	1	is treating	5
6	doctor	1	is treating	6
7	doctor	1	is treating	7
8	doctor	1	is treating	9
9	doctor	1	is treating	10
10	doctor	1	is treating	11
11	doctor	1	is treating	12
12	doctor	1	is treating	14
13	doctor	1	is treating	15
14	doctor	1	is treating	17
15	doctor	1	is treating	18
16	doctor	1	is treating	19
17	doctor	1	is treating	20
18	doctor	1	is treating	21
19	doctor	1	is treating	24
20	doctor	1	is treating	25
21	doctor	1	is treating	27
22	doctor	1	is treating	28
23	doctor	1	is treating	30
24	doctor	1	is treating	31
25	doctor	1	is treating	41
26	doctor	1	is treating	42
27	doctor	1	is treating	43
28	doctor	1	is treating	44
29	doctor	1	is treating	51
30	doctor	1	is treating	54
31	doctor	1	is treating	59
32	doctor	1	is treating	60
33	doctor	2	is treating	0
34	doctor	2	is treating	8
35	doctor	2	is treating	33
36	doctor	2	is treating	35
37	doctor	4	is treating	0
38	doctor	4	is treating	38
39	doctor	4	is treating	50
40	doctor	4	is treating	53
41	doctor	6	is treating	22
42	doctor	6	is treating	49
43	doctor	6	is treating	57
44	doctor	8	is treating	8
45	doctor	8	is treating	23
46	doctor	8	is treating	63
47	doctor	9	is treating	37
48	doctor	9	is treating	40
49	doctor	9	is treating	47
50	doctor	11	is treating	16
51	doctor	11	is treating	26
52	doctor	11	is treating	50
53	doctor	12	is treating	13
54	doctor	12	is treating	34
55	doctor	12	is treating	47
56	doctor	16	is treating	45
57	doctor	18	is treating	29
58	doctor	23	is treating	29
59	doctor	23	is treating	32
60	doctor	25	is treating	36
61	doctor	27	is treating	39
62	doctor	30	is treating	58
63	doctor	33	is treating	0
64	doctor	33	is treating	46
65	doctor	33	is treating	47
66	doctor	33	is treating	49
67	doctor	33	is treating	62
68	doctor	38	is treating	56
69	doctor	39	is treating	48
70	doctor	39	is treating	50
71	doctor	39	is treating	64
72	doctor	42	is treating	52
73	doctor	47	is treating	55
74	doctor	49	is treating	64

[Un | Semi] structured Data

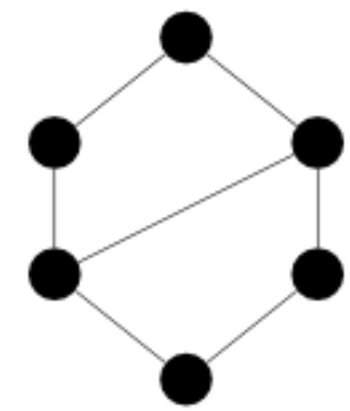
Some data is unstructured or semi-structured.

Graph Structured Data

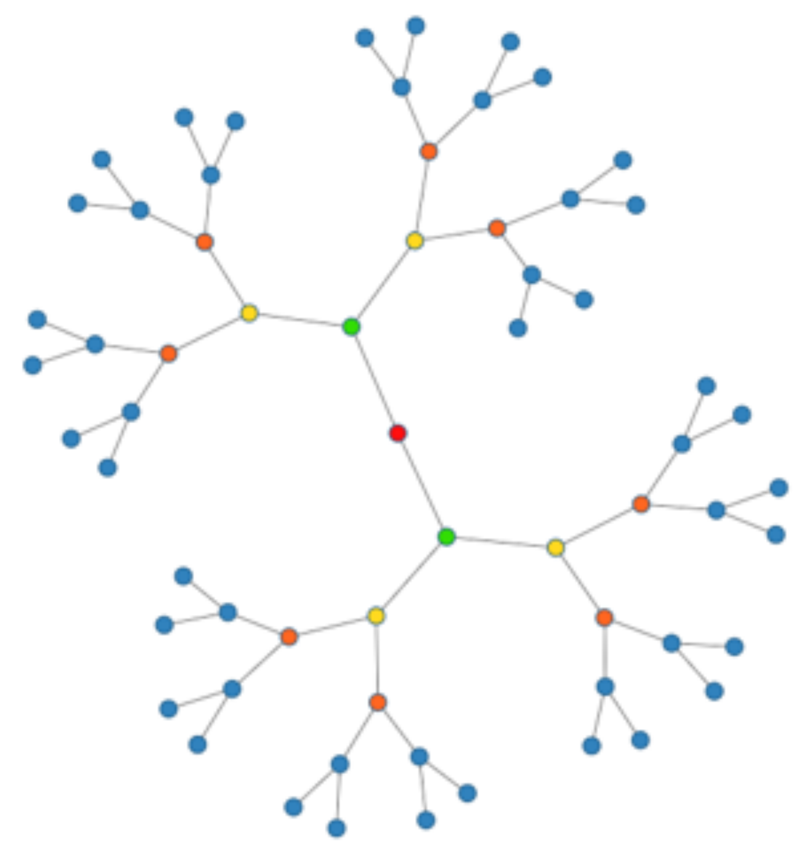
Virtually all data (structured, unstructured, and semi-structured) can be modeled as a network.



A network, or **graph**, is a collection of dots (“vertices") and lines (“edges”) connecting those dots.



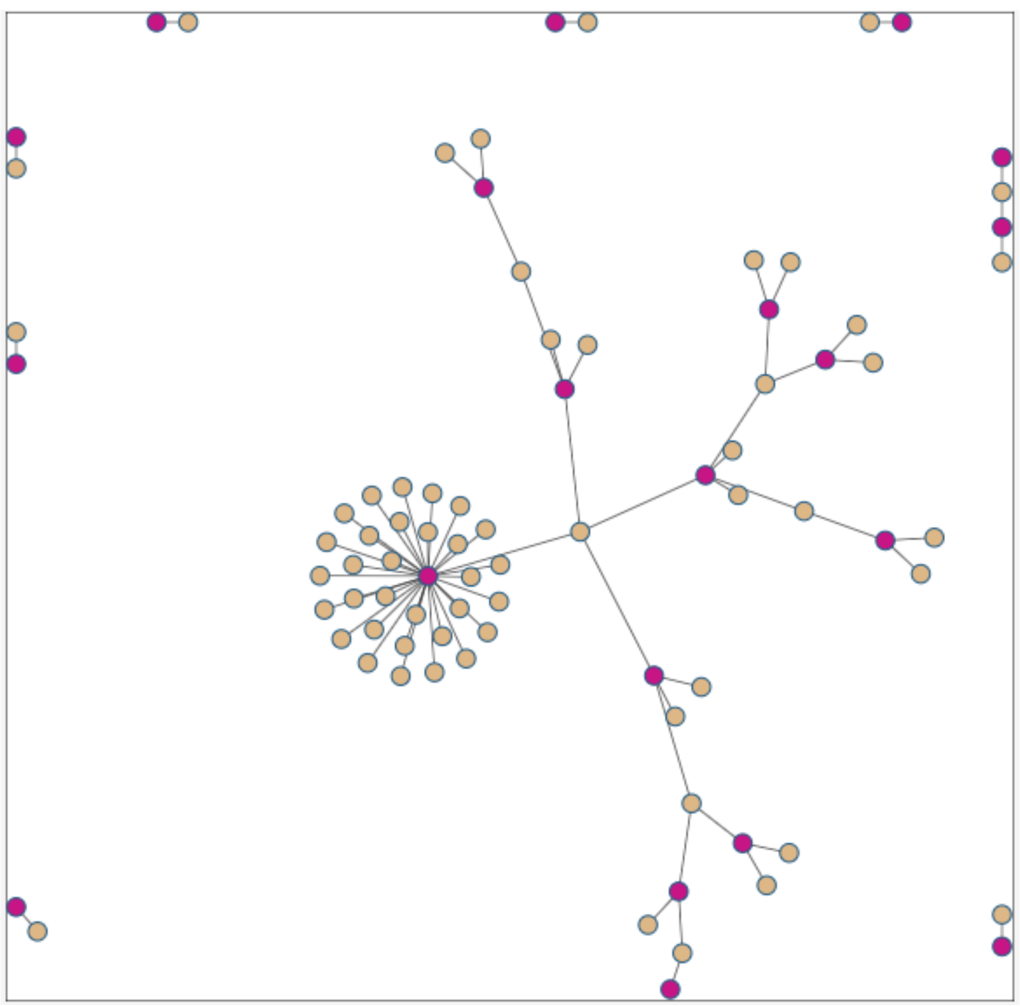
A graph with six (6) vertices and seven (7) edges



A graph with 63 vertices and 62 edges

Graph Structured Data

Graphs can show multi-dimensional relationships that are difficult to see in other data models.



	doctor unknown	did integer	is_treating unknown	pid integer
1	doctor	1	is treating	0
2	doctor	1	is treating	2
3	doctor	1	is treating	3
4	doctor	1	is treating	4
5	doctor	1	is treating	5
6	doctor	1	is treating	6
7	doctor	1	is treating	7
8	doctor	1	is treating	9
9	doctor	1	is treating	10
10	doctor	1	is treating	11
11	doctor	1	is treating	12
12	doctor	1	is treating	14
13	doctor	1	is treating	15
14	doctor	1	is treating	17
15	doctor	1	is treating	18
16	doctor	1	is treating	19
17	doctor	1	is treating	20
18	doctor	1	is treating	21
19	doctor	1	is treating	24
20	doctor	1	is treating	25
21	doctor	1	is treating	27
22	doctor	1	is treating	28
23	doctor	1	is treating	30
24	doctor	1	is treating	31
25	doctor	1	is treating	41
26	doctor	1	is treating	42
27	doctor	1	is treating	43
28	doctor	1	is treating	44
29	doctor	1	is treating	51
30	doctor	1	is treating	54
31	doctor	1	is treating	59
32	doctor	1	is treating	60
33	doctor	2	is treating	0
34	doctor	2	is treating	8
35	doctor	2	is treating	33
36	doctor	2	is treating	35
37	doctor	4	is treating	0
38	doctor	4	is treating	38
39	doctor	4	is treating	50
40	doctor	4	is treating	53
41	doctor	6	is treating	22
42	doctor	6	is treating	49
43	doctor	6	is treating	57
44	doctor	8	is treating	8
45	doctor	8	is treating	23
46	doctor	8	is treating	63
47	doctor	9	is treating	37
48	doctor	9	is treating	40
49	doctor	9	is treating	47
50	doctor	11	is treating	16
51	doctor	11	is treating	26
52	doctor	11	is treating	50
53	doctor	12	is treating	13
54	doctor	12	is treating	34
55	doctor	12	is treating	47
56	doctor	16	is treating	45
57	doctor	18	is treating	29
58	doctor	23	is treating	29
59	doctor	23	is treating	32
60	doctor	25	is treating	36
61	doctor	27	is treating	39
62	doctor	30	is treating	58
63	doctor	33	is treating	0
64	doctor	33	is treating	46
65	doctor	33	is treating	47
66	doctor	33	is treating	49
67	doctor	33	is treating	62
68	doctor	38	is treating	56
69	doctor	39	is treating	48
70	doctor	39	is treating	50
71	doctor	39	is treating	64
72	doctor	42	is treating	52
73	doctor	47	is treating	55
74	doctor	49	is treating	64

The same data set shown at left as a graph, and at right as tables of rows and columns.

Big Data Challenges

Big Data is always **Growing**
- more sources
- more products and services
- more messages
- more transactions

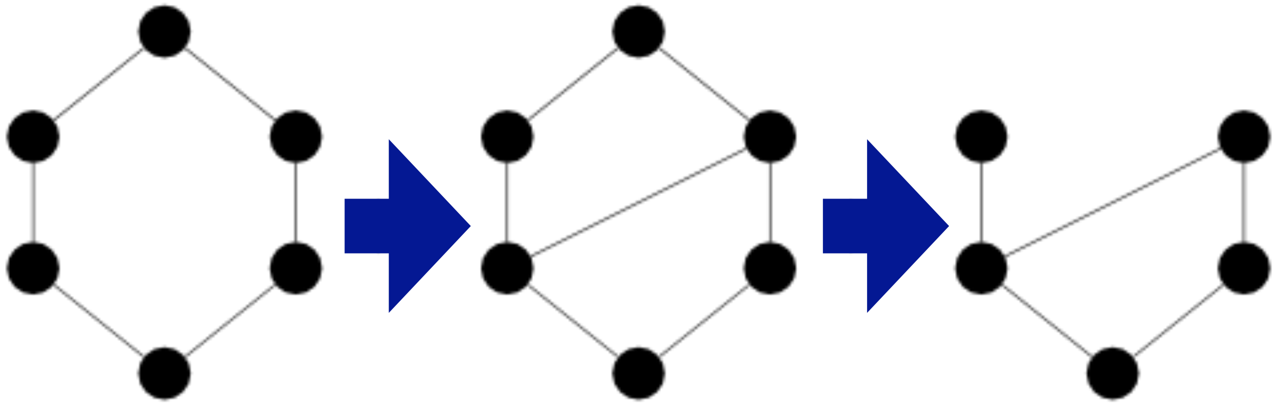
Big Data is always **Changing**
- adding, removing, and modifying connections

In other words . . .
Big Data is always **Evolving**

Q: How do we gain insight from evolving networks?

Dynamic Graphs

A: We gain insight from evolving networks by treating them like Dynamic Graphs where vertices (dots) model entities and edges (lines) model relationships between entities.



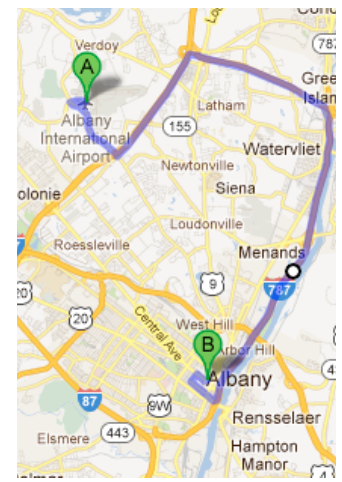
The evolution of a network can be modeled as a **series of graphs** that represent that network at different points in time.

Dynamic Graphs are Everywhere

Understanding dynamic graph evolution enables applications in many areas, among them:

- Social and Business networks
- National Security
- Marketing
- Transportation
- Pharmacology
- Communications networks
- Financial networks
- Social Contact analysis
- Sensor networks

How do we gain insight?



So Many Systems

So many tools.

Which can provide insight into evolving networks?

So Many Systems

Non-relational systems do not effectively address these challenges.

NoSQL schemes, **XML** files, **object** repositories, **document** collections, and **key-value** stores are too unstructured to be reliably queried and often require task-specific code to support accurate analysis.

None of these approaches embrace graph theoretic models. As a result, graph analytics requires custom programming.

Many Systems

Relational systems do not effectively address these challenges.

Relational database systems require breaking down graph structures into separate entities for vertices and edges.

Graph analysis in this scheme involves costly join and self-join operations (the vertices table to itself, through the edge table) for every “hop” from a source vertex to a destination.

A Few Systems

Current graph systems do not effectively address these challenges.

Today's **graph** systems are designed for **one operation** on **one graph** at a time.

Pregel (Google), Trinity (Microsoft), Giraph (open source), Neo4j (open source), GPS (Stanford), and others store and analyze one graph at a time, thus missing the complexity and subtlety inherent in network evolution.

No Systems?

It gets worse.

None of the aforementioned systems take advantage of commonalities in their data storage and processing. With petabyte-scale data, duplication is a deal-breaker.

If only there were a dynamic graph database that could conveniently and efficiently store series of large graphs while supporting diverse operators for queries, aggregates, indexing, and filtering.

G* The Dynamic Graph Database

- Conveniently and efficiently stores series of large graphs.
- Compresses graphs based on their commonalities.
- Supports diverse operators for graph, aggregate, indexing, and filtering.
- Accelerates complex queries on graphs using operators that share common results across graphs.

National Science Foundation
WHERE DISCOVERIES BEGIN

Korea Institute of Science and Technology Information
www.kisti.re.kr

G* The Dynamic Graph Database

Summary

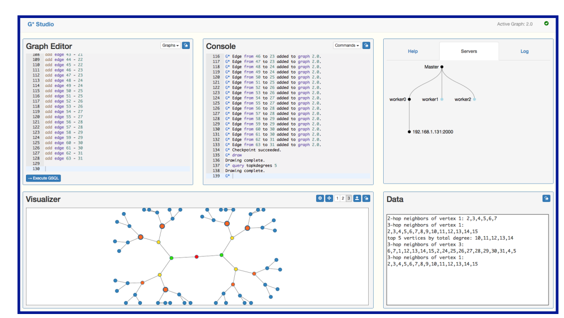
G* is a dynamic graph database system with many features.

- graph distribution
  - multi-core scale up
  - multi-server scale out
- deduplicated disk storage for very large graphs
- in-memory compact indexing
- shared computation
- easy use of sophisticated parallel graph-theoretic queries
- integrates with Relational databases and other stores

National Science Foundation
WHERE DISCOVERIES BEGIN

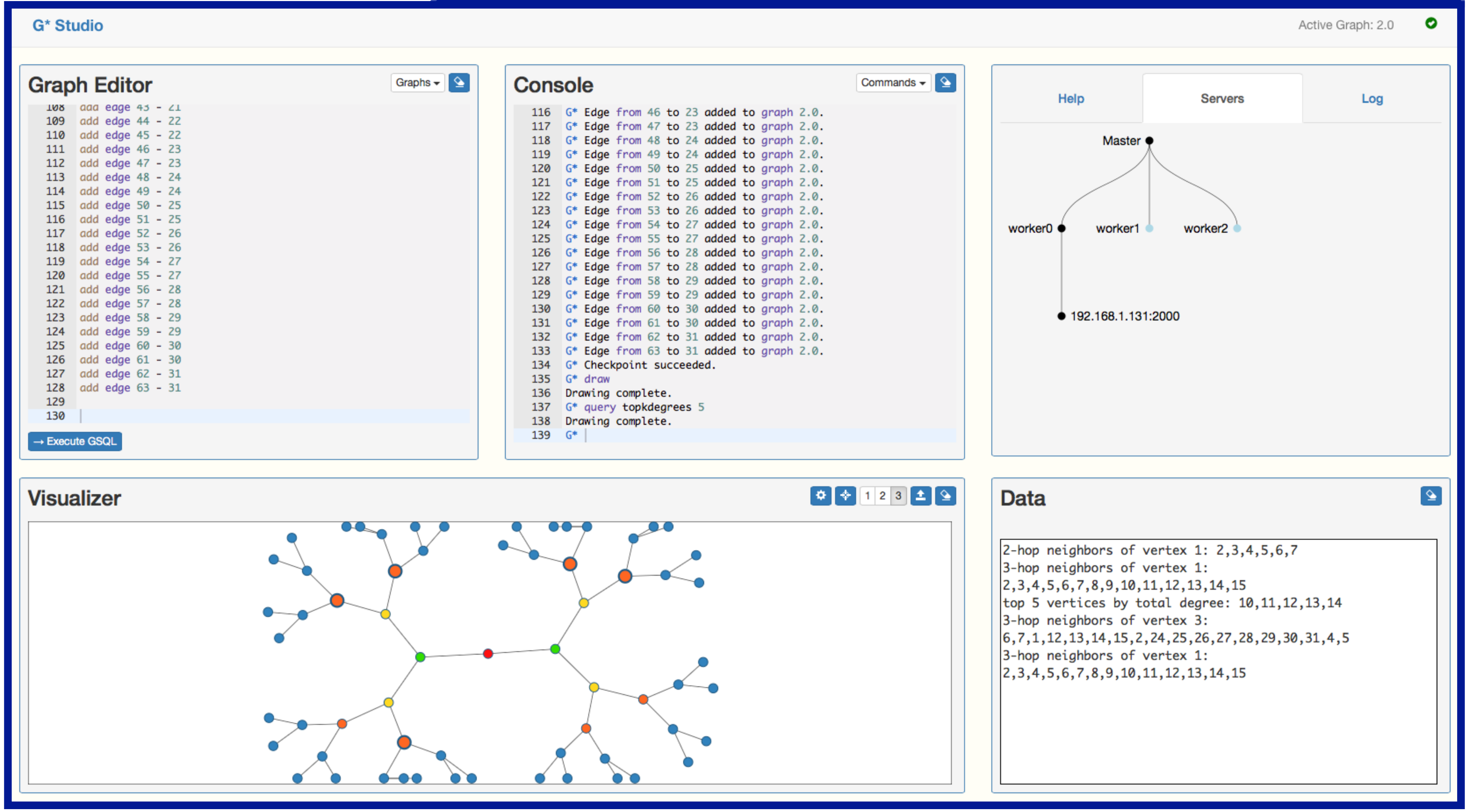
Analyzing evolving graphs enables applications in many areas.

- social media analysis
- network traffic threat assessment
- fraud detection
- marketing
- transportation
- epidemiology
- pharmacology
- . . .and many other areas



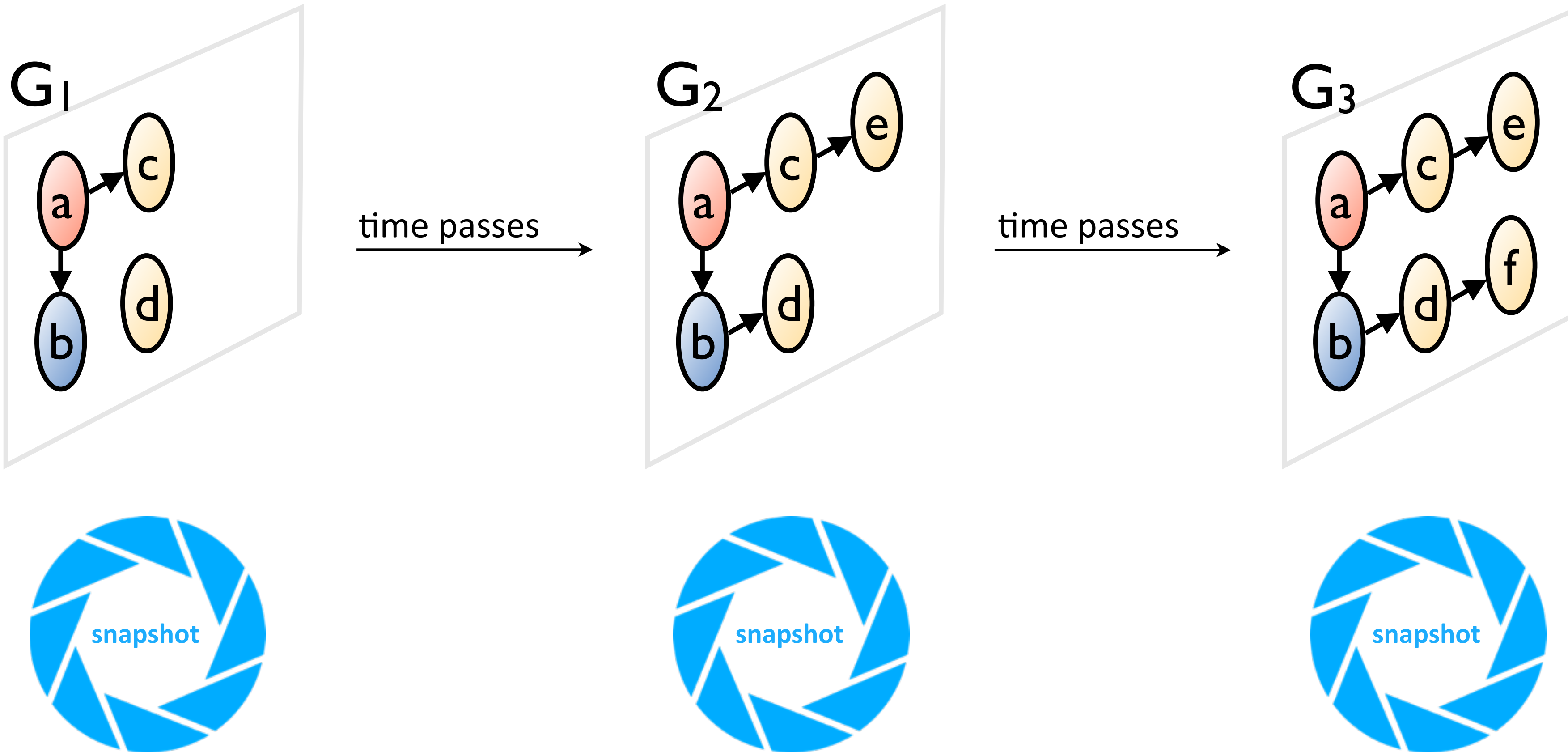
G* The Dynamic Graph Database

Browser Application



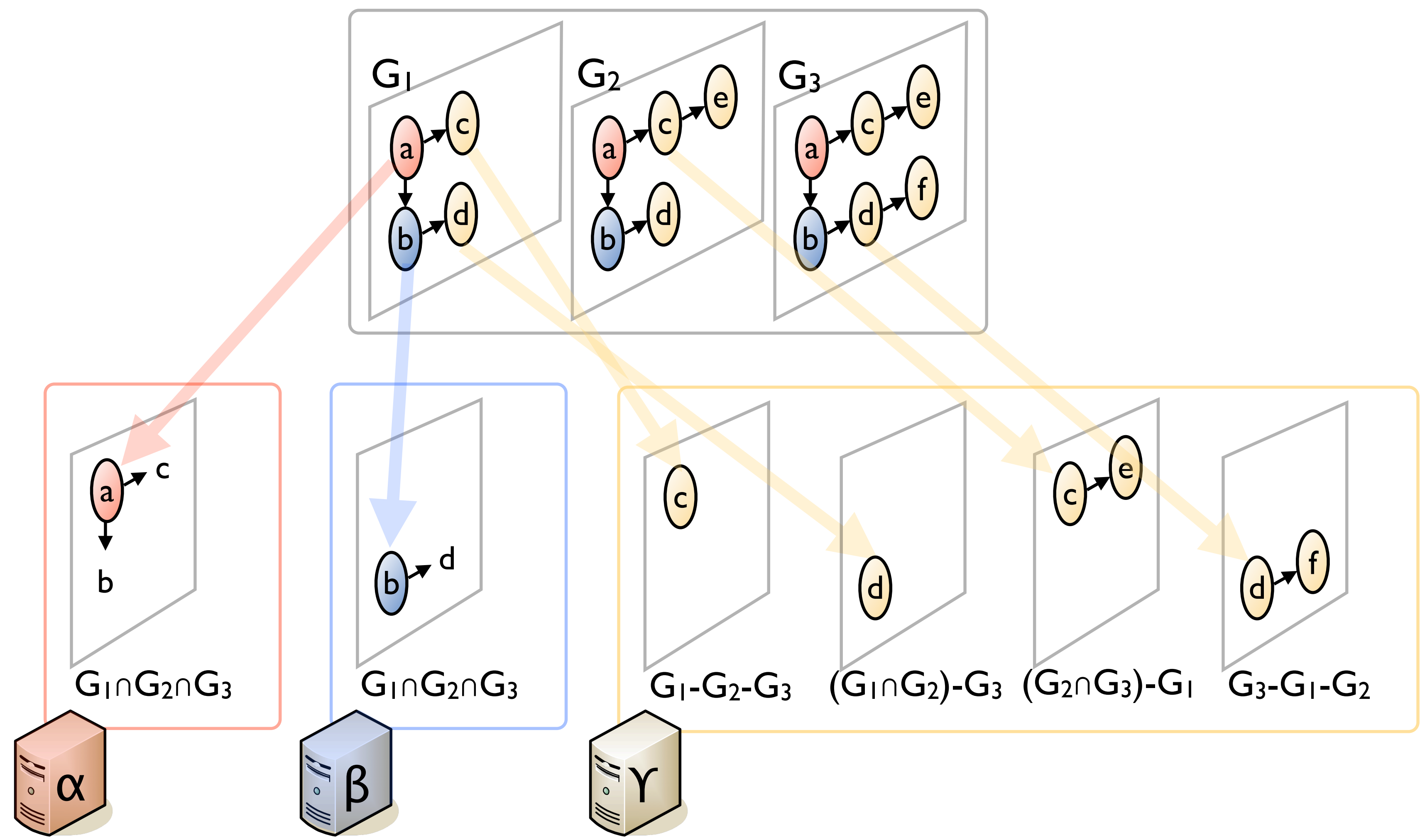
G* The Dynamic Graph Database

Snapshot Management



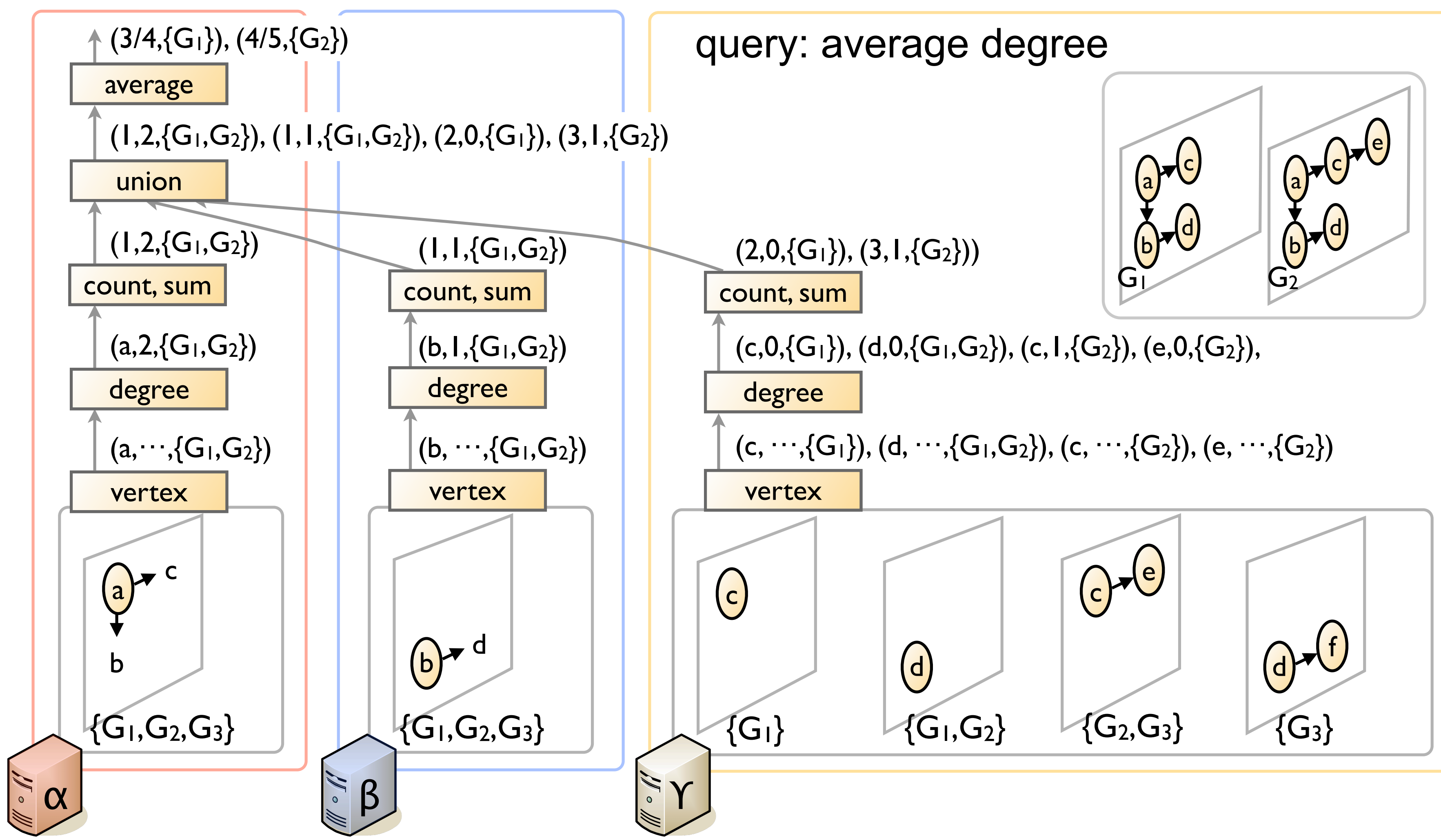
G* The Dynamic Graph Database

Deduplicated Graph Distribution



G* The Dynamic Graph Database

Parallel Graph Query Execution



G* The Dynamic Graph Database

Integrates with Other Systems

PostgreSQL

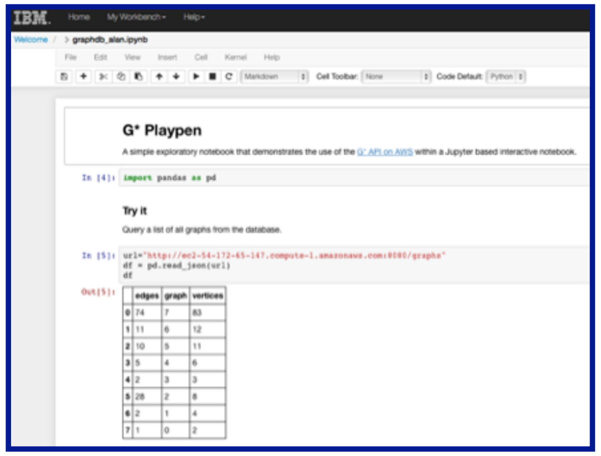
imshealth
INTELLIGENCE APPLIED.

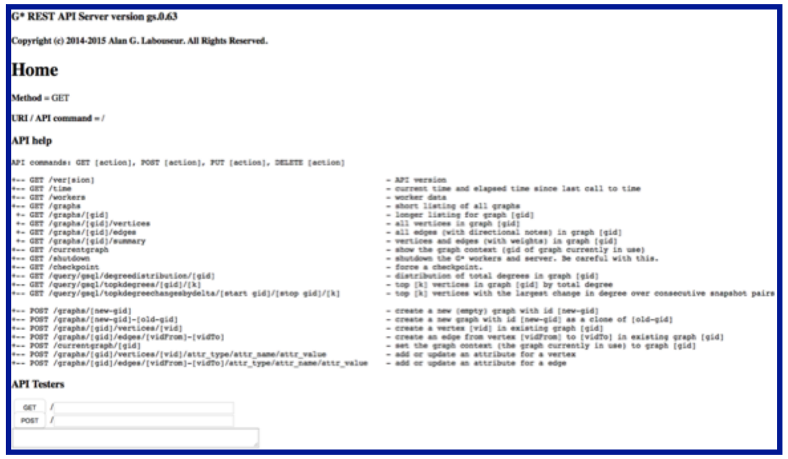
ORACLE

Microsoft® SQL Server®

Your Custom Application

iOS

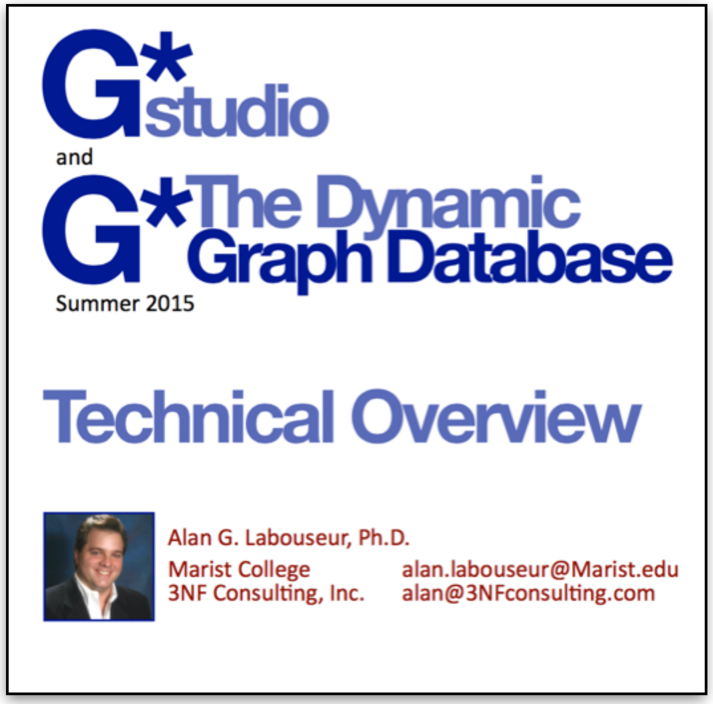


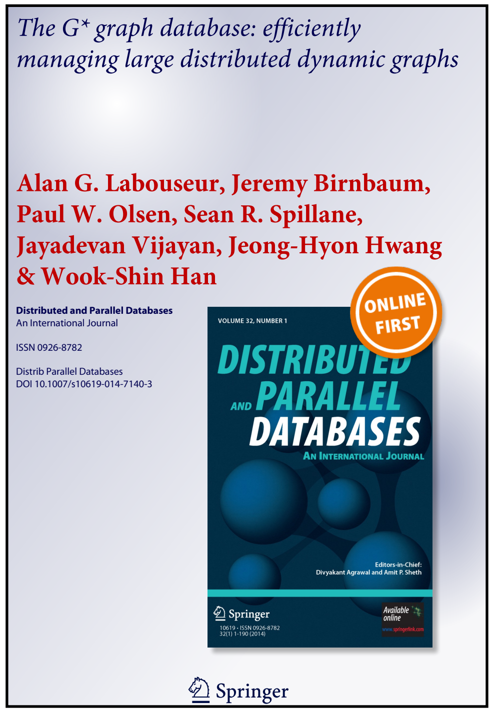


G* The Dynamic Graph Database

There's more to this story.

- Technical Overview
- Journal and conference publications
- Research and consulting available
- Try it live on 3NFconsulting.com





Contact

Alan G. Labouseur, Ph.D.

phone 1-845-440-1102

web www.Labouseur.com
www.3NFconsulting.com
www.marist.edu/compscimath/faculty.html

e-mail alan@Labouseur.com
alan@3NFconsulting.com
alan.labouseur@Marist.edu