CC5212-1

Procesamiento Masivo de Datos

Otoño 2019

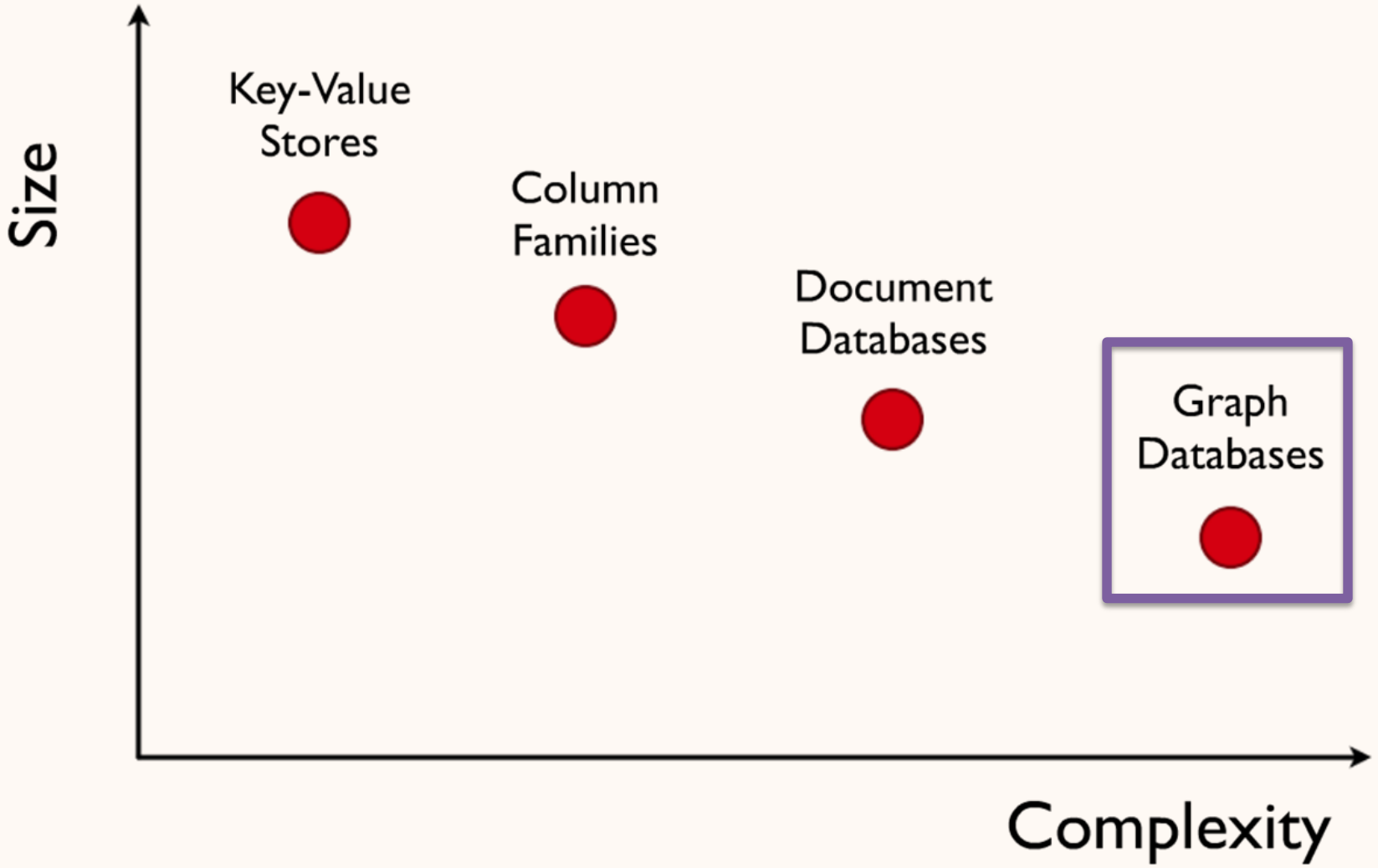
Lecture 11

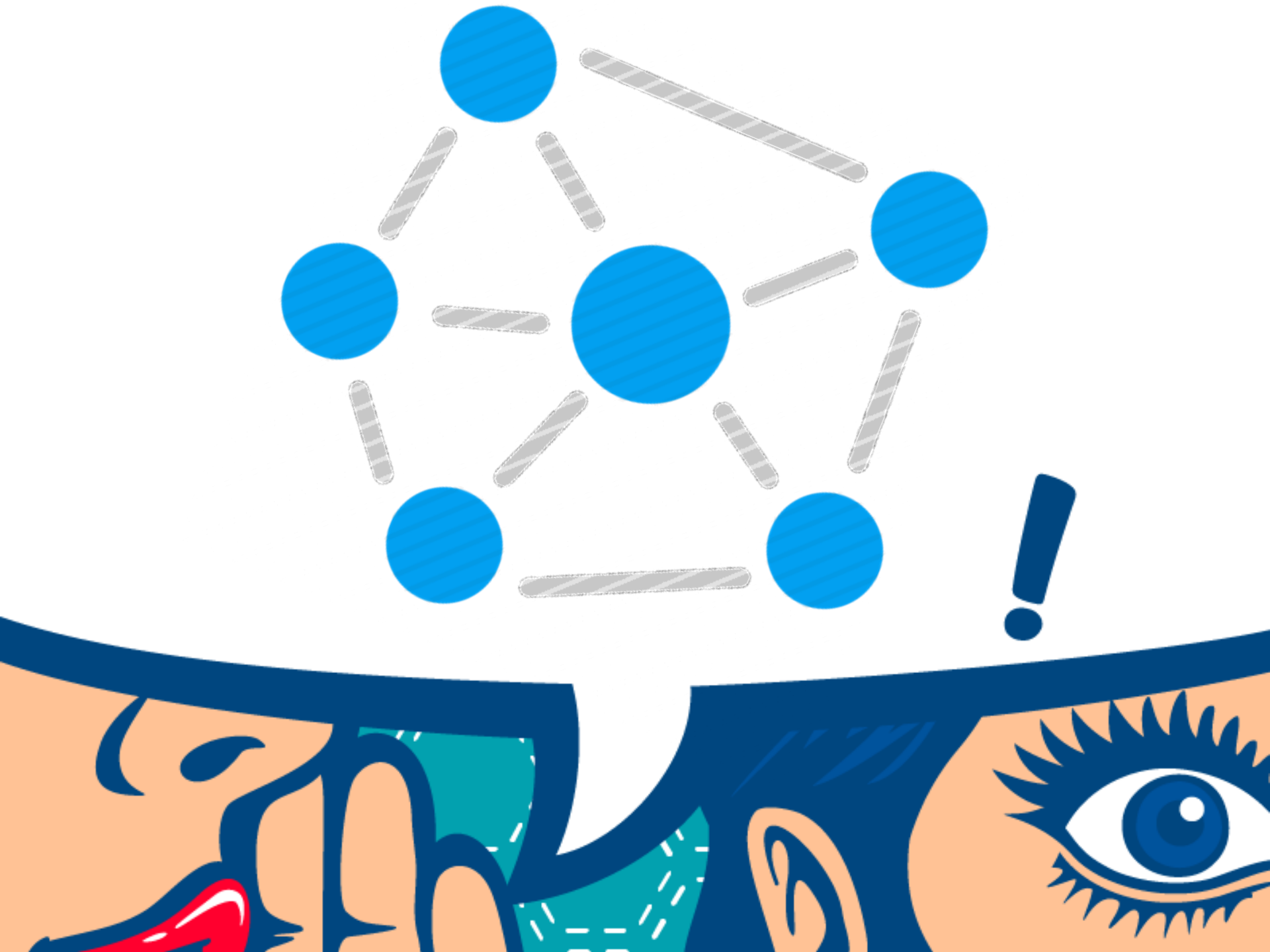
NoSQL: Neo4J

Aidan Hogan

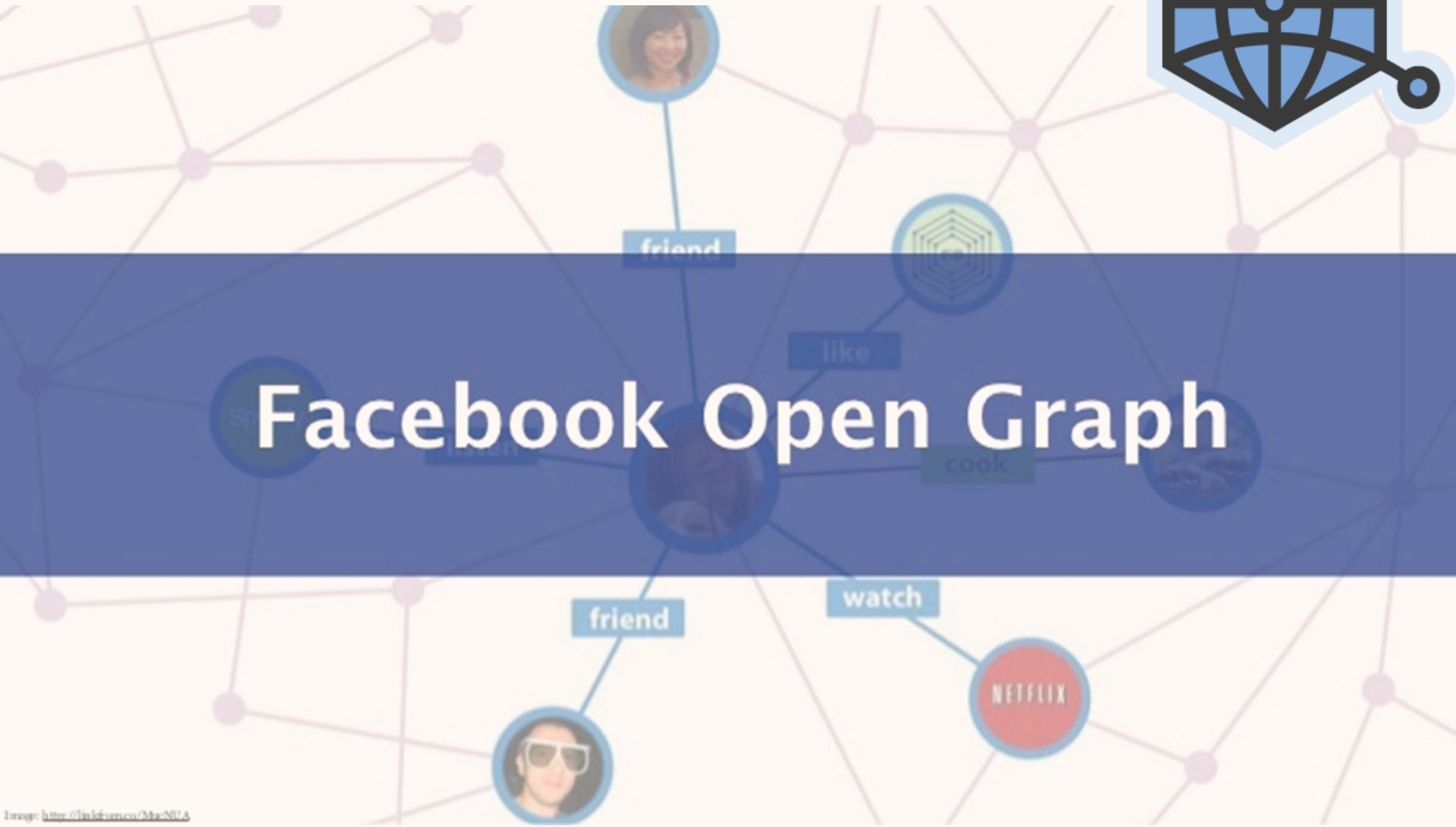
aidhog@gmail.com

NoSQL

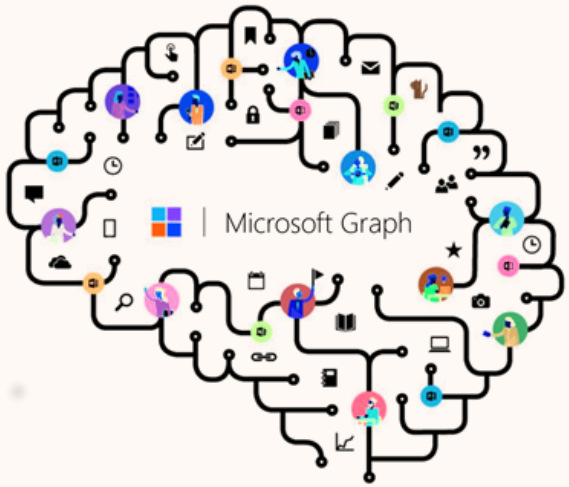

Size
Key-Value Stores
Column Families
Document Databases
Graph Databases
Complexity




Google
Knowledg
e
Graph

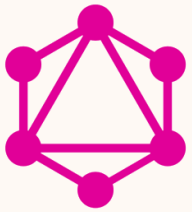


Facebook Open Graph



graph.microsoft.com

Thinking in Graphs



GraphQL

It's Graphs All the Way Down *

> With GraphQL, you model your business domain as a graph

Graphs are powerful tools for modeling many real-world phenomena because they resemble our natural mental models and verbal descriptions of the underlying process. With GraphQL, you model your business domain as a graph by defining a schema; within your schema, you define different types of nodes and how they connect/relate to one another. On the client, this creates a pattern similar to Object-Oriented Programming: types that reference other types. On the server, since GraphQL only defines the interface, you have the freedom to use it with any backend (new or legacy!).

Shared Language

> Naming things is a hard but important part of building intuitive APIs

Think of your GraphQL schema as an expressive shared language for your team and your users. To build a good schema, examine the everyday language you use to describe your business. For example, let's try to describe an email app in plain english:

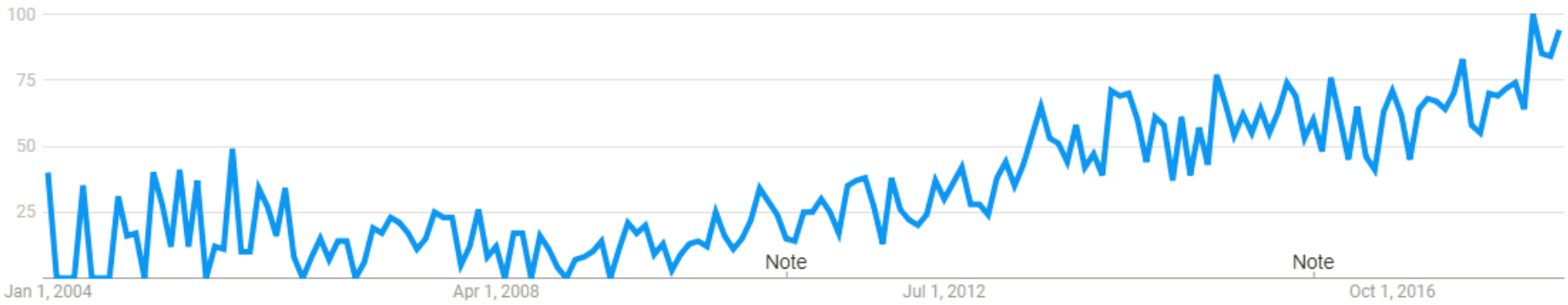

Amazon Neptune
Fast, reliable graph database built for the cloud
Get started with Amazon Neptune

Graph database
Topic

+ Compare

United States | 2004 - present | All categories | Web Search

Interest over time



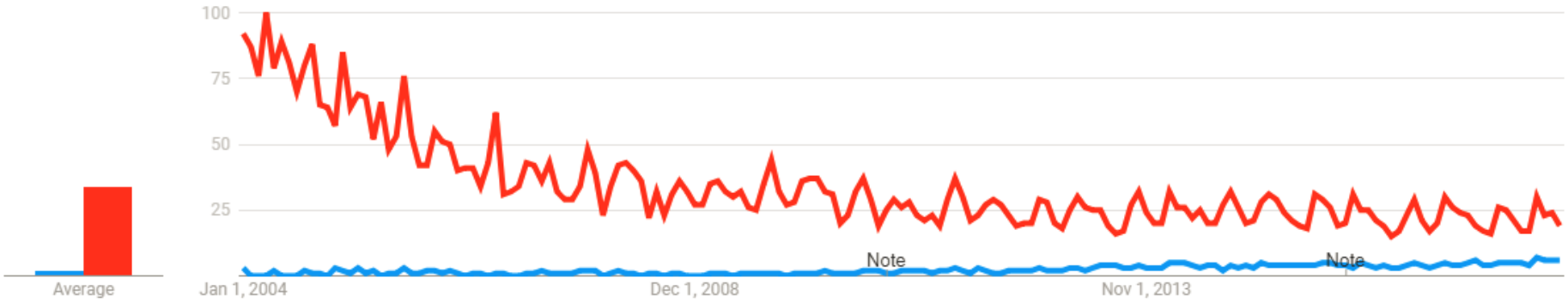
Graph database
Topic

Relational database
Topic

+ Add comparison

United States | 2004 - present | All categories | Web Search

Interest over time

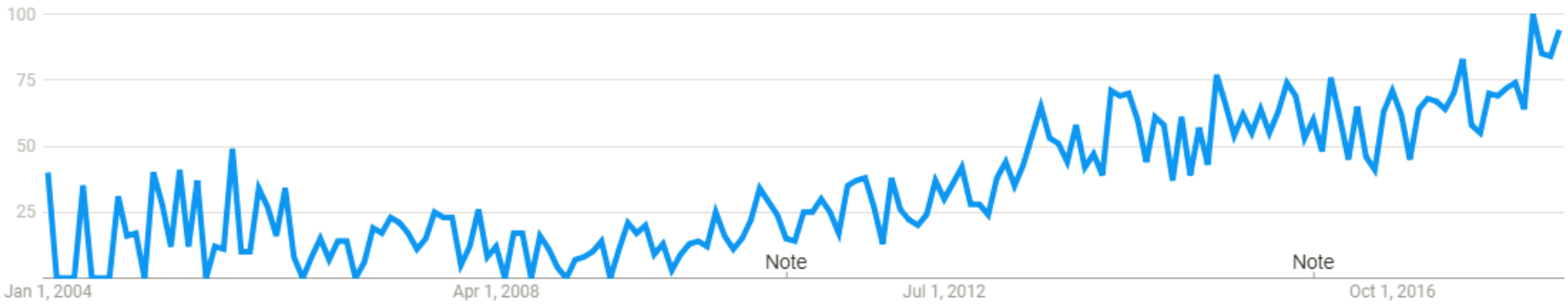


Graph database
Topic

+ Compare

United States | 2004 - present | All categories | Web Search

Interest over time



347 systems in ranking, May 2019

Rank May 2019	Rank Apr 2019	Rank May 2018	DBMS	Database Model	Score May 2019	Score Apr 2019	Score May 2018
1.	1.	1.	Oracle	Relational, Multi-model	1285.55	+5.61	-4.87
2.	2.	2.	MySQL	Relational, Multi-model	1218.96	+3.82	-4.38
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	1072.19	+12.23	-13.66
4.	4.	4.	PostgreSQL	Relational, Multi-model	478.89	+0.17	+77.99
5.	5.	5.	MongoDB	Document	408.07	+6.10	+65.96
6.	6.	6.	IBM Db2	Relational, Multi-model	174.44	-1.61	-11.17
7.	↑ 8.	↑ 9.	Elasticsearch	Search engine, Multi-model	148.62	+2.62	+18.18
8.	↓ 7.	↓ 7.	Redis	Key-value, Multi-model	148.40	+2.03	+13.06
9.	9.	↓ 8.	Microsoft Access	Relational	143.78	-0.87	+10.67
10.	↑ 11.	10.	Cassandra	Wide column	125.72	+2.11	+7.89
11.	↓ 10.	11.	SQLite	Relational	122.90	-1.32	+7.44
12.	12.	↑ 14.	MariaDB	Relational, Multi-model	86.52	+1.29	+21.53
13.	13.	13.	Splunk	Search engine	85.24	+2.15	+20.15
14.	↑ 15.	↑ 18.	Hive	Relational	77.90	+3.19	+20.93
15.	↓ 14.	↓ 12.	Teradata	Relational	76.04	+0.69	+1.63
16.	16.	↓ 15.	Solr	Search engine	60.80	+0.57	-0.72
17.	17.	17.	HBase	Wide column	59.77	+1.11	-0.18
18.	18.	↑ 19.	FileMaker	Relational	58.51	+0.09	+3.84
19.	19.	↑ 21.	Amazon DynamoDB	Multi-model	55.93	-0.08	+11.74
20.	↑ 21.	20.	SAP HANA	Relational, Multi-model	55.74	+0.39	+7.37
21.	↓ 20.	↓ 16.	SAP Adaptive Server	Relational	55.44	-0.36	-6.07
22.	22.	22.	Neo4j	Graph	51.03	+1.54	+10.45
23.	23.	↑ 24.	Couchbase	Document	34.67	-1.61	+2.26
24.	↑ 25.	↓ 23.	Memcached	Key-value	28.90	+0.17	-4.66
25.	↓ 24.	↑ 26.	Microsoft Azure SQL Database	Relational, Multi-model	28.77	-0.02	+3.56

http://db-engines.com/en/ranking

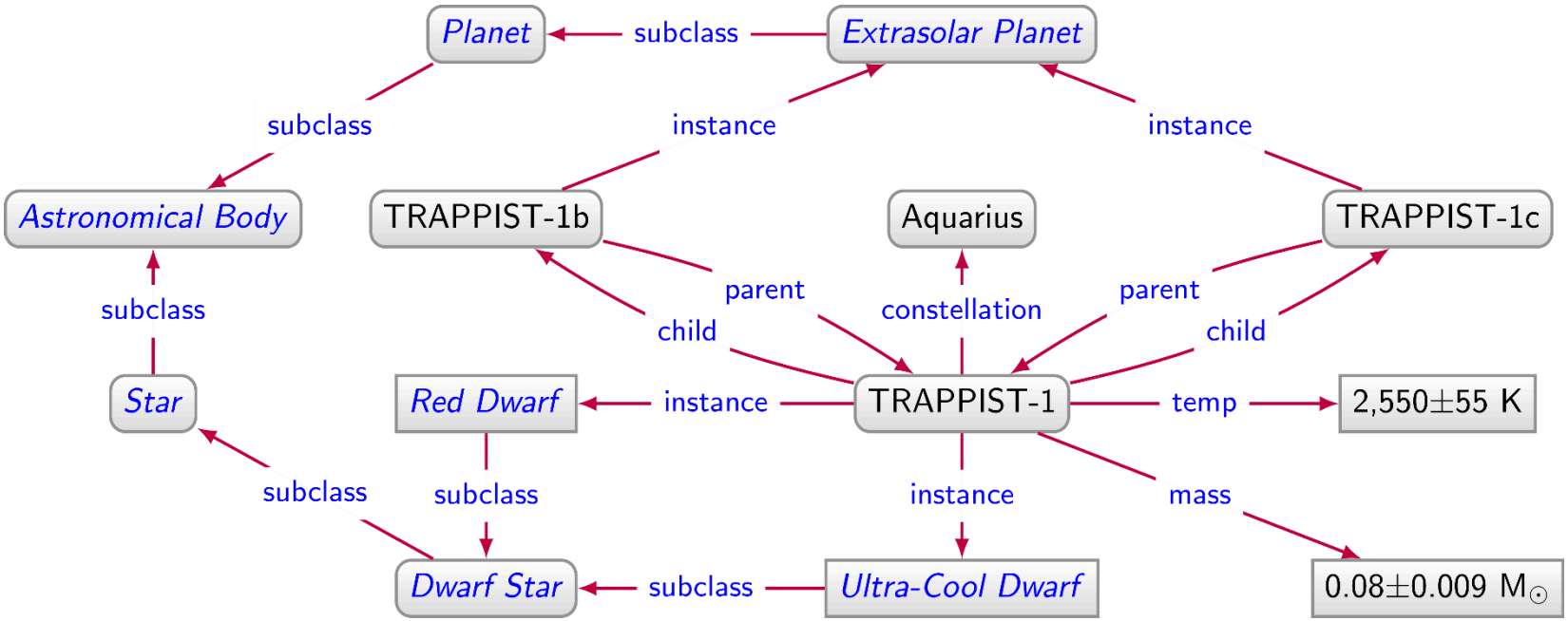
☐ include secondary database models

32 systems in ranking, June 2019

Rank Jun 2019	Rank May 2019	Rank Jun 2018	DBMS	Database Model	Score Jun 2019	Score May 2019	Score Jun 2018
1.	1.	1.	Neo4j	Graph	49.56	-1.48	+7.58
2.	2.	2.	Microsoft Azure Cosmos DB	Multi-model	28.25	+0.65	+9.05
3.	3.	3.	OrientDB	Multi-model	5.59	-0.78	+0.25
4.	4.	4.	ArangoDB	Multi-model	4.57	-0.22	+1.05
5.	5.	5.	Virtuoso	Multi-model	3.11	-0.21	+1.33
6.	6.	↑ 11.	JanusGraph	Graph	1.55	-0.07	+1.19
7.	7.	7.	Amazon Neptune	Multi-model	1.24	-0.09	+0.57
8.	↑ 10.	↑ 10.	GraphDB	Multi-model	1.09	+0.05	+0.69
9.	↓ 8.	↓ 6.	Giraph	Graph	1.08	-0.10	+0.12
10.	↑ 11.	↓ 8.	AllegroGraph	Multi-model	0.93	+0.02	+0.37
11.	↓ 9.	↑ 21.	Dgraph	Graph	0.89	-0.15	+0.77
12.	↑ 13.	↑ 15.	TigerGraph	Graph	0.72	+0.02	+0.55
13.	↓ 12.	↓ 9.	Stardog	Multi-model	0.72	-0.03	+0.22
14.	14.	↓ 13.	Sqrrl	Multi-model	0.59	-0.01	+0.30
15.	15.	↑ 18.	Blazegraph	Multi-model	0.56	0.00	+0.44
16.	16.	↓ 12.	Graph Engine	Multi-model	0.53	-0.01	+0.23
17.	17.	↓ 14.	InfiniteGraph	Graph	0.38	0.00	+0.19
18.	18.	↑ 20.	FaunaDB	Multi-model	0.36	-0.03	+0.24
19.	19.	19.	FlockDB	Graph	0.27	+0.00	+0.16
20.	↑ 22.	↑ 22.	InfoGrid	Graph	0.26	+0.04	+0.16
21.	↓ 20.	↑ 24.	AgensGraph	Multi-model	0.24	-0.02	+0.20
22.	↓ 21.	↑ 28.	AnzoGraph	Multi-model	0.21	-0.02	+0.21
23.	23.	↓ 17.	HyperGraphDB	Graph	0.21	0.00	+0.07
24.	24.	↑ 27.	GRAKN.AI	Multi-model	0.21	+0.04	+0.20
25.	↑ 26.	↓ 16.	Sparksee	Graph	0.13	+0.00	-0.04

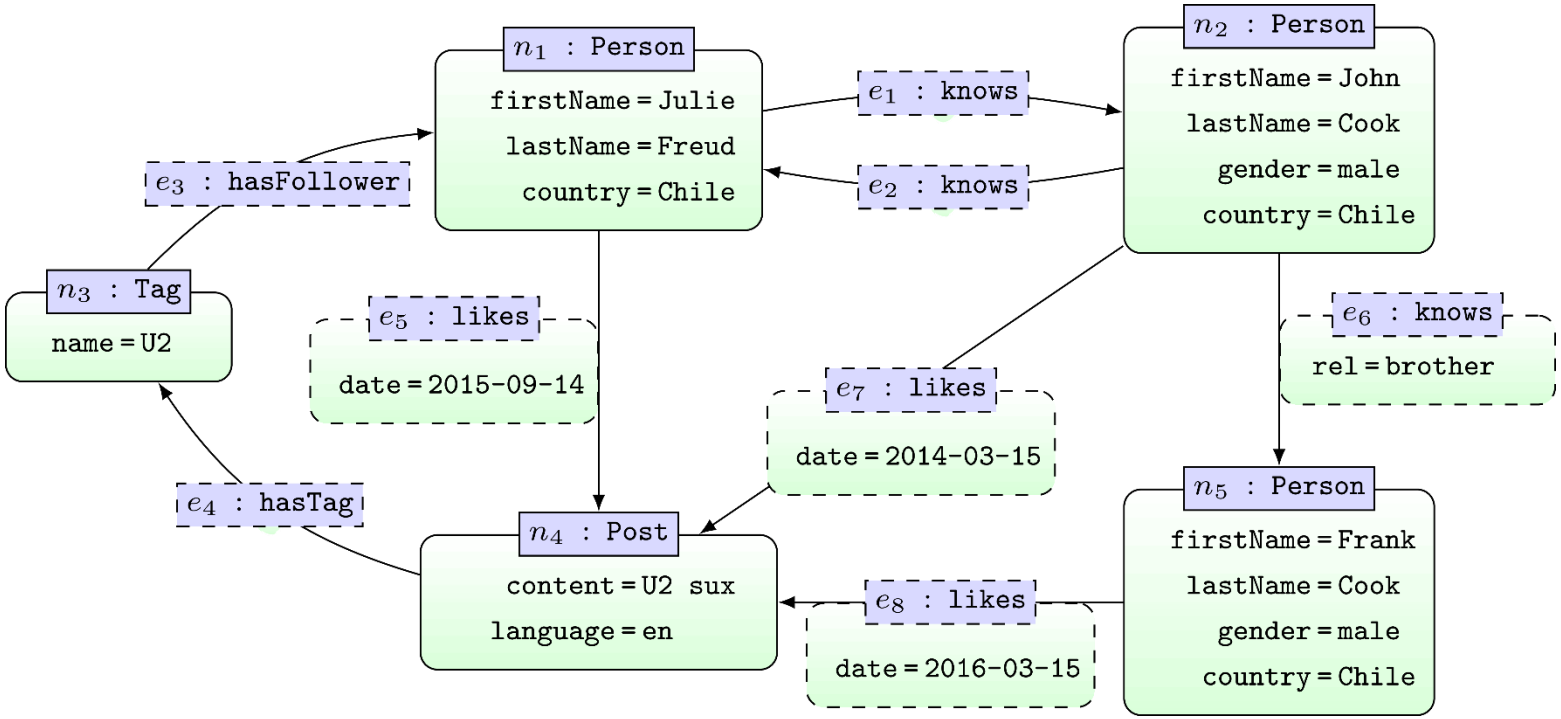
What is a Graph Database?

Directed Edge-labelled Graph




```
SELECT ?const (COUNT(DISTINCT ?body) AS ?num)
WHERE {
  ?body :instance/:subclass* :AstronomicalBody .
  ?body :parent?/:constellation ?const .
}
GROUP BY ?const
ORDER BY DESC(?num)
```

Property Graph




```
MATCH (x1:Person {firstName:"Julie"})-[:knows*]->(x2:Person)
MATCH (x2)-[:likes]->()-[:hasTag]->()-[:hasFollower]->(x1)
RETURN x2.firstName
```

Why do we need Graph Databases?

Why do we need Graph Databases?

Flexibility

Relational Databases …

Relational Databases ...

Debit

account	comment	date	time	amount	total	id
7873698669	Initial deposit	2020-21-01	20:02:02	300000	300000	TRCXGU8JSHD
7873698669	C0°0°L Designs	2020-02-06	09:15:33	50000	325000	TRCCIA2J8AO

Credit

account	comment	date	time	amount	total	id
7873698669	Electricity	2020-02-02	20:00:01	8200	291800	TRCJASJDA9A
7873698669	Heat	2020-02-02	20:00:02	600	291200	TRC81KAQWAS
7873698669	Moviestar	2020-02-02	20:00:03	16200	275000	TRCK8J7JA8D
7873698669	ATM	2020-02-08	16:05:02	100000	225000	TRCPM8A45AD

Account

number	rut	type	total_clp	total_usd
7873698669	32.000.273-K	Current	225000	344,94

Client

rut	name	phone	address
32.000.273-K	Kelvin	+56976698463	Campo de Hielo Sur, Depto 273

Exchange

c1	c2	value
CLP	USD	0,0001533
USD	CLP	652,2750000

Planets / Relational Database

Planets / Relational Database

Planet
name
Mercury
Venus
Earth
Mars
Jupiter
Saturn
Uranus
Neptune
Pluto

Planets / Relational Database

Planet

name	**dist**
Mercury	
Venus	
Earth	1.00
Mars	
Jupiter	
Saturn	
Uranus	
Neptune	
Pluto	

Planets / Relational Database

Planet

name	**dist**
Mercury	0.39
Venus	0.72
Earth	1.00
Mars	1.52
Jupiter	
Saturn	
Uranus	
Neptune	
Pluto	49.31

Planets / Relational Database

Planet

name	dist	radius
Mercury	0.39	0.38
Venus	0.72	
Earth	1.00	1.00
Mars	1.52	0.53
Jupiter		10.97
Saturn	9.54	
Uranus	19.19	3.98
Neptune		
Pluto	49.31	

Planets / Relational Database

Planet

name	dist	radius	grav	days	years	temp	ring
Mercury	0.39	0.38	2.8	58.646	0.241	440	false
Venus	0.72	0.95	8.9	-243.019	0.615	730	false
Earth	1.00	1.00	9.8	0.997	1.000	288	false
Mars	1.52	0.53	3.7	1.026	1.880	186	false
Jupiter	5.20	10.97	22.9	0.414	11.862	152	true
Saturn	9.54	9.14	9.1	0.444	29.447	134	true
Uranus	19.19	3.98	7.8	-0.719	84.017	76	true
Neptune	30.07	3.86	11.0	0.671	164.791	53	true
Pluto	49.31	0.19	0.063	6.39	248.000	44	false

Planets / Relational Database

Planet

name	dist	radius	grav	days	years	temp	ring
Mercury	0.39	0.38	2.8	58.646	0.241	440	false
Venus	0.72	0.95	8.9	-243.019	0.615	730	false
Earth	1.00	1.00	9.8	0.997	1.000	288	false
Mars	1.52	0.53	3.7	1.026	1.880	186	false
Jupiter	5.20	10.97	22.9	0.414	11.862	152	true
Saturn	9.54	9.14	9.1	0.444	29.447	134	true
Uranus	19.19	3.98	7.8	-0.719	84.017	76	true
Neptune	30.07	3.86	11.0	0.671	164.791	53	true
Pluto	49.31	0.19	0.063	6.39	248.000	44	false

Planets / Relational Database

Planet

name	dist	radius	grav	days	years	temp	ring	moon
Mercury	0.39	0.38	2.8	58.646	0.241	440	false	⊥
Venus	0.72	0.95	8.9	-243.019	0.615	730	false	⊥
Earth	1.00	1.00	9.8	0.997	1.000	288	false	Luna
Mars	1.52	0.53	3.7	1.026	1.880	186	false	Phobos, Deimos
Jupiter	5.20	10.97	22.9	0.414	11.862	152	true	Callisto, Ganymede, ...
Saturn	9.54	9.14	9.1	0.444	29.447	134	true	Titan, Rhea, ...
Uranus	19.19	3.98	7.8	-0.719	84.017	76	true	Oberon, Titania, ...
Neptune	30.07	3.86	11.0	0.671	164.791	53	true	Triton, ...
Pluto	49.31	0.19	0.063	6.39	248.000	44	false	Charon

Planets / Relational Database

Planet

name	dist	radius	grav	days	years	temp	ring
Mercury	0.39	0.38	2.8	58.646	0.241	440	false
Venus	0.72	0.95	8.9	-243.019	0.615	730	false
Earth	1.00	1.00	9.8	0.997	1.000	288	false
Mars	1.52	0.53	3.7	1.026	1.880	186	false
Jupiter	5.20	10.97	22.9	0.414	11.862	152	true
Saturn	9.54	9.14	9.1	0.444	29.447	134	true
Uranus	19.19	3.98	7.8	-0.719	84.017	76	true
Neptune	30.07	3.86	11.0	0.671	164.791	53	true
Pluto	49.31	0.19	0.063	6.39	248.000	44	false

Moon

name	planet
Ganimedes	Jupiter
Calisto	Jupiter
Europa	Jupiter
Io	Jupiter
Titan	Saturn
Triton	Neptune
Luna	Terra
Oberon	Uranus
Charon	Pluto
...	...

Planets / Relational Database

Planet

name	dist	radius	grav	days	years	temp	ring
Mercury	0.39	0.38	2.8	58.646	0.241	440	false
Venus	0.72	0.95	8.9	-243.019	0.615	730	false
Earth	1.00	1.00	9.8	0.997	1.000	288	false
Mars	1.52	0.53	3.7	1.026	1.880	186	false
Jupiter	5.20	10.97	22.9	0.414	11.862	152	true
Saturn	9.54	9.14	9.1	0.444	29.447	134	true
Uranus	19.19	3.98	7.8	-0.719	84.017	76	true
Neptune	30.07	3.86	11.0	0.671	164.791	53	true
Pluto	49.31	0.19	0.063	6.39	248.000	44	false

Moon

name	planet	discoverer	year
Ganimedes	Jupiter	Galileo Galilei	1610
Calisto	Jupiter	Galileo Galilei	1610
Europa	Jupiter	Galileo Galilei	1610
Io	Jupiter	Galileo Galilei	1610
Titan	Saturn	Christiaan Huygens	1655
Triton	Neptune	William Lassell	1846
Luna	Terra	$\perp$	$\perp$
Oberon	Uranus	William Herschel	1787
Charon	Pluto	$\perp$	1978
...	...	...	...

Planets / Relational Database

Planet

name	dist	radius	grav	days	years	temp	ring
Mercury	0.39	0.38	2.8	58.646	0.241	440	false
Venus	0.72	0.95	8.9	-243.019	0.615	730	false
Earth	1.00	1.00	9.8	0.997	1.000	288	false
Mars	1.52	0.53	3.7	1.026	1.880	186	false
Jupiter	5.20	10.97	22.9	0.414	11.862	152	true
Saturn	9.54	9.14	9.1	0.444	29.447	134	true
Uranus	19.19	3.98	7.8	-0.719	84.017	76	true
Neptune	30.07	3.86	11.0	0.671	164.791	53	true
Pluto	49.31	0.19	0.063	6.39	248.000	44	false

Moon

name	planet
Ganimedes	Jupiter
Calisto	Jupiter
Europa	Jupiter
Io	Jupiter
Titan	Saturn
Triton	Neptune
Luna	Terra
Oberon	Uranus
Charon	Pluto
...	...

MoonDiscoverer

name	discoverer
Ganimedes	Galileo Galilei
Calisto	Galileo Galilei
Europa	Galileo Galilei
Io	Galileo Galilei
Titan	Christiaan Huygens
Triton	William Lassell
Oberon	William Herschel
...	...

MoonDiscYear

name	year
Ganimedes	1610
Calisto	1610
Europa	1610
Io	1610
Titan	1655
Triton	1846
Oberon	1787
Charon	1978
...	...

Planets / Relational Database

Planet

name	dist	radius	grav	days	years	temp	ring
Mercury	0.39	0.38	2.8	58.646	0.241	440	false
Venus	0.72	0.95	8.9	-243.019	0.615	730	false
Earth	1.00	1.00	9.8	0.997	1.000	288	false
Mars	1.52	0.53	3.7	1.026	1.880	186	false
Jupiter	5.20	10.97	22.9	0.414	11.862	152	true
Saturn	9.54	9.14	9.1	0.444	29.447	134	true
Uranus	19.19	3.98	7.8	-0.719	84.017	76	true
Neptune	30.07	3.86	11.0	0.671	164.791	53	true
Pluto	49.31	0.19	0.063	6.39	248.000	44	false

Moon

name	planet
Ganimedes	Jupiter
Calisto	Jupiter
Europa	Jupiter
Io	Jupiter
Titan	Saturn
Triton	Neptune
Luna	Terra
Oberon	Uranus
Charon	Pluto
...	...

MoonDiscoverer

name	discoverer
Ganimedes	Galileo Galilei
Calisto	Galileo Galilei
Europa	Galileo Galilei
Io	Galileo Galilei
Titan	Christiaan Huygens
Triton	William Lassell
Oberon	William Herschel
...	...

MoonDiscYear

name	year
Ganimedes	1610
Calisto	1610
Europa	1610
Io	1610
Titan	1655
Triton	1846
Oberon	1787
Charon	1978
...	...

Planets / Relational Database

Planet

name	dist	radius	grav	days	years	temp	ring
Mercury	0.39	0.38	2.8	58.646	0.241	440	false
Venus	0.72	0.95	8.9	-243.019	0.615	730	false
Earth	1.00	1.00	9.8	0.997	1.000	288	false
Mars	1.52	0.53	3.7	1.026	1.880	186	false
Jupiter	5.20	10.97	22.9	0.414	11.862	152	true
Saturn	9.54	9.14	9.1	0.444	29.447	134	true
Uranus	19.19	3.98	7.8	-0.719	84.017	76	true
Neptune	30.07	3.86	11.0	0.671	164.791	53	true
Pluto	49.31	0.19	0.063	6.39	248.000	44	false

Moon

name	P.name
Ganimedes	Jupiter
Calisto	Jupiter
Europa	Jupiter
Io	Jupiter
Titan	Saturn
Triton	Neptune
Luna	Earth
Oberon	Uranus
Charon	Pluto
...	...

MoonDiscoverer

name	discoverer
Ganimedes	Galileo Galilei
Calisto	Galileo Galilei
Europa	Galileo Galilei
Io	Galileo Galilei
Titan	Christiaan Huygens
Triton	William Lassell
Oberon	William Herschel
...	...

MoonDiscYear

name	year
Ganimedes	1610
Calisto	1610
Europa	1610
Io	1610
Titan	1655
Triton	1846
Oberon	1787
Charon	1978
...	...

Planets / Relational Database

Planet

name	dist	radius	grav	days	years	temp	ring
Mercury	0.39	0.38	2.8	58.646	0.241	440	false
Venus	0.72	0.95	8.9	-243.019	0.615	730	false
Earth	1.00	1.00	9.8	0.997	1.000	288	false
Mars	1.52	0.53	3.7	1.026	1.880	186	false
Jupiter	5.20	10.97	22.9	0.414	11.862	152	true
Saturn	9.54	9.14	9.1	0.444	29.447	134	true
Uranus	19.19	3.98	7.8	-0.719	84.017	76	true
Neptune	30.07	3.86	11.0	0.671	164.791	53	true
Pluto	49.31	0.19	0.063	6.39	248.000	44	false

Moon

name	P.name
Ganimedes	Jupiter
Calisto	Jupiter
Europa	Jupiter
Io	Jupiter
Titan	Saturn
Triton	Neptune
Luna	Earth
Oberon	Uranus
Charon	Pluto
...	...

MoonDiscoverer

name	discoverer
Ganimedes	Galileo Galilei
Calisto	Galileo Galilei
Europa	Galileo Galilei
Io	Galileo Galilei
Titan	Christiaan Huygens
Triton	William Lassell
Oberon	William Herschel
...	...

MoonDiscYear

name	year
Ganimedes	1610
Calisto	1610
Europa	1610
Io	1610
Titan	1655
Triton	1846
Oberon	1787
Charon	1978
...	...

Planets / Relational Database

Planet

name	dist	radius	grav	days	years	temp	ring
Mercury	0.39	0.38	2.8	58.646	0.241	440	false
Venus	0.72	0.95	8.9	-243.019	0.615	730	false
Earth	1.00	1.00	9.8	0.997	1.000	288	false
Mars	1.52	0.53	3.7	1.026	1.880	186	false
Jupiter	5.20	10.97	22.9	0.414	11.862	152	true
Saturn	9.54	9.14	9.1	0.444	29.447	134	true
Uranus	19.19	3.98	7.8	-0.719	84.017	76	true
Neptune	30.07	3.86	11.0	0.671	164.791	53	true

DwarfPlanet

name	dist	radius	grav	days	years	temp	ring
Pluto	49.31	0.19	0.063	6.39	248.000	44	false

Moon

name	P.name
Ganimedes	Jupiter
Calisto	Jupiter
Europa	Jupiter
Io	Jupiter
Titan	Saturn
Triton	Neptune
Luna	Earth
Oberon	Uranus
Charon	Pluto
...	...

MoonDiscoverer

name	discoverer
Ganimedes	Galileo Galilei
Calisto	Galileo Galilei
Europa	Galileo Galilei
Io	Galileo Galilei
Titan	Christiaan Huygens
Triton	William Lassell
Oberon	William Herschel
...	...

MoonDiscYear

name	year
Ganimedes	1610
Calisto	1610
Europa	1610
Io	1610
Titan	1655
Triton	1846
Oberon	1787
Charon	1978
...	...

Planets / Relational Database

Planet

name	dist	radius	grav	days	years	temp	ring
Mercury	0.39	0.38	2.8	58.646	0.241	440	false
Venus	0.72	0.95	8.9	-243.019	0.615	730	false
Earth	1.00	1.00	9.8	0.997	1.000	288	false
Mars	1.52	0.53	3.7	1.026	1.880	186	false
Jupiter	5.20	10.97	22.9	0.414	11.862	152	true
Saturn	9.54	9.14	9.1	0.444	29.447	134	true
Uranus	19.19	3.98	7.8	-0.719	84.017	76	true
Neptune	30.07	3.86	11.0	0.671	164.791	53	true

DwarfPlanet

name	dist	radius	grav	days	years	temp	ring
Pluto	49.31	0.19	0.063	6.39	248.000	44	false

Moon

name	P.name
Ganimedes	Jupiter
Calisto	Jupiter
Europa	Jupiter
Io	Jupiter
Titan	Saturn
Triton	Neptune
Luna	Earth
Oberon	Uranus
Charon	Pluto
...	...

MoonDiscoverer

name	discoverer
Ganimedes	Galileo Galilei
Calisto	Galileo Galilei
Europa	Galileo Galilei
Io	Galileo Galilei
Titan	Christiaan Huygens
Triton	William Lassell
Oberon	William Herschel
...	...

MoonDiscYear

name	year
Ganimedes	1610
Calisto	1610
Europa	1610
Io	1610
Titan	1655
Triton	1846
Oberon	1787
Charon	1978
...	...

Planets / Relational Database

Planet

name	dist	radius	grav	days	years	temp	ring
Mercury	0.39	0.38	2.8	58.646	0.241	440	false
Venus	0.72	0.95	8.9	-243.019	0.615	730	false
Earth	1.00	1.00	9.8	0.997	1.000	288	false
Mars	1.52	0.53	3.7	1.026	1.880	186	false
Jupiter	5.20	10.97	22.9	0.414	11.862	152	true
Saturn	9.54	9.14	9.1	0.444	29.447	134	true
Uranus	19.19	3.98	7.8	-0.719	84.017	76	true
Neptune	30.07	3.86	11.0	0.671	164.791	53	true

DwarfPlanet

name	dist	radius	grav	days	years	temp	ring
Pluto	49.31	0.19	0.063	6.39	248.000	44	false

Moon

name	parent
Ganimedes	Jupiter
Calisto	Jupiter
Europa	Jupiter
Io	Jupiter
Titan	Saturn
Triton	Neptune
Luna	Earth
Oberon	Uranus
Charon	Pluto
...	...

MoonDiscoverer

name	discoverer
Ganimedes	Galileo Galilei
Calisto	Galileo Galilei
Europa	Galileo Galilei
Io	Galileo Galilei
Titan	Christiaan Huygens
Triton	William Lassell
Oberon	William Herschel
...	...

MoonDiscYear

name	year
Ganimedes	1610
Calisto	1610
Europa	1610
Io	1610
Titan	1655
Triton	1846
Oberon	1787
Charon	1978
...	...

Planets / Relational Database

Planet

name	dist	radius	grav	days	years	temp	ring
Mercury	0.39	0.38	2.8	58.646	0.241	440	false
Venus	0.72	0.95	8.9	-243.019	0.615	730	false
Earth	1.00	1.00	9.8	0.997	1.000	288	false
Mars	1.52	0.53	3.7	1.026	1.880	186	false
Jupiter	5.20	10.97	22.9	0.414	11.862	152	true
Saturn	9.54	9.14	9.1	0.444	29.447	134	true
Uranus	19.19	3.98	7.8	-0.719	84.017	76	true
Neptune	30.07	3.86	11.0	0.671	164.791	53	true

DwarfPlanet

name	dist	radius	grav	days	years	temp	ring
Pluto	49.31	0.19	0.063	6.39	248.000	44	false

Moon

name	parent
Ganimedes	Jupiter
Calisto	Jupiter
Europa	Jupiter
Io	Jupiter
Titan	Saturn
Triton	Neptune
Luna	Earth
Oberon	Uranus
Charon	Pluto
...	...

MoonDiscoverer

name	discoverer
Ganimedes	Galileo Galilei
Calisto	Galileo Galilei
Europa	Galileo Galilei
Io	Galileo Galilei
Titan	Christiaan Huygens
Triton	William Lassell
Oberon	William Herschel
...	...

MoonDiscYear

name	year
Ganimedes	1610
Calisto	1610
Europa	1610
Io	1610
Titan	1655
Triton	1846
Oberon	1787
Charon	1978
...	...

Planets / Relational Database

Planet

name	dist	radius	grav	days	years	temp	ring
Mercury	0.39	0.38	2.8	58.646	0.241	440	false
Venus	0.72	0.95	8.9	-243.019	0.615	730	false
Earth	1.00	1.00	9.8	0.997	1.000	288	false
Mars	1.52	0.53	3.7	1.026	1.880	186	false
Jupiter	5.20	10.97	22.9	0.414	11.862	152	true
Saturn	9.54	9.14	9.1	0.444	29.447	134	true
Uranus	19.19	3.98	7.8	-0.719	84.017	76	true
Neptune	30.07	3.86	11.0	0.671	164.791	53	true

(ノ°□°)ノ ┻━┻

DwarfPlanet

name	dist	radius	grav	days	years	temp	ring
Pluto	49.31	0.19	0.063	6.39	248.000	44	false

Moon

name	parent
Ganimedes	Jupiter
Calisto	Jupiter
Europa	Jupiter
Io	Jupiter
Titan	Saturn
Triton	Neptune
Luna	Earth
Oberon	Uranus
Charon	Pluto
...	...

MoonDiscoverer

name	discoverer
Ganimedes	Galileo Galilei
Calisto	Galileo Galilei
Europa	Galileo Galilei
Io	Galileo Galilei
Titan	Christiaan Huygens
Triton	William Lassell
Oberon	William Herschel
...	...

MoonDiscYear

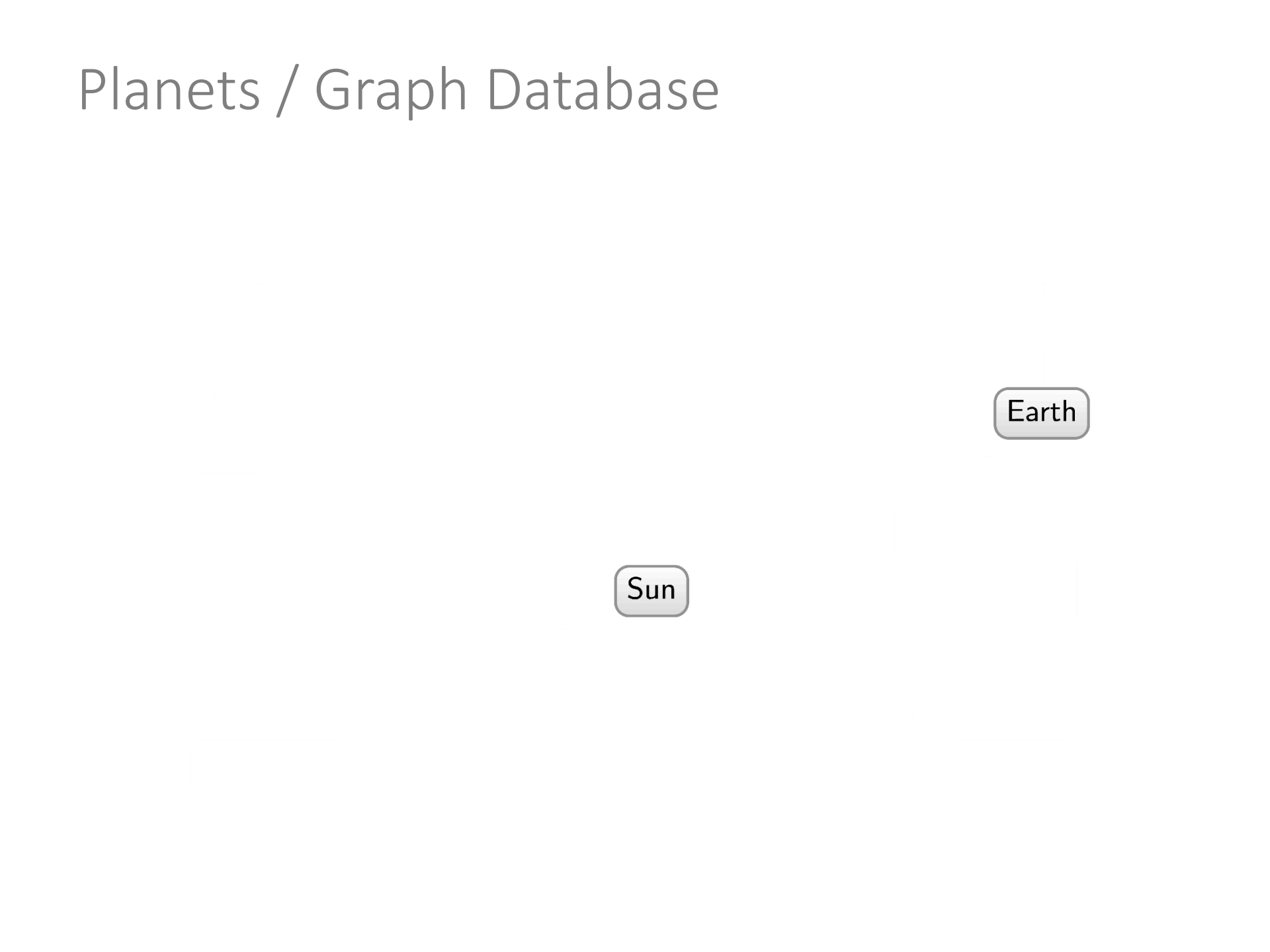
name	year
Ganimedes	1610
Calisto	1610
Europa	1610
Io	1610
Titan	1655
Triton	1846
Oberon	1787
Charon	1978
...	...

Planets / Graph Database

Planets / Graph Database

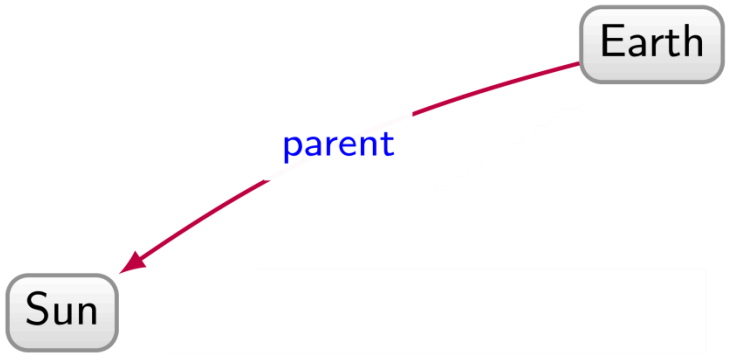
Planets / Graph Database

Planets / Graph Database

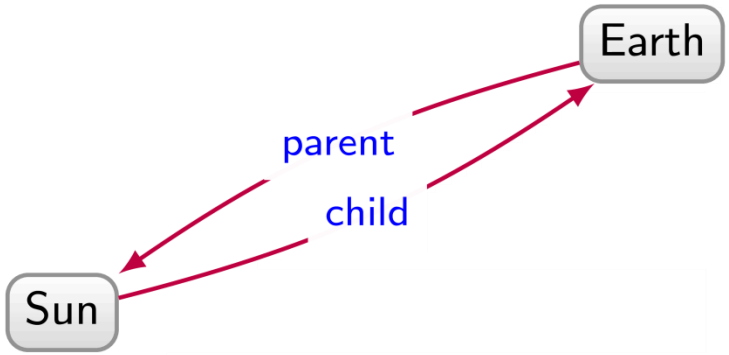


Planets / Graph Database

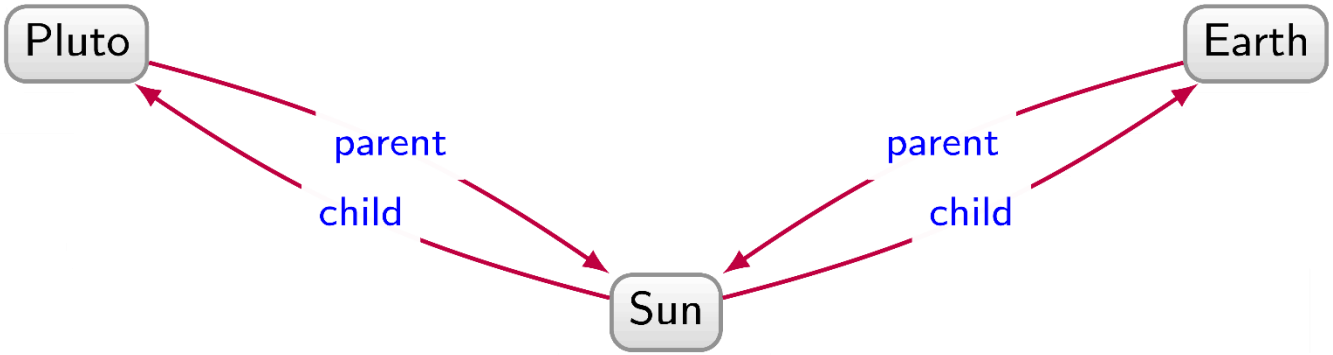
Planets / Graph Database



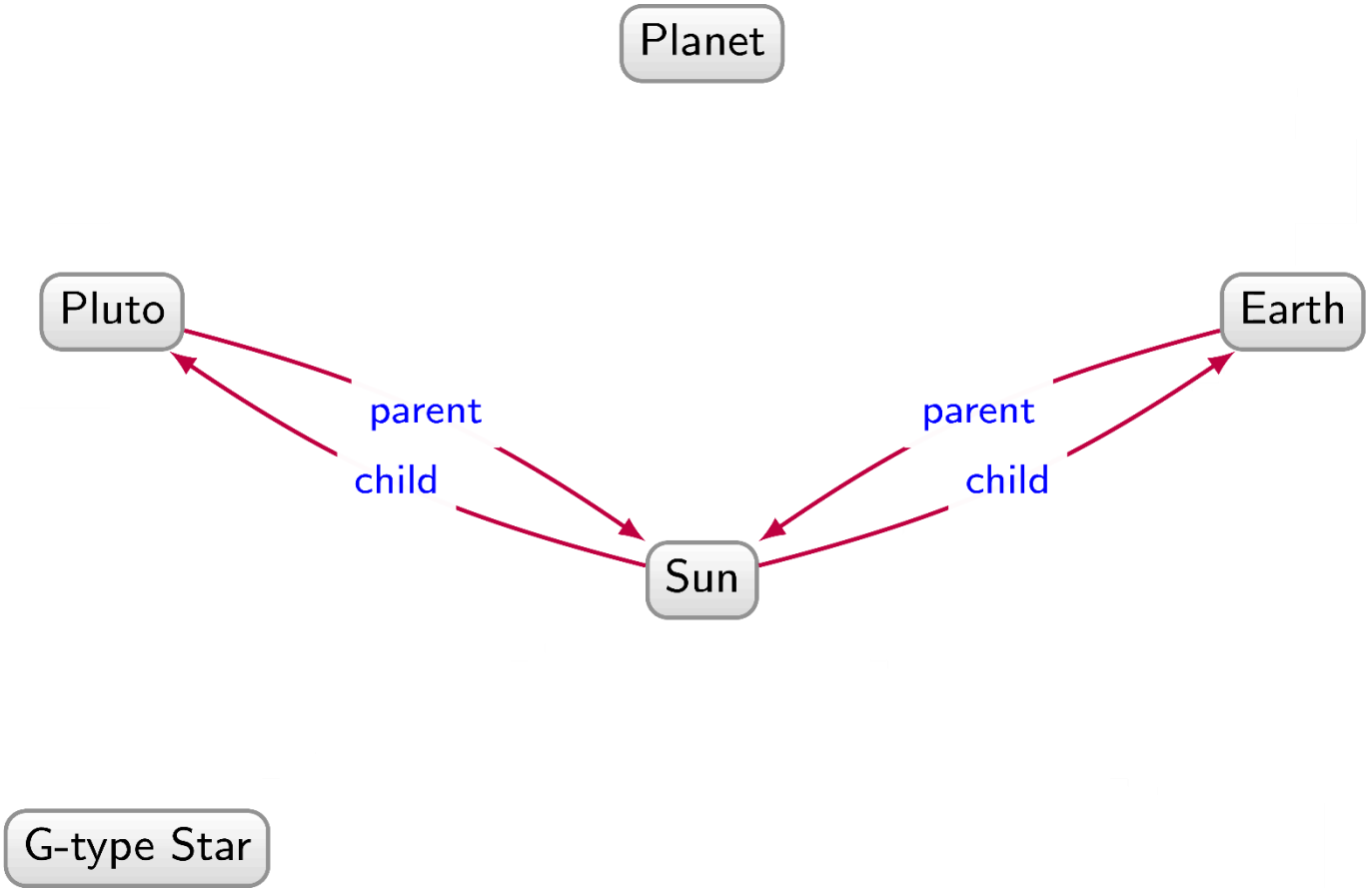
Planets / Graph Database



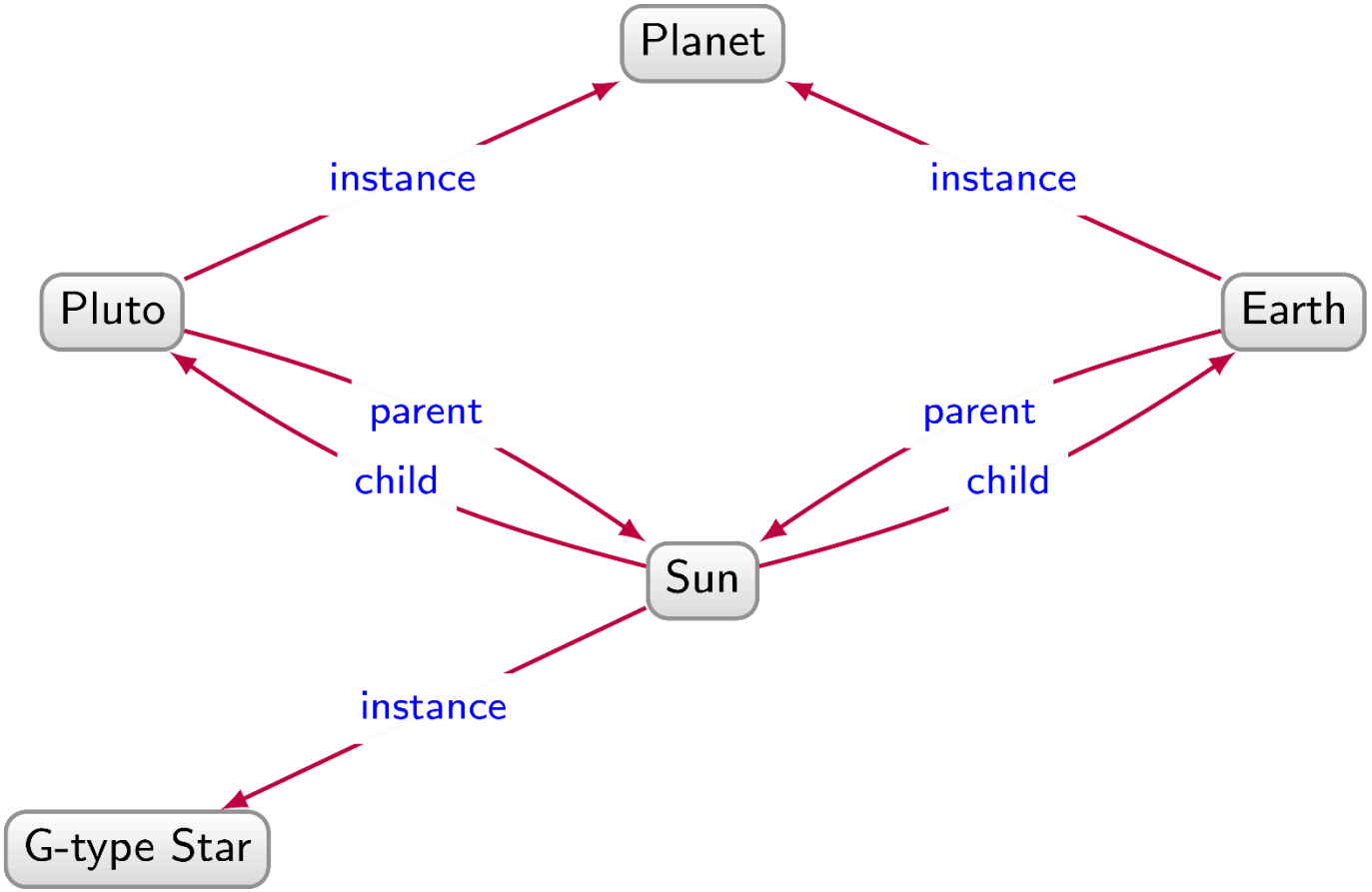
Planets / Graph Database



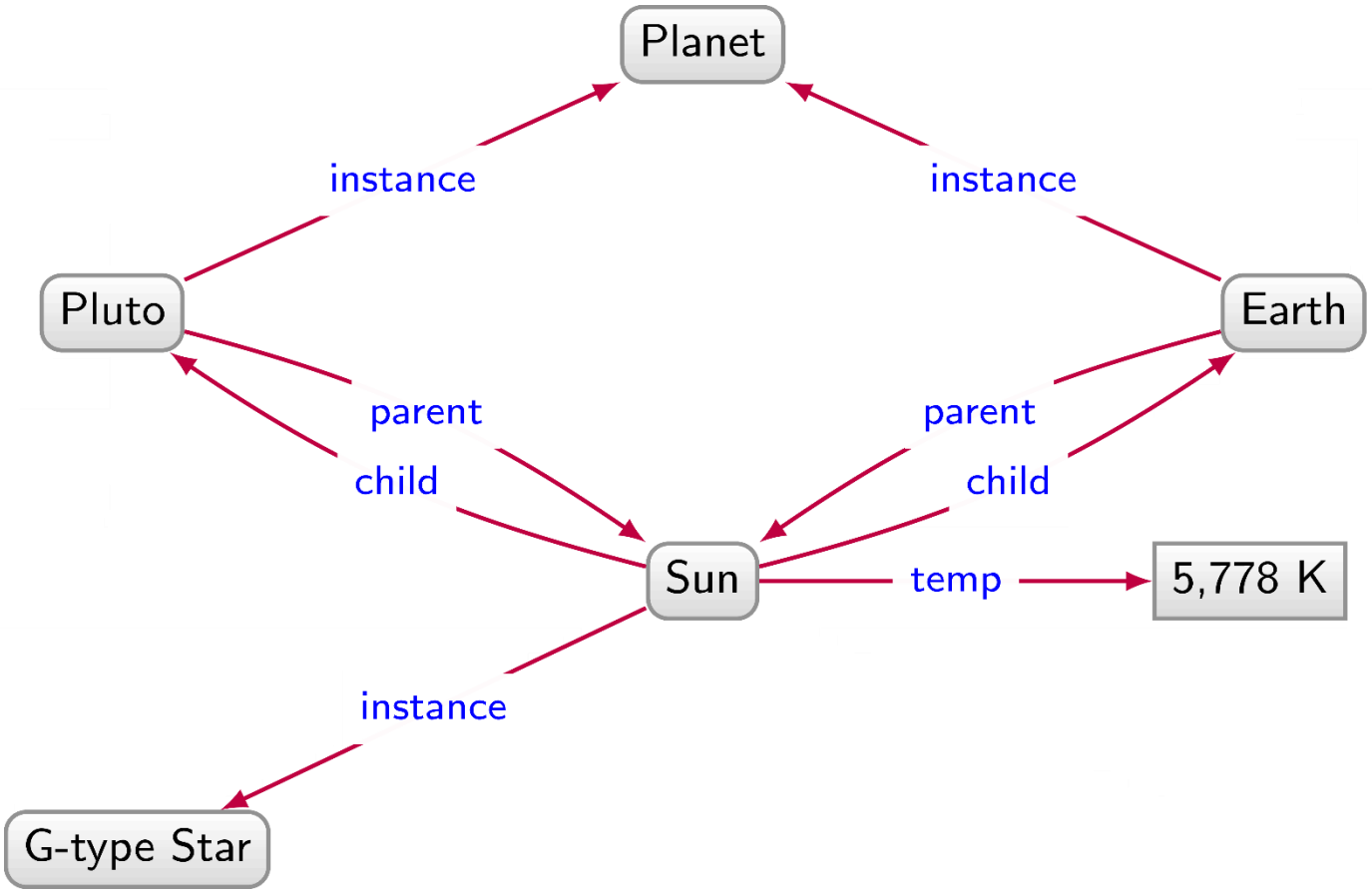
Planets / Graph Database



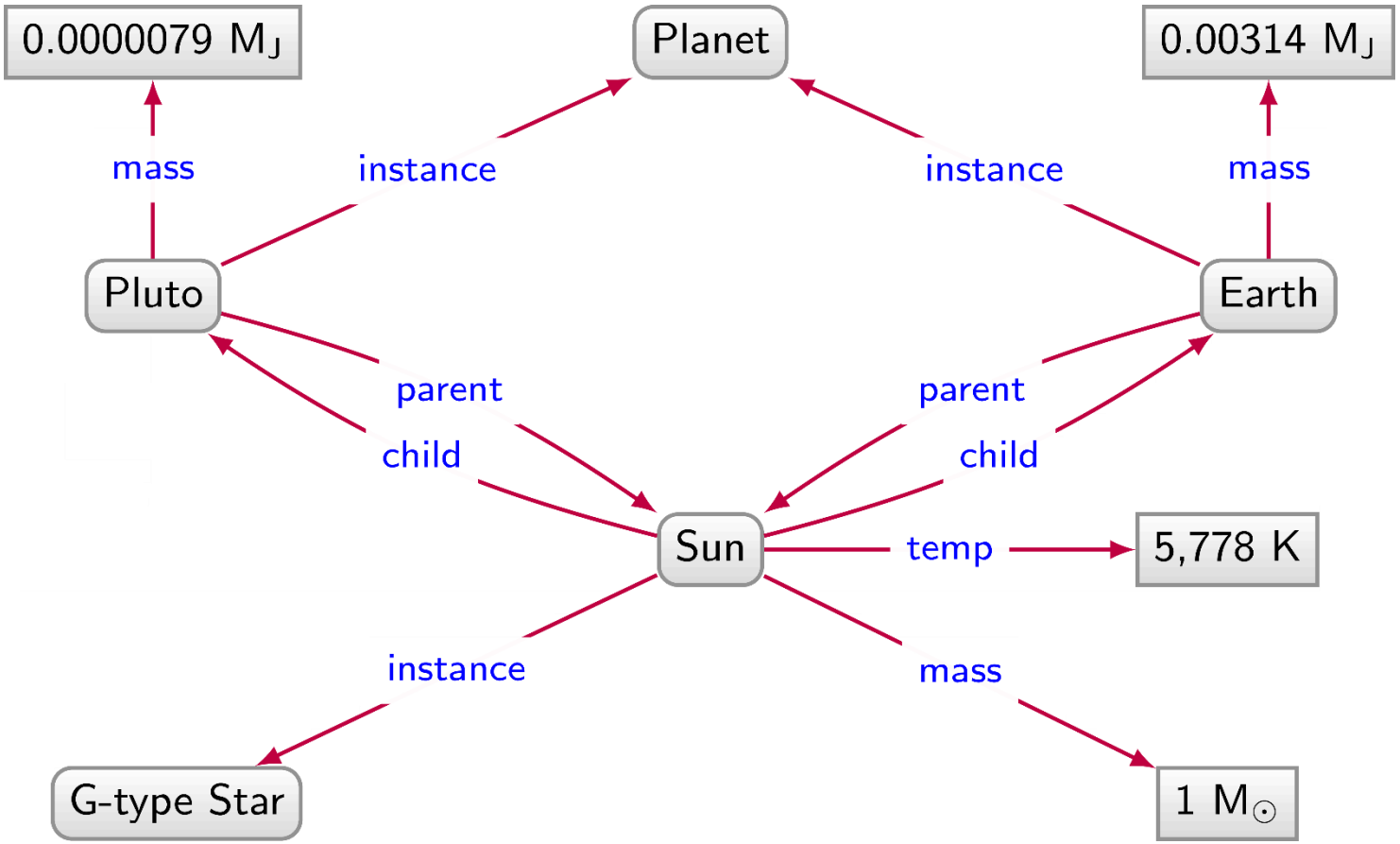
Planets / Graph Database



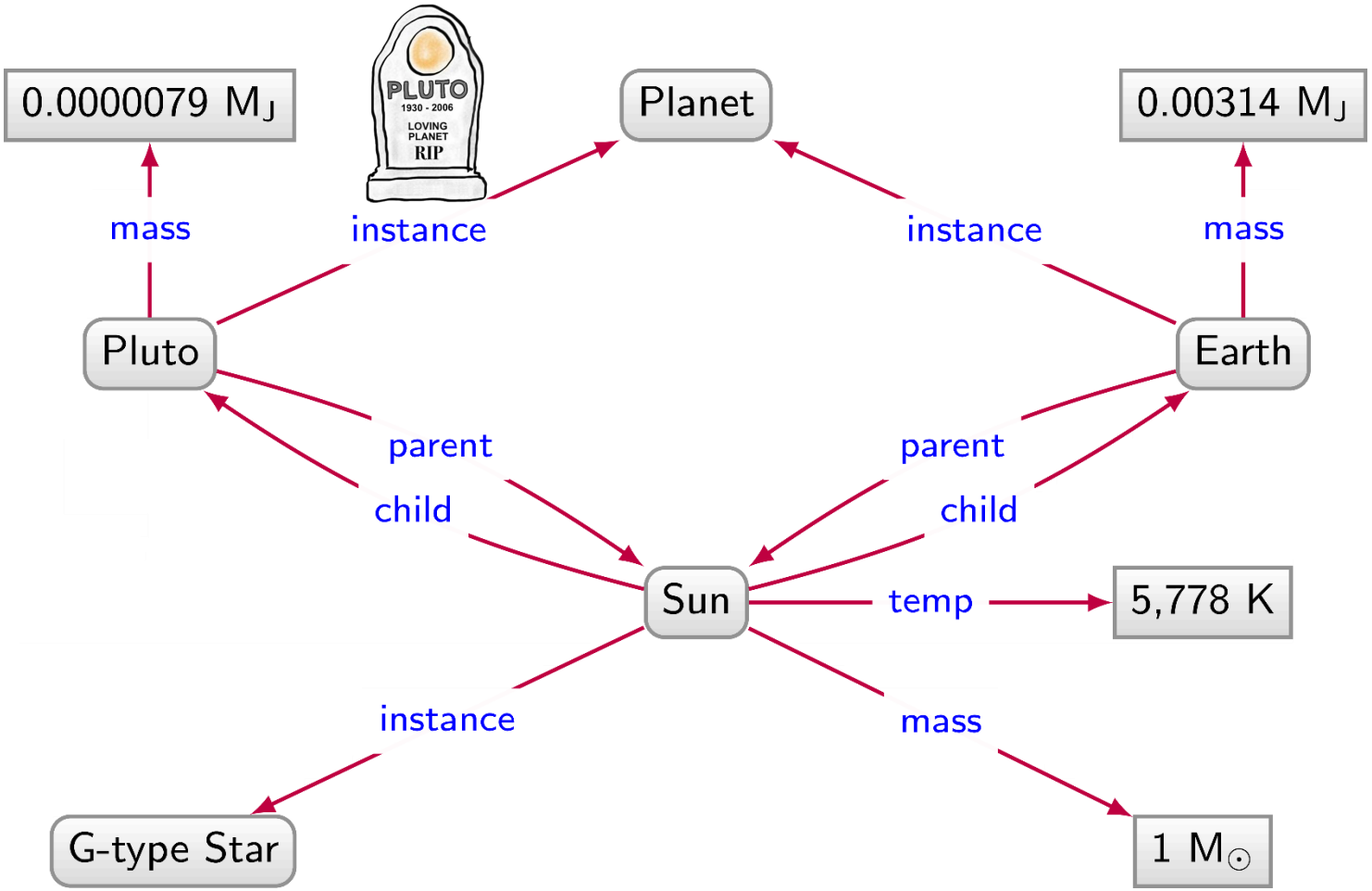
Planets / Graph Database



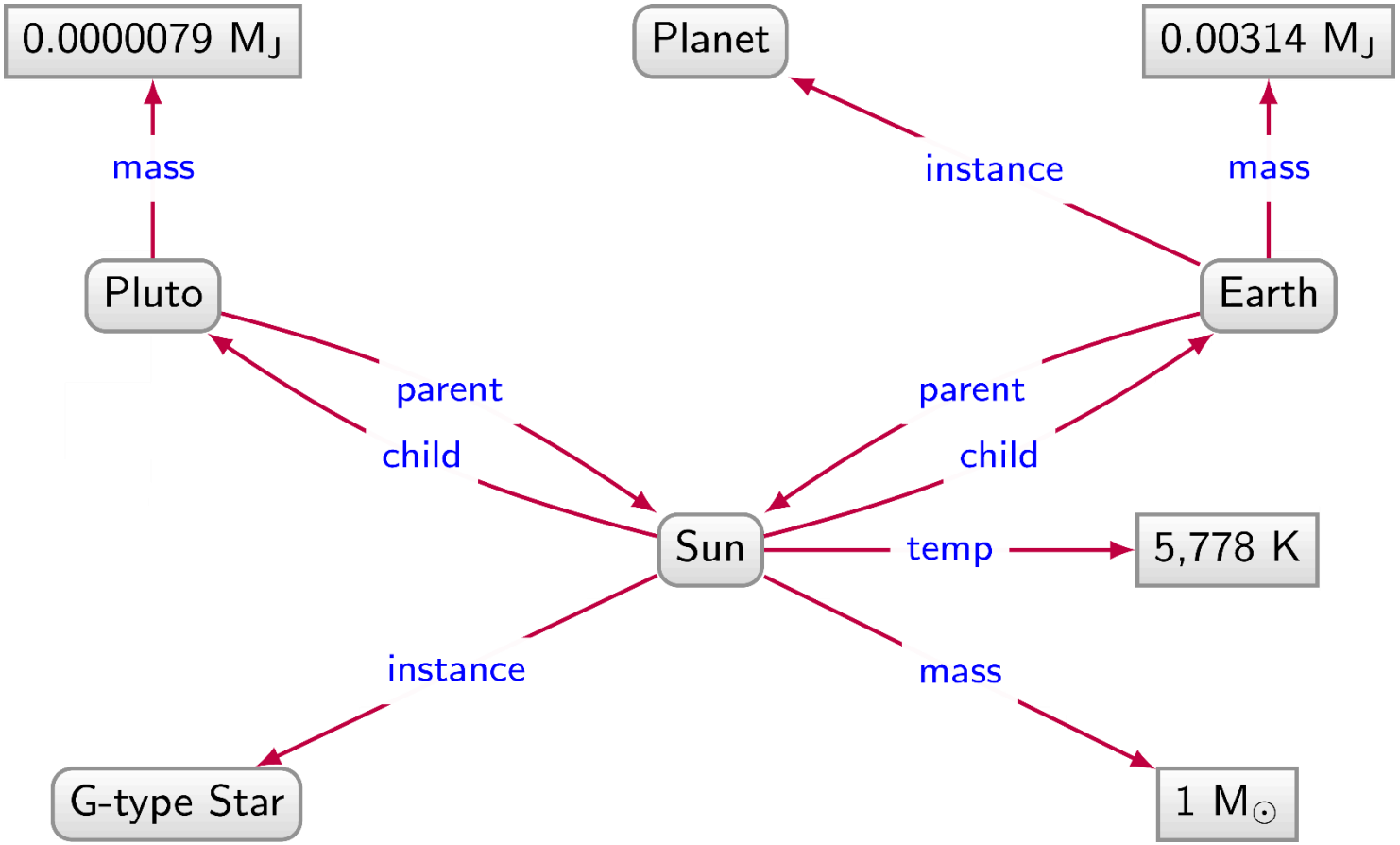
Planets / Graph Database



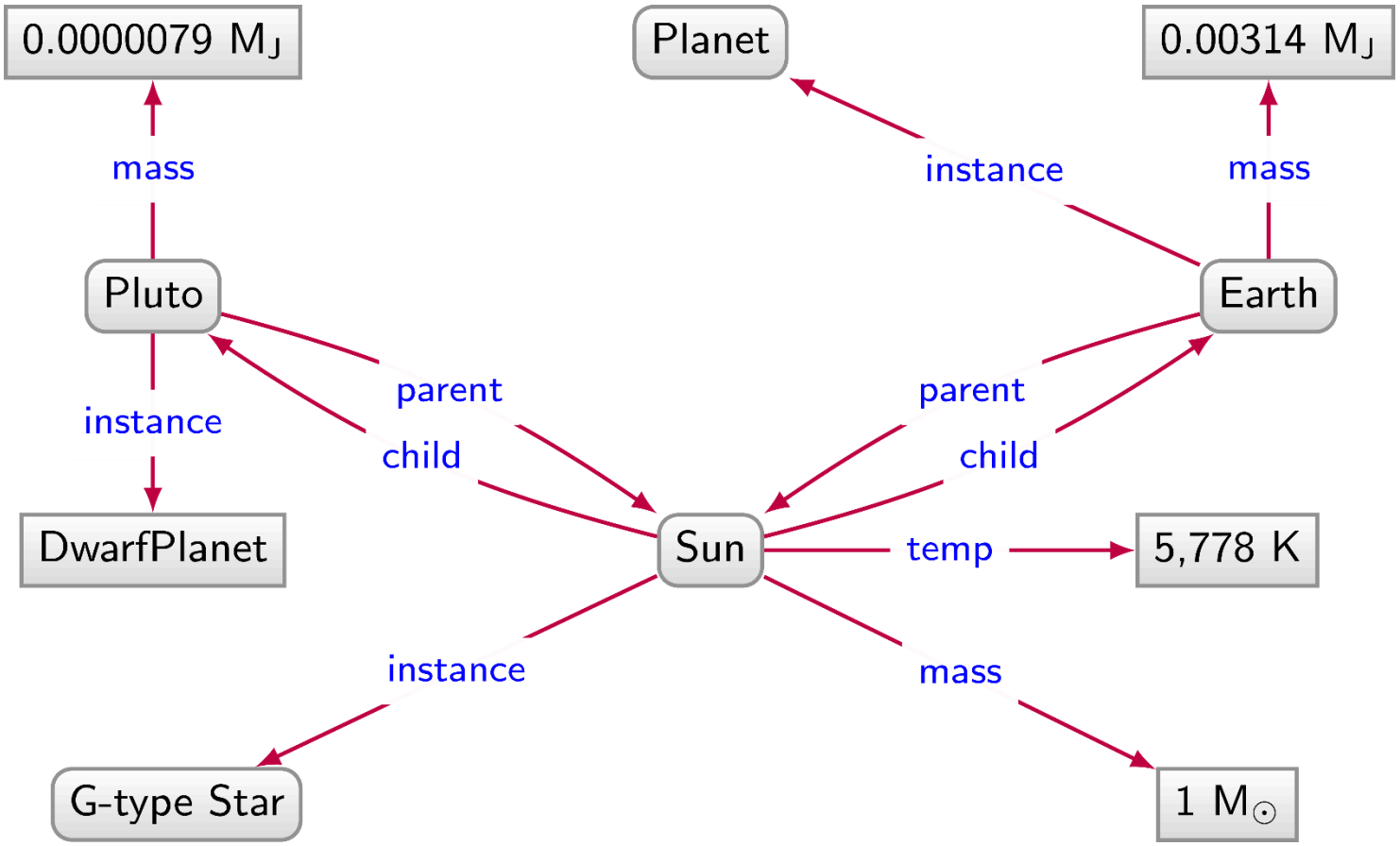
Planets / Graph Database



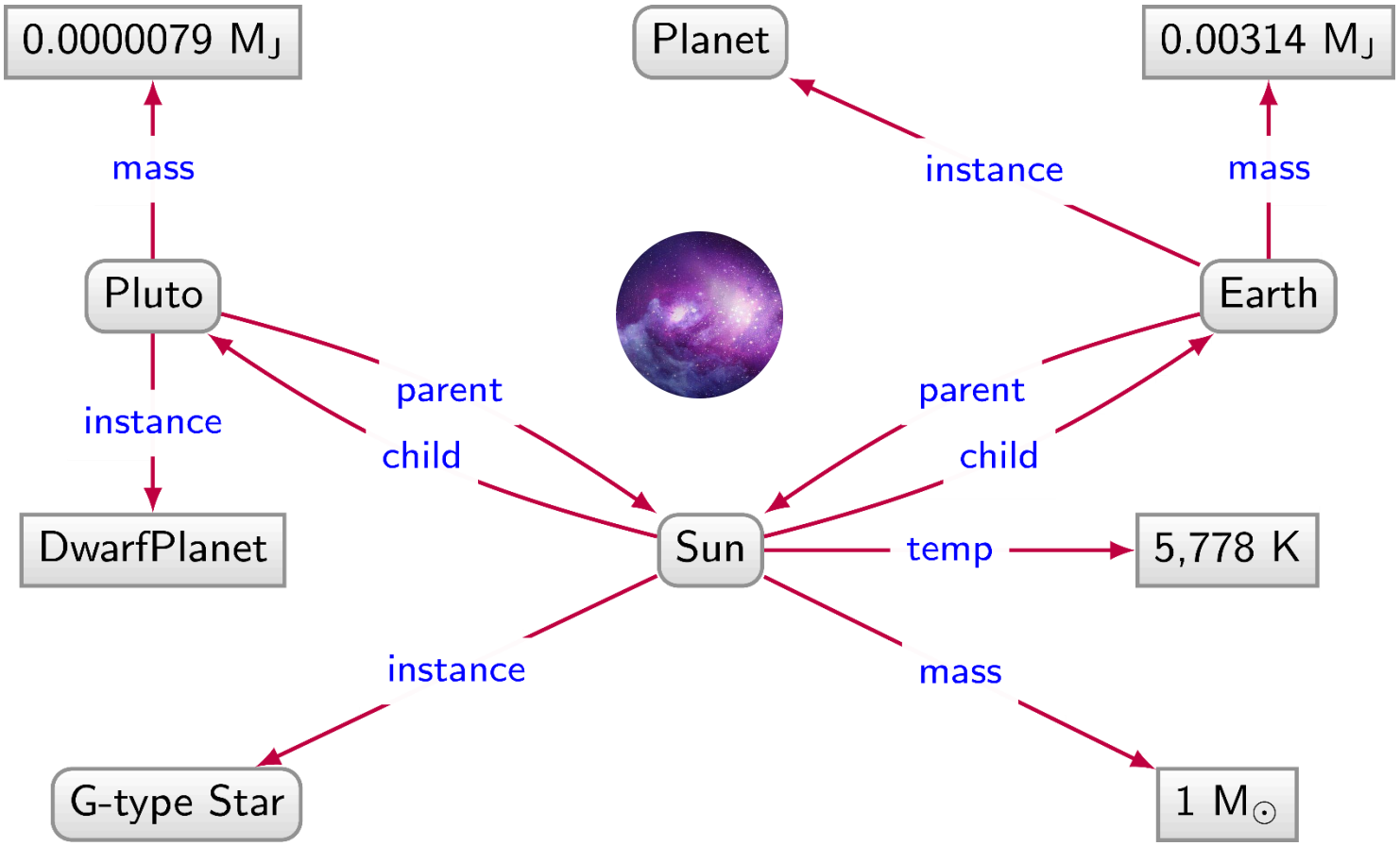
Planets / Graph Database



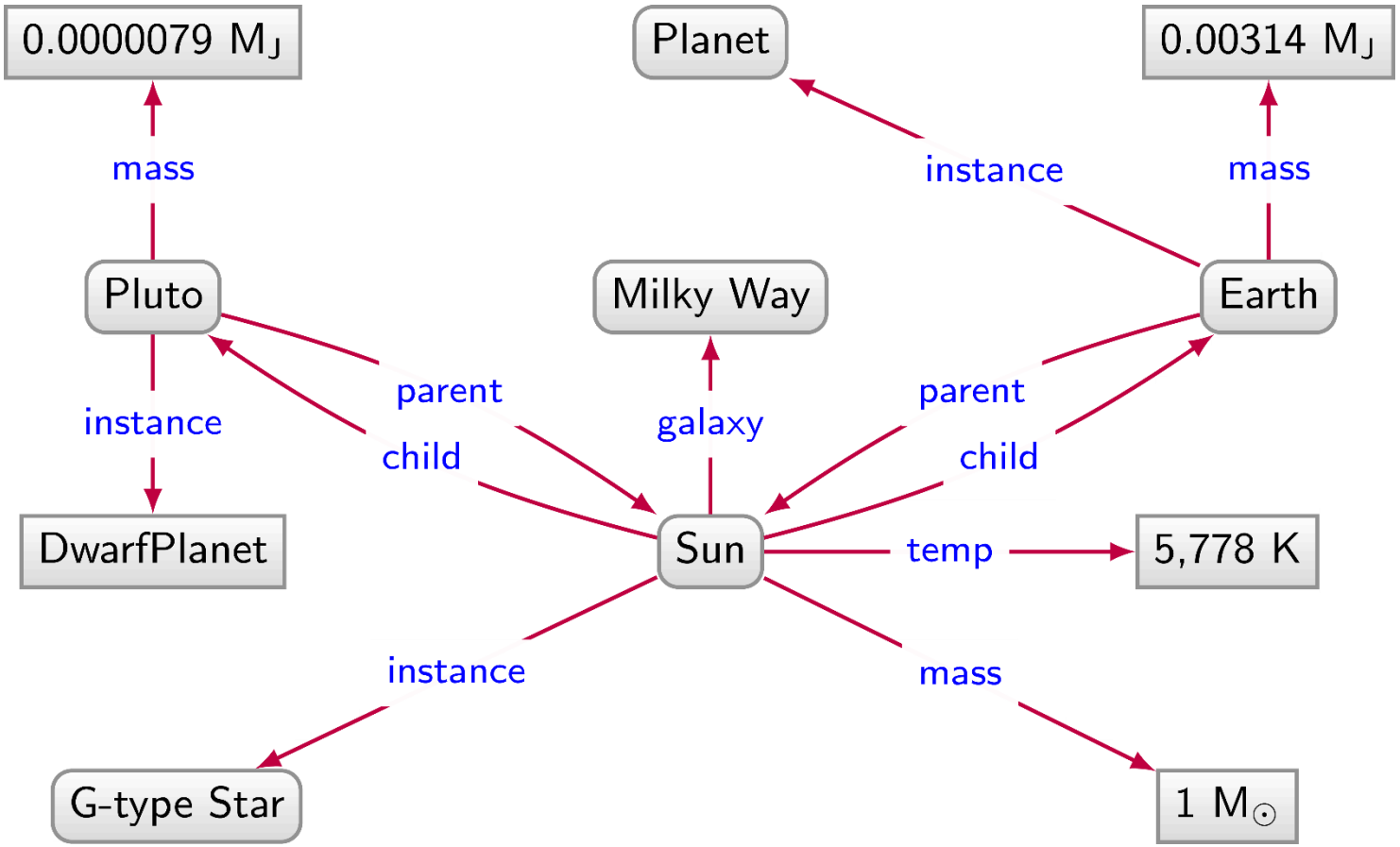
Planets / Graph Database



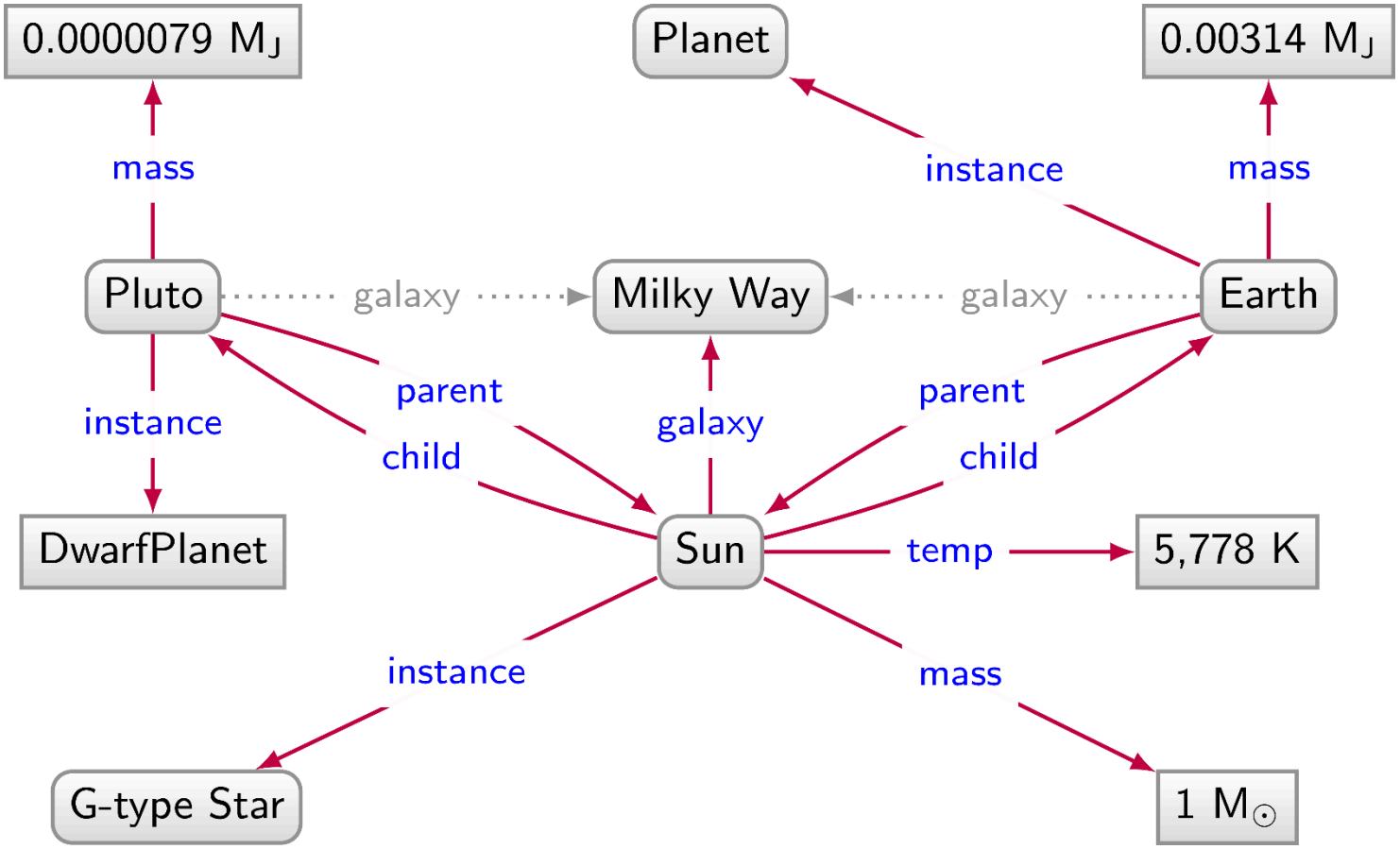
Planets / Graph Database



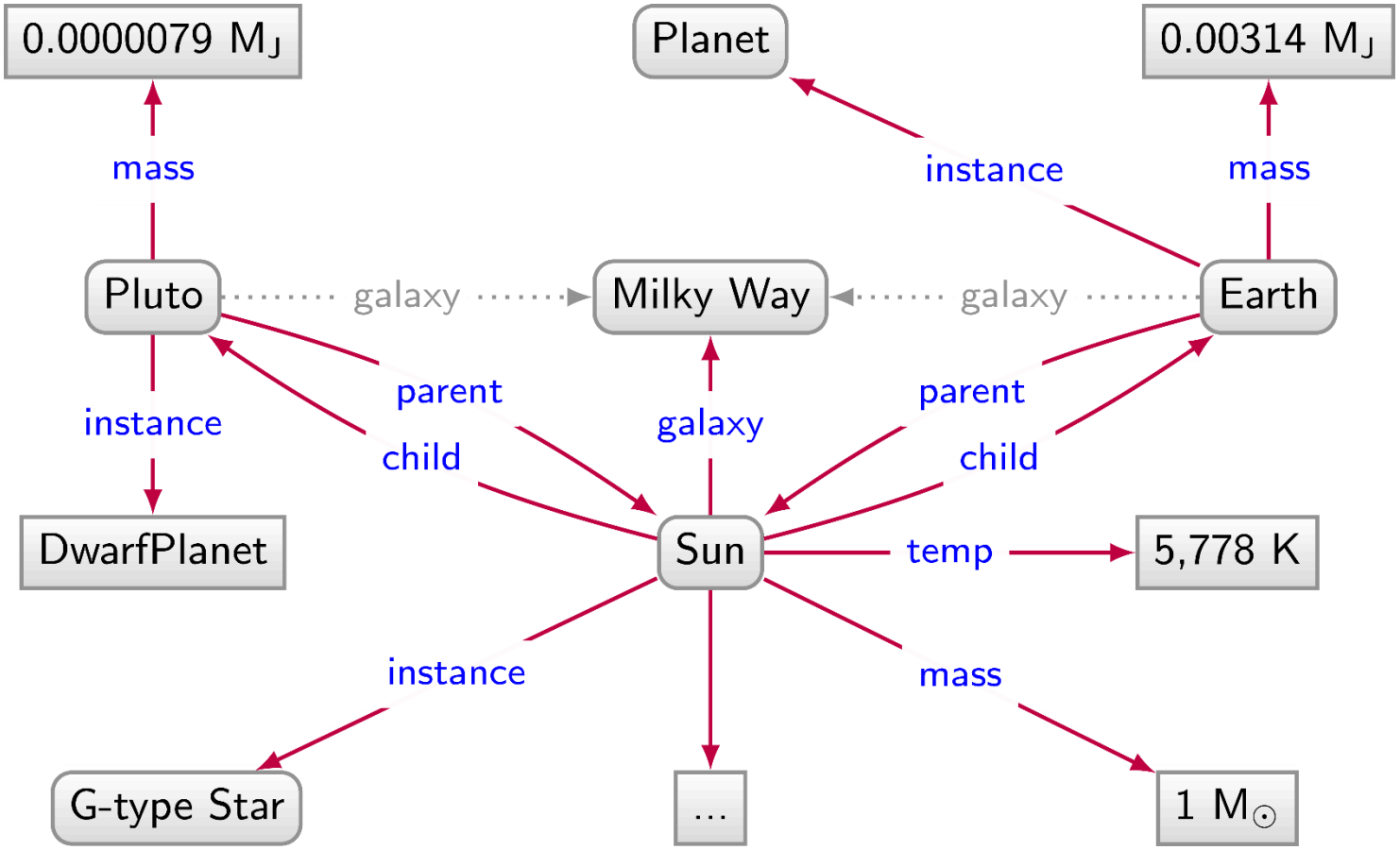
Planets / Graph Database



Planets / Graph Database



Planets / Graph Database

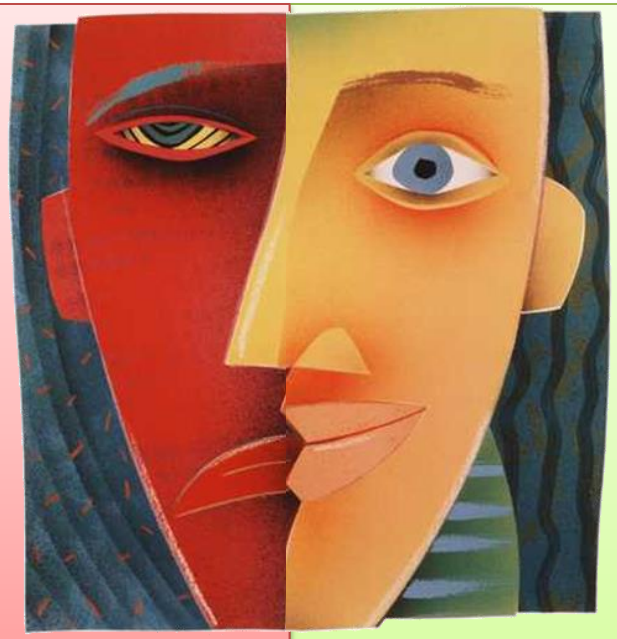


Relational databases: Pros and Cons

Planet

name	dist	radius	grav	days	years	temp	ring
Mercury	0.39	0.38	2.8	58.646	0.241	440	false
Venus	0.72	0.95	8.9	-243.019	0.615	730	false

We have to impose a
structure (schema)
from the start



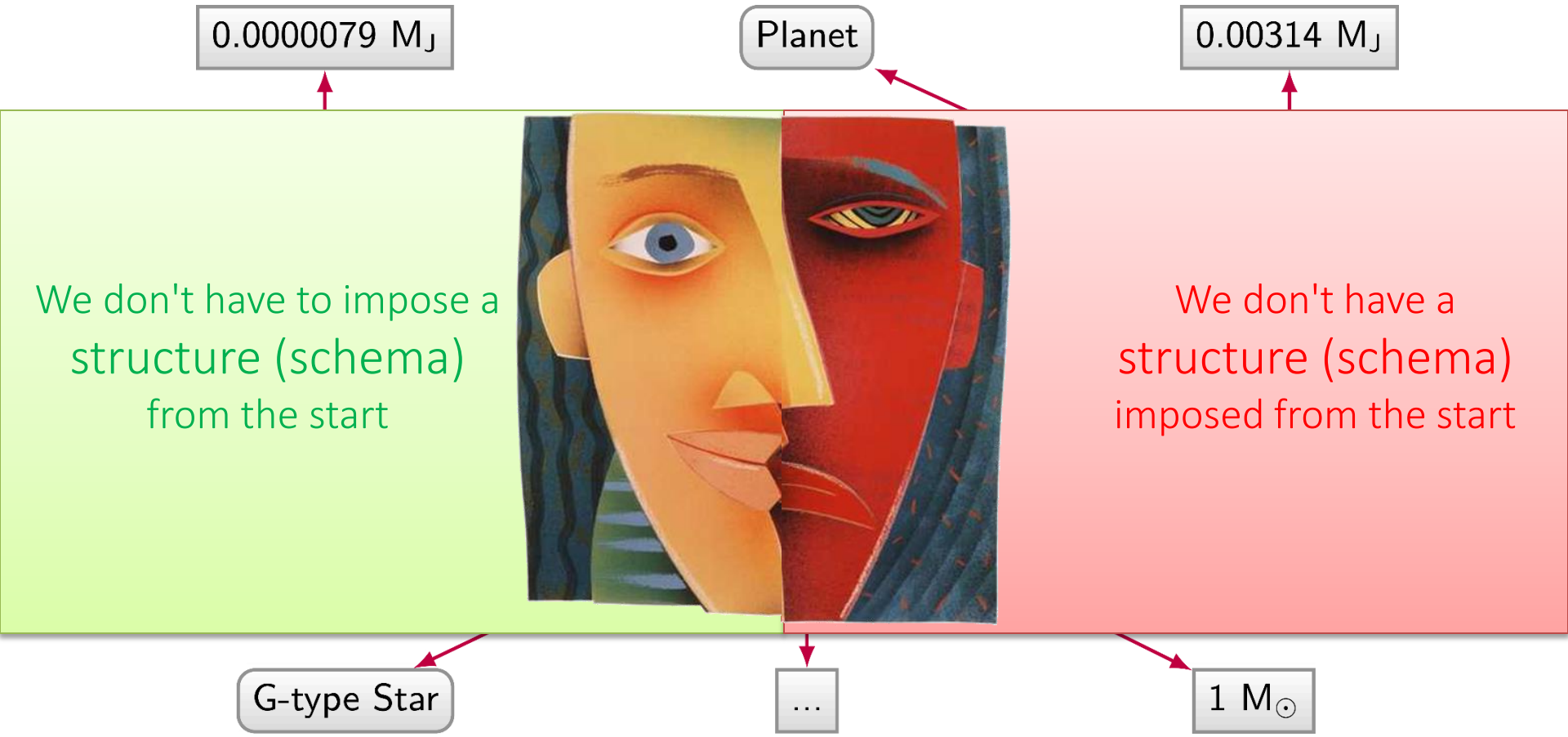
We have a
structure (schema)
imposed from the start

Europa	Jupiter
Io	Jupiter
Titan	Saturn
Triton	Neptune
Luna	Earth
Oberon	Uranus
Charon	Pluto
...	...

Europa	Galileo Galilei
Io	Galileo Galilei
Titan	Christiaan Huygens
Triton	William Lassell
Oberon	William Herschel
...	...

Europa	1610
Io	1610
Titan	1655
Triton	1846
Oberon	1787
Charon	1978
...	...

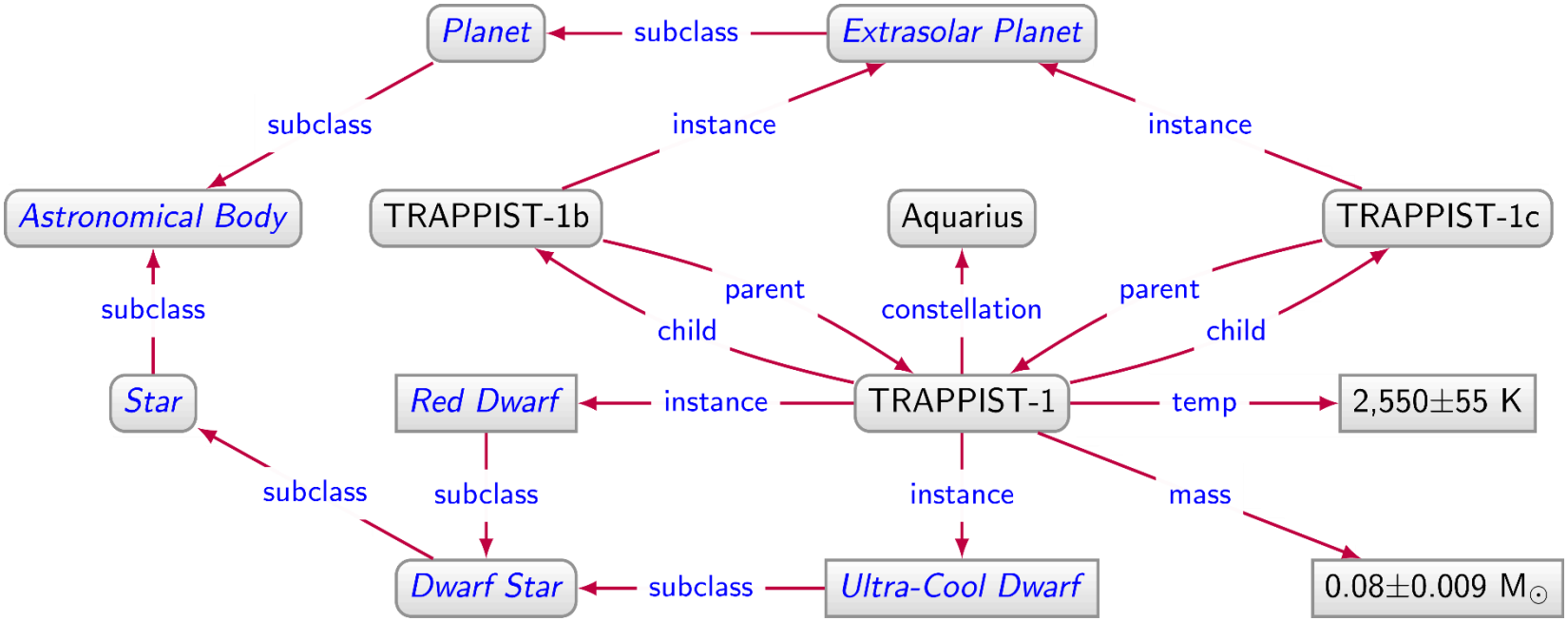
Graph Databases: Pros and Cons



Why do we need Graph Databases?

Path Queries

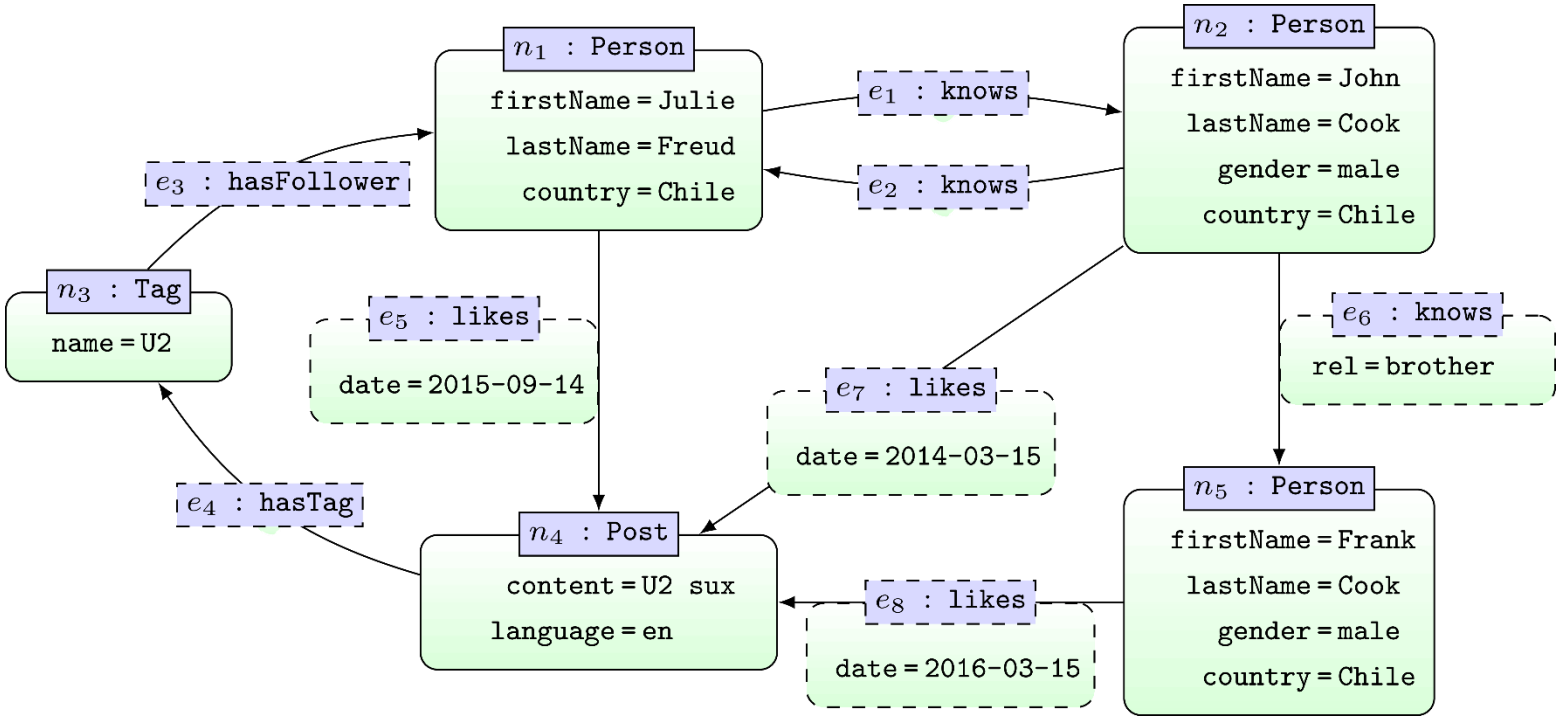
Directed Edge-labelled Graph




```
SELECT ?const (COUNT(DISTINCT ?body) AS ?num)
WHERE {
  ?body :instance/:subclass* :AstronomicalBody .
  ?body :parent?/:constellation ?const .
}
GROUP BY ?const
ORDER BY DESC(?num)
```

?const	?num
:Aquarius	3

Property Graph



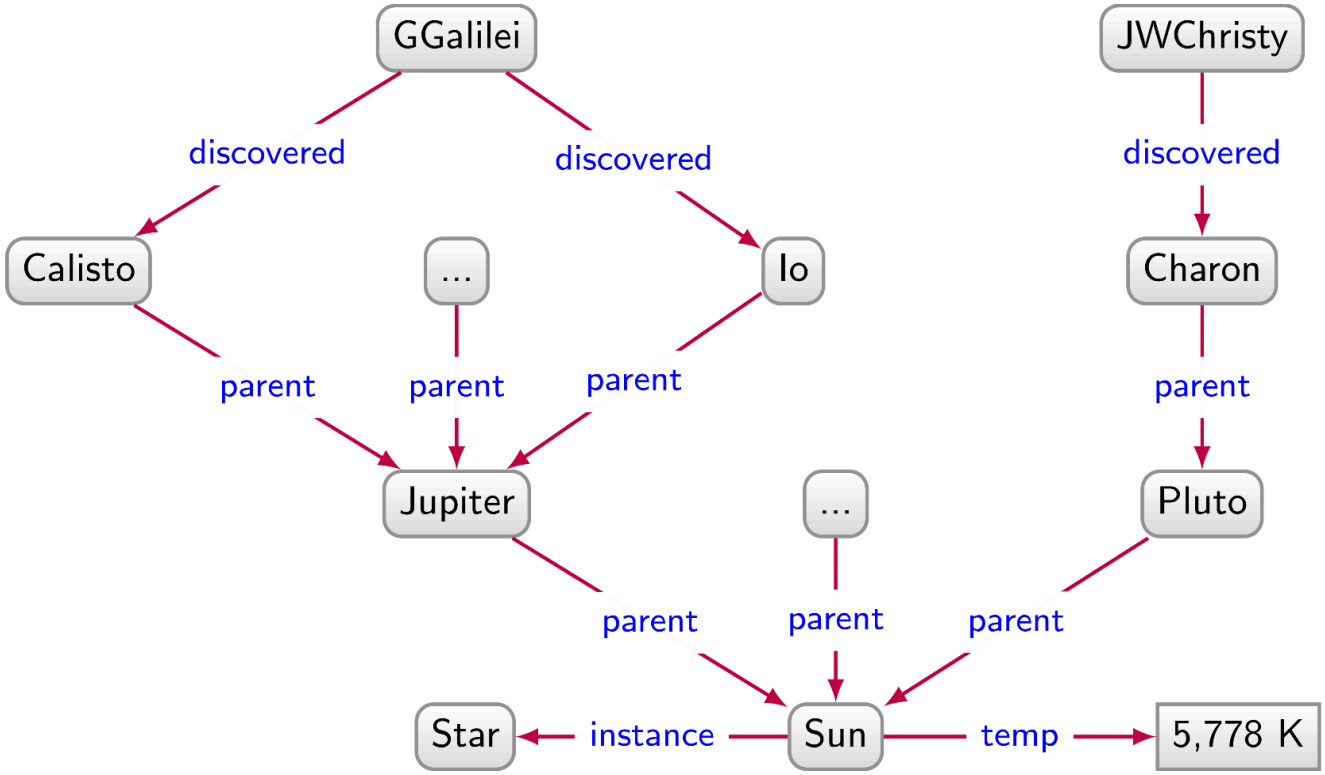

```
MATCH (x1:Person {firstName:"Julie"})-[:knows*]->(x2:Person)
MATCH (x2)-[:likes]->()-[:hasTag]->()-[:hasFollower]->(x1)
RETURN x2.firstName
```

???

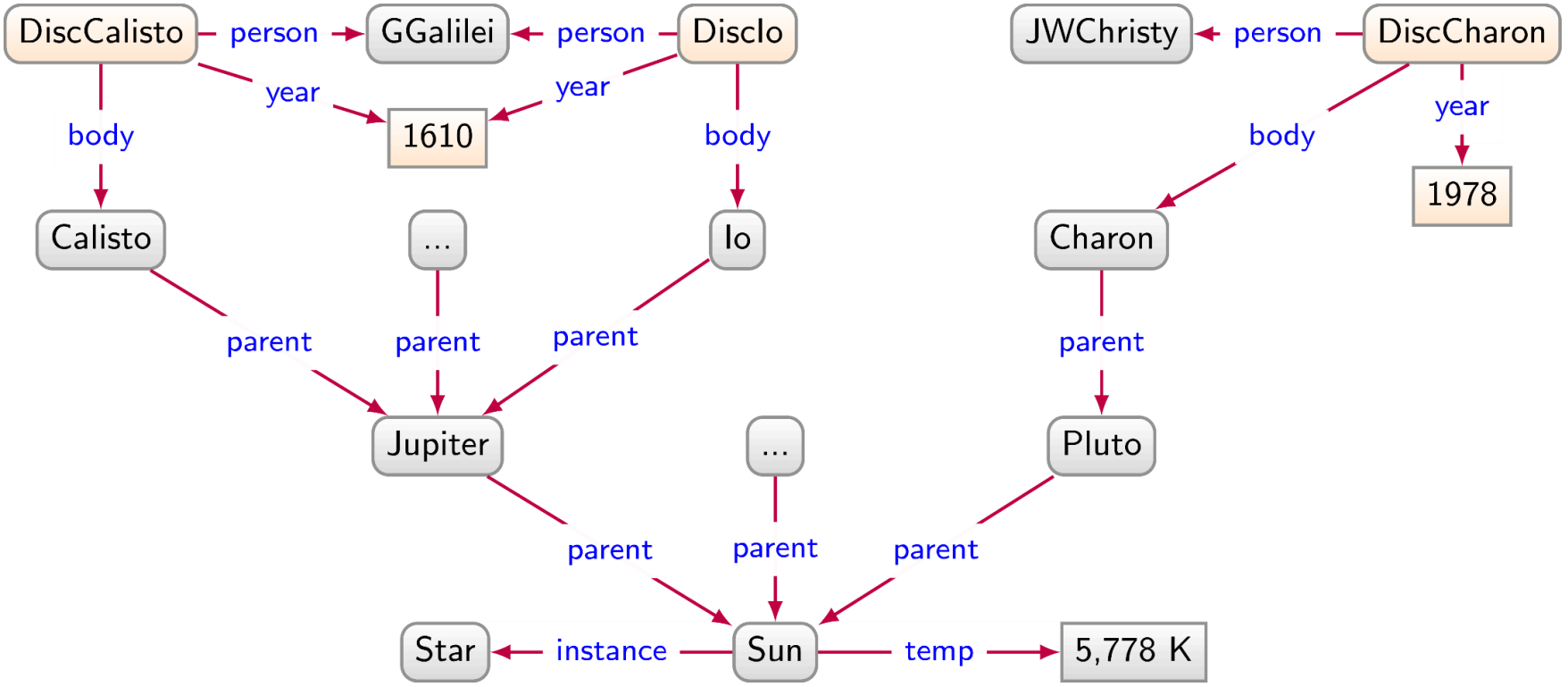
Why do we need Property Graphs?

Directed Labelled Graph

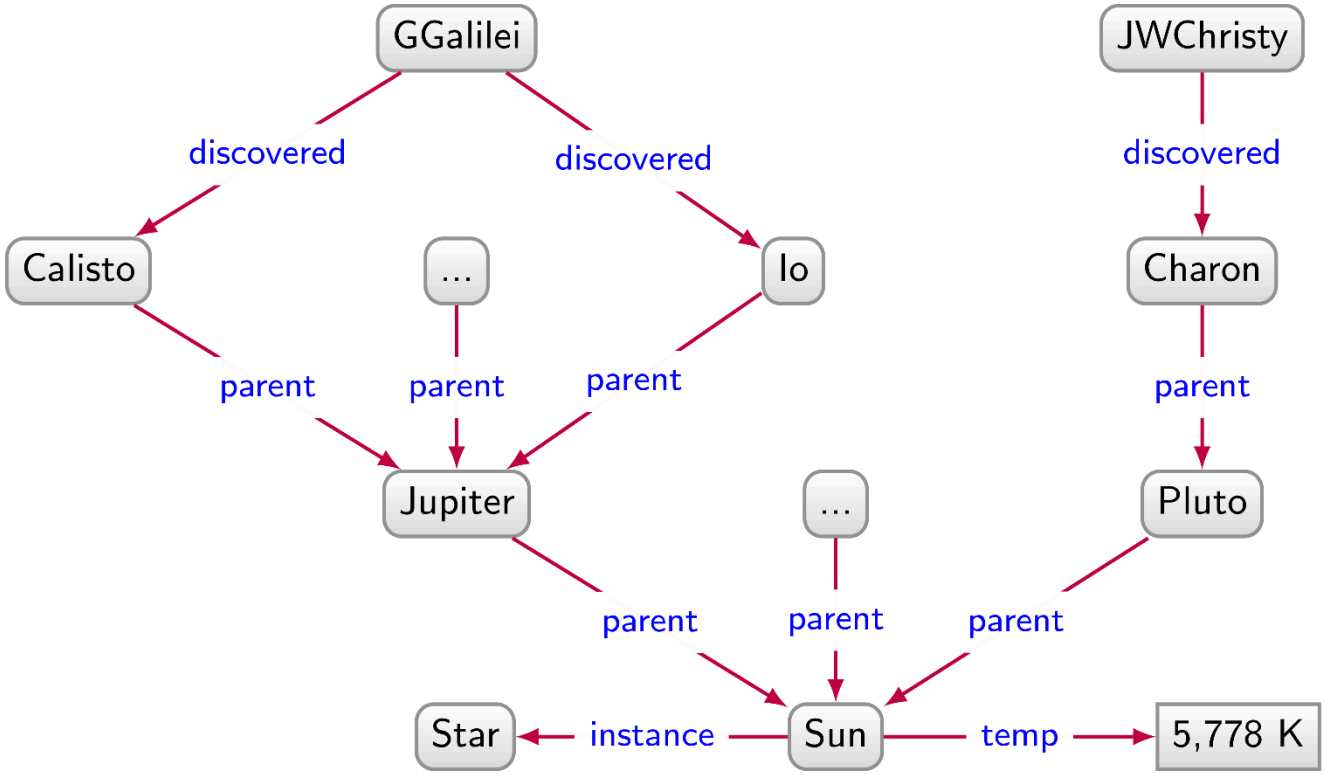
How can we say that Galileo Galilei discovered Calisto and Io in 1610 while James W. Christy discovered Charon in 1978?



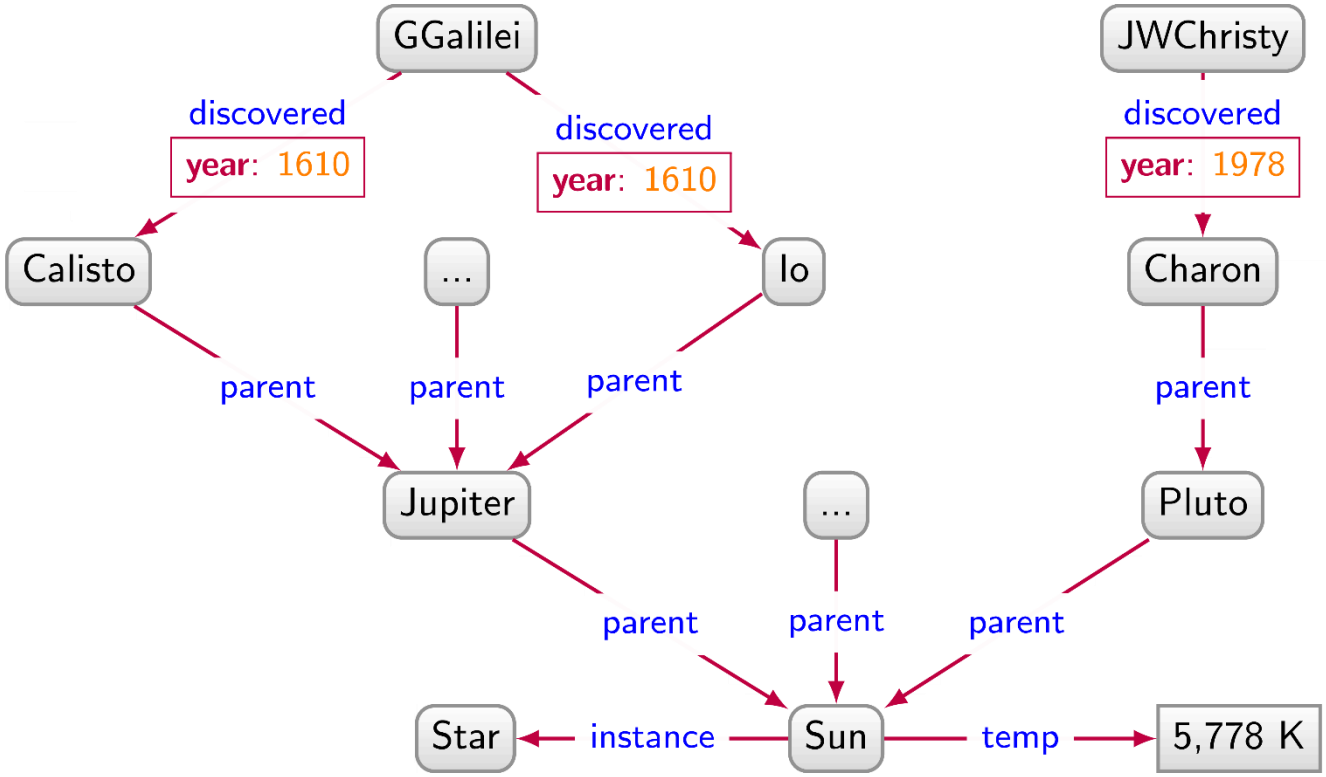
Directed Labelled Graph



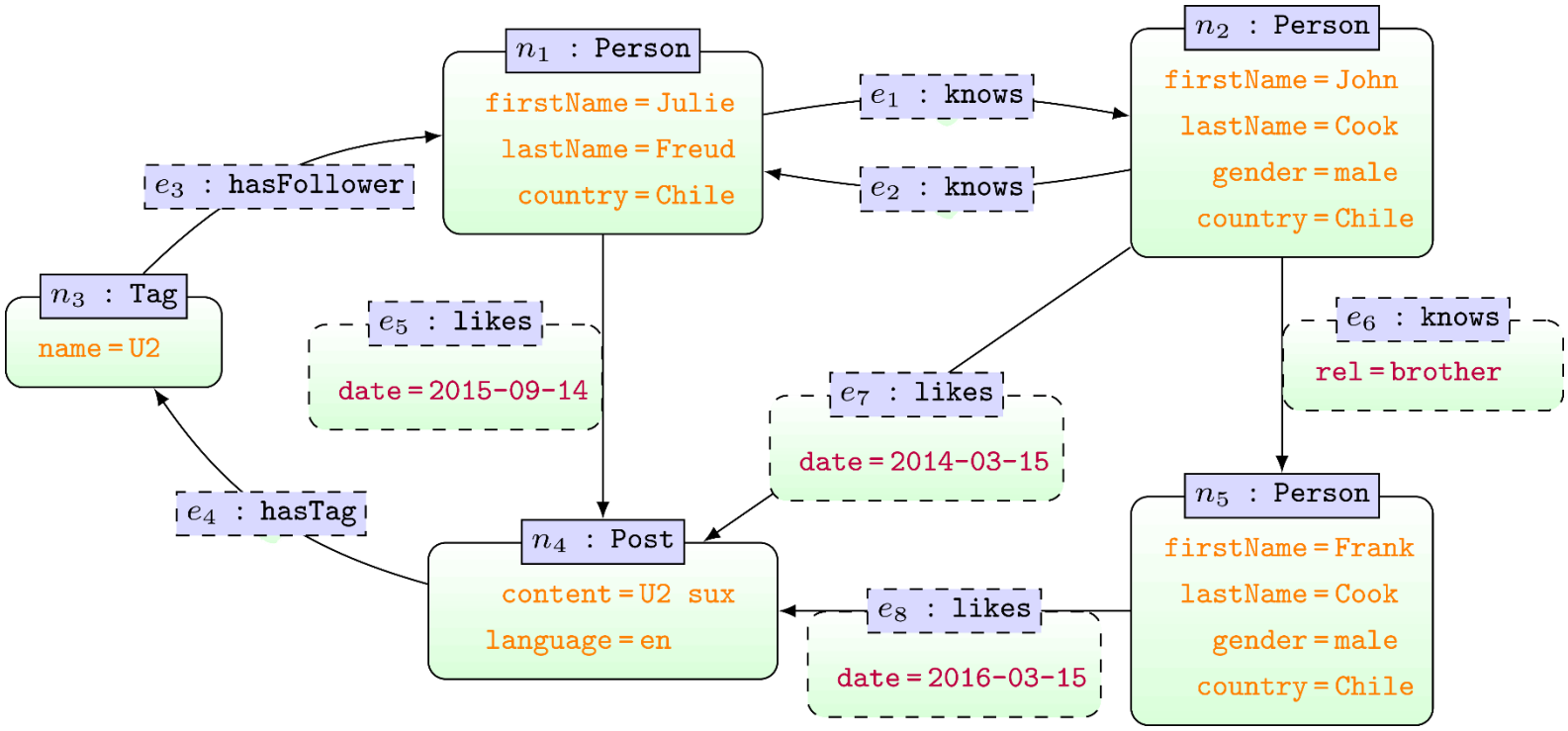
Wouldn't it have been nice to simply ...



Wouldn't it have been nice to simply ...

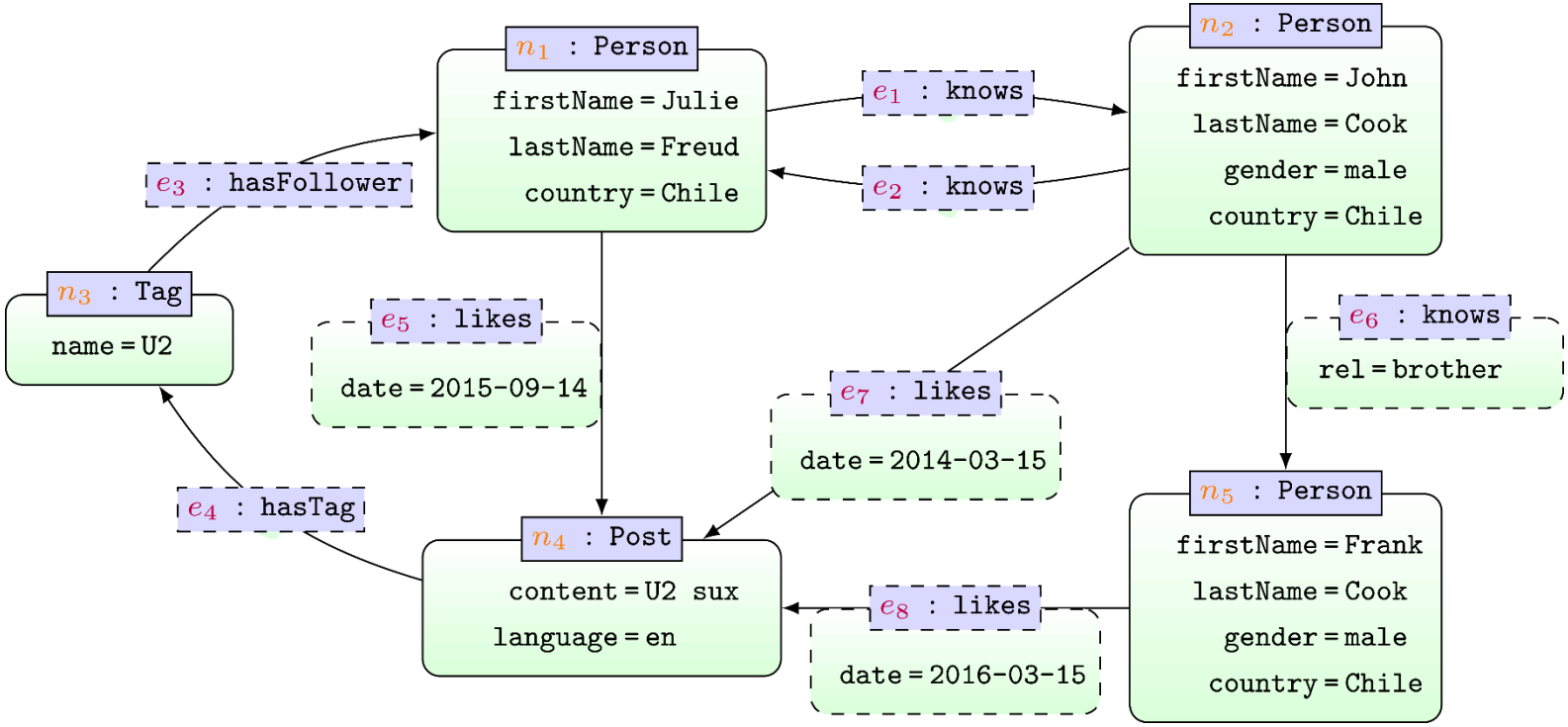


Property Graphs ...



... attributes on nodes and edges

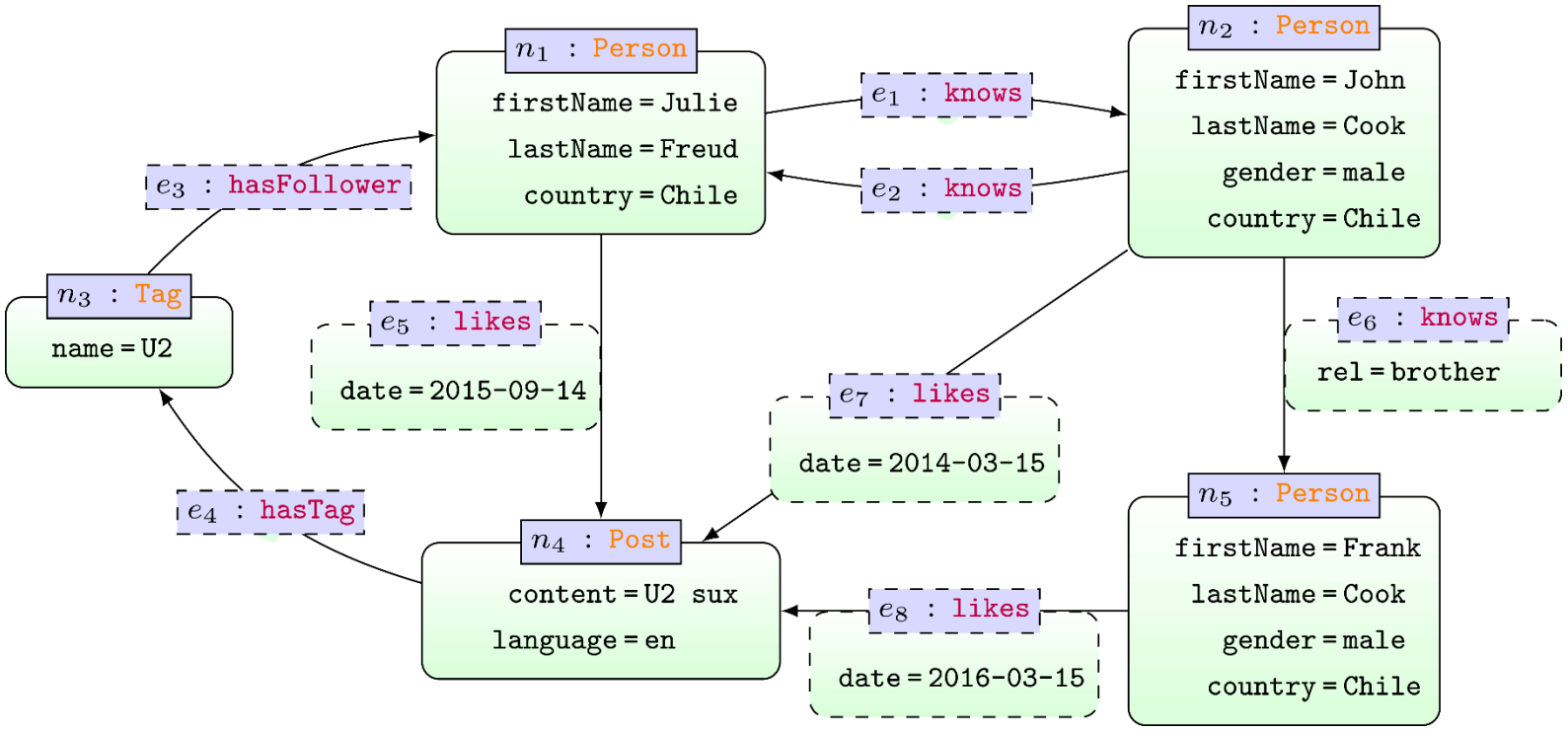
Property Graphs



... attributes on nodes and edges

... IDs on nodes and edges

Property Graphs



... attributes on nodes and edges

... IDs on nodes and edges

... labels on nodes and edges

Popular Graph Databases

347 systems in ranking, May 2019

Rank May 2019	Rank Apr 2019	Rank May 2018	DBMS	Database Model	Score May 2019	Score Apr 2019	Score May 2018
1.	1.	1.	Oracle	Relational, Multi-model	1285.55	+5.61	-4.87
2.	2.	2.	MySQL	Relational, Multi-model	1218.96	+3.82	-4.38
3.	3.	3.	Microsoft SQL Server	Relational, Multi-model	1072.19	+12.23	-13.66
4.	4.	4.	PostgreSQL	Relational, Multi-model	478.89	+0.17	+77.99
5.	5.	5.	MongoDB	Document	408.07	+6.10	+65.96
6.	6.	6.	IBM Db2	Relational, Multi-model	174.44	-1.61	-11.17
7.	↑ 8.	↑ 9.	Elasticsearch	Search engine, Multi-model	148.62	+2.62	+18.18
8.	↓ 7.	↓ 7.	Redis	Key-value, Multi-model	148.40	+2.03	+13.06
9.	9.	↓ 8.	Microsoft Access	Relational	143.78	-0.87	+10.67
10.	↑ 11.	10.	Cassandra	Wide column	125.72	+2.11	+7.89
11.	↓ 10.	11.	SQLite	Relational	122.90	-1.32	+7.44
12.	12.	↑ 14.	MariaDB	Relational, Multi-model	86.52	+1.29	+21.53
13.	13.	13.	Splunk	Search engine	85.24	+2.15	+20.15
14.	↑ 15.	↑ 18.	Hive	Relational	77.90	+3.19	+20.93
15.	↓ 14.	↓ 12.	Teradata	Relational	76.04	+0.69	+1.63
16.	16.	↓ 15.	Solr	Search engine	60.80	+0.57	-0.72
17.	17.	17.	HBase	Wide column	59.77	+1.11	-0.18
18.	18.	↑ 19.	FileMaker	Relational	58.51	+0.09	+3.84
19.	19.	↑ 21.	Amazon DynamoDB	Multi-model	55.93	-0.08	+11.74
20.	↑ 21.	20.	SAP HANA	Relational, Multi-model	55.74	+0.39	+7.37
21.	↓ 20.	↓ 16.	SAP Adaptive Server	Relational	55.44	-0.36	-6.07
22.	22.	22.	Neo4j	Graph	51.03	+1.54	+10.45
23.	23.	↑ 24.	Couchbase	Document	34.67	-1.61	+2.26
24.	↑ 25.	↓ 23.	Memcached	Key-value	28.90	+0.17	-4.66
25.	↓ 24.	↑ 26.	Microsoft Azure SQL Database	Relational, Multi-model	28.77	-0.02	+3.56

http://db-engines.com/en/ranking

☐ include secondary database models

32 systems in ranking, June 2019

Rank Jun 2019	Rank May 2019	Rank Jun 2018	DBMS	Database Model	Score Jun 2019	Score May 2019	Score Jun 2018
1.	1.	1.	Neo4j	Graph	49.56	-1.48	+7.58
2.	2.	2.	Microsoft Azure Cosmos DB	Multi-model	28.25	+0.65	+9.05
3.	3.	3.	OrientDB	Multi-model	5.59	-0.78	+0.25
4.	4.	4.	ArangoDB	Multi-model	4.57	-0.22	+1.05
5.	5.	5.	Virtuoso	Multi-model	3.11	-0.21	+1.33
6.	6.	↑ 11.	JanusGraph	Graph	1.55	-0.07	+1.19
7.	7.	7.	Amazon Neptune	Multi-model	1.24	-0.09	+0.57
8.	↑ 10.	↑ 10.	GraphDB	Multi-model	1.09	+0.05	+0.69
9.	↓ 8.	↓ 6.	Giraph	Graph	1.08	-0.10	+0.12
10.	↑ 11.	↓ 8.	AllegroGraph	Multi-model	0.93	+0.02	+0.37
11.	↓ 9.	↑ 21.	Dgraph	Graph	0.89	-0.15	+0.77
12.	↑ 13.	↑ 15.	TigerGraph	Graph	0.72	+0.02	+0.55
13.	↓ 12.	↓ 9.	Stardog	Multi-model	0.72	-0.03	+0.22
14.	14.	↓ 13.	Sqrrl	Multi-model	0.59	-0.01	+0.30
15.	15.	↑ 18.	Blazegraph	Multi-model	0.56	0.00	+0.44
16.	16.	↓ 12.	Graph Engine	Multi-model	0.53	-0.01	+0.23
17.	17.	↓ 14.	InfiniteGraph	Graph	0.38	0.00	+0.19
18.	18.	↑ 20.	FaunaDB	Multi-model	0.36	-0.03	+0.24
19.	19.	19.	FlockDB	Graph	0.27	+0.00	+0.16
20.	↑ 22.	↑ 22.	InfoGrid	Graph	0.26	+0.04	+0.16
21.	↓ 20.	↑ 24.	AgensGraph	Multi-model	0.24	-0.02	+0.20
22.	↓ 21.	↑ 28.	AnzoGraph	Multi-model	0.21	-0.02	+0.21
23.	23.	↓ 17.	HyperGraphDB	Graph	0.21	0.00	+0.07
24.	24.	↑ 27.	GRAKN.AI	Multi-model	0.21	+0.04	+0.20
25.	↑ 26.	↓ 16.	Sparksee	Graph	0.13	+0.00	-0.04

Neo4j

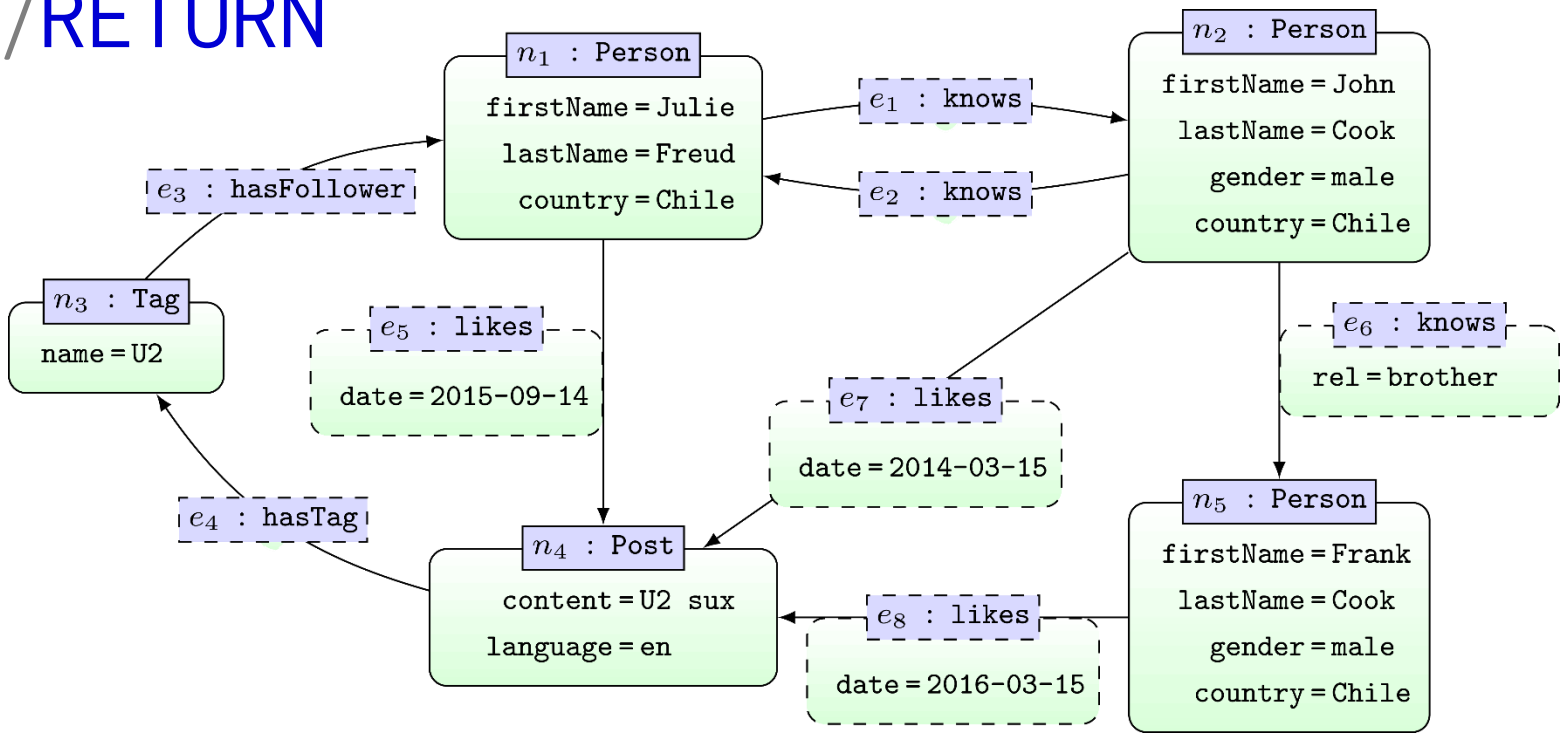
Neo4j Graph Database

- Data Model: Property Graphs
- Query Language: Cypher
- Scripting Language: Gremlin
- Licence: Open Source (Single Machine)
  Commercial (Cluster Edition)

Cypher:
MATCH/RETURN

https://neo4j.com/docs/developer-manual/3.4/cypher/clauses/match/

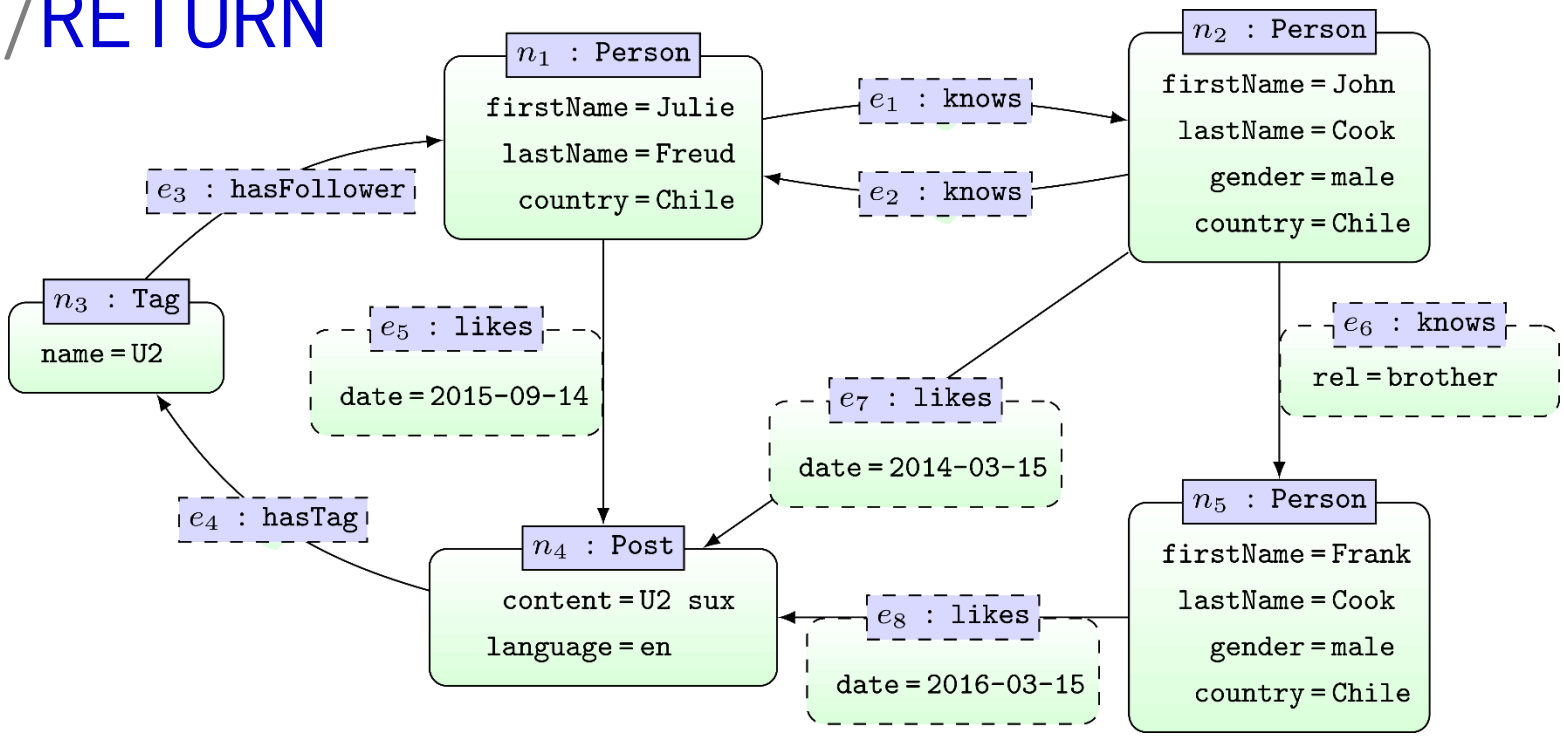
MATCH/RETURN




```
MATCH (x:Post)
RETURN x
```

x
(:Post {content: "U2 sux", language: "en"})

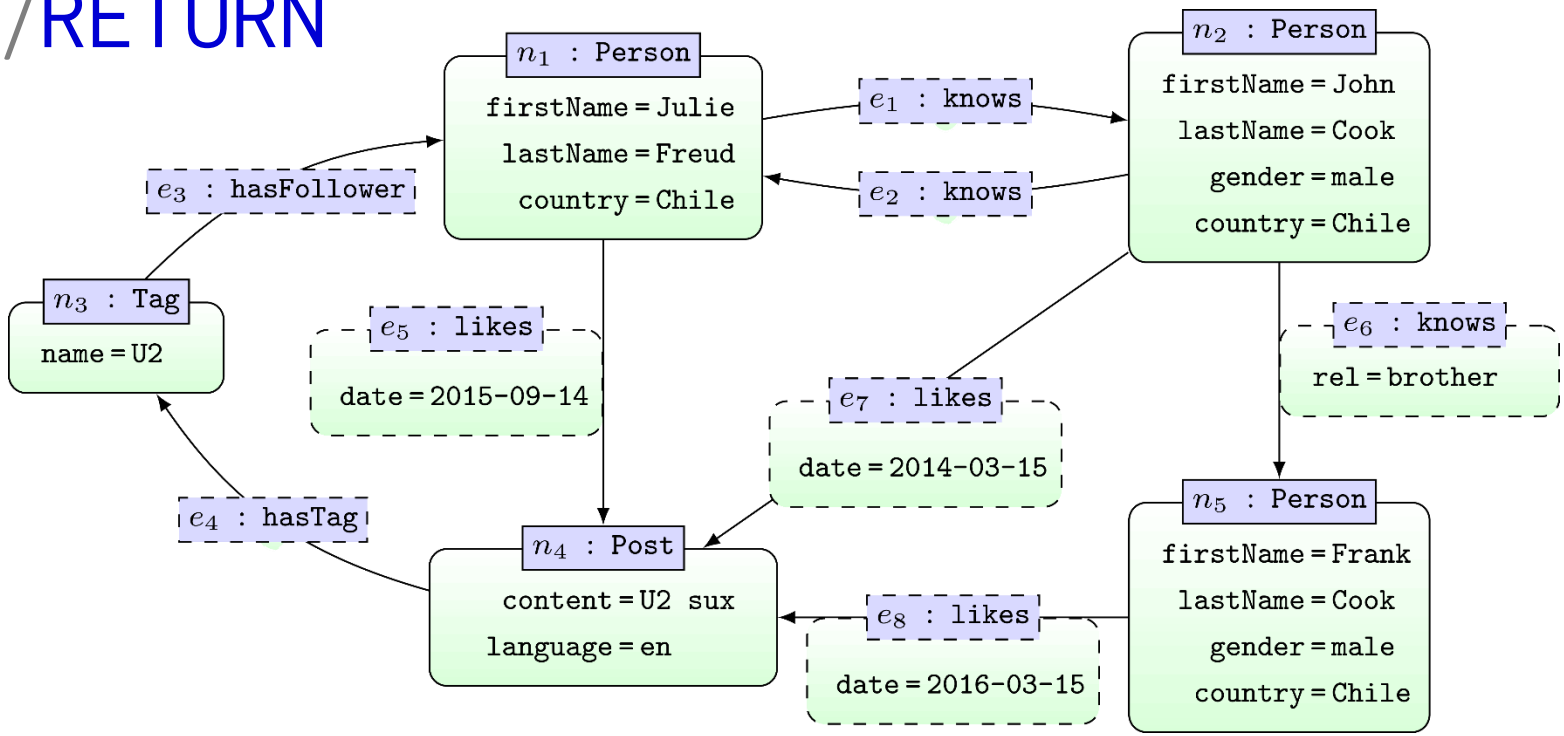
MATCH/RETURN




```
MATCH (x:Person)
RETURN x.firstName
```

x.firstName
Julie
John
Frank

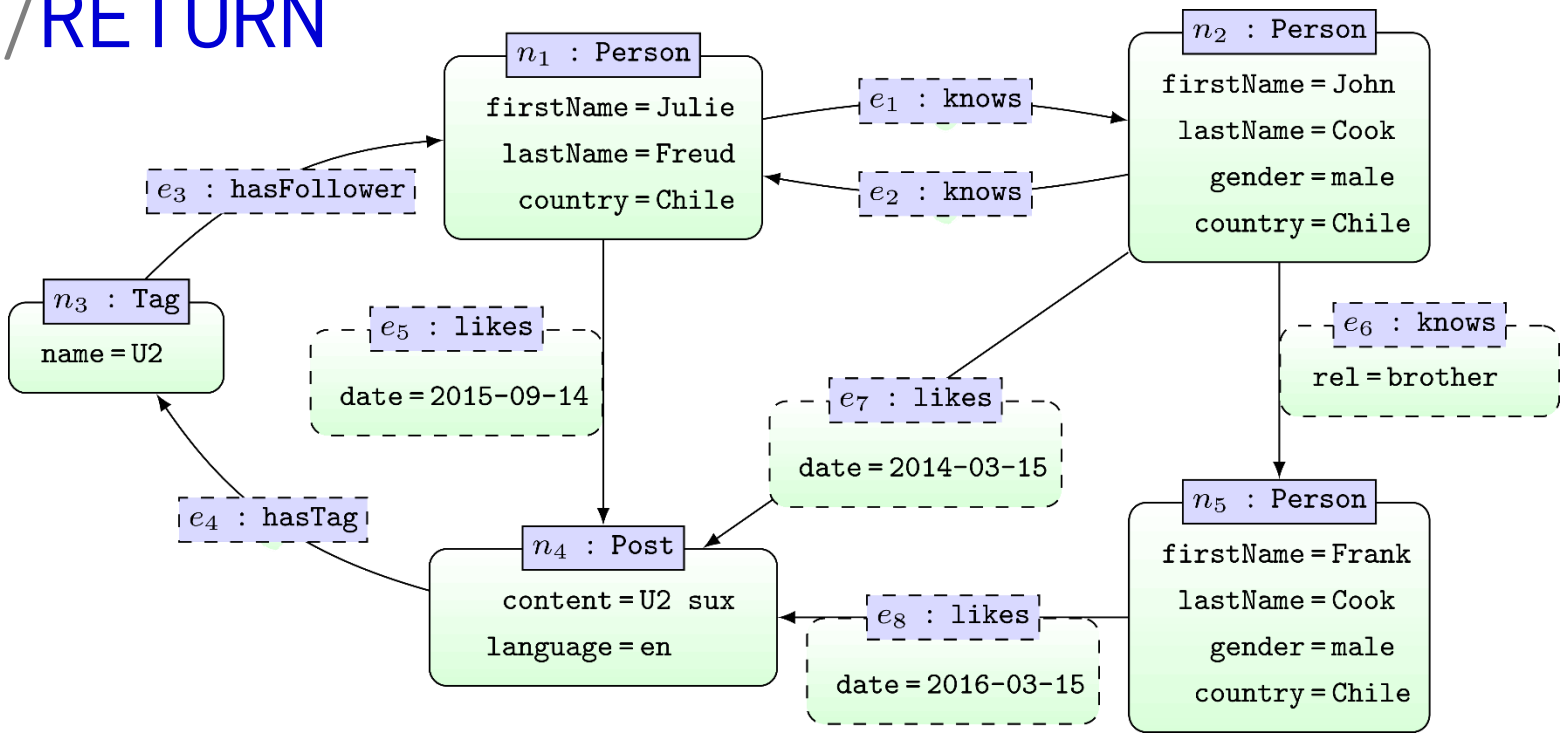
MATCH/RETURN




```
MATCH (x:Person {gender: "male", lastName: "Cook"})
RETURN x.firstName
```

x.firstName
John
Frank

MATCH/RETURN

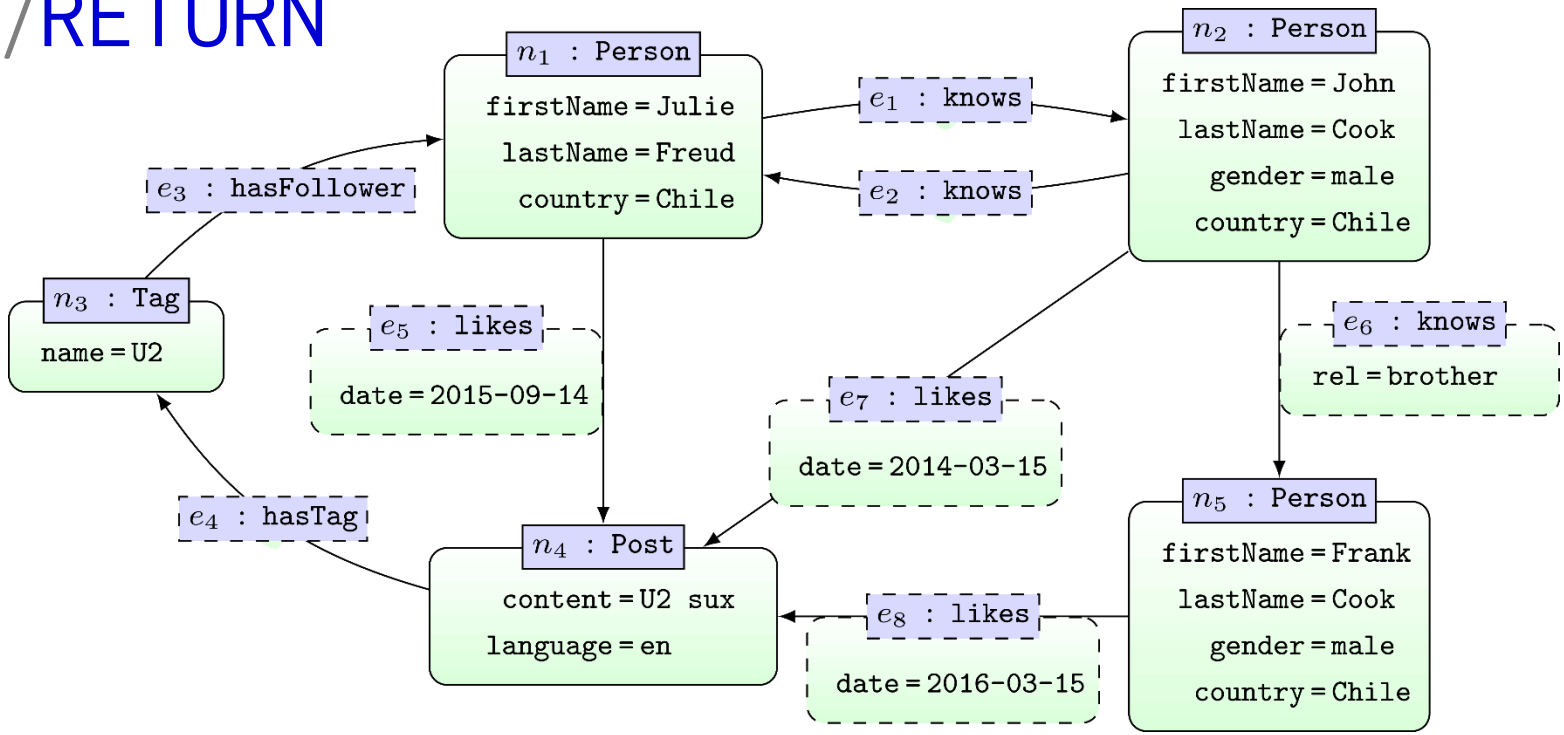



```
MATCH (x:Person)
RETURN x.firstName,x.gender
```

x.firstName	x.gender
Julie	
John	male
Frank	male

... matching nodes returned with blank attributes

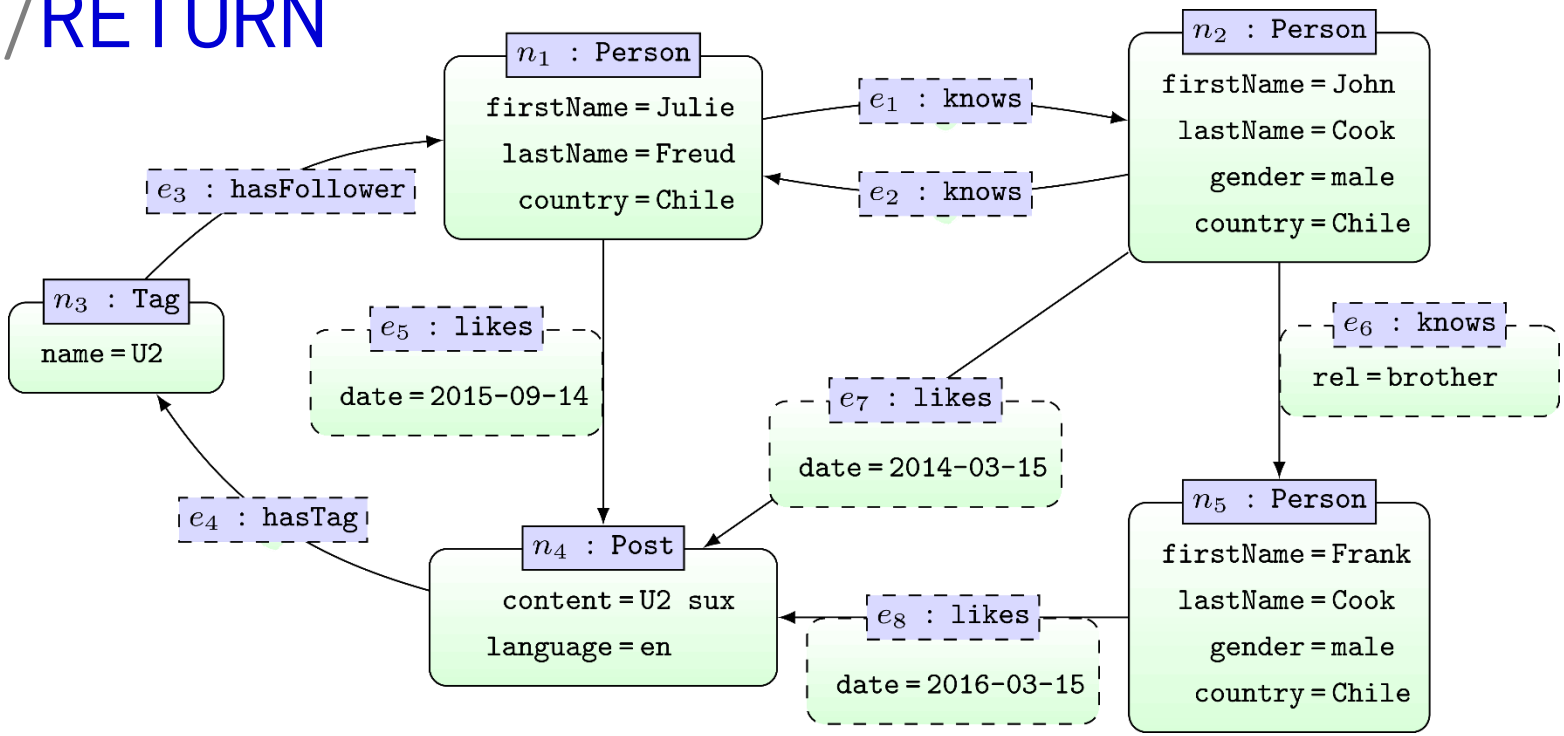
MATCH/RETURN




```
MATCH (x)
RETURN x.firstName,x.gender
```

x.firstName	x.gender
Julie	
John	male
Frank	male

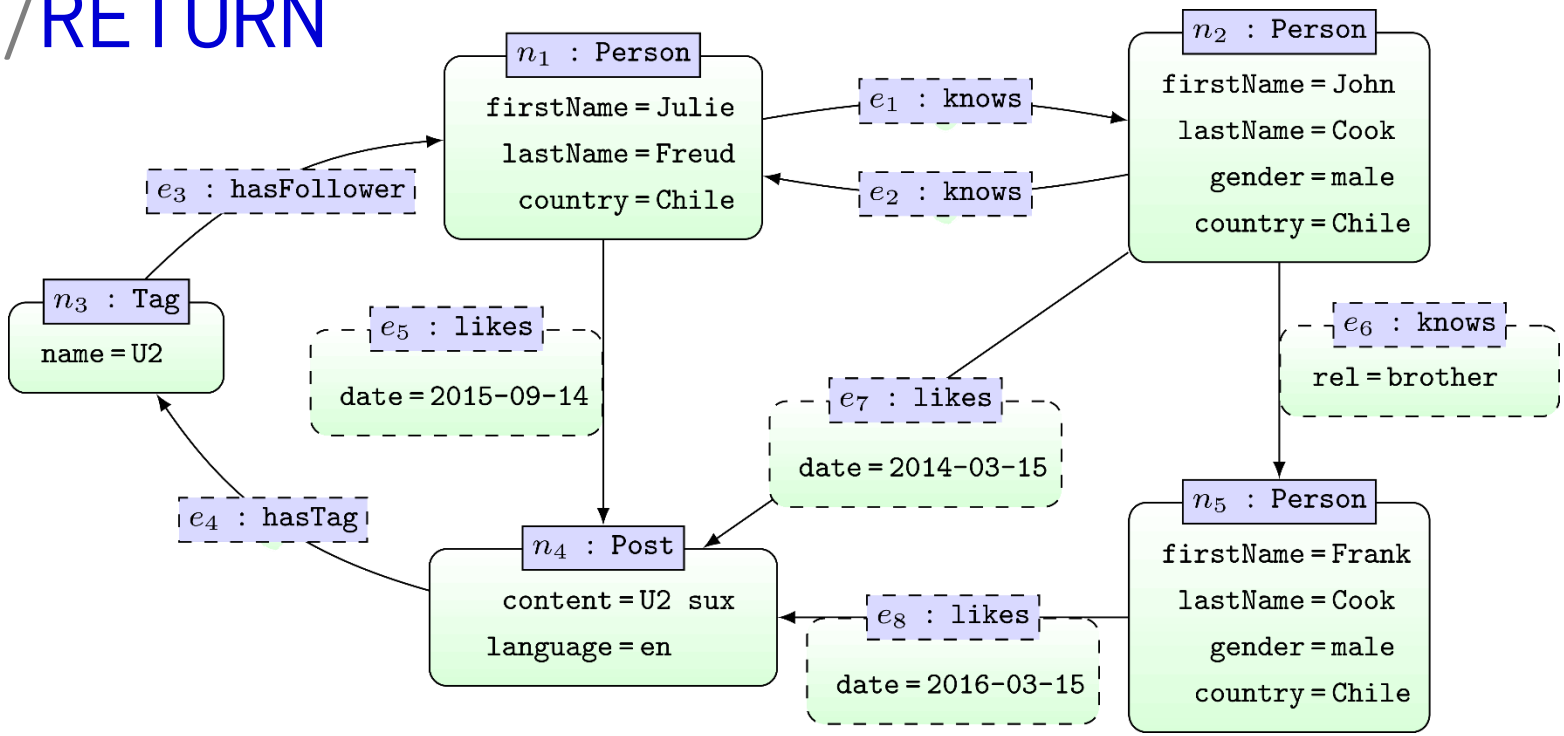
MATCH/RETURN




```
MATCH (:Person)-->(x:Person)
RETURN x.firstName
```

x.firstName
Julie
John
Frank

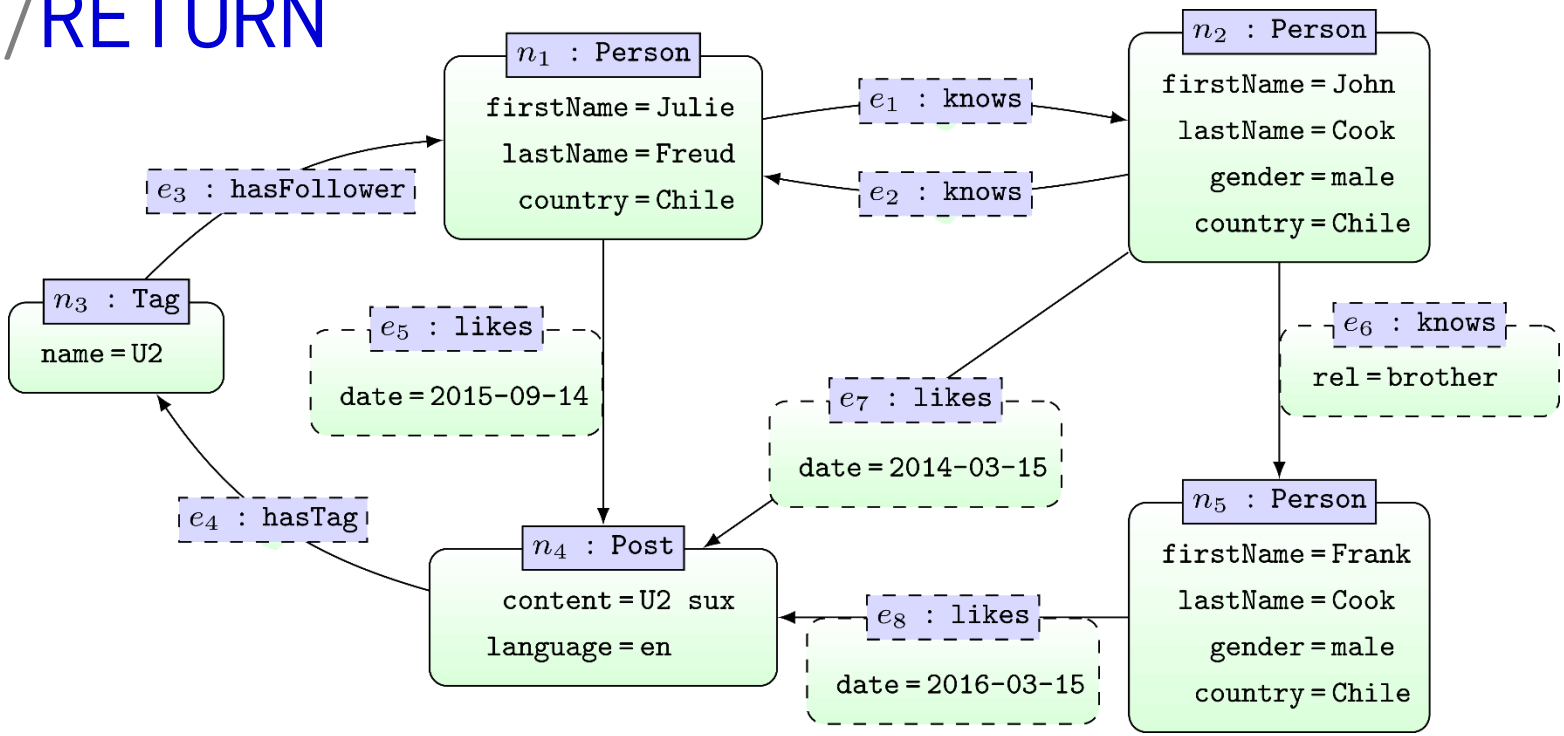
MATCH/RETURN




```
MATCH (x:Person)-->(:Person)
RETURN x.firstName
```

x.firstName
Julie
John

MATCH/RETURN

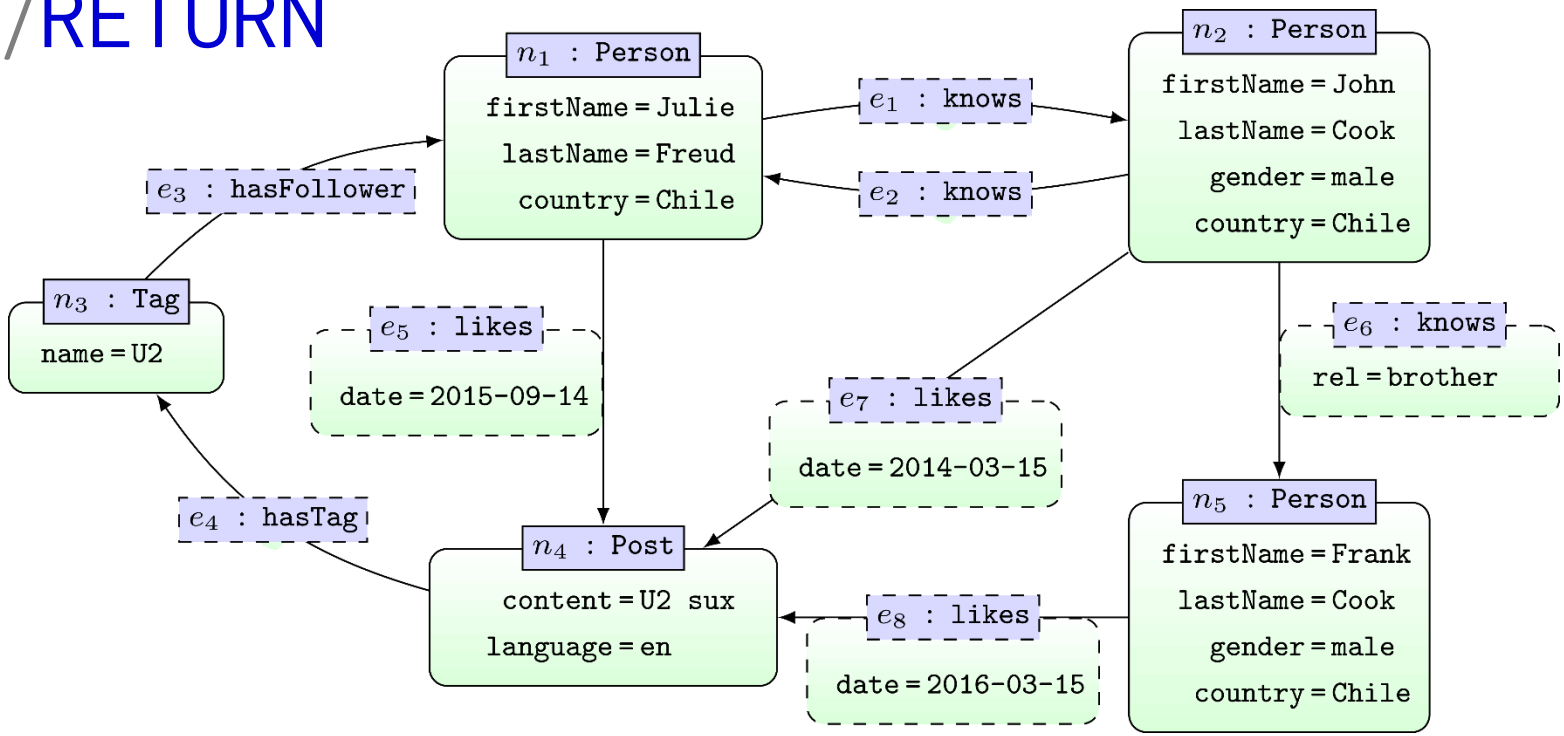



```
MATCH (x:Person)-->()
RETURN x.firstName
```

x.firstName
Julie
Julie
John
John
John
Frank

... multiplicity of results corresponds to number of matches

MATCH/RETURN

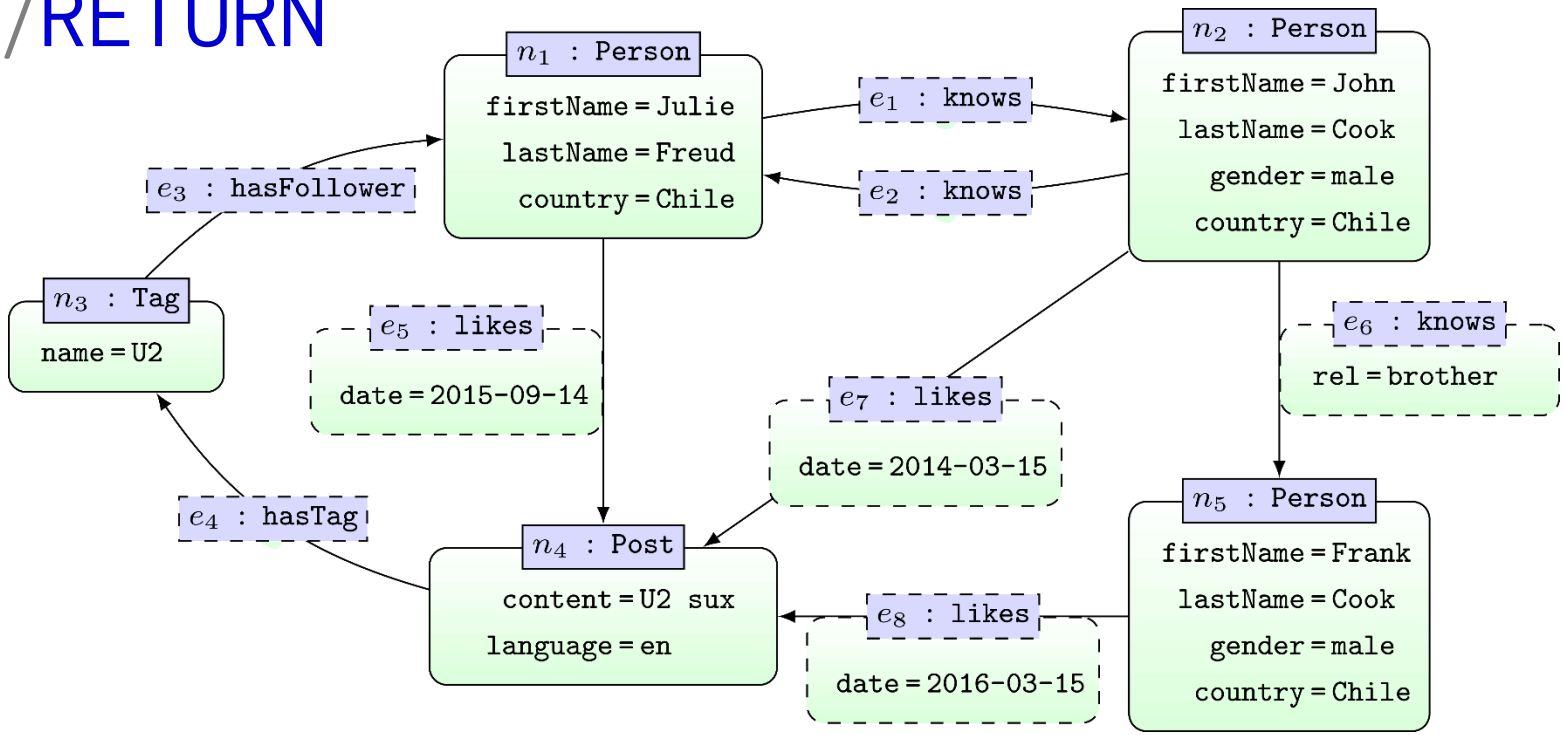



```
MATCH (x:Person)-->()
RETURN DISTINCT x.firstName
```

x.firstName
Julie
John
Frank

... RETURN DISTINCT removes duplicates

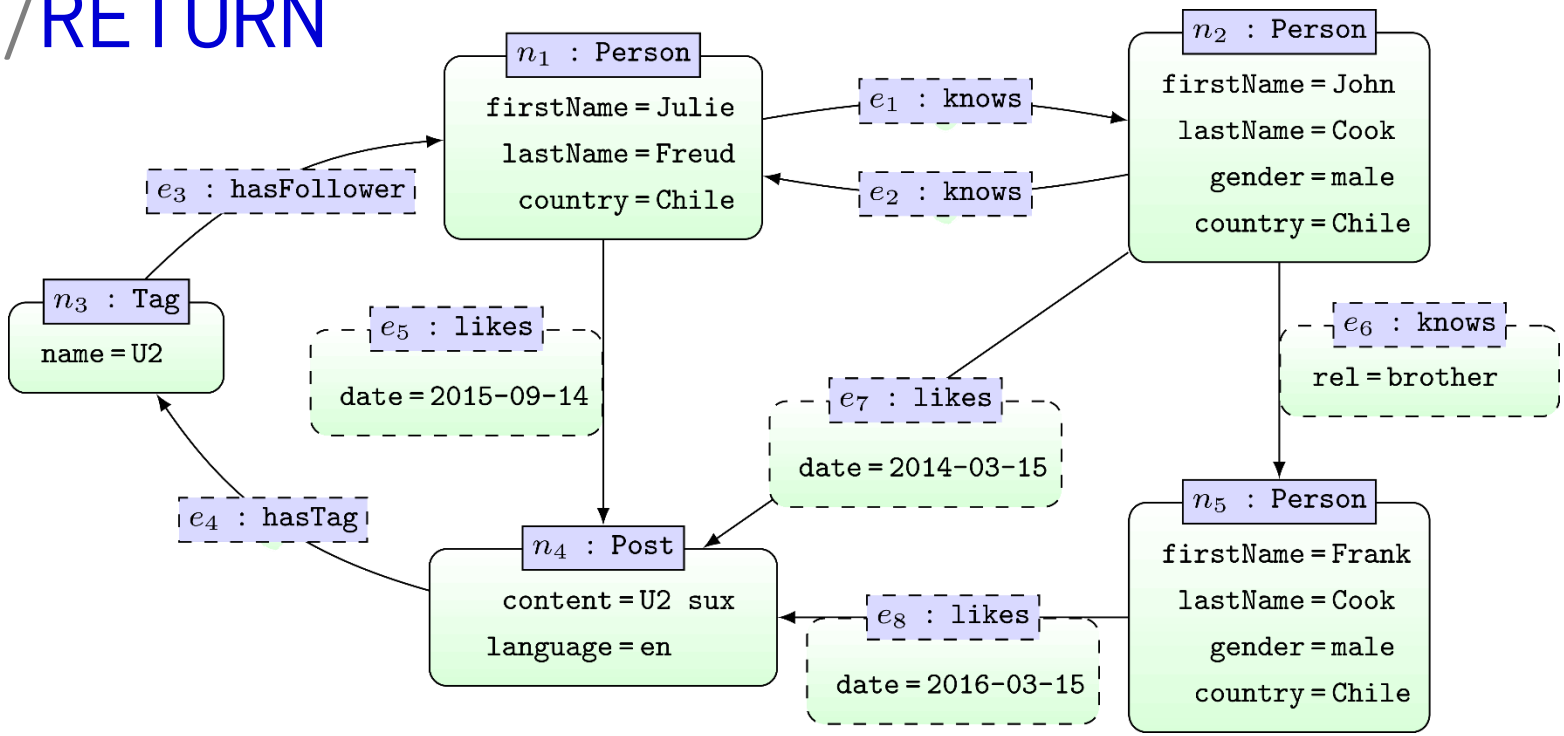
MATCH/RETURN




```
MATCH (x1:Person)-->(x2:Person)
RETURN x1.firstName,x2.firstname
```

x1.firstName	x2.firstName
Julie	John
John	Julie
John	Frank

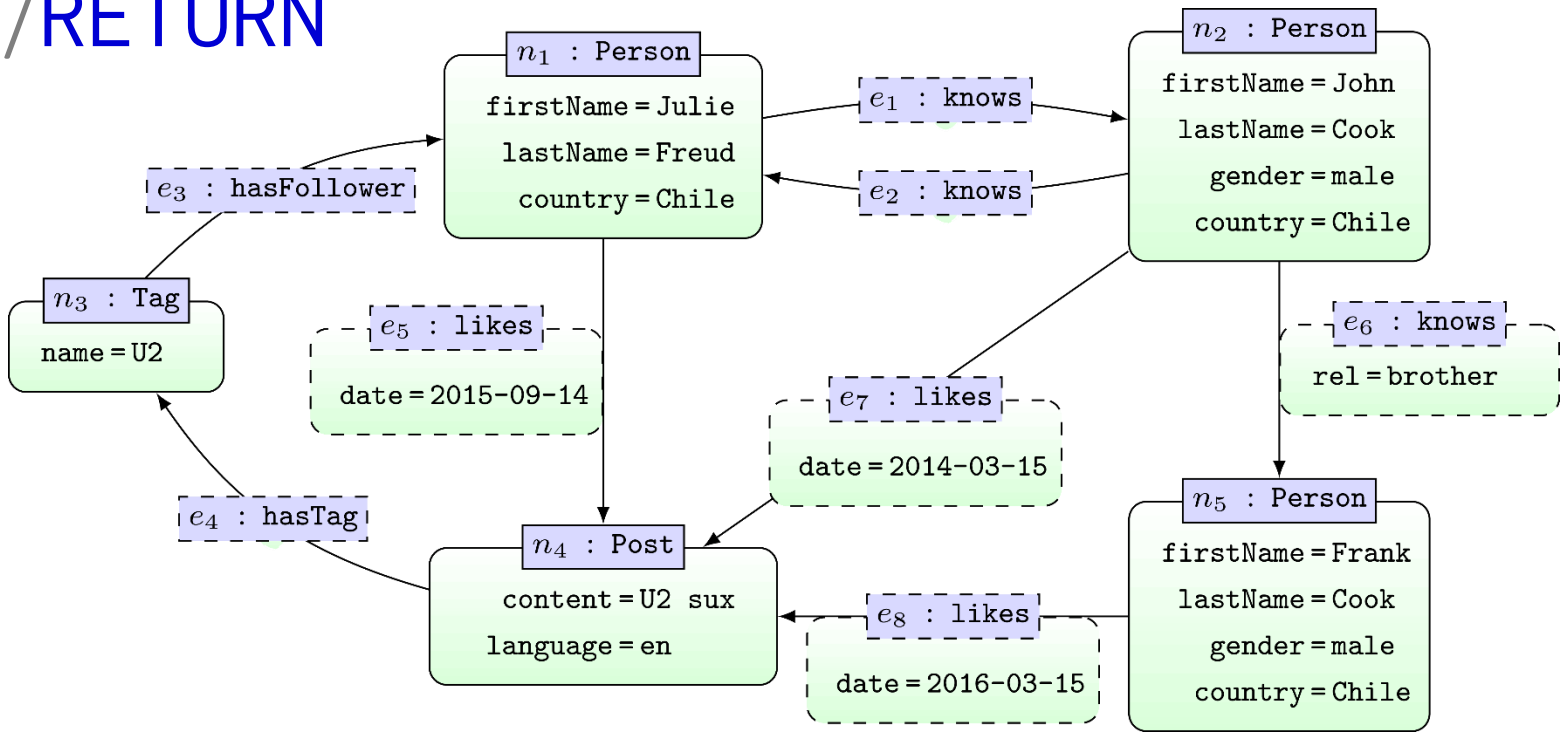
MATCH/RETURN




```
MATCH (x1:Person)-[r]->(x2:Person)
RETURN x1.firstName,x2.firstName,r.rel
```

x1.firstName	x2.firstName	r.rel
Julie	John	
John	Julie	
John	Frank	brother

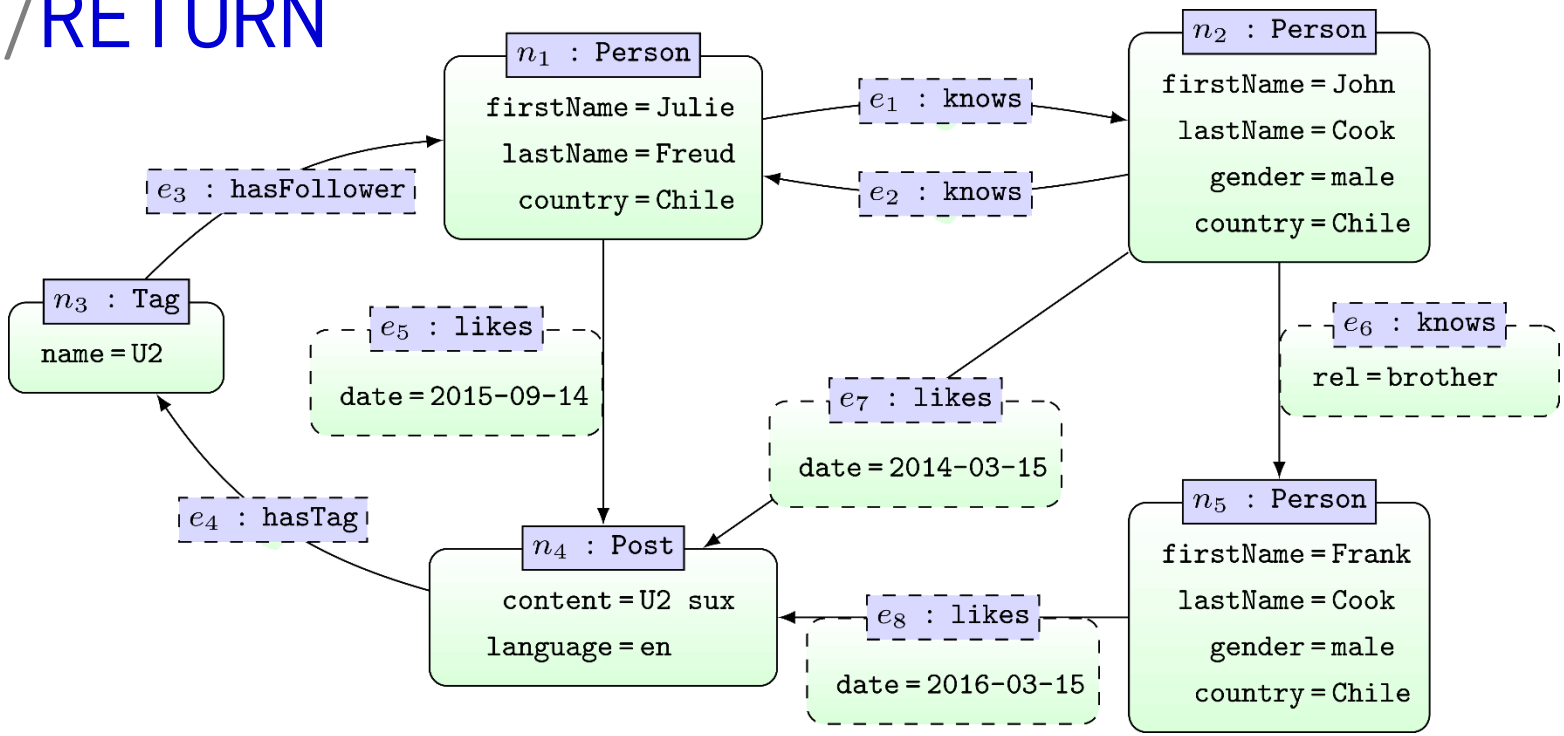
MATCH/RETURN




```
MATCH (x1:Person)-[r]->(x2:Person)
RETURN r
```

r
[:knows]
[:knows]
[:knows {rel: "brother"}]

MATCH/RETURN

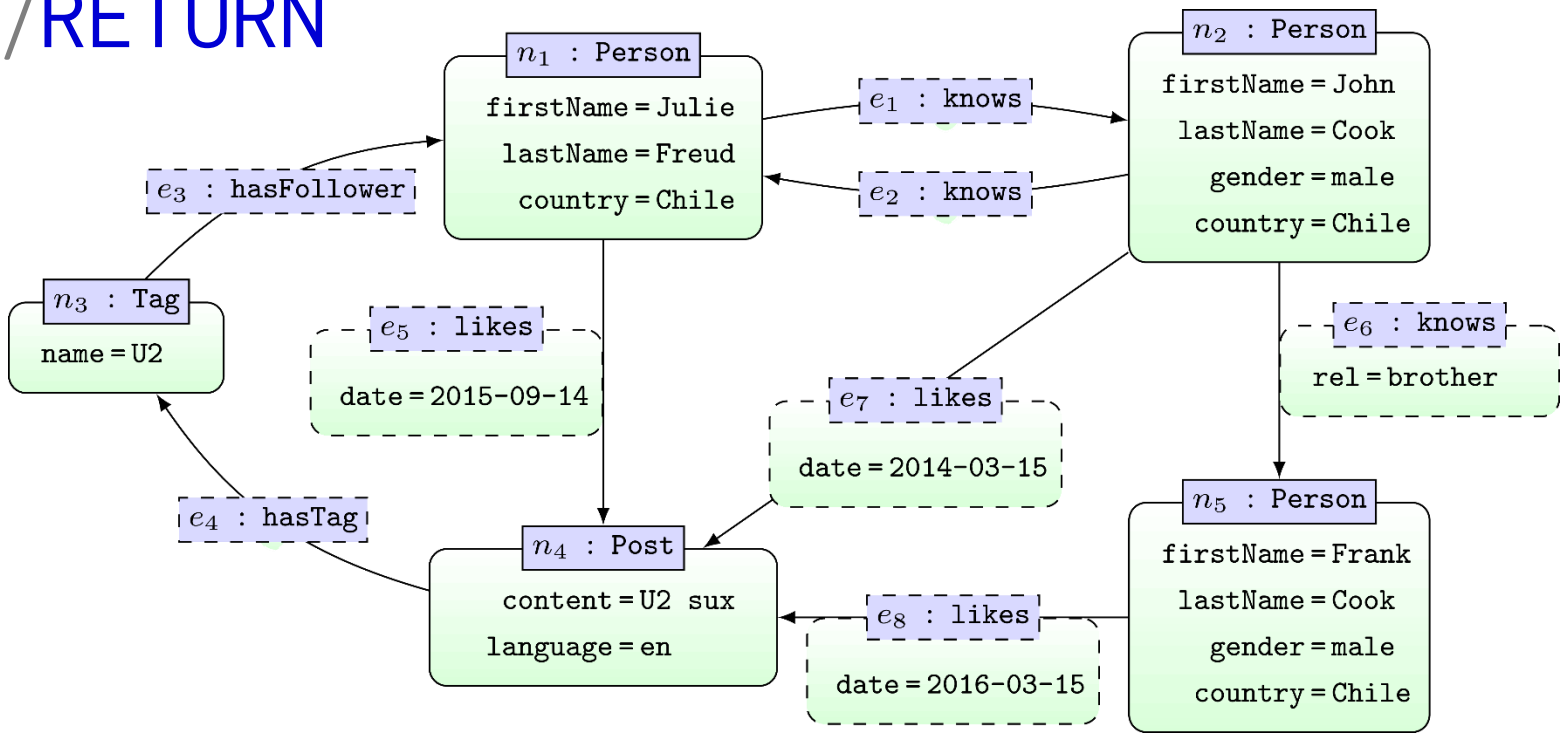



```
MATCH ()<-[:knows]-(y)-[:knows]->()
RETURN y.firstName
```

y.firstName
John
John

... MATCH will not match the same edge twice

MATCH/RETURN

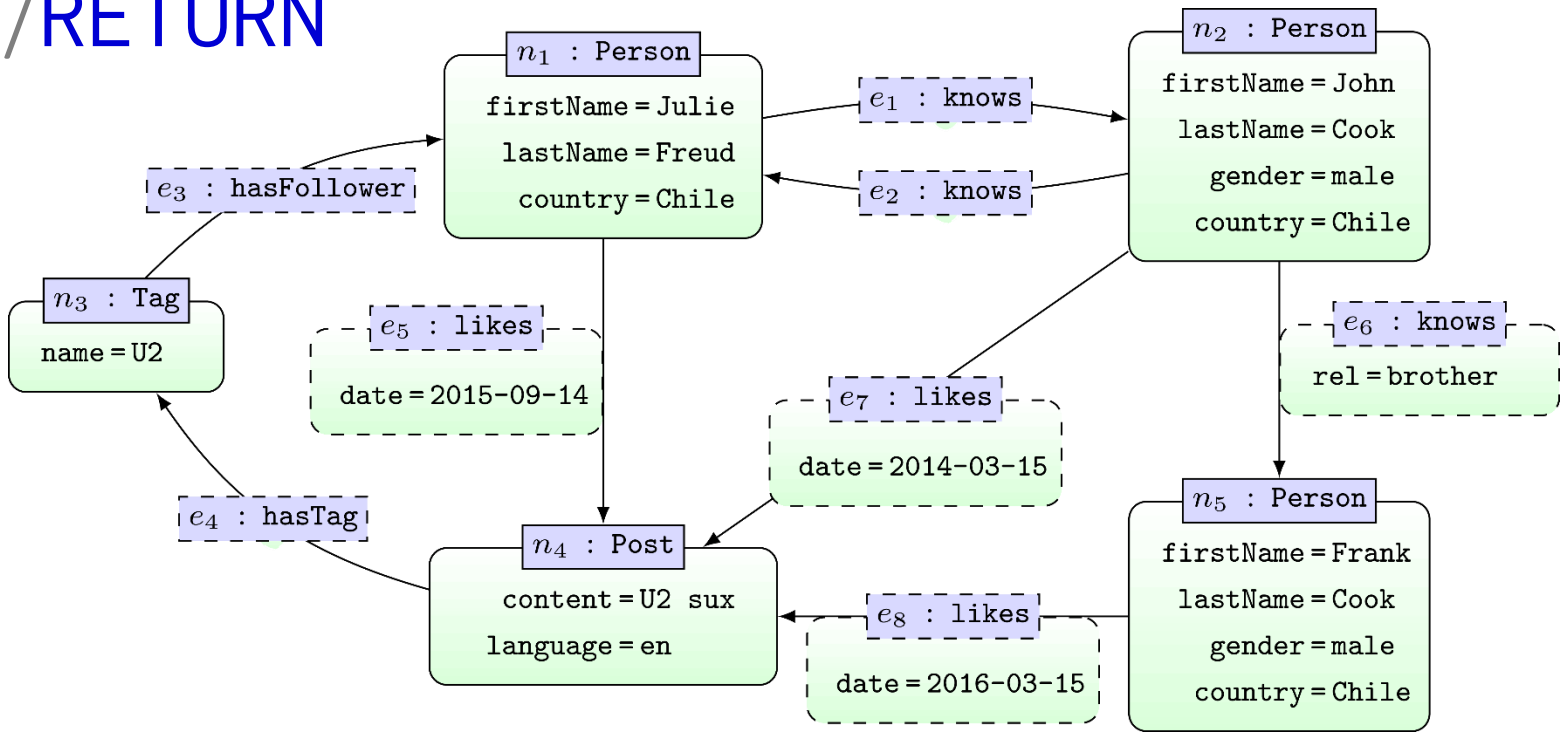



```
MATCH ()-[:knows]->(y)-[:knows]->()
RETURN y.firstName
```

y.firstName
Julie
John
John

... MATCH will match same node twice

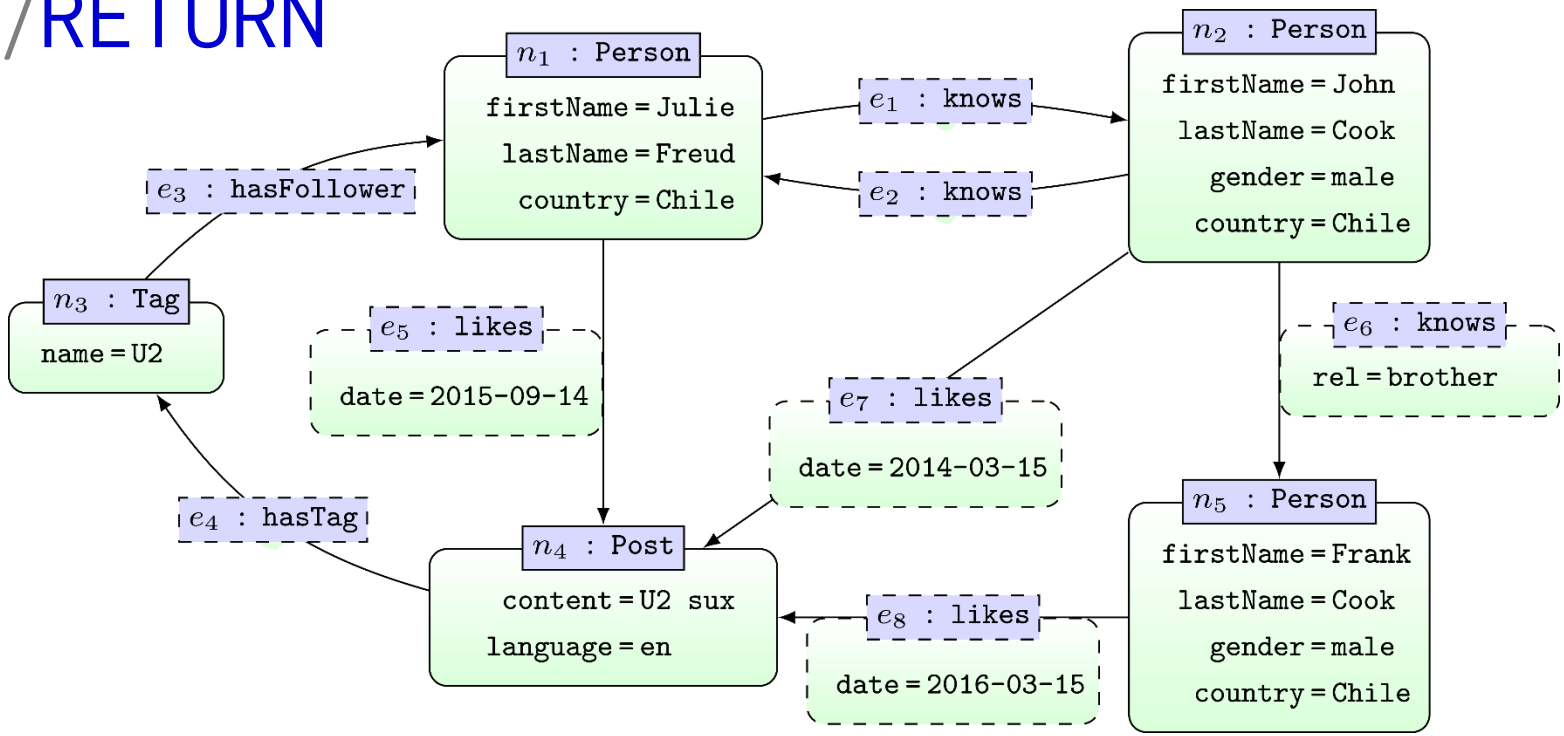
MATCH/RETURN




```
MATCH (x:Person)-->()-->()-->(x)
RETURN x.firstName
```

x.firstName
Julie

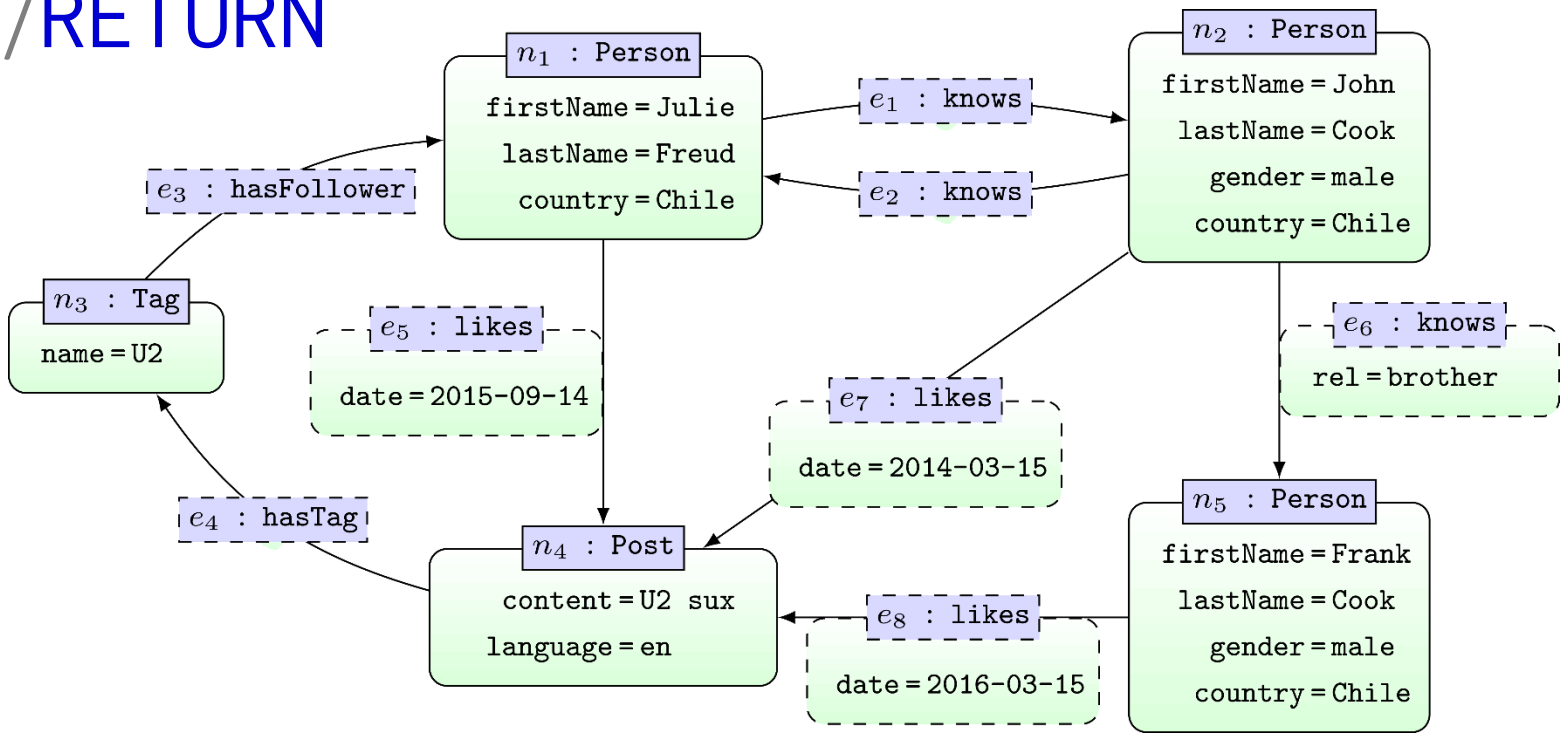
MATCH/RETURN




```
MATCH (x)-->(y)-->(x)
RETURN x.firstName
```

x.firstName
Julie
John

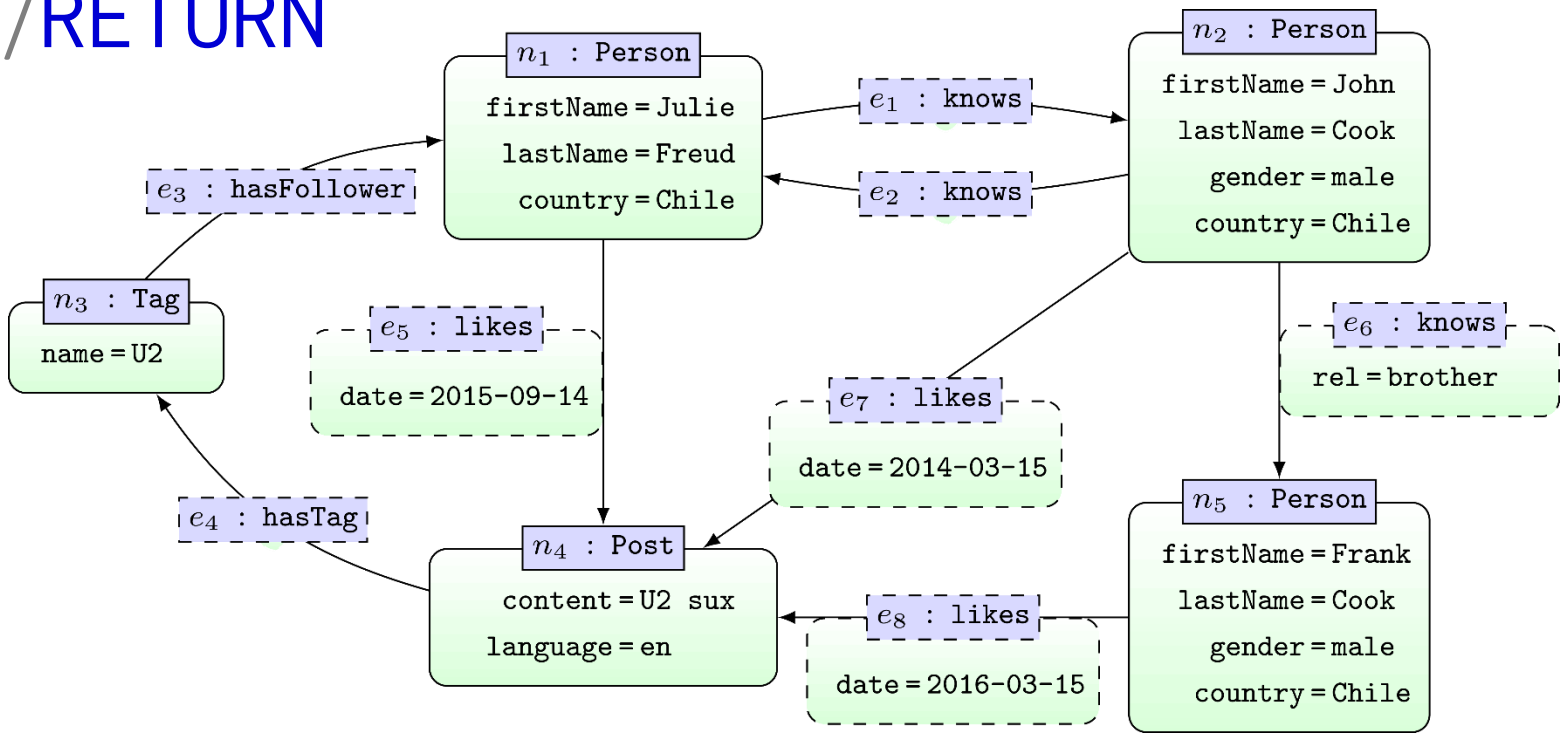
MATCH/RETURN




```
MATCH (x)-->(y)-->(x)-->(y)
RETURN x.firstName
```

x.firstName

MATCH/RETURN

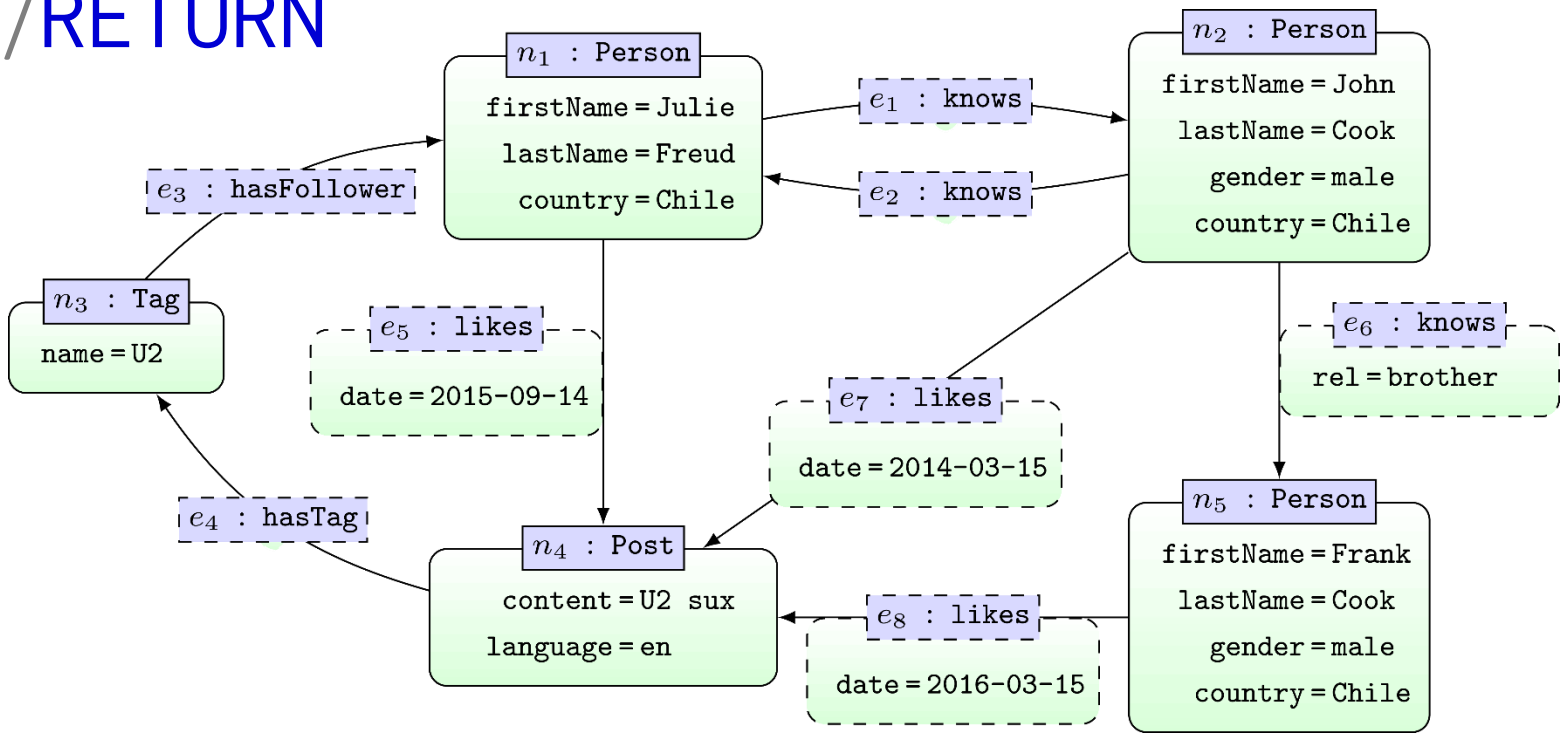



```
MATCH (x1)-[:likes]->(y)<-[:likes]-(x2)
RETURN x1.firstName AS n1, x2.firstName AS n2
```

n1	n2
Julie	John
John	Julie
John	Frank
Frank	John
Frank	Julie
Julie	Frank

... AS renames columns in results

MATCH/RETURN

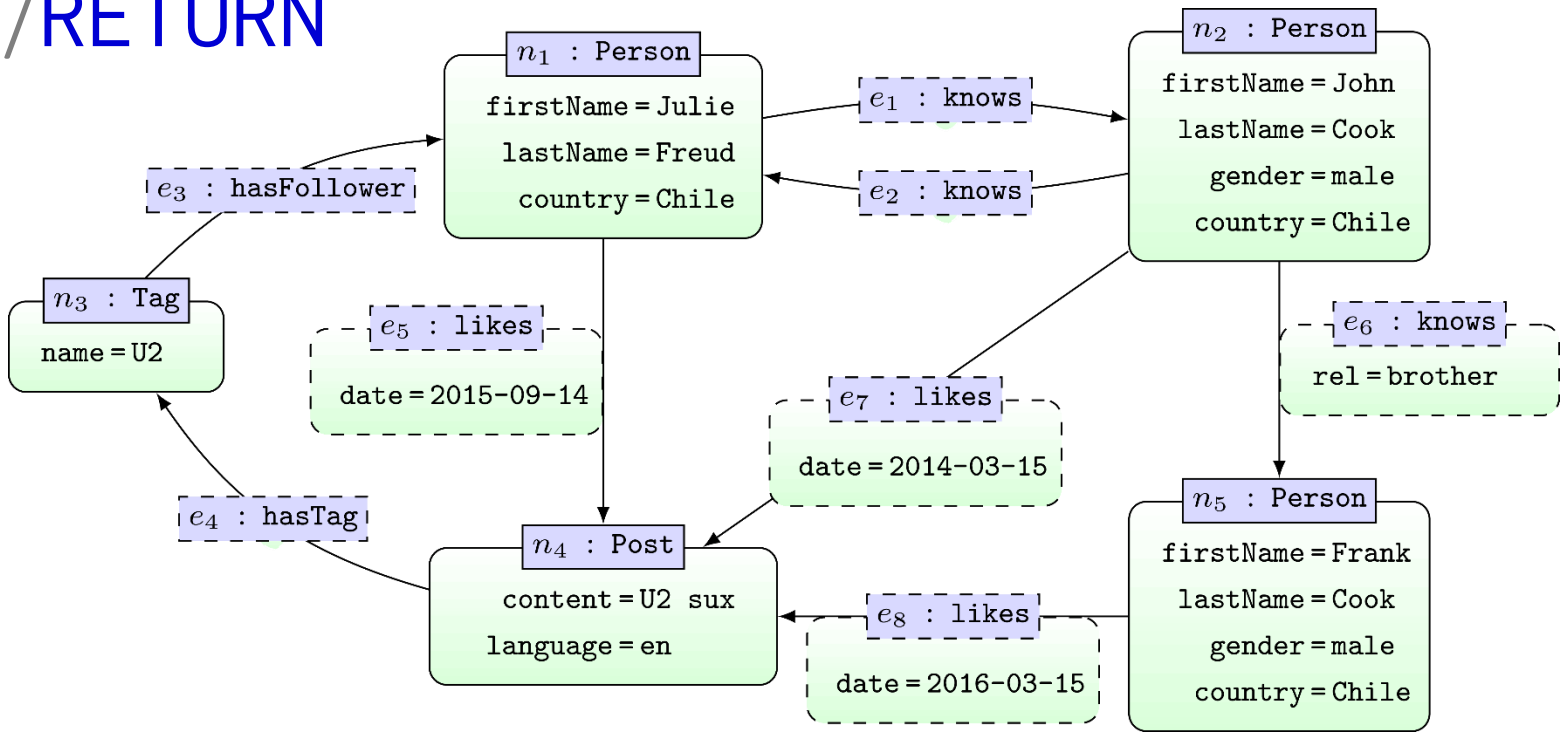



```
MATCH (x1)-[:likes]->(y)
MATCH (y)<-[:likes]-(x2)
RETURN x1.firstName AS n1, x2.firstName AS n2
```

n1	n2
Julie	John
John	Julie
Julie	Julie
John	Frank
...	...

... use multiple MATCH to match same edge multiple times

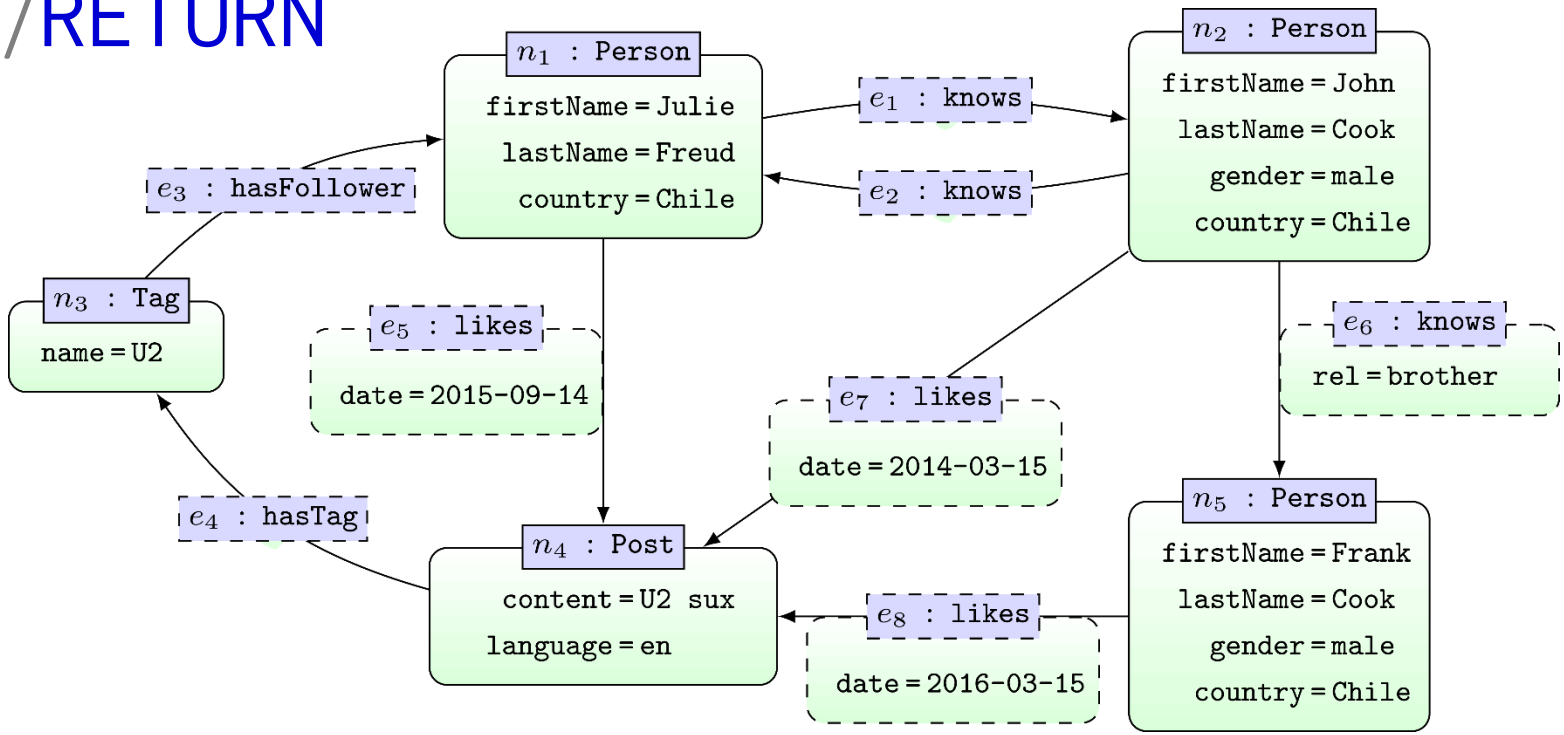
MATCH/RETURN




```
MATCH (x1)-[:likes]->(y)-[:hasTag]->(z),
   (x2)-[:likes]->(y)
RETURN z.name
```

z.name
U2
U2
U2
U2
U2
U2

MATCH/RETURN



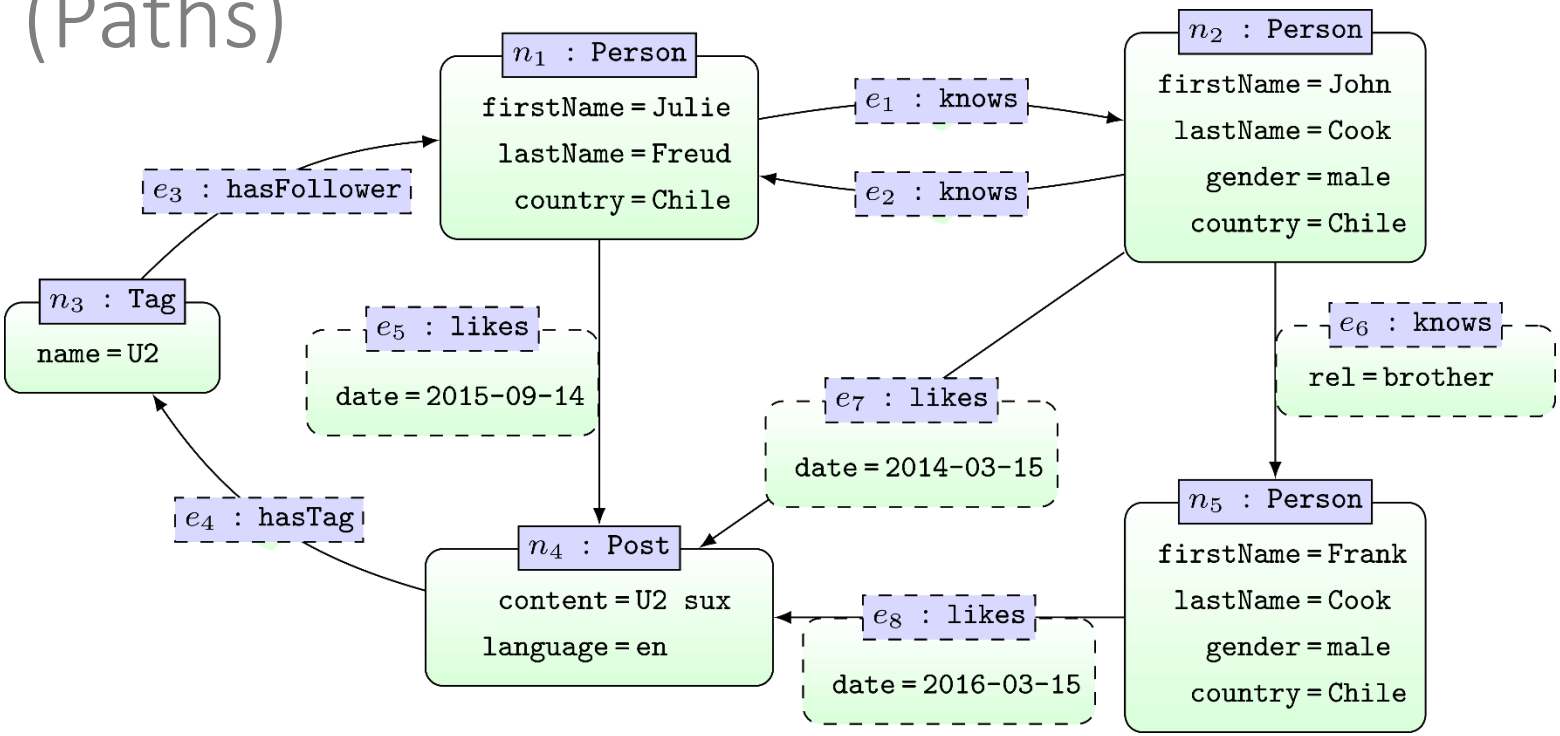

```
MATCH (x1)-[:likes]->(y)-[:hasTag]->(z)
MATCH (x2)-[:likes]->(y)
RETURN z.name
```

z.name
U2
U2
U2
U2
U2
U2
U2
U2
U2

Cypher:
MATCH (Paths)

https://neo4j.com/docs/developer-manual/3.4/cypher/clauses/where/

MATCH (Paths)

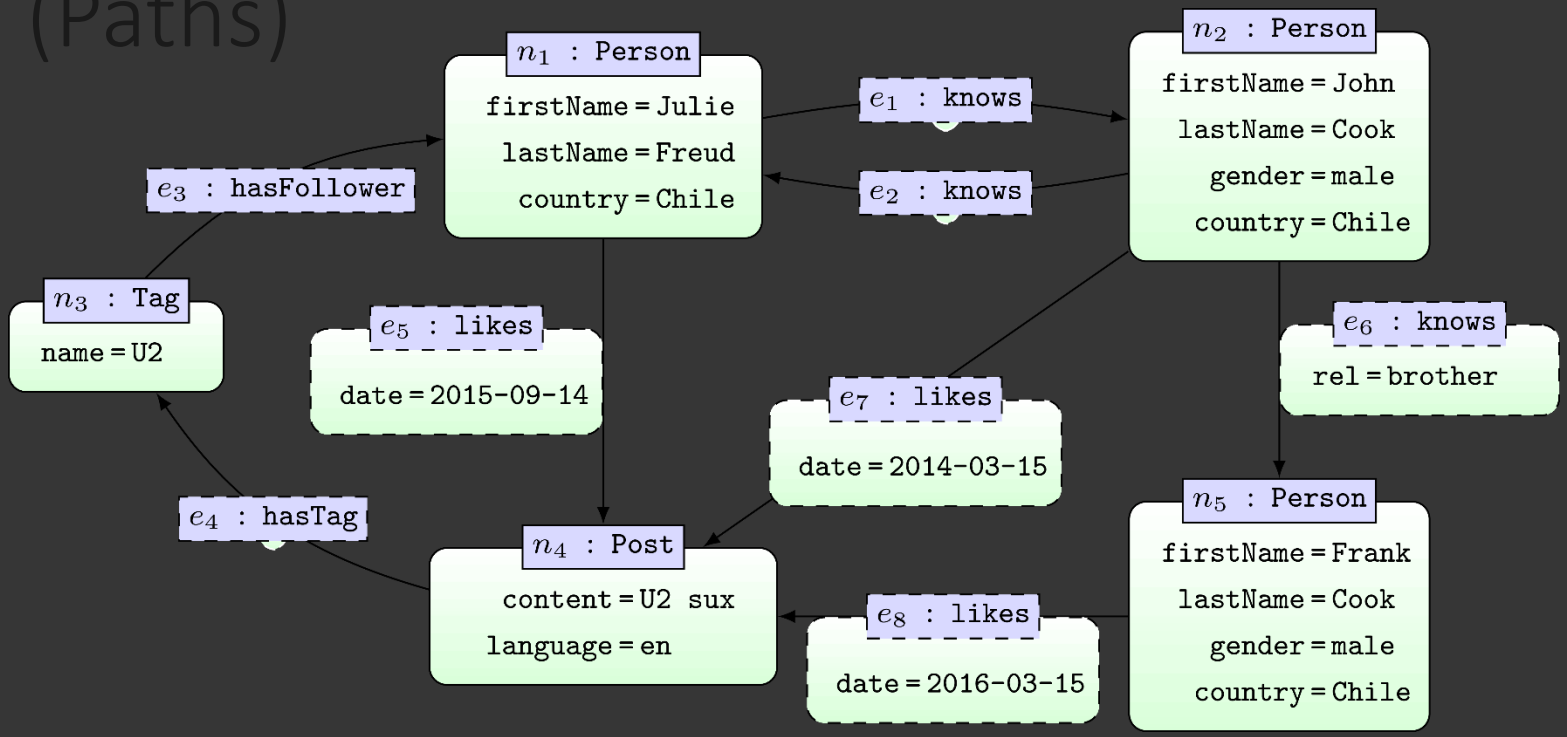



```
MATCH (x1)-[:knows*]->(x2)
RETURN x1.firstName, x2.firstName
```

x1.firstName	x2.firstName
Julie	John
Julie	Frank
Julie	Julie
John	Julie
John	John
John	Frank
John	Frank

... paths visit each edge at most once!

MATCH (Paths)



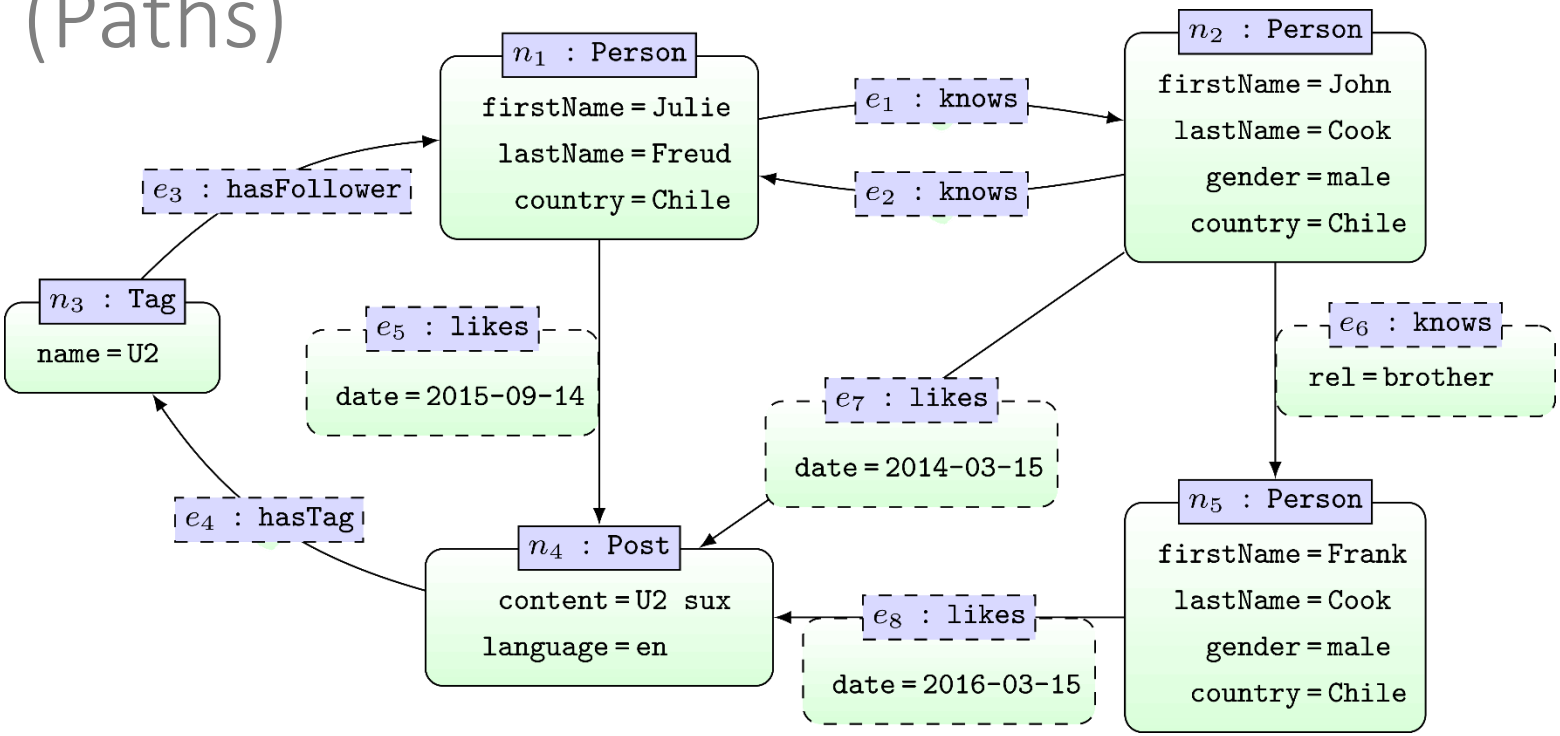

```
MATCH (x1)-[:knows*]->(x2)
RETURN x1.firstName, x2.firstName
```

x1.firstName	x2.firstName
Julie	John
Julie	Frank
Julie	Julie
John	John
John	Frank
John	Frank

Otherwise ... ... we could have infinite paths!

... paths visit each edge at most once!

MATCH (Paths)

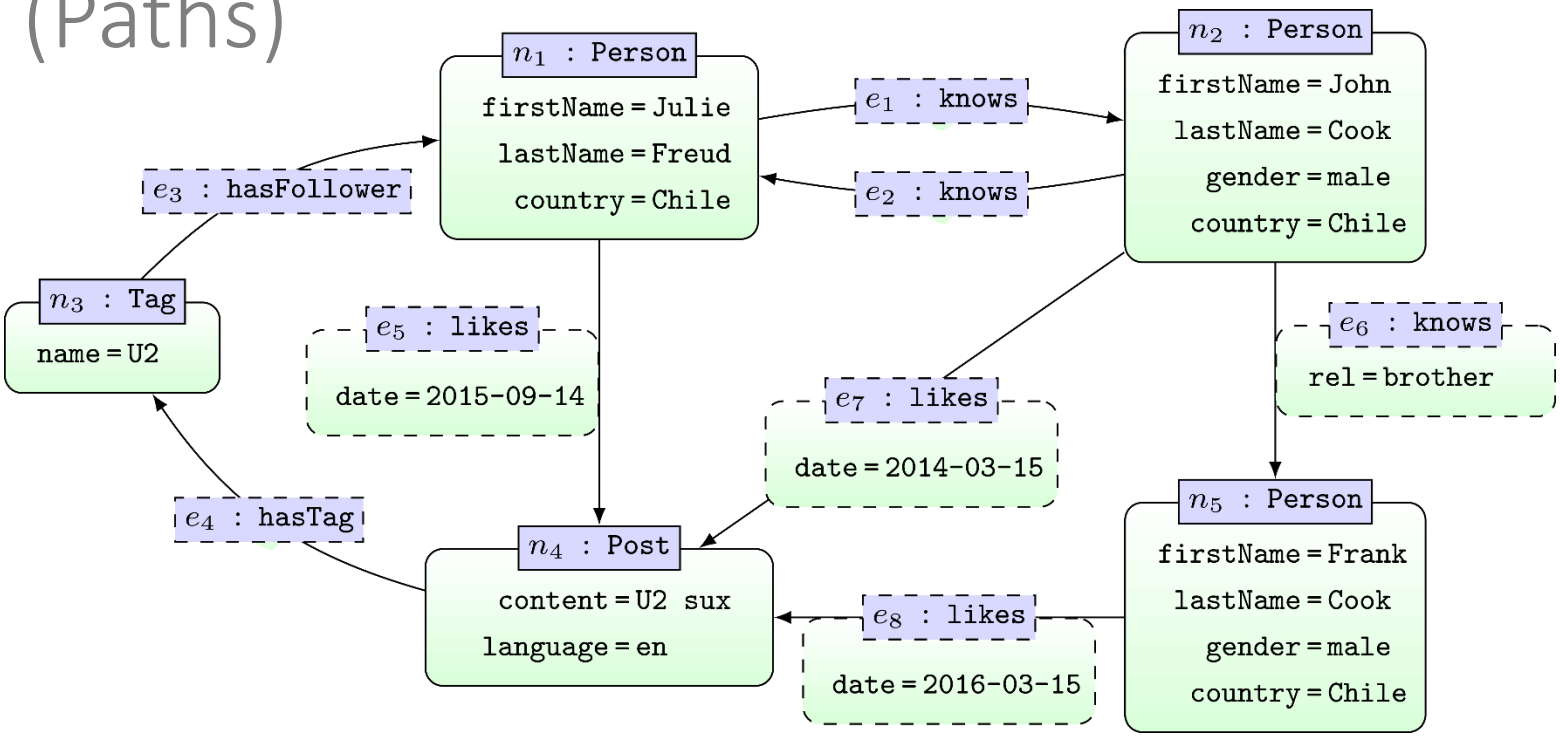



```
MATCH (x1)-[:knows*3]->(x2)
RETURN x1.firstName, x2.firstName
```

x1.firstName	x2.firstName
John	Frank

... can set minimum path length (no. of nodes visited)

MATCH (Paths)

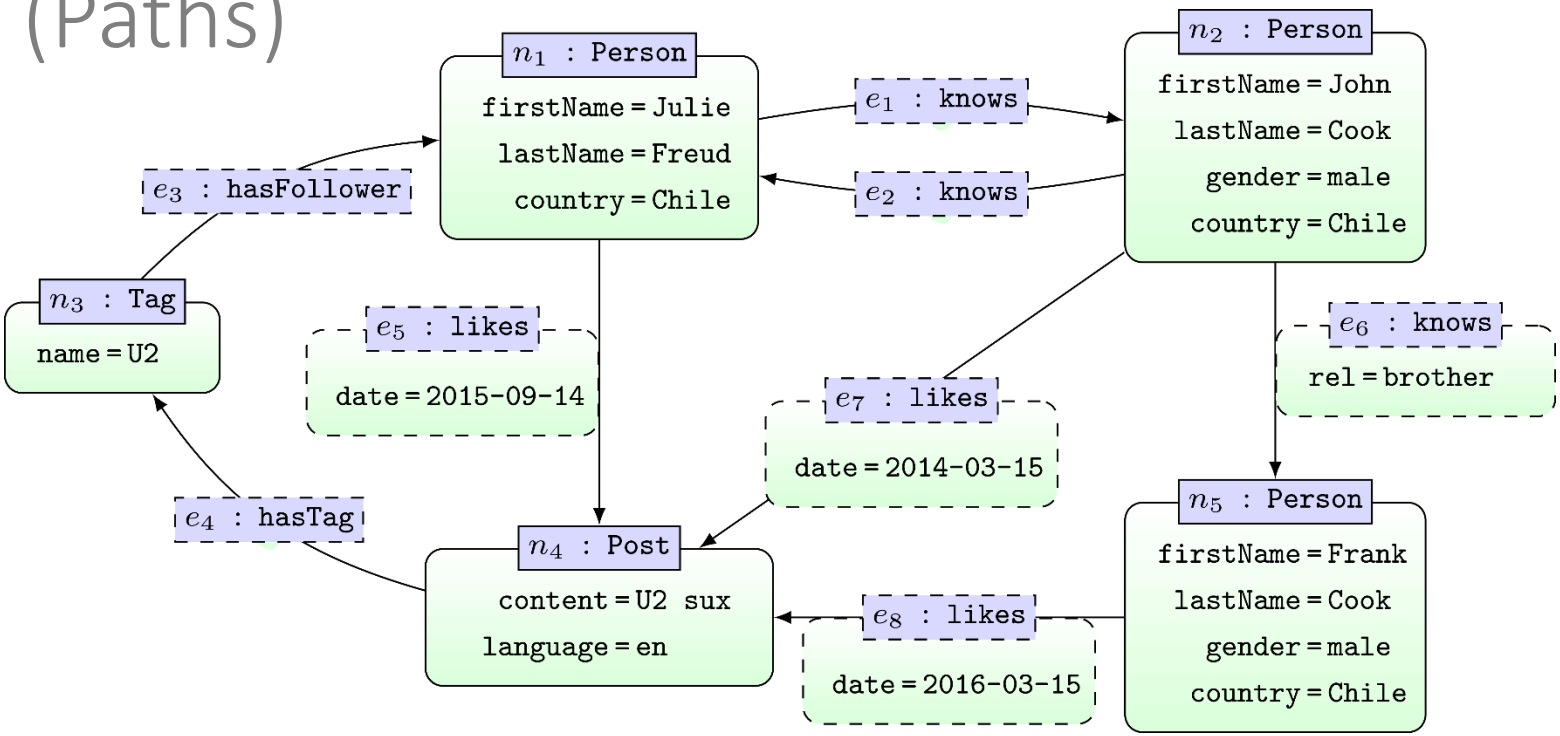



```
MATCH (x1)-[:knows*2..3]->(x2)
RETURN x1.firstName, x2.firstName
```

x1.firstName	x2.firstName
Julie	Frank
Julie	Julie
John	Frank
John	John

... or range of path length

MATCH (Paths)

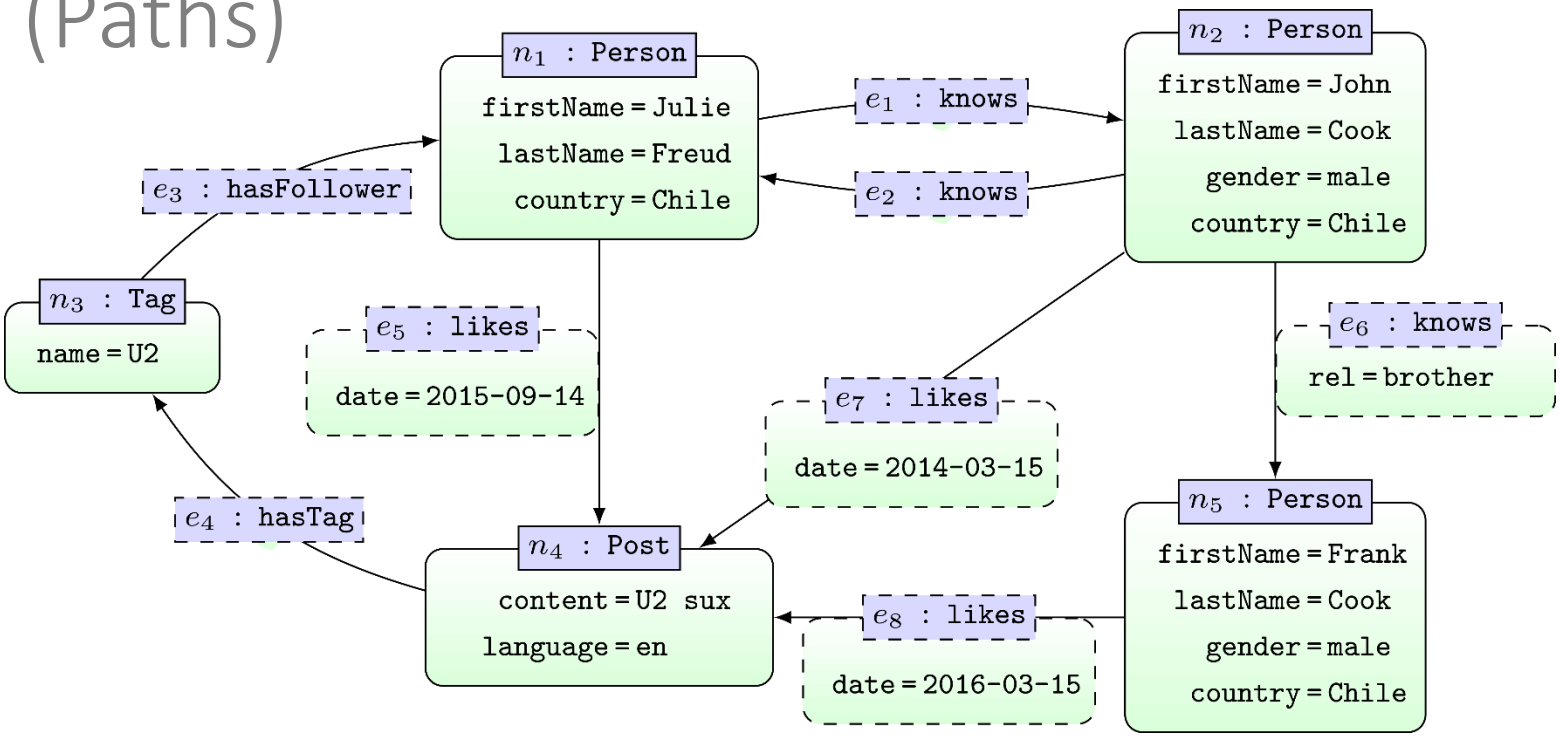



```
MATCH (x1)-[:knows*..2]->(x2)
RETURN x1.firstName, x2.firstName
```

x1.firstName	x2.firstName
Julie	John
Julie	Frank
Julie	Julie
John	Julie
John	John
John	Frank

... or maximum path length

MATCH (Paths)

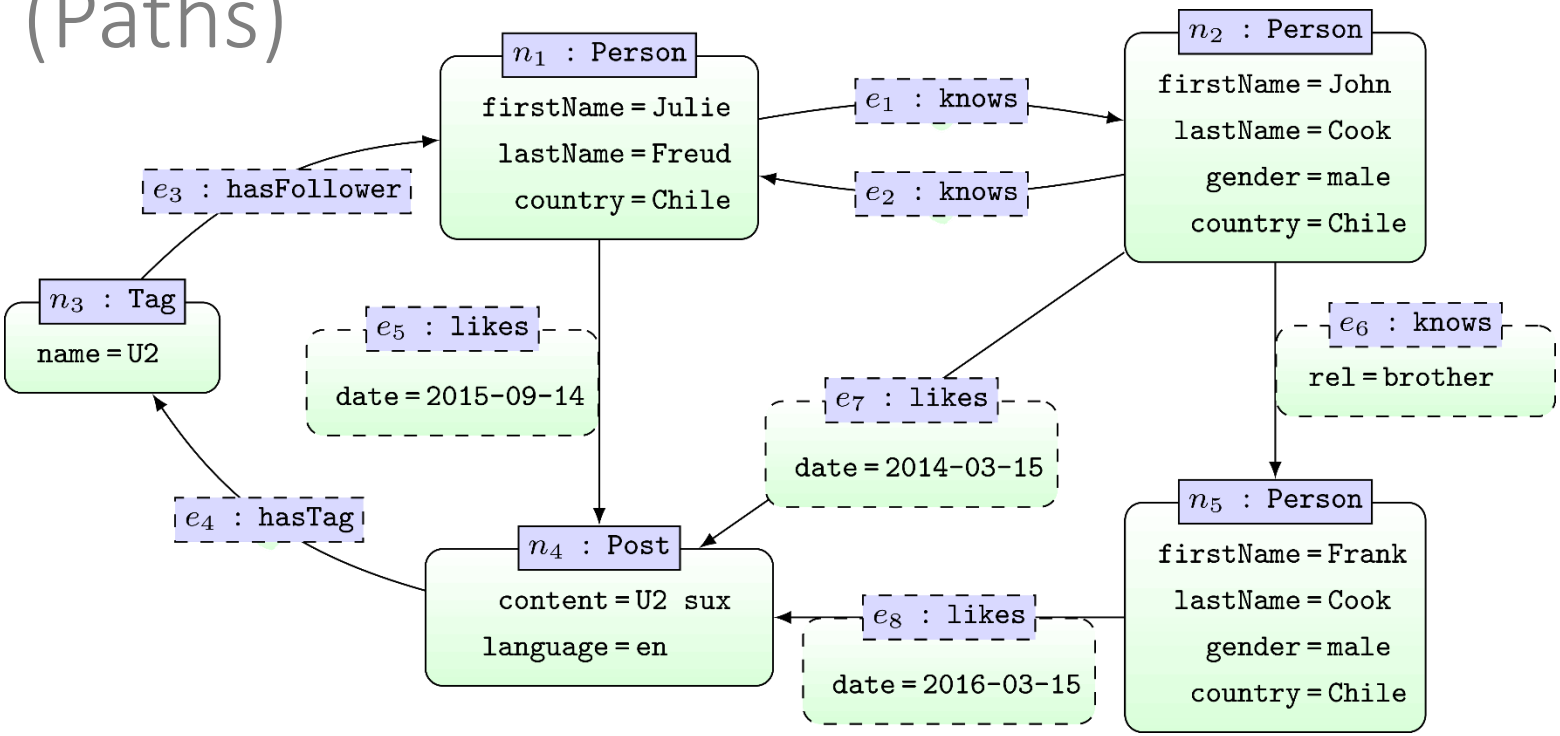



```
MATCH (x1)-[:knows*0..1]->(x2)
RETURN x1.firstName, x2.firstName
```

x1.firstName	x2.firstName
Julie	Julie
Julie	John
John	John
John	Julie
John	Frank
Frank	Frank

... 0-length path is the node itself; will match any node

MATCH (Paths)

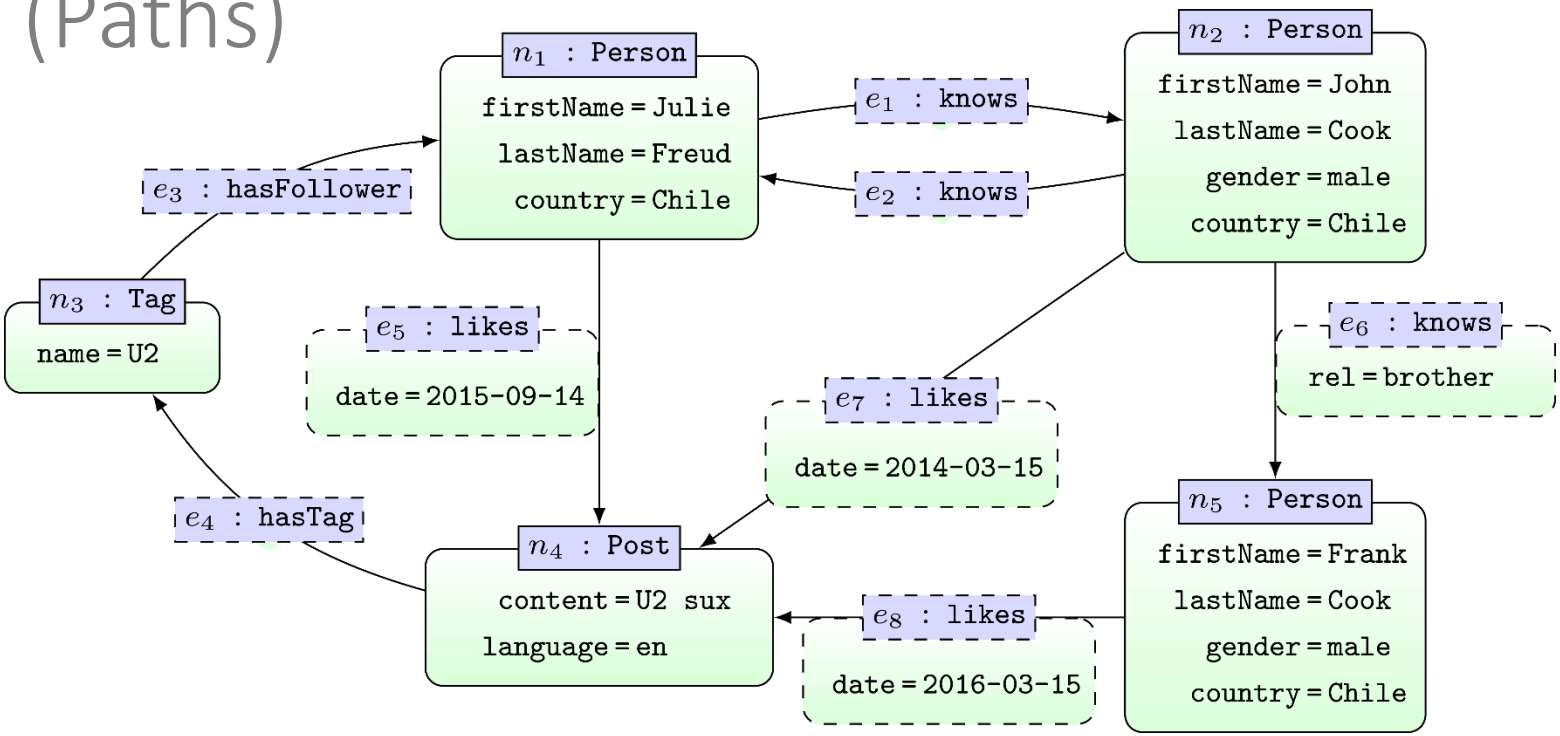



```
MATCH p = (x1)-[:knows*3]->(x2)
RETURN p
```

p
(:Person {firstName:"John",...})-[:knows]->(:Person {firstName:"Julie",...})-[:knows]->(:Person {firstName:"John",...})-[:knows rel:"brother"]->(:Person {firstName:"Frank",...})

... can return a full path

MATCH (Paths)

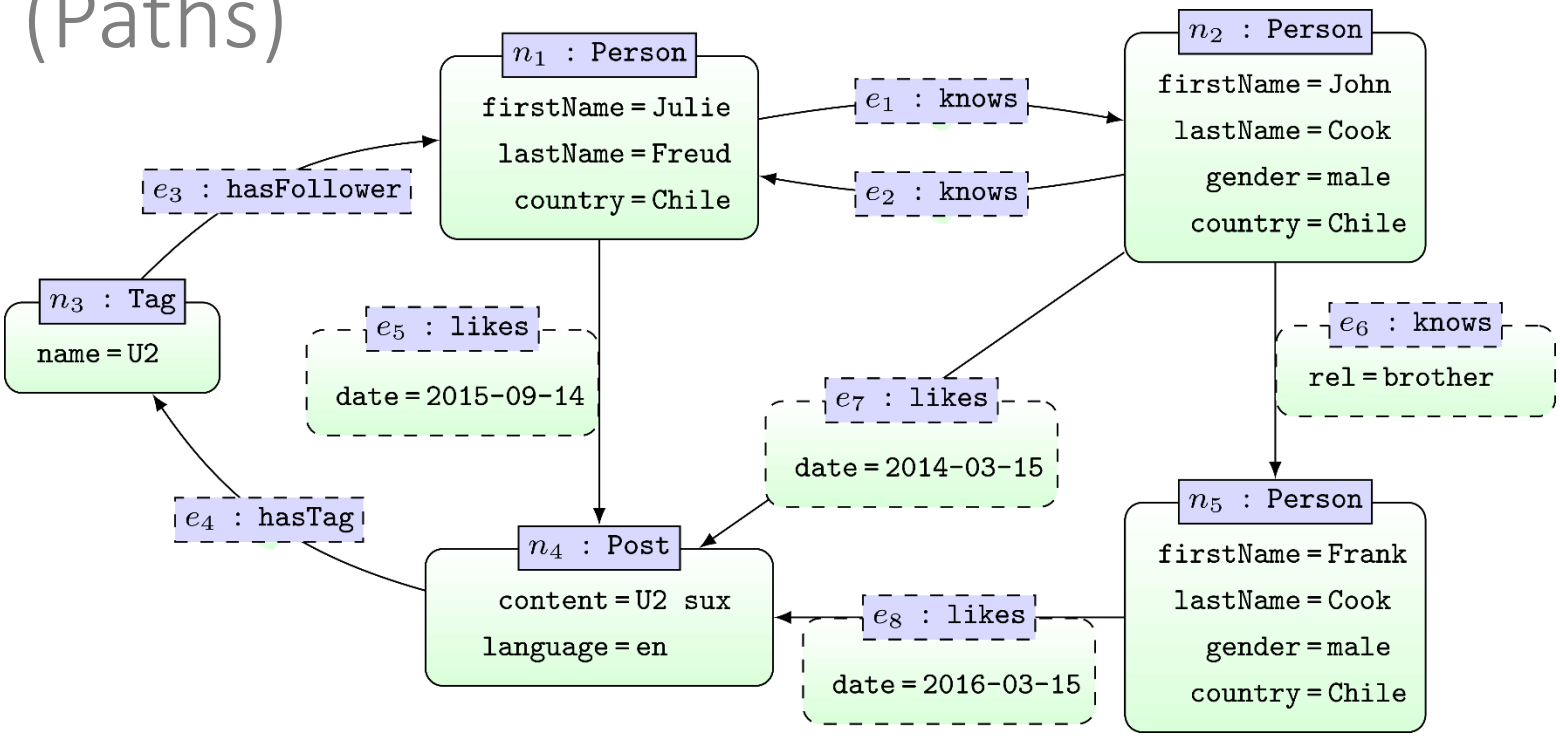



```
MATCH p = (x1)-[:knows*3]->(x2)
RETURN p
```

p
(:Person {firstName:"John",...})-[:knows]->(:Person {firstName:"Julie",...})-[:knows]->(:Person {firstName:"John",...})-[:knows rel:"brother"]->(:Person {firstName:"Frank",...})

... can return a full path

MATCH (Paths)

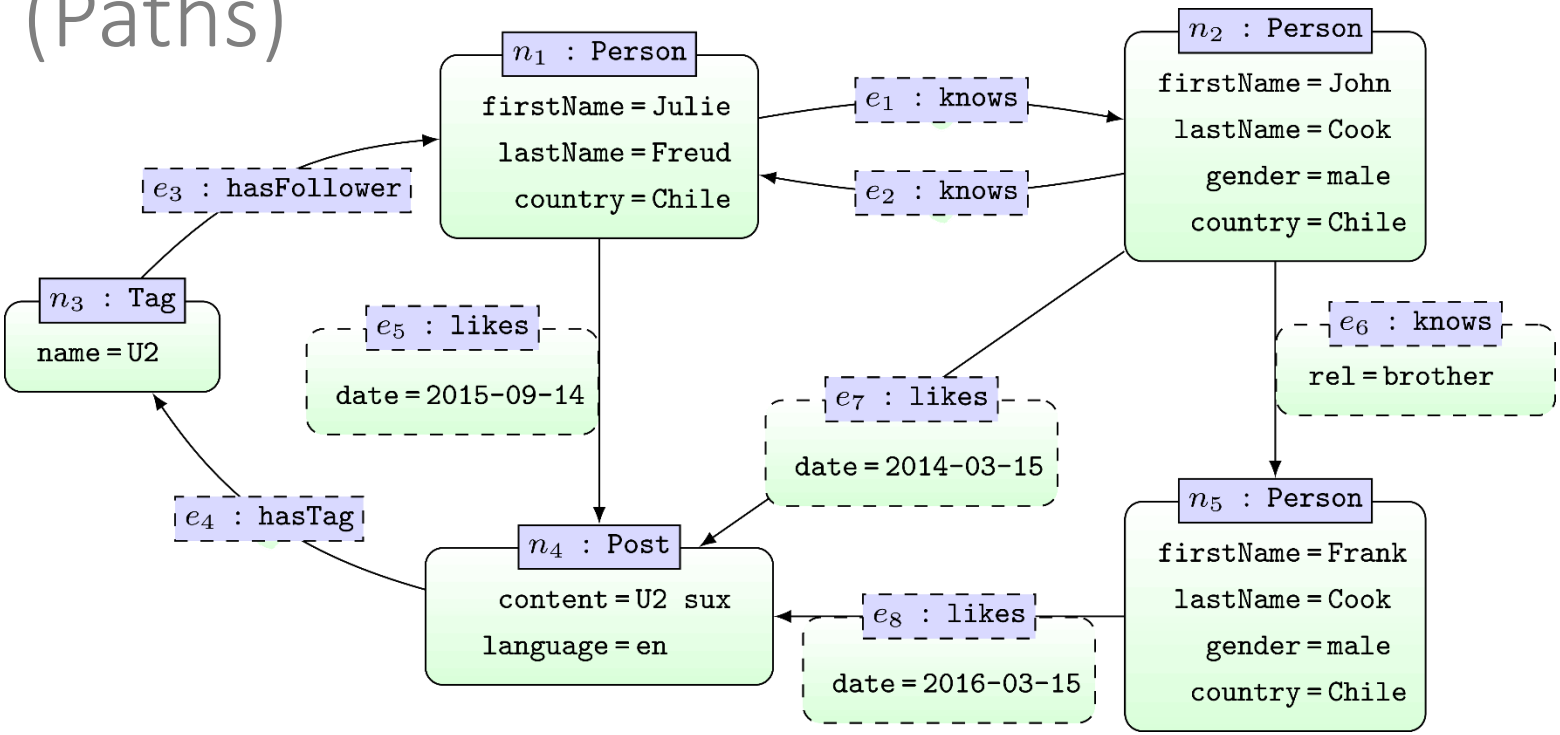



```
MATCH p = (x1)-[:knows*3]->(x2)
RETURN p
```

p
$n_2 \cdot e_2 \cdot n_1 \cdot e_1 \cdot n_2 \cdot e_6 \cdot n_5$

... can return a full path

MATCH (Paths)

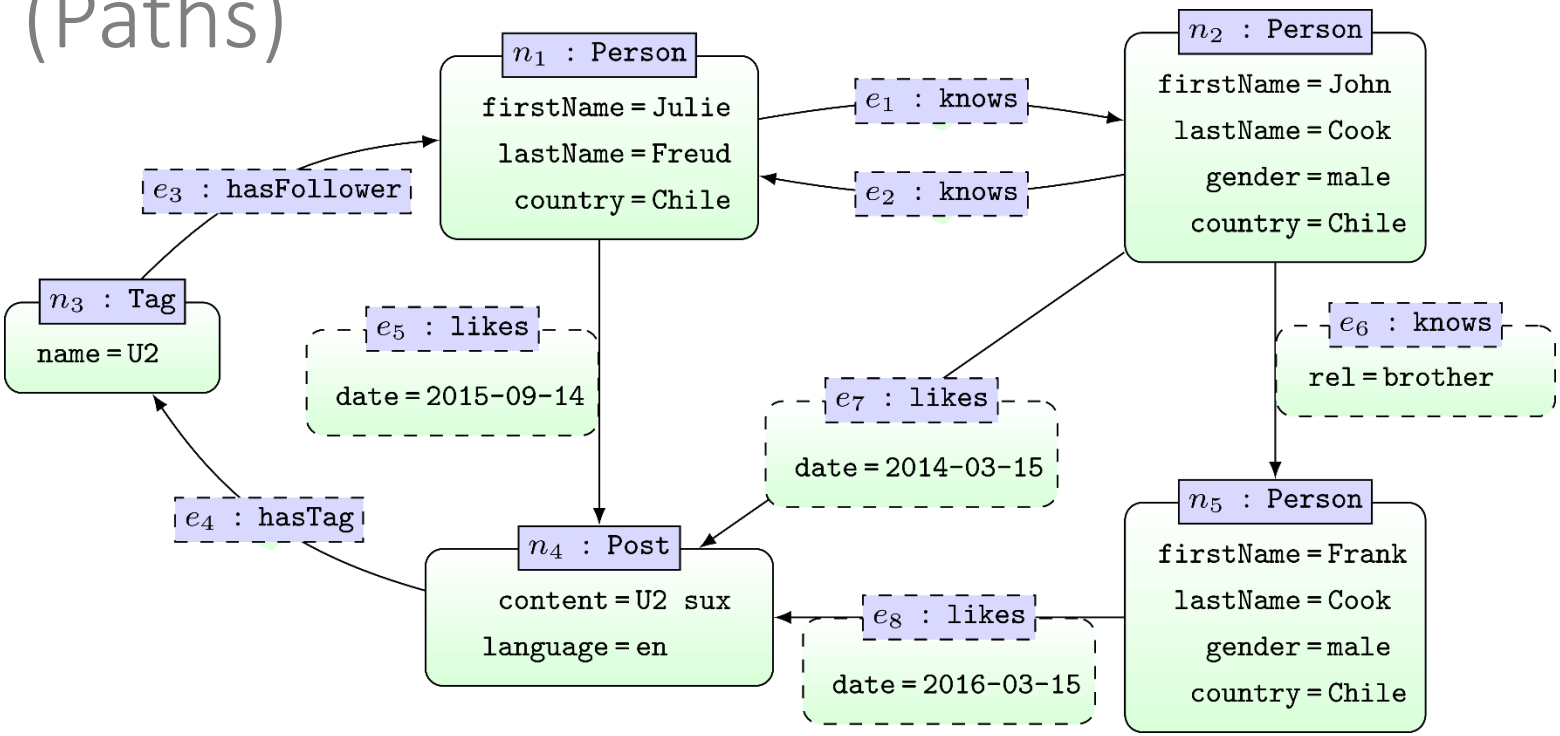



```
MATCH (f {firstName: 'Frank'}),
  (j {firstName: 'Julie'}),
  p = shortestPath((f)-[*]-(j))
RETURN p
```

p
$n_5 \cdot e_6 \cdot n_2 \cdot e_2 \cdot n_1$

... returns any shortest path (matching criteria)

MATCH (Paths)

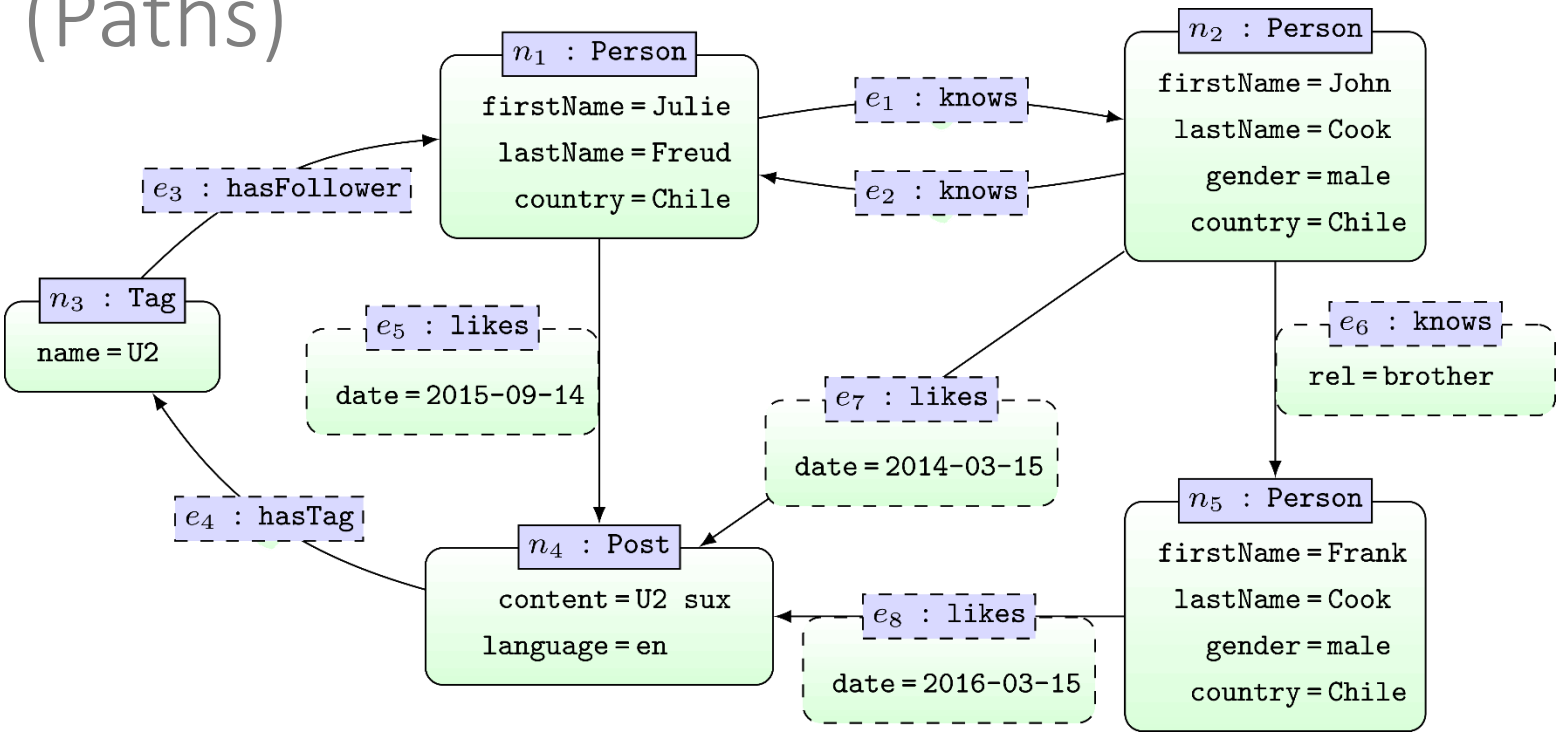



```
MATCH (f {firstName: 'Frank'}),
  (j {firstName: 'Julie'}),
  p = allShortestPaths((f)-[*]-(j))
RETURN p
```

p
$n_5 \cdot e_6 \cdot n_2 \cdot e_2 \cdot n_1$
$n_5 \cdot e_6 \cdot n_2 \cdot e_1 \cdot n_1$
$n_5 \cdot e_8 \cdot n_4 \cdot e_5 \cdot n_1$

... returns all shortest paths (matching criteria)

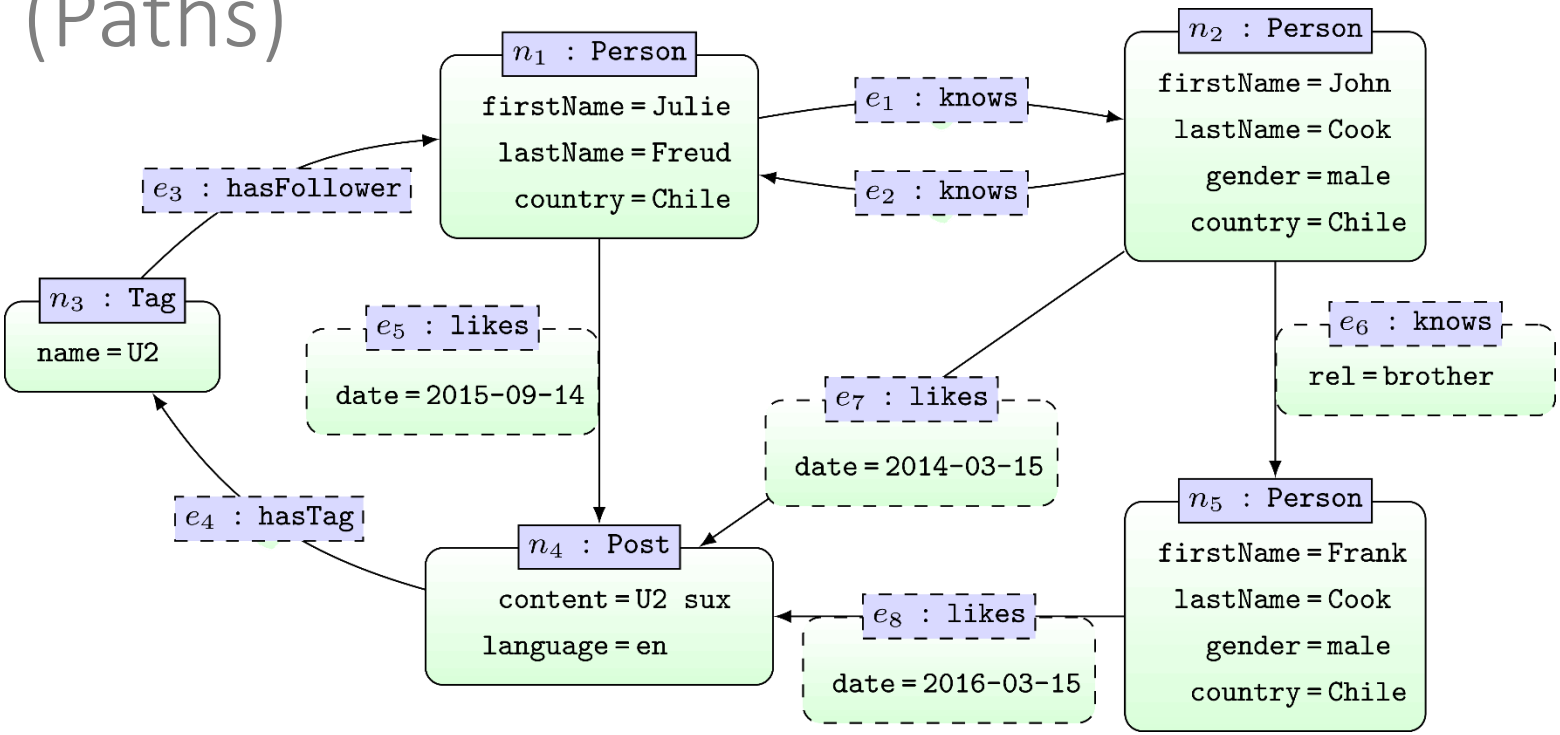
MATCH (Paths)




```
MATCH (f {firstName: 'Frank'}),
  (j {firstName: 'Julie'}),
  p = shortestPath((f)-[*]->(j))
RETURN p
```

p
$n_5 \cdot e_8 \cdot n_4 \cdot e_4 \cdot n_3 \cdot e_3 \cdot n_1$

MATCH (Paths)

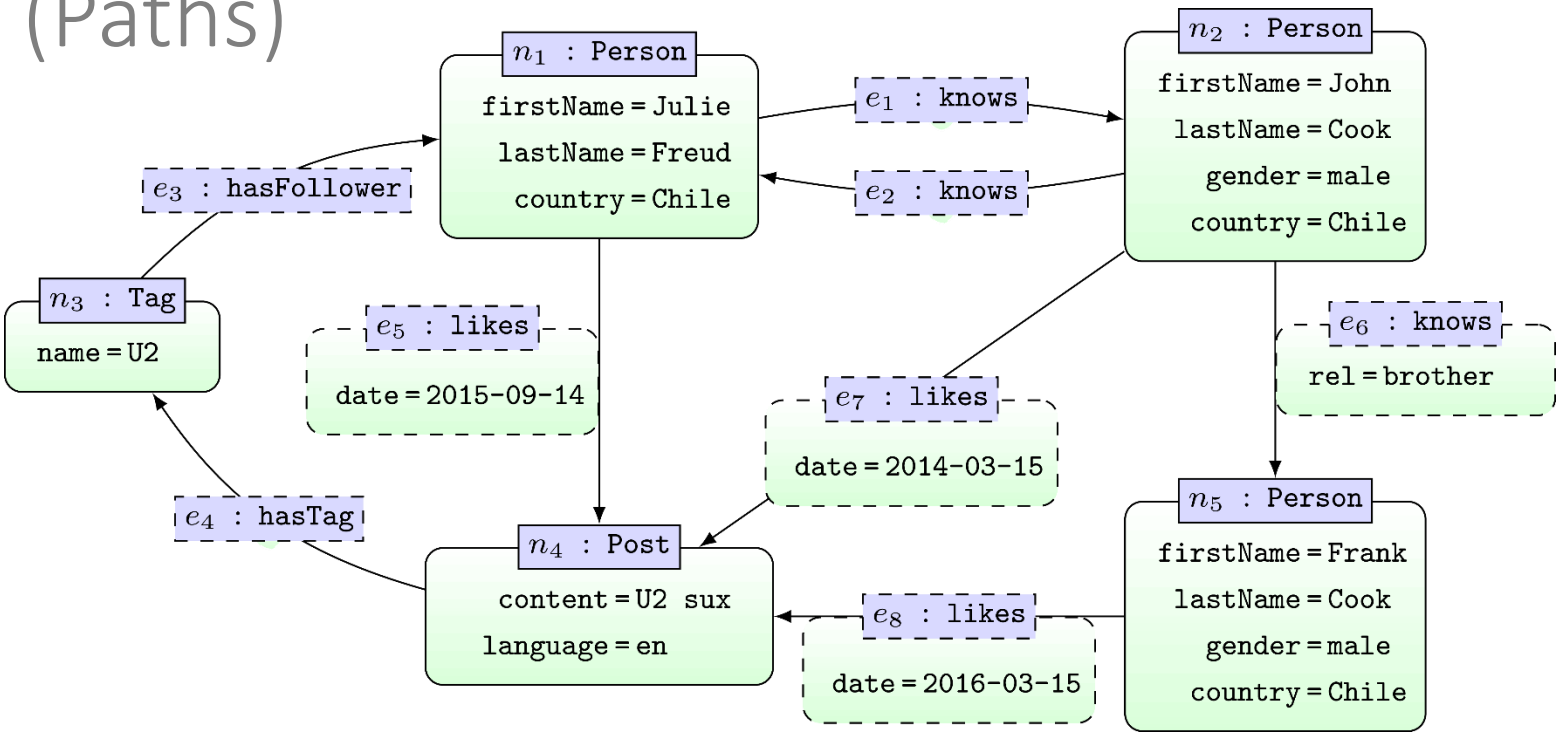



```
MATCH (c1 {lastName: 'Cook'}),
  (c2 {lastName: 'Freud'}),
  p = shortestPath((c1)-[*]->(c2))
RETURN p
```

p
$n_5 \cdot e_8 \cdot n_4 \cdot e_4 \cdot n_3 \cdot e_3 \cdot n_1$
$n_2 \cdot e_2 \cdot n_1$

... returns a shortest path for each matching pair of nodes

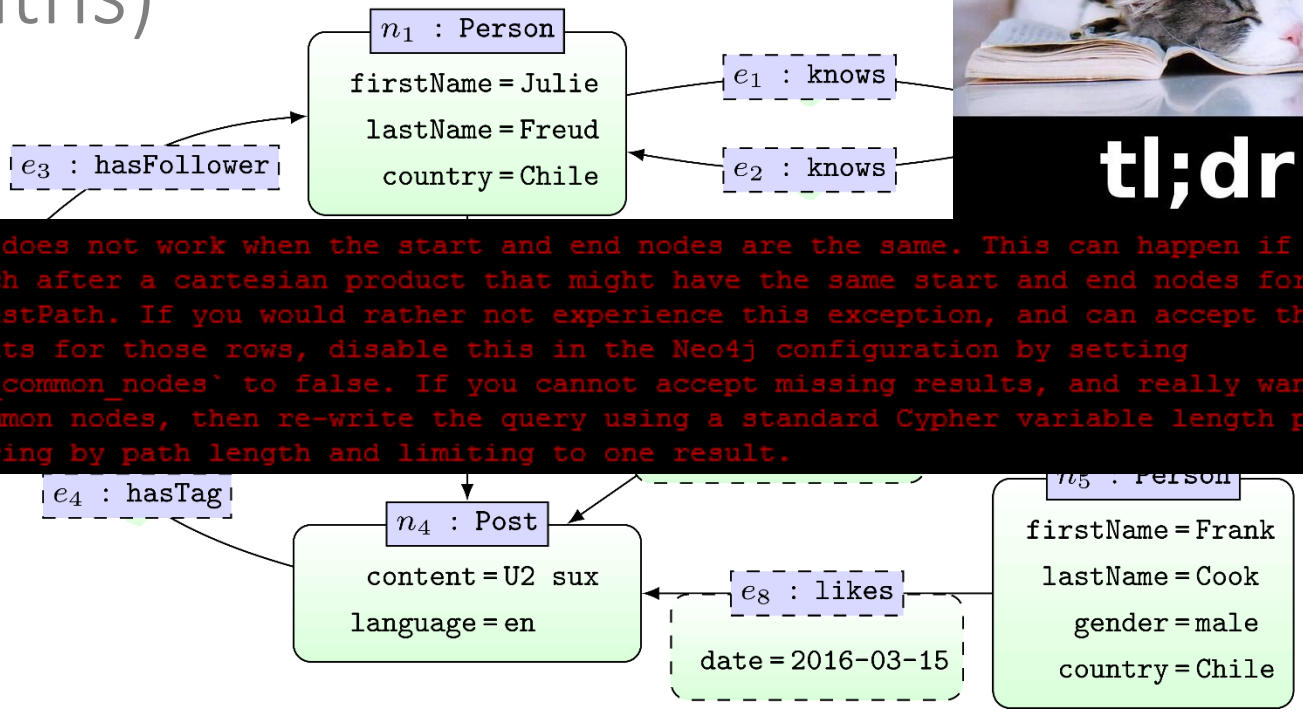
MATCH (Paths)




```
MATCH (c1 {lastName: 'Cook'}),
  (c2 {lastName: 'Cook'}),
  p = shortestPath((c1)-[*]->(c2))
RETURN p
```

MATCH (Paths)

tl;dr




```
The shortest path algorithm does not work when the start and end nodes are the same. This can happen if you
perform a shortestPath search after a cartesian product that might have the same start and end nodes for some
of the rows passed to shortestPath. If you would rather not experience this exception, and can accept the
possibility of missing results for those rows, disable this in the Neo4j configuration by setting
`cypher.forbid_shortestpath_common_nodes` to false. If you cannot accept missing results, and really want the
shortestPath between two common nodes, then re-write the query using a standard Cypher variable length pattern
expression followed by ordering by path length and limiting to one result.
```

```
MATCH (c1 {lastName: 'Cook'}),
  (c2 {lastName: 'Cook'}),
  p = shortestPath((c1)-[*]->(c2))
RETURN p
```

Cypher:

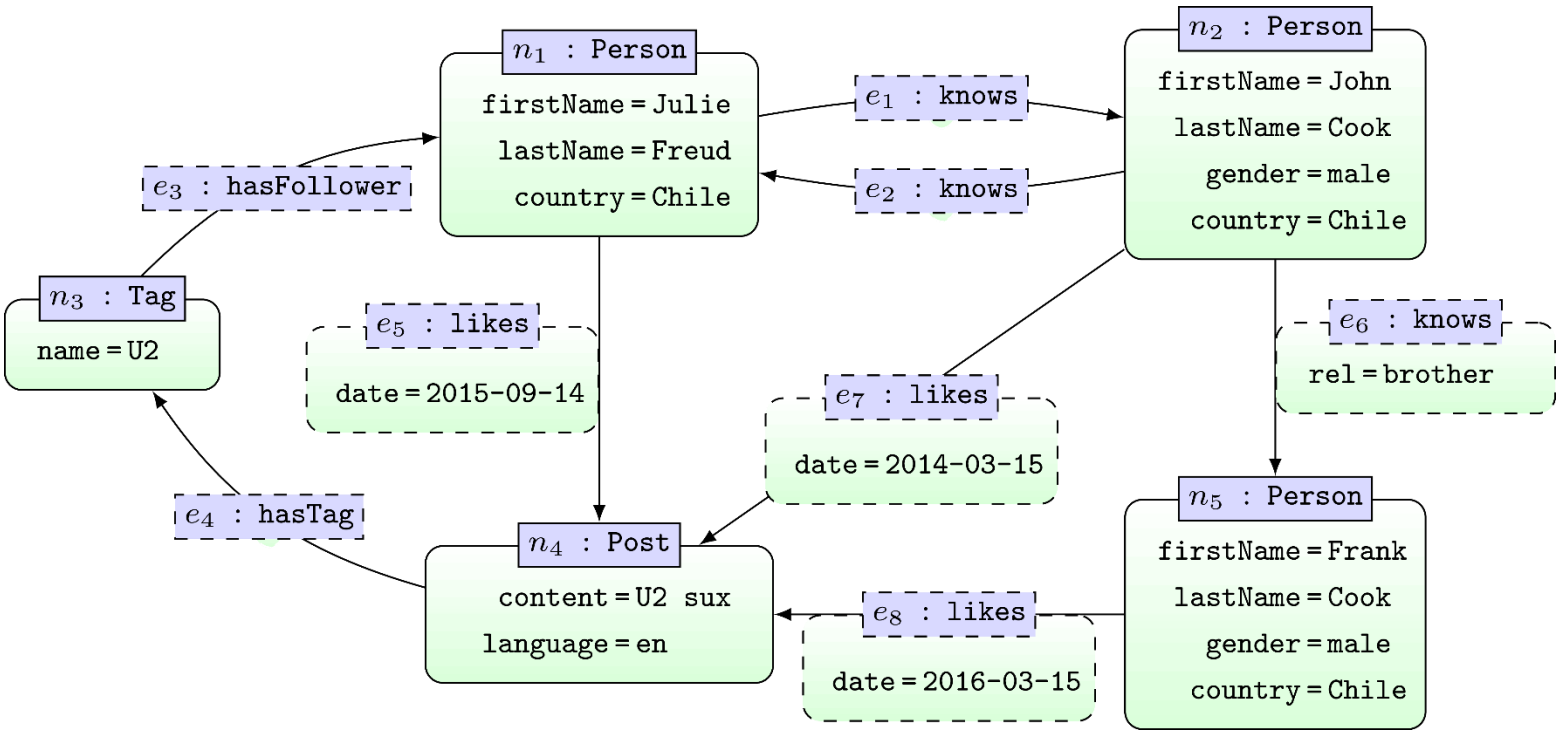
WHERE

https://neo4j.com/docs/developer-manual/3.4/cypher/clauses/where/

WHERE

- Boolean:
  - `AND`, `OR`, `XOR`, `NOT`
- (In)equalities:
  - `<`, `>`, `<>`, `<=`, `>=`
- Exists attribute property:
  - `EXISTS`
- Boolean:
  - `STARTS WITH`, `ENDS WITH`, `CONTAINS`, `=~` (Regex)

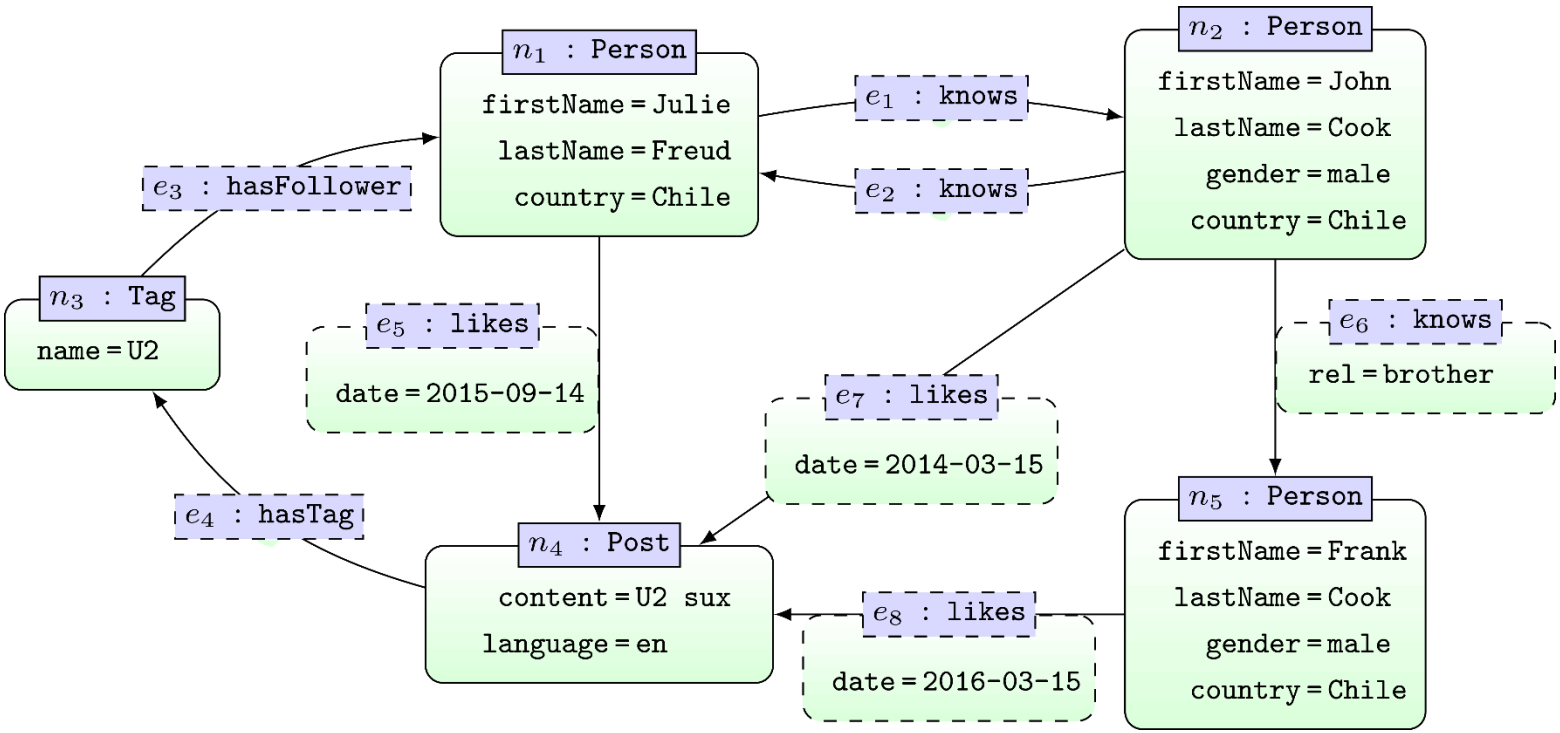
WHERE




```
MATCH (x)-[r:likes]->(y:Post)
WHERE r.date > '2010-01-01' AND r.date < '2015-01-01'
RETURN x.firstName
```

x.firstName
John

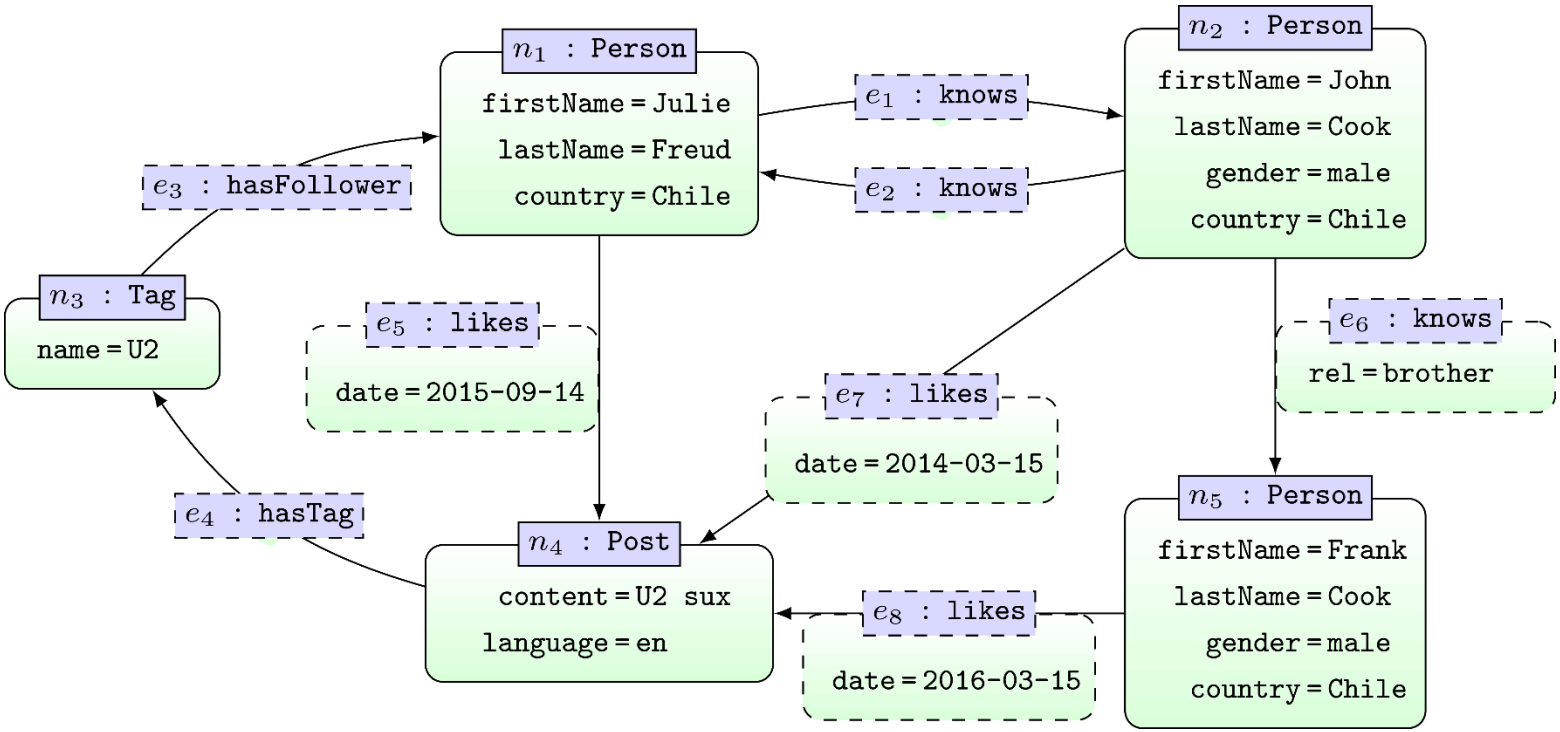
WHERE




```
MATCH (x)
WHERE EXISTS(x.gender)
RETURN x.firstName
```

x.firstName
John
Frank

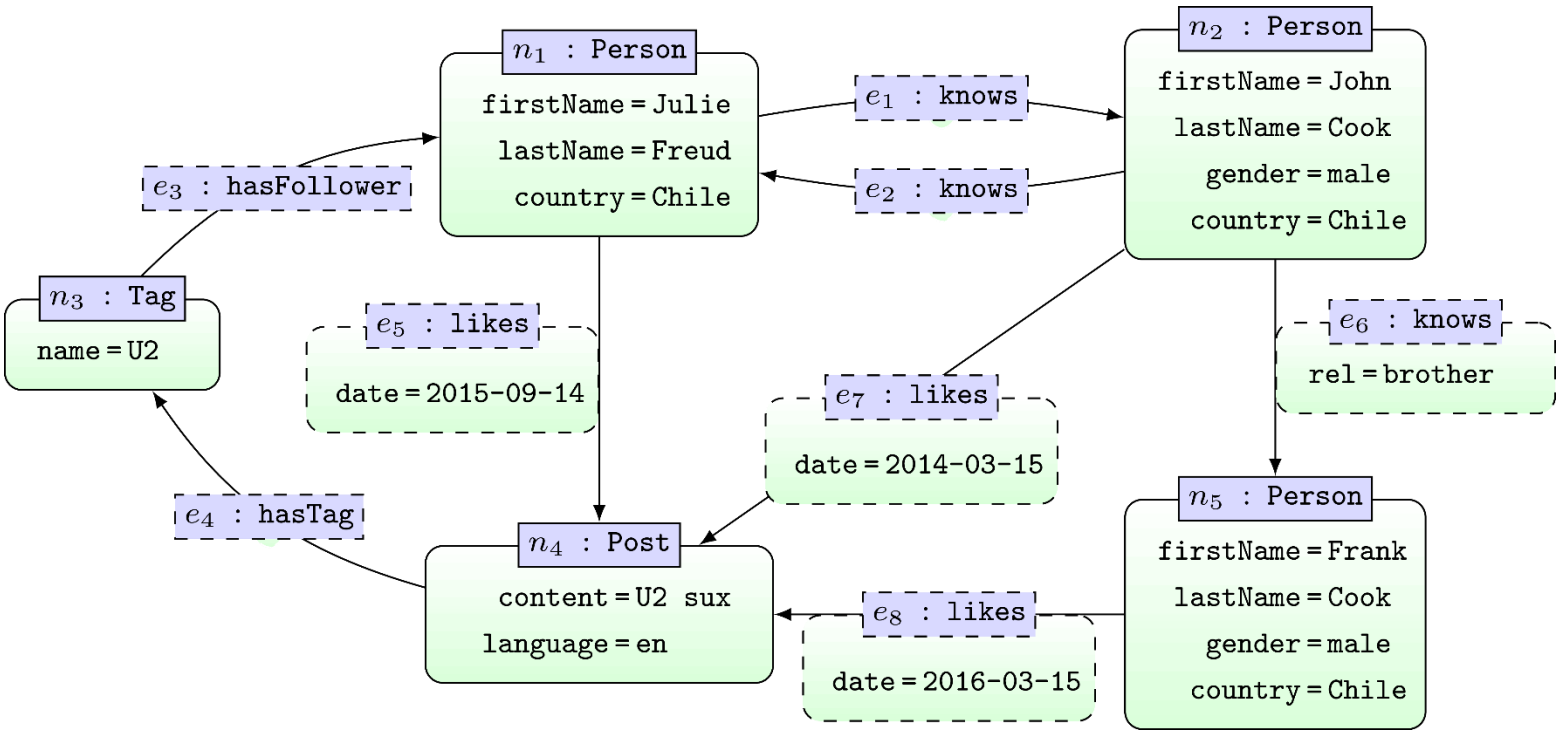
WHERE




```
MATCH (x)
WHERE x.firstName STARTS WITH 'J'
RETURN x.firstName
```

x.firstName
John
Julie

WHERE




```
MATCH (x)
WHERE x.name =~ '.*[0-9]'
RETURN x.name
```

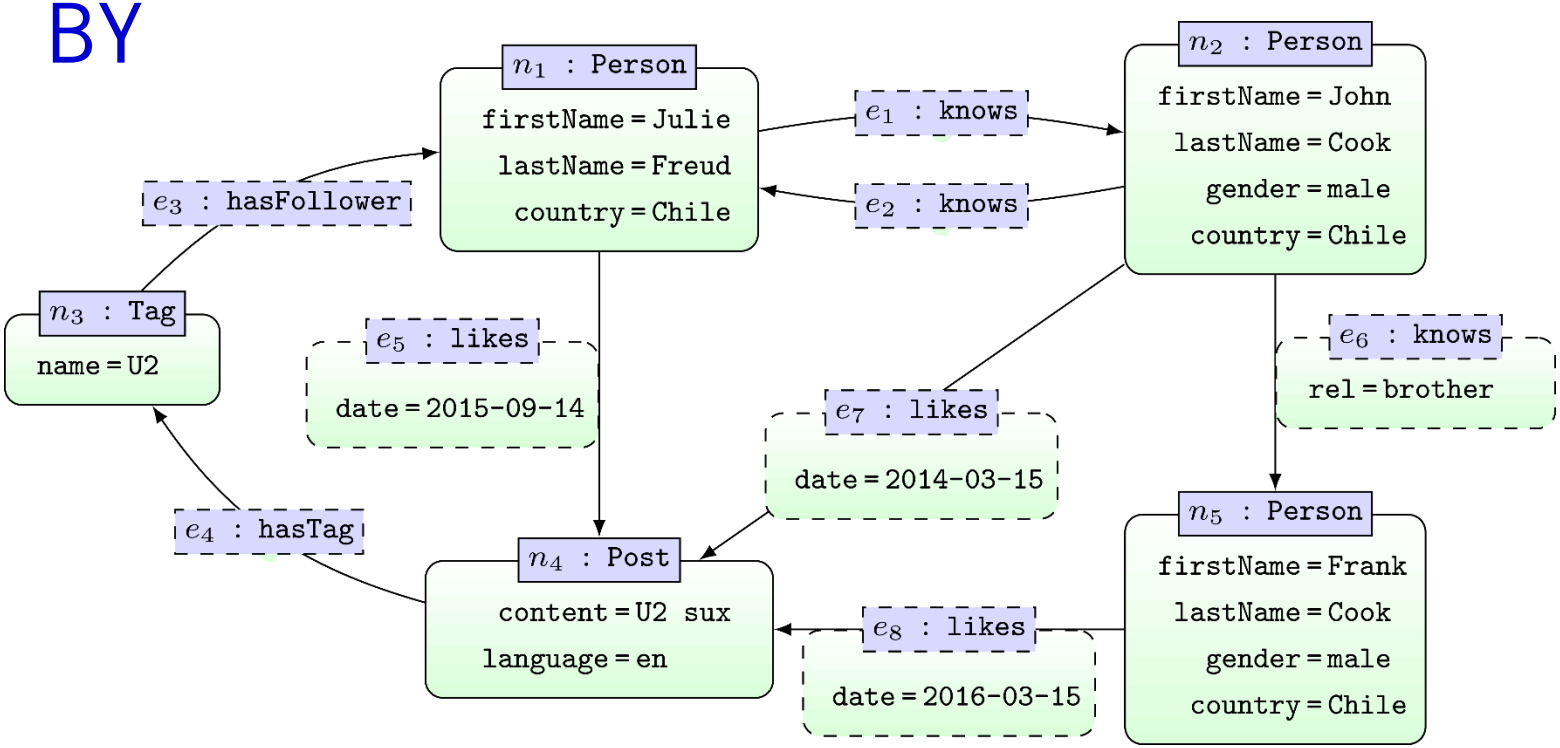
x.name
U2

Cypher:

ORDER BY/SKIP/LIMIT

https://neo4j.com/docs/developer-manual/3.4/cypher/clauses/order-by/
https://neo4j.com/docs/developer-manual/3.4/cypher/clauses/skip/
https://neo4j.com/docs/developer-manual/3.4/cypher/clauses/limit/

ORDER BY



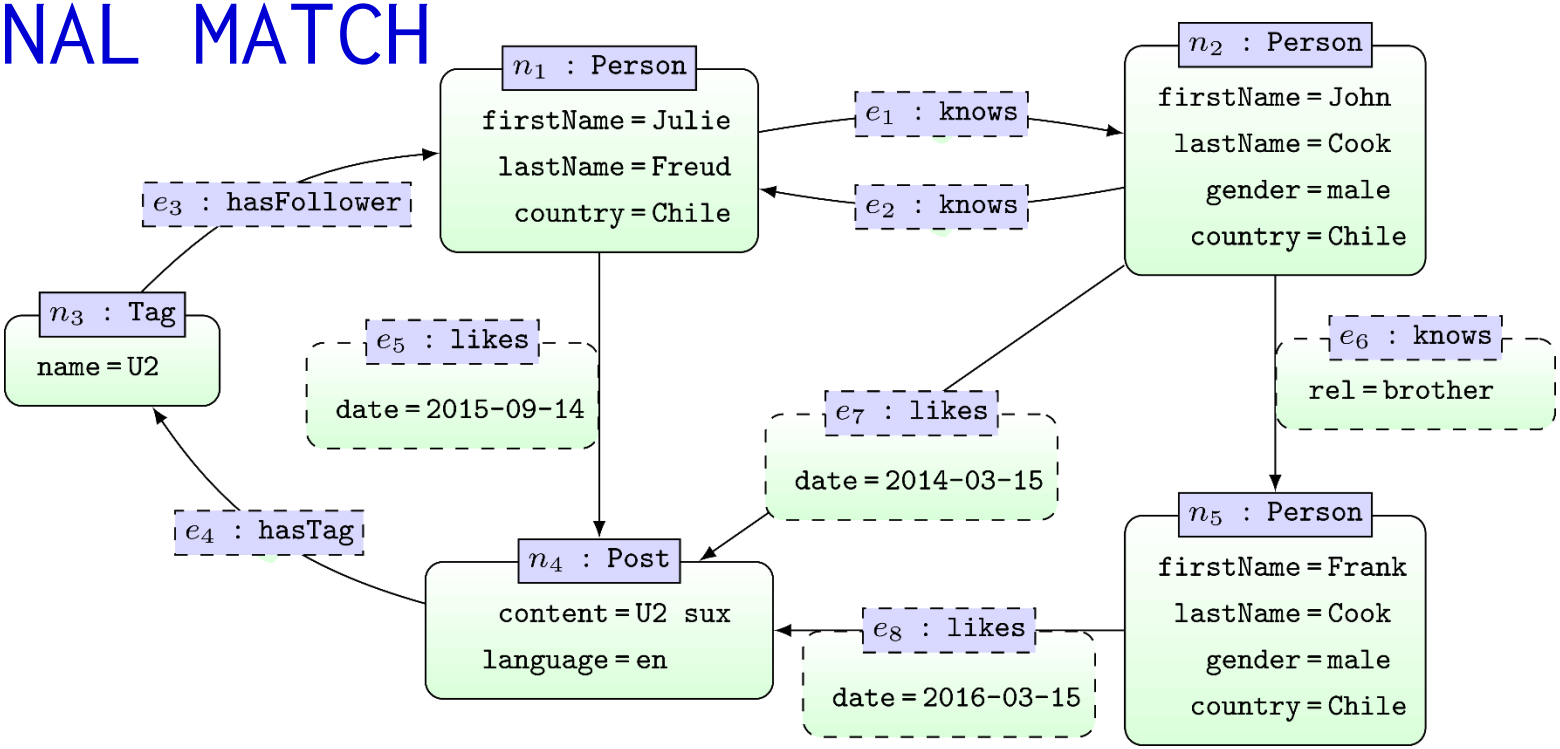

```
MATCH ()-[r:likes]->(p:Post)
RETURN r.date, p.content, p.language
ORDER BY p.content, r.date DESC
SKIP 1
LIMIT 1
```

r.date	p.content	p.language
2015-09-14	U2 sux	en

Cypher:
OPTIONAL MATCH

https://neo4j.com/docs/developer-manual/3.4/cypher/clauses/optional-match/

OPTIONAL MATCH




```
MATCH (x1)-[:knows]->(x2)
OPTIONAL MATCH (y)-[:hasFollower]->(x1)
RETURN x1.firstName,y.name
```

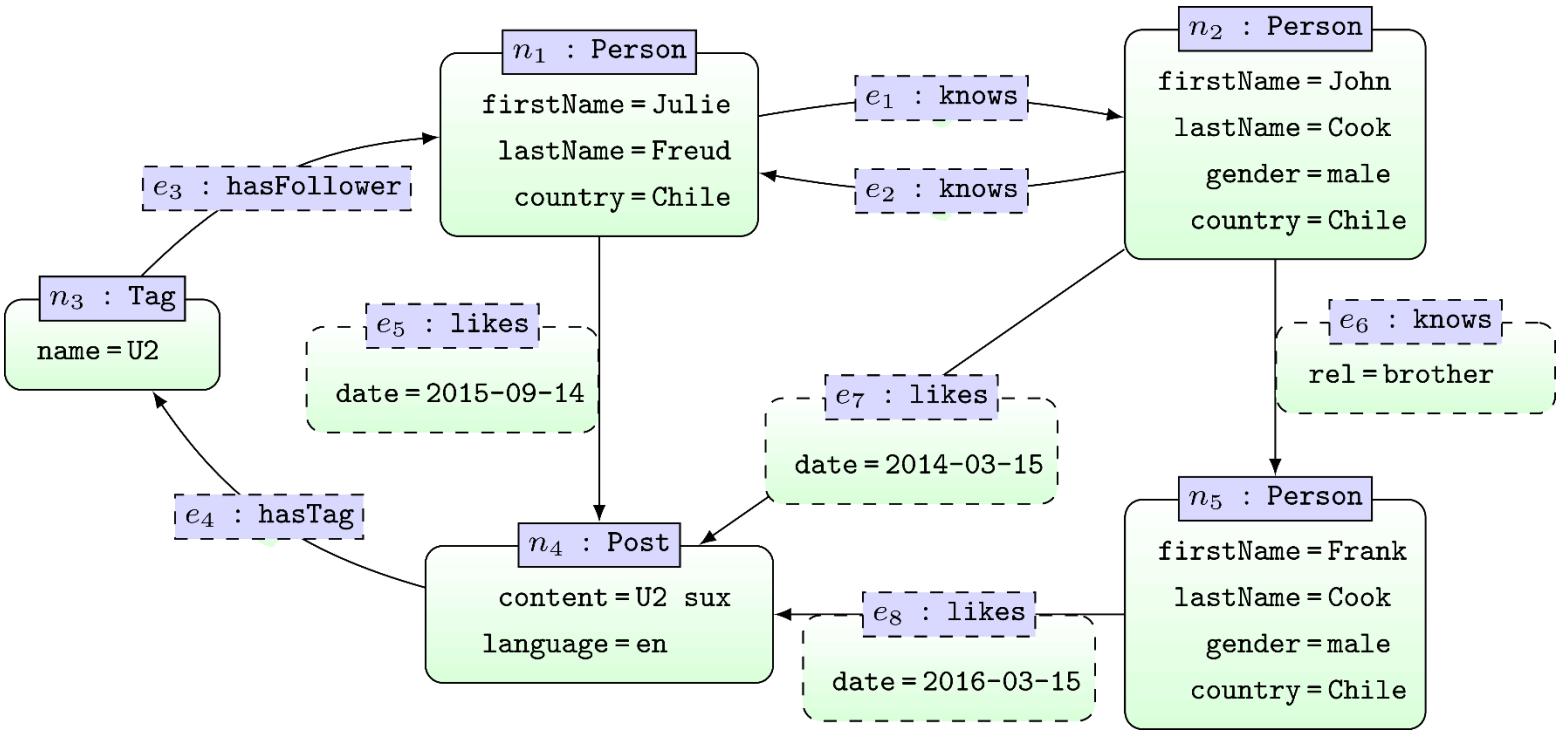
x1.firstName	y.name
Julie	U2
John	
John	

... OPTIONAL MATCH acts like a left join

Cypher:
UNION (ALL)

https://neo4j.com/docs/developer-manual/3.4/cypher/clauses/optional-match/

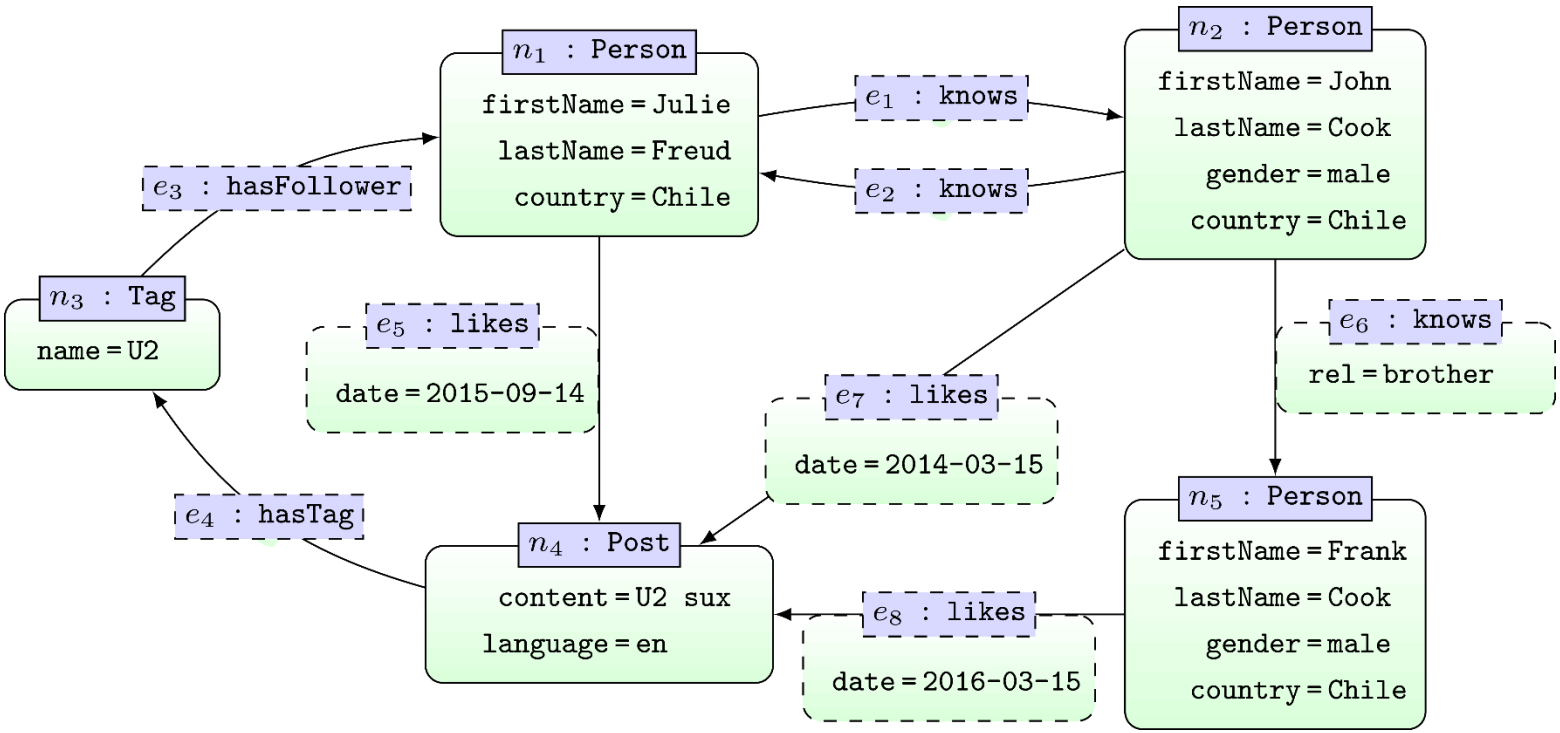
UNION




```
MATCH (x1)-[:knows]->(x2)
RETURN x1.firstName
UNION
MATCH (x1)-[:knows]->(x2)
RETURN x2.firstName
```

... column names have to be the same in the UNION

UNION

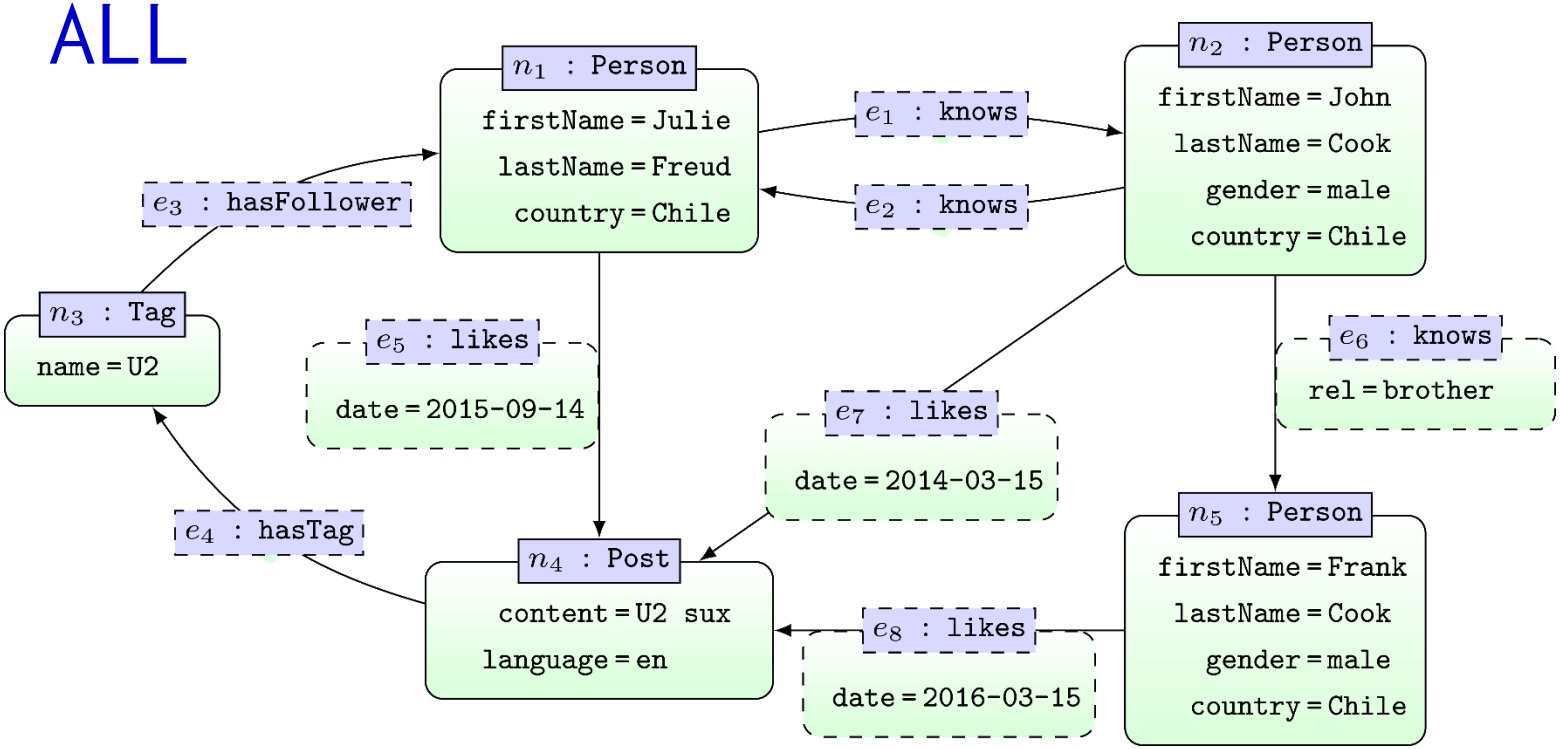



```
MATCH (x1)-[:knows]->(x2)
RETURN x1.firstName
UNION
MATCH (x2)-[:knows]->(x1)
RETURN x1.firstName
```

x1.firstName
Julie
John
Frank

... UNION applies set union

UNION ALL




```
MATCH (x1)-[:knows]->(x2)
RETURN x1.firstName
UNION ALL
MATCH (x2)-[:knows]->(x1)
RETURN x1.firstName
```

x1.firstName
Julie
Julie
John
John
John
Frank

... UNION ALL applies bag union

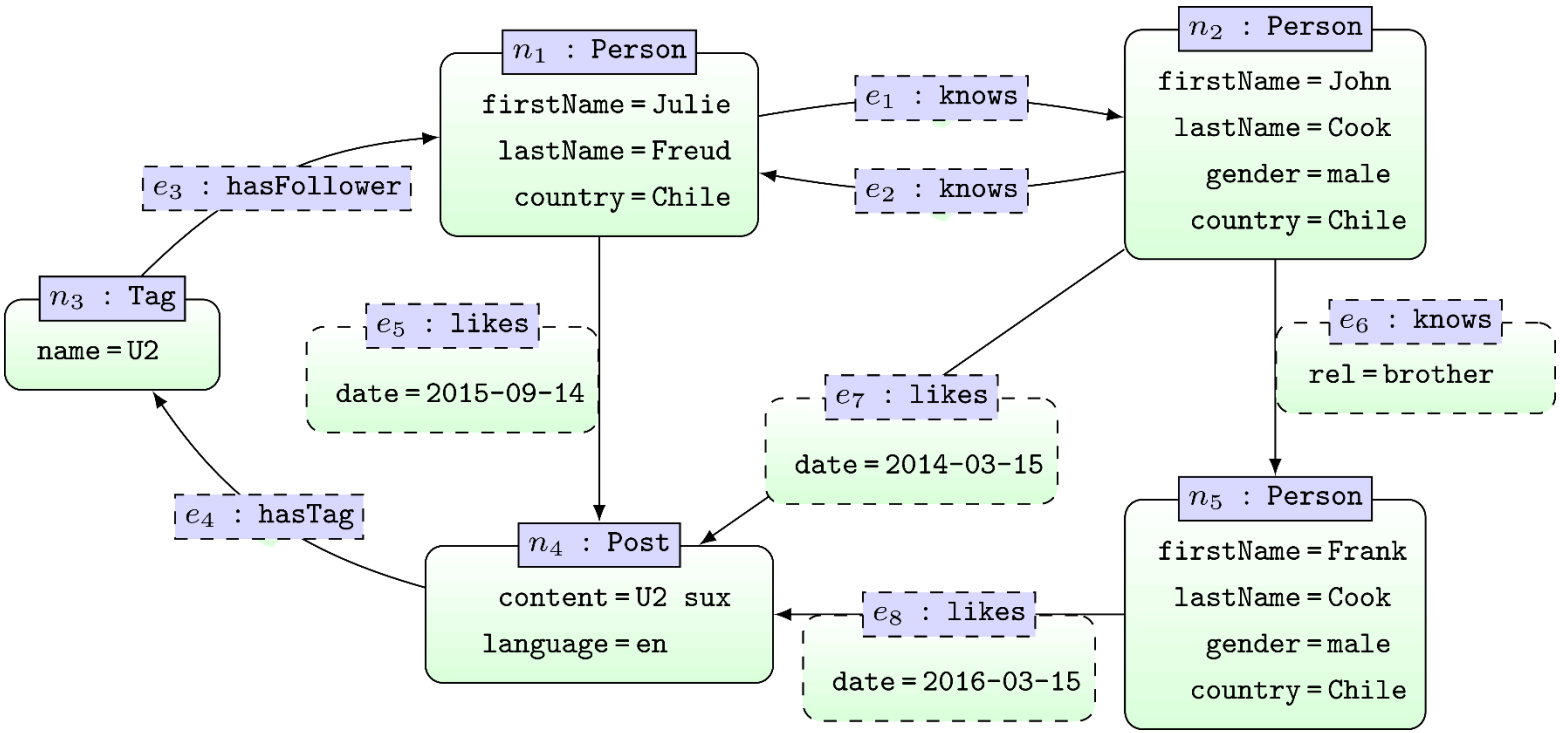
Cypher: Aggregation

https://neo4j.com/docs/developer-manual/current/cypher/functions/aggregating/

Aggregation

- `count`
- `max`/`min`
- `avg`
- `percentileCont`/`percentileDisc`
  - Computes percentile of some value w.r.t. some list
  - (*continuous*: interpolates / *discrete*: rounds)
- `stDev`/`stDevP`
  - Computes standard deviation (*sample*/*population*)

COUNT

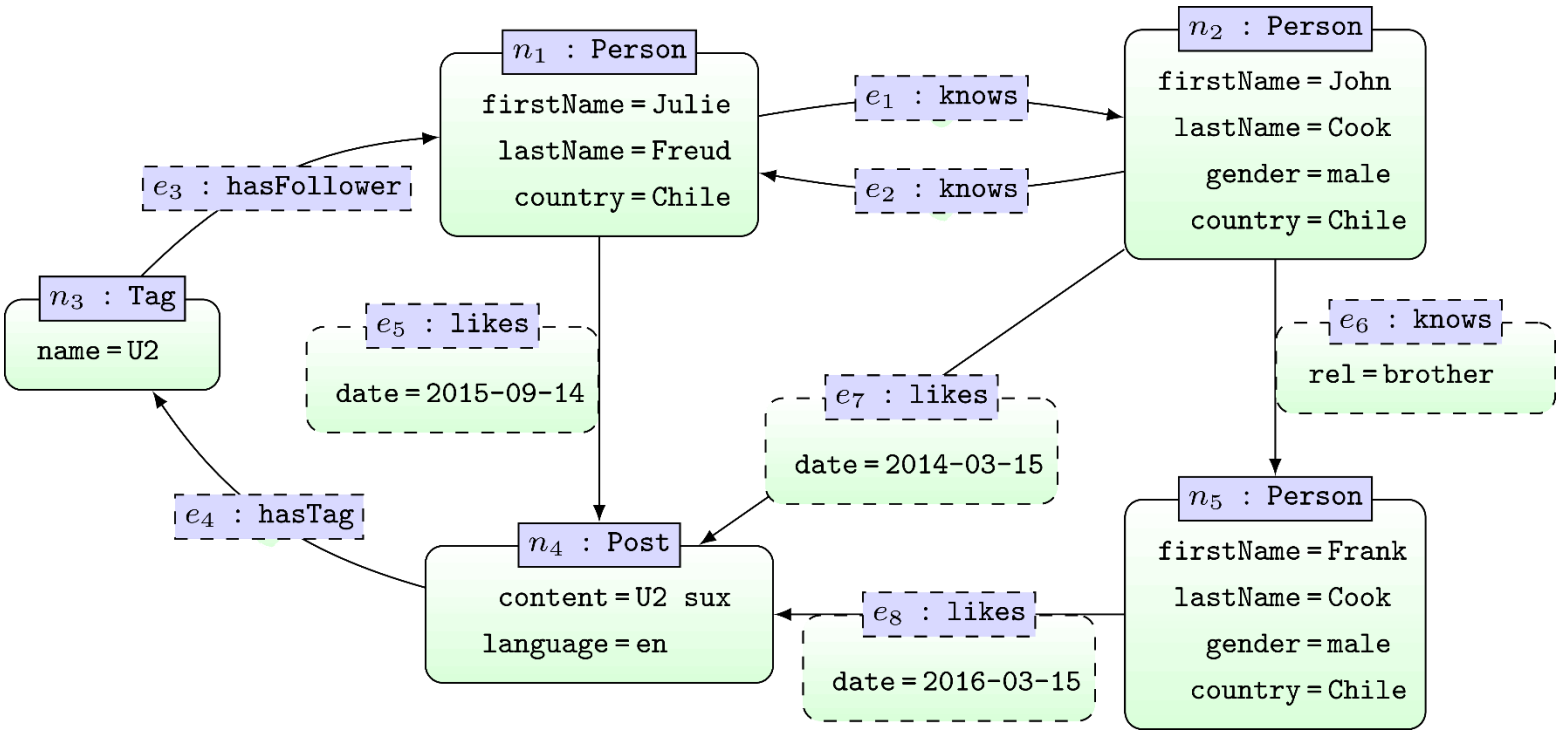



```
MATCH (x1)-[:knows*]->(x2)
RETURN x1.firstName, count(x2)
```

x1.firstName	count(x2)
Julie	3
John	4

... GROUP BY implicit on non-aggregated columns

COUNT

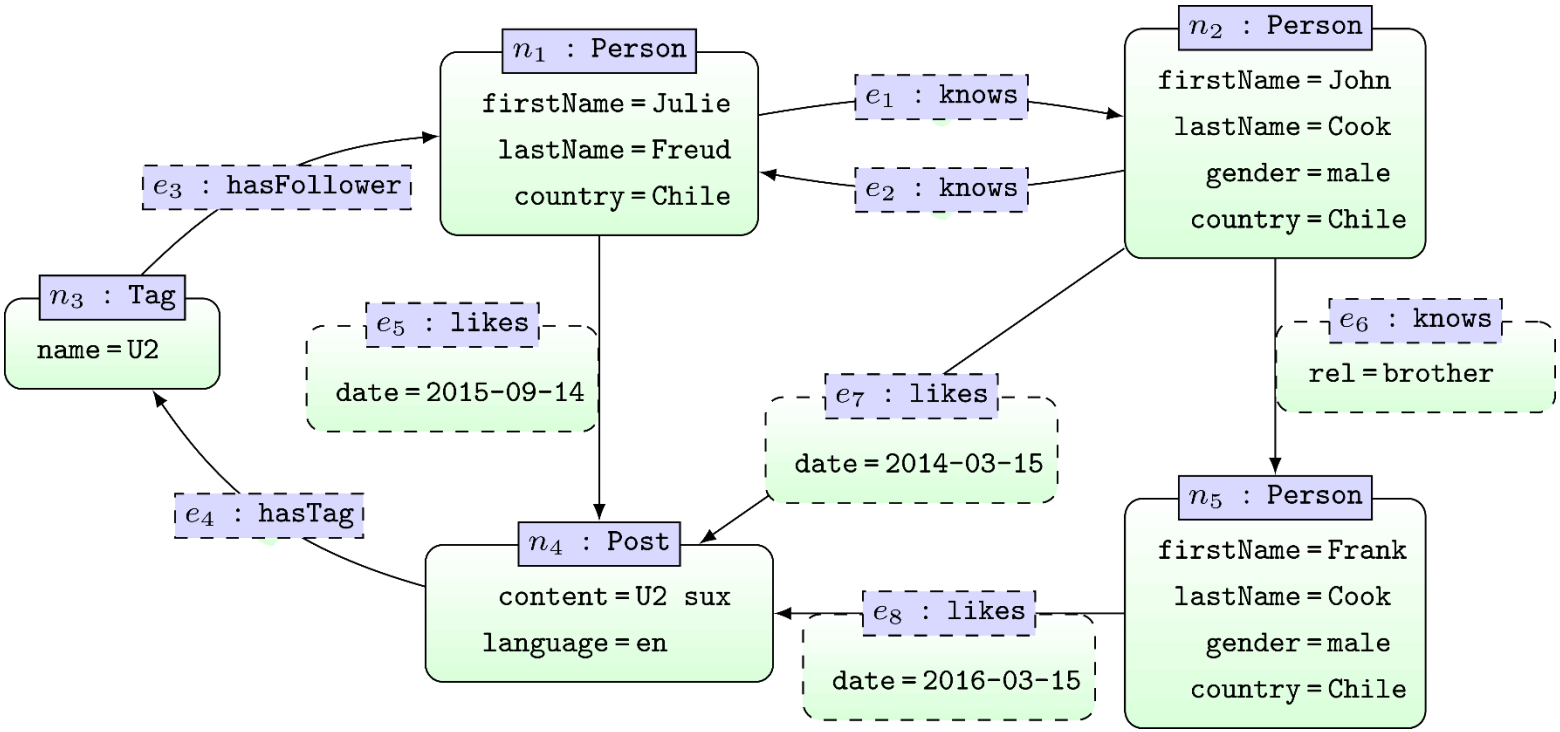



```
MATCH (x1)-[:knows*]->(x2)
RETURN DISTINCT x1.firstName, count(x2)
```

x1.firstName	count(x2)
Julie	3
John	4

... removes duplicate results, not count arguments

COUNT




```
MATCH (x1)-[:knows*]->(x2)
RETURN x1.firstName, count(distinct x2)
```

x1.firstName	count(x2)
Julie	3
John	3

... removes duplicate count arguments

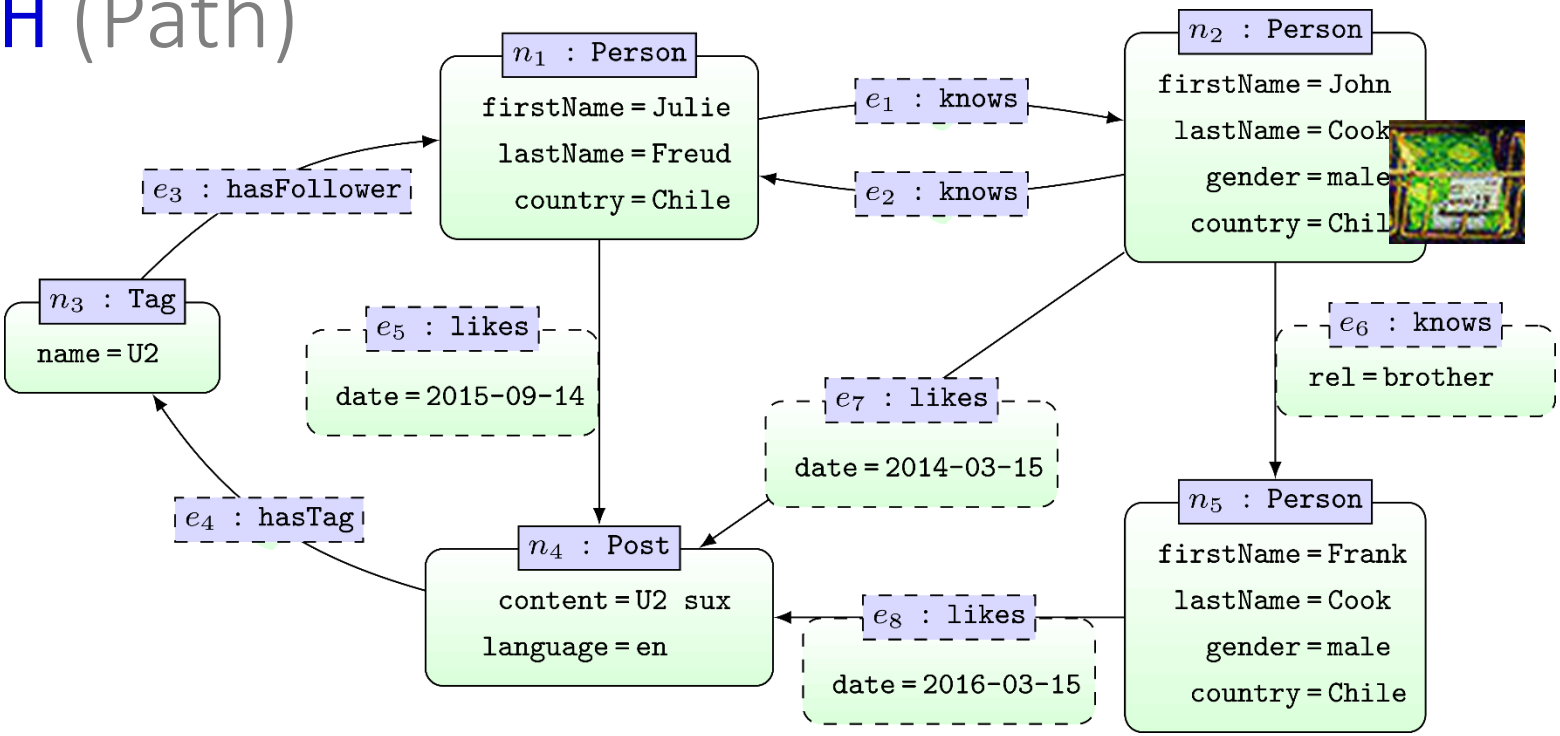
Cypher: Other Functions

https://neo4j.com/docs/developer-manual/current/cypher/functions/

https://neo4j.com/docs/developer-manual/current/cypher/functions/

https://neo4j.com/docs/developer-manual/current/cypher/functions/

LENGTH (Path)




```
MATCH (f {firstName: 'Frank'}),
  (j {firstName: 'Julie'}),
  p = shortestPath((f)-[*]->(j))
RETURN length(p)
```

length(p)
3

https://neo4j.com/docs/developer-manual/current/cypher/functions/

Cypher: Update Graphs

CREATE/REMOVE/...

Update graphs

- CREATE nodes and relationships
  - https://neo4j.com/docs/developer-manual/current/cypher/clauses/create/
- DELETE nodes and relationships
  - https://neo4j.com/docs/developer-manual/current/cypher/clauses/delete/
- DETACH DELETE nodes with relationships
  - https://neo4j.com/docs/developer-manual/current/cypher/clauses/delete/
- SET update labels and attributes
  - https://neo4j.com/docs/developer-manual/current/cypher/clauses/set/
- REMOVE remove labels and attributes
  - https://neo4j.com/docs/developer-manual/current/cypher/clauses/remove/

Update graphs

- Create the nodes we've seen

```
CREATE (:Person { firstName:'Julie', lastName:'Freud', country:'Chile' });
CREATE (:Person { firstName:'John', lastName:'Cook', country:'Chile', gender:'male' });
CREATE (:Tag { name:'U2' });
CREATE (:Post { content:'U2 sux', language:'en' });
CREATE (:Person { firstName:'Frank', lastName:'Cook', country:'Chile', gender:'male' });
```

- Create the edges (sample) we've seen

```
MATCH  (n1 { firstName:'Julie' }),(n2 { firstName:'John' }),(n3:Tag),(n4:Post),(n5 { firstName:'Frank' })
CREATE (n1)-[e1:knows]->(n2)
CREATE (n2)-[e2:knows]->(n1)
CREATE (n3)-[e3:hasFollower]->(n1)
CREATE (n4)-[e4:hasTag]->(n3)
CREATE (n1)-[e5:likes { date:'2015-09-14'}]->(n4)
CREATE (n2)-[e6:knows { rel:'brother'}]->(n5)
CREATE (n2)-[e7:likes { date:'2014-03-15'}]->(n4)
CREATE (n5)-[e8:likes { date:'2016-03-15'}]->(n4);...
```

- Drop all nodes and edges

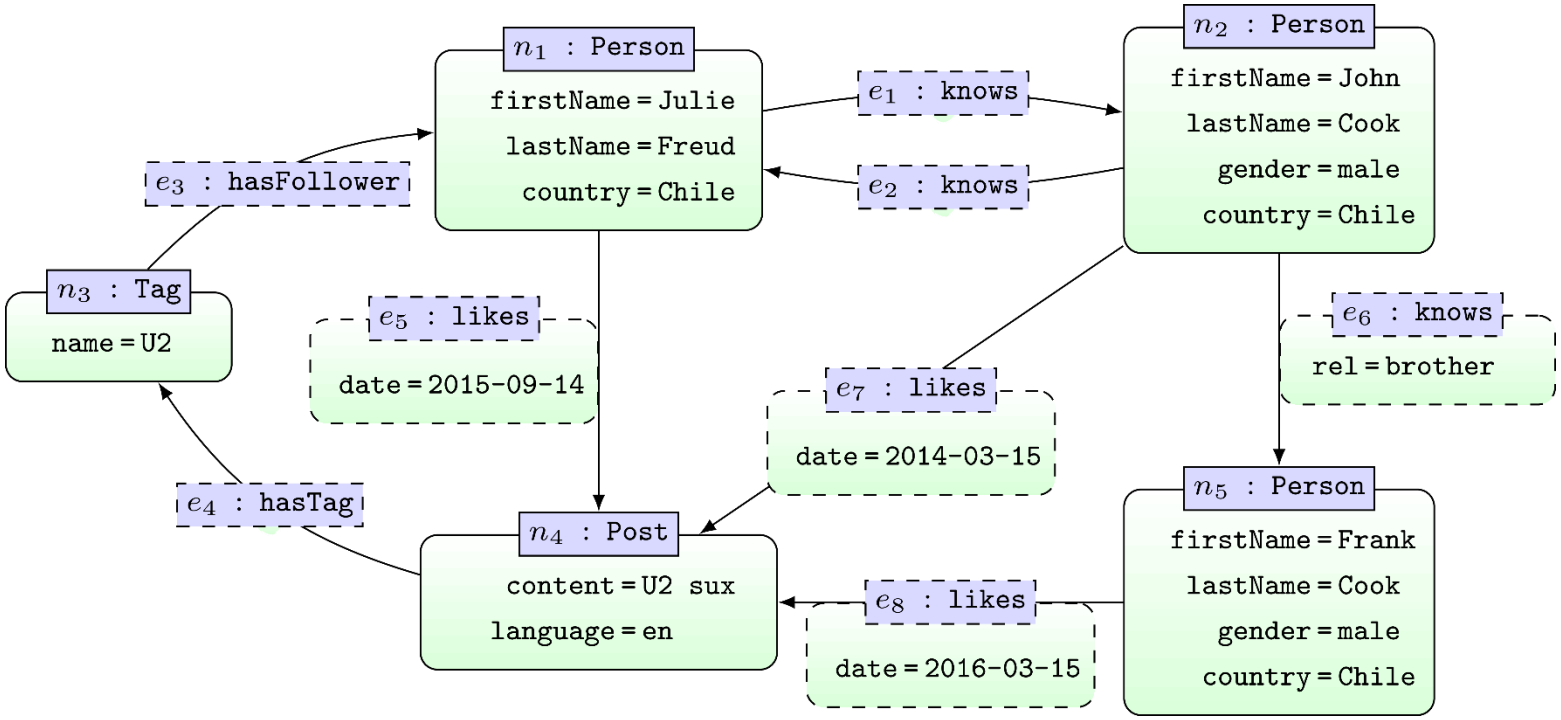
```
MATCH (n) DETACH DELETE n;
```

/Core of Cypher
/Part of Neo4j

Neo4j Graph Database

- Data Model: Property Graphs
- Query Language: Cypher
- Scripting Language: Gremlin
- Licence: Open Source (Single Machine)
  Commercial (Cluster Edition)

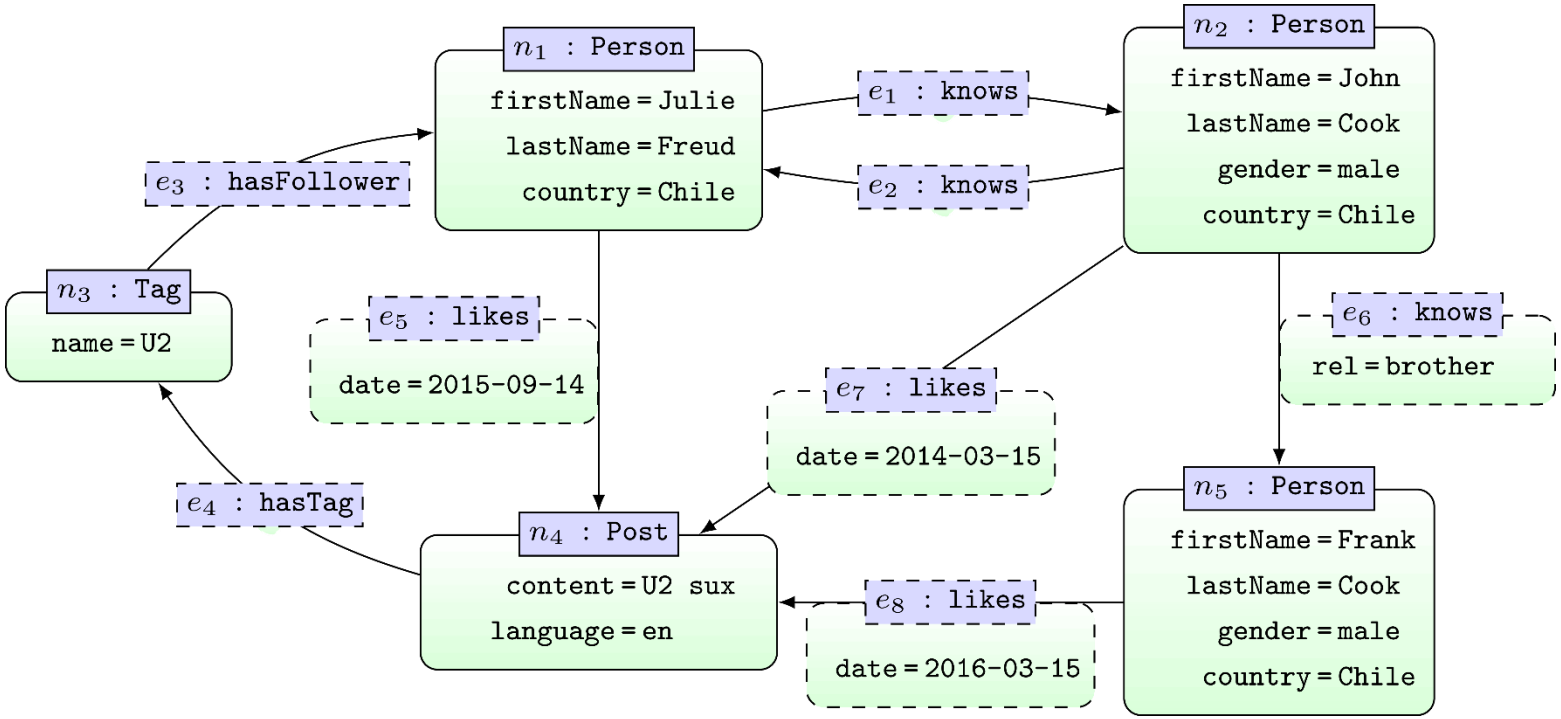
Property Graph: Cypher




```
MATCH (x1:Person {firstName:"Julie"})-[:knows*]->(x2:Person)
MATCH (x2)-[:likes]->()-[:hasTag]->()-[:hasFollower]->(x1)
RETURN x2.firstName
```

x2.firstName
Julie
John
Frank

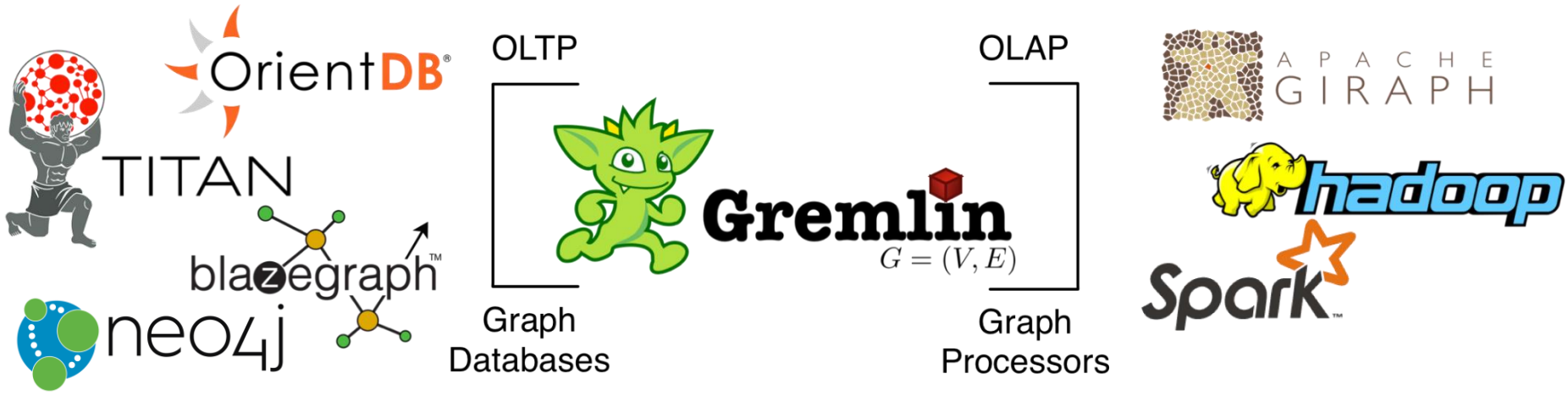
Property Graph: Gremlin



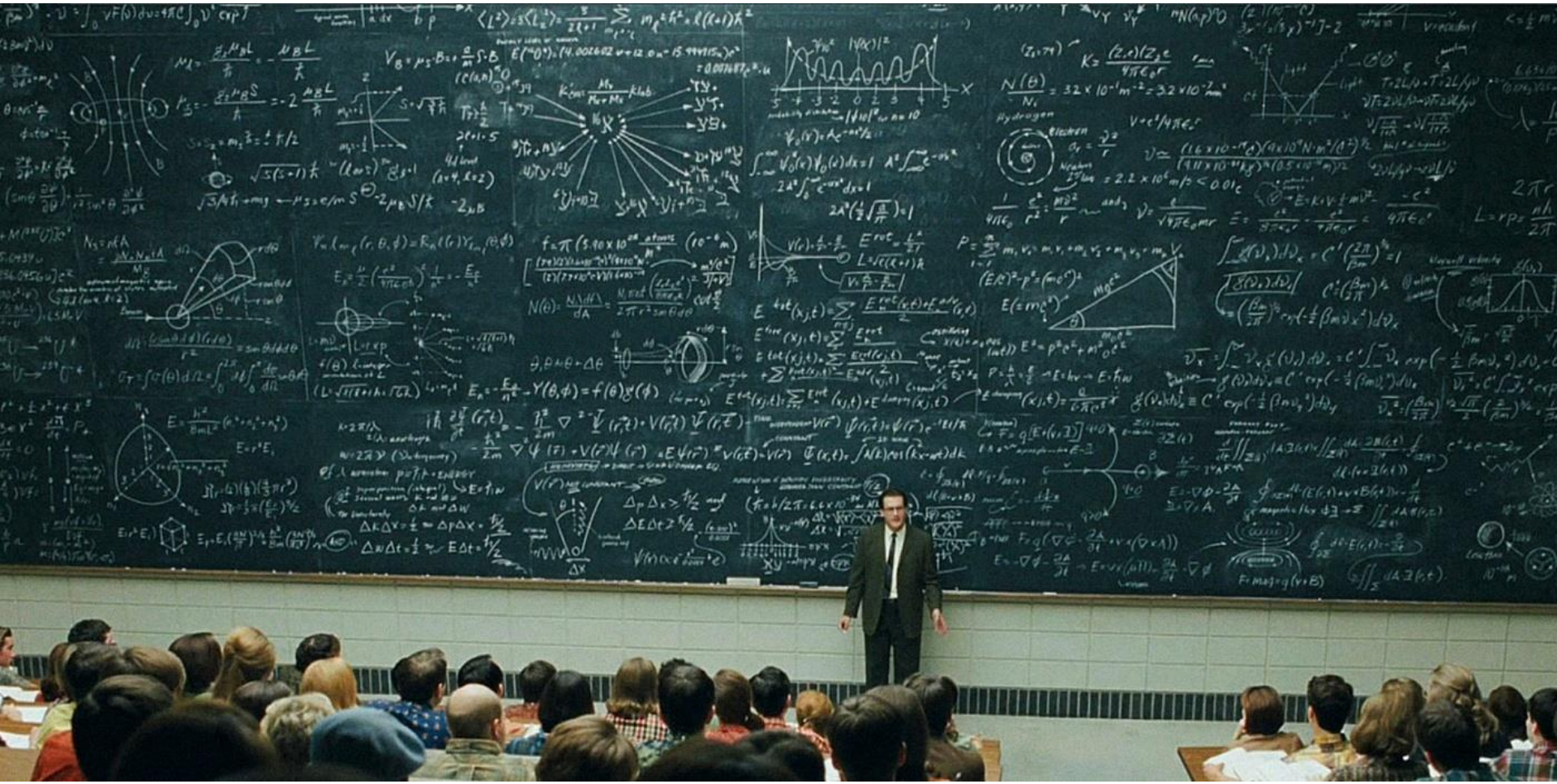

```
n1 = g.v.filter{firstName:"Julie"}.inV.outE('knows').loop()
n2 = n1.call().more.functions().until('done')
```

x2.firstName
Julie
John
Frank

Gremlin: Graph Queries + Processing



Graph Processing: Abstraction (in brief)

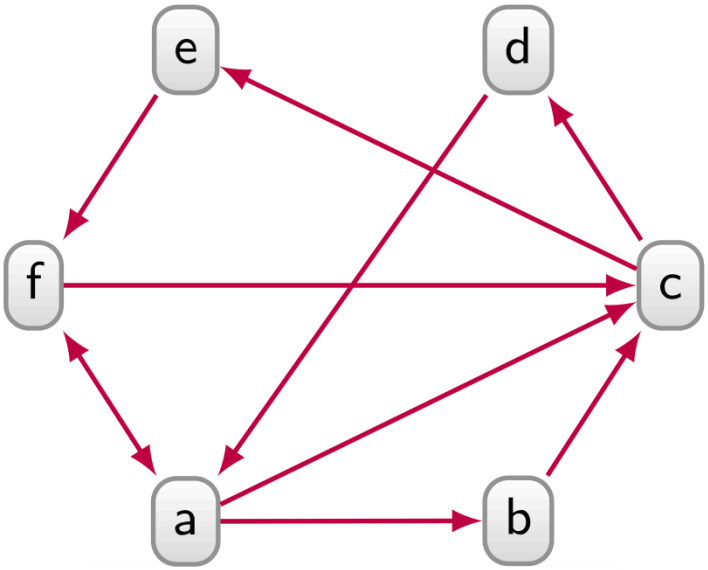


Graph Parallel Computation

Aka. Systolic Computation,
Asynchronous Actor Model,
Vertex-Centric Computation,
...

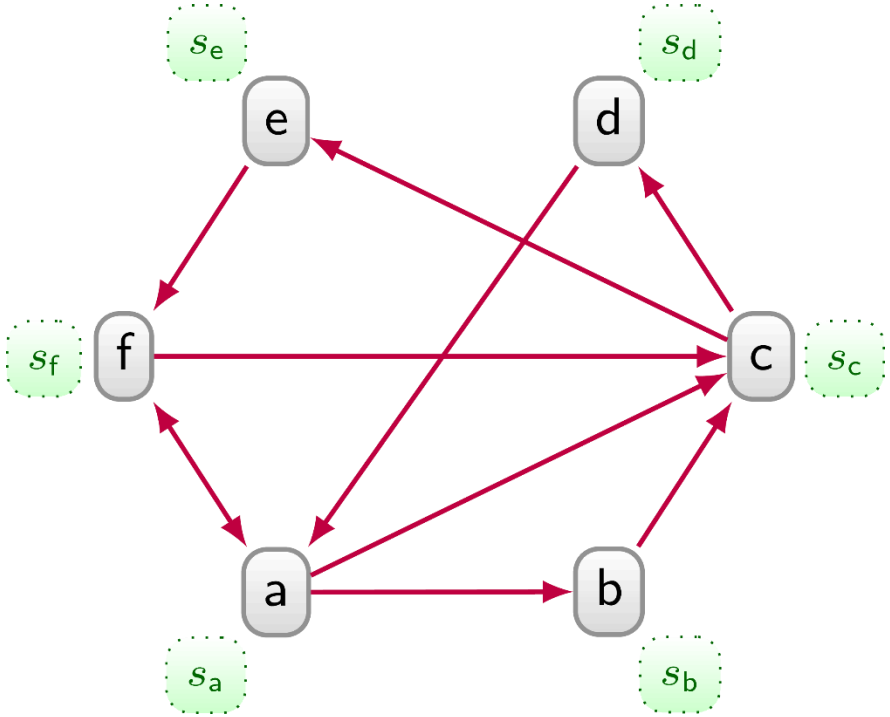
Graph Parallel Computation

- Graph



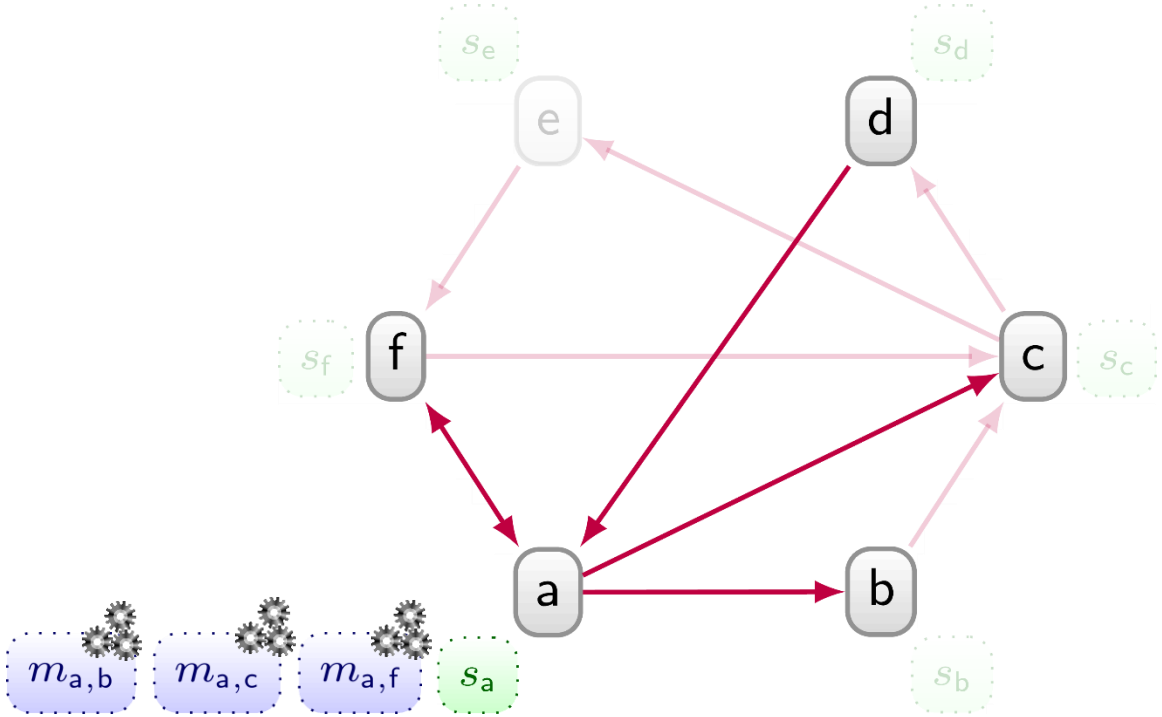
Graph Parallel Computation

- Graph
- Vertexes have state



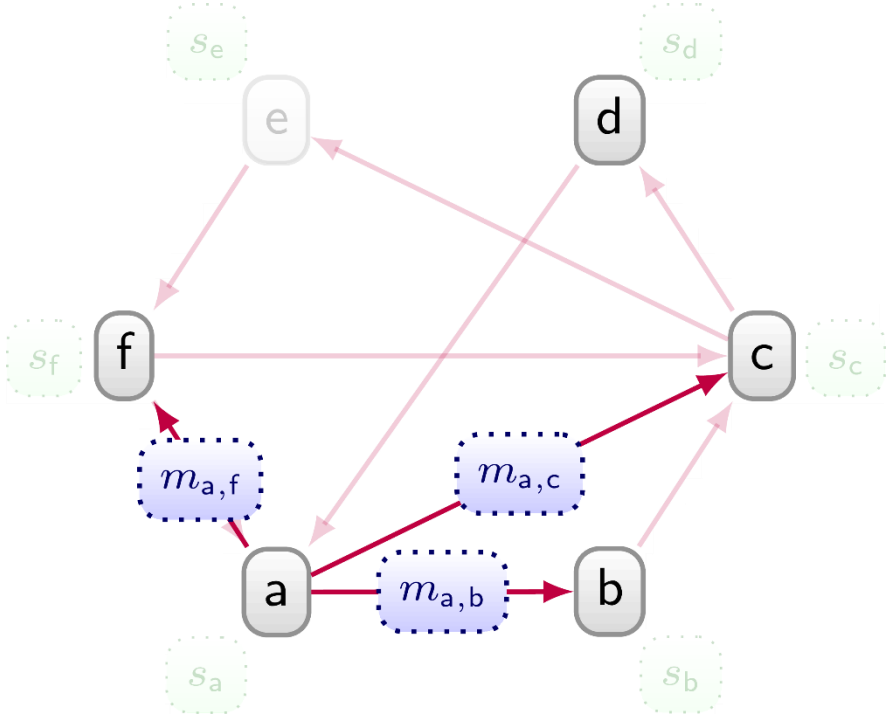
Graph Parallel Computation

- Graph
- Vertexes have state
- Vertexes compute messages from state and neighbours



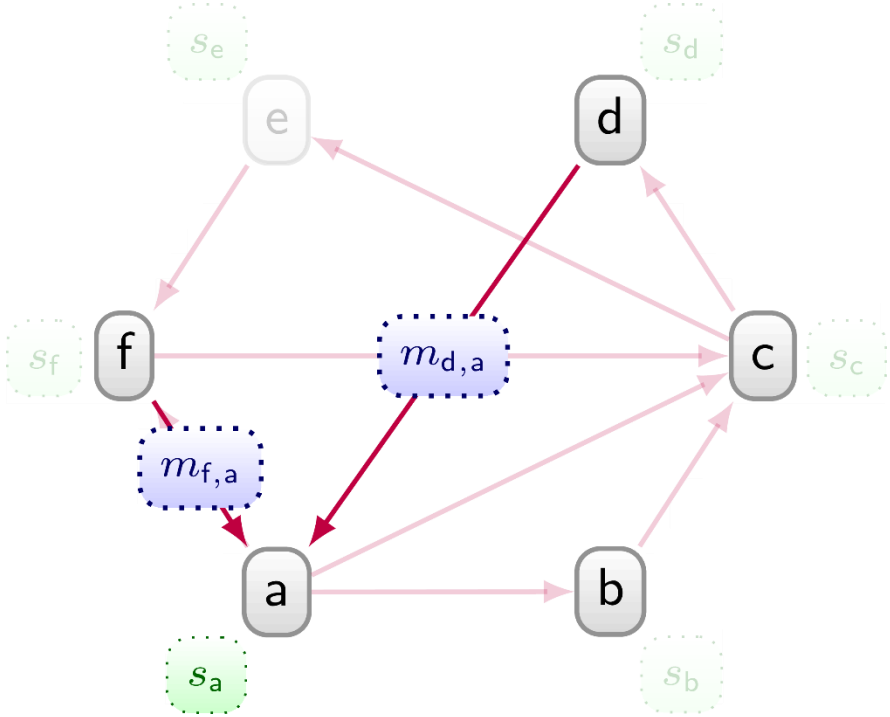
Graph Parallel Computation

- Graph
- Vertexes have state
- Vertexes compute messages from state and neighbours
- Vertexes send messages to neighbours



Graph Parallel Computation

- Graph
- Vertexes have state
- Vertexes compute messages from state and neighbours
- Vertexes send messages to neighbours
- Vertexes receive messages and compute new state

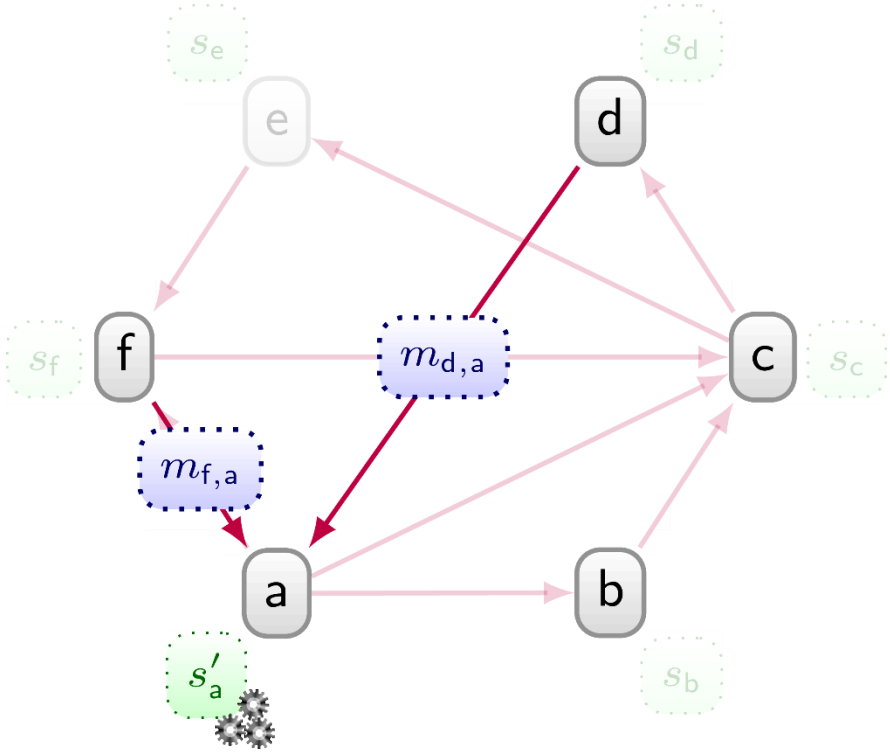


Graph Parallel Computation

- Graph
- Vertexes have state
- Vertexes compute messages from state and neighbours
- Vertexes send messages to neighbours
- Vertexes receive messages and compute new state

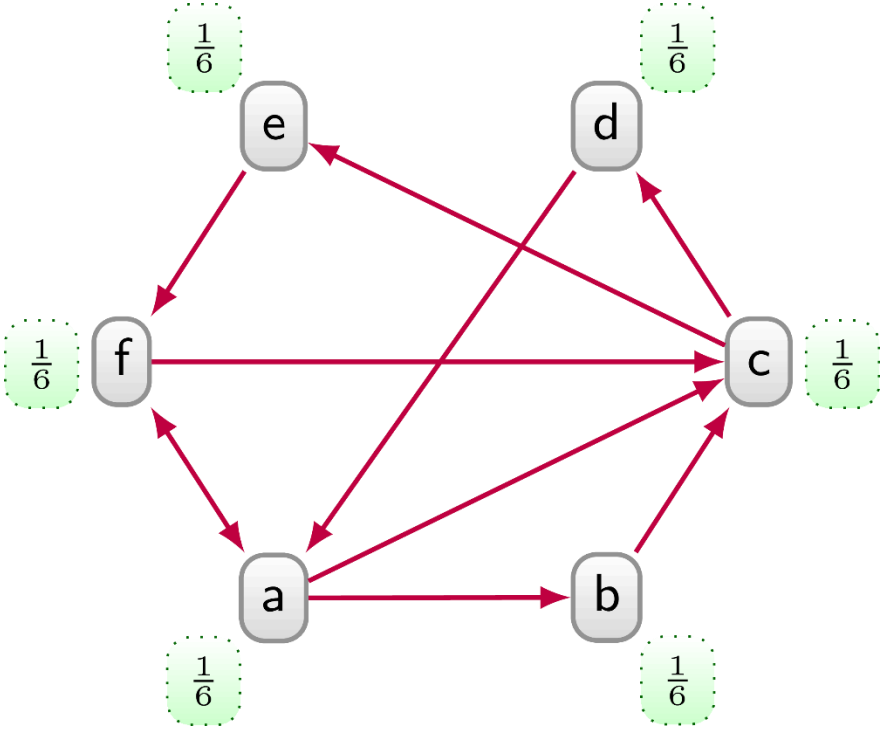
  ... recursively

  ... (in parallel)



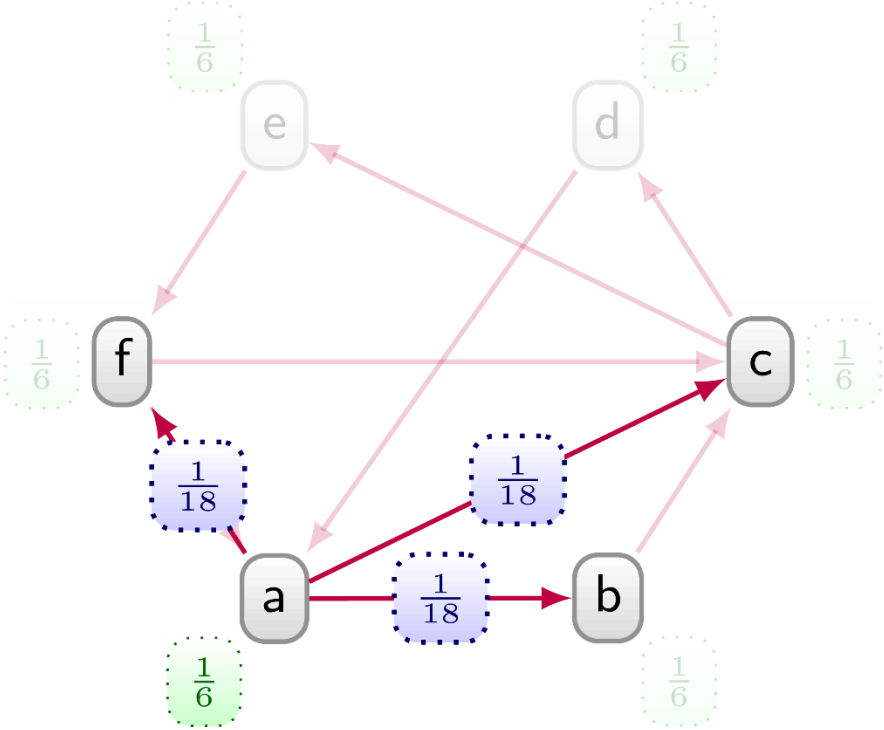
Example: PageRank (simplified)

$\text{PageRank}(G)$:



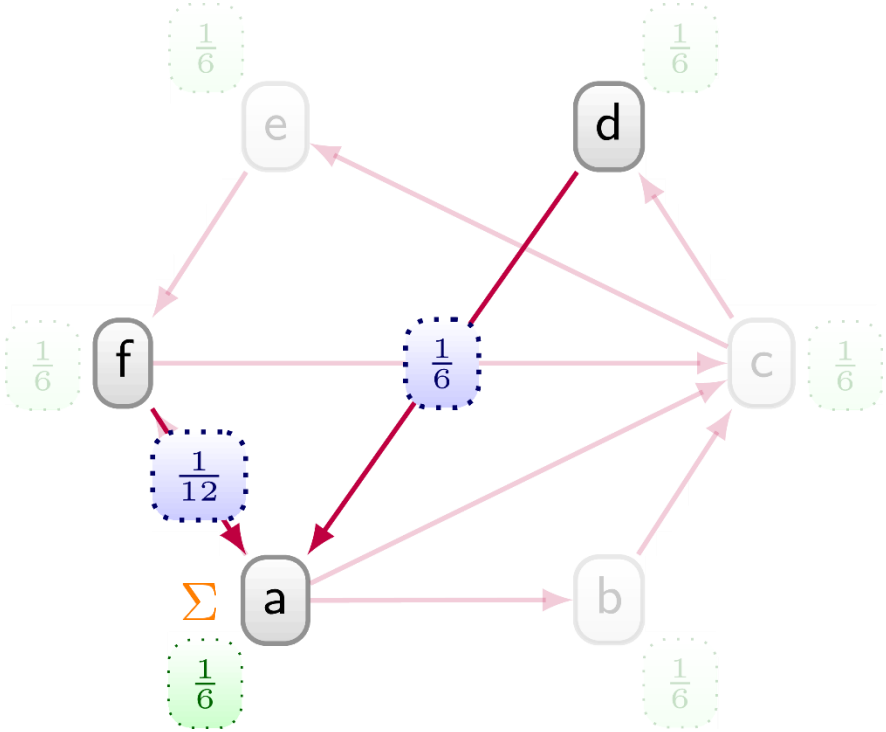
Example: PageRank (simplified)

$\mathrm{PageRank}(G)$:



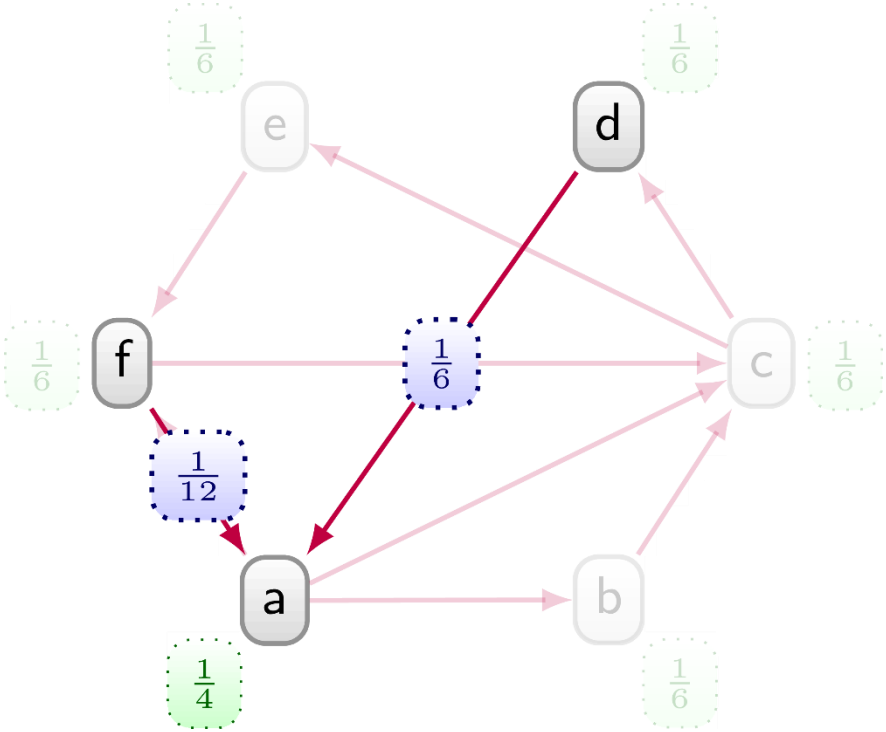
Example: PageRank (simplified)

$\mathrm{PageRank}(G)$:



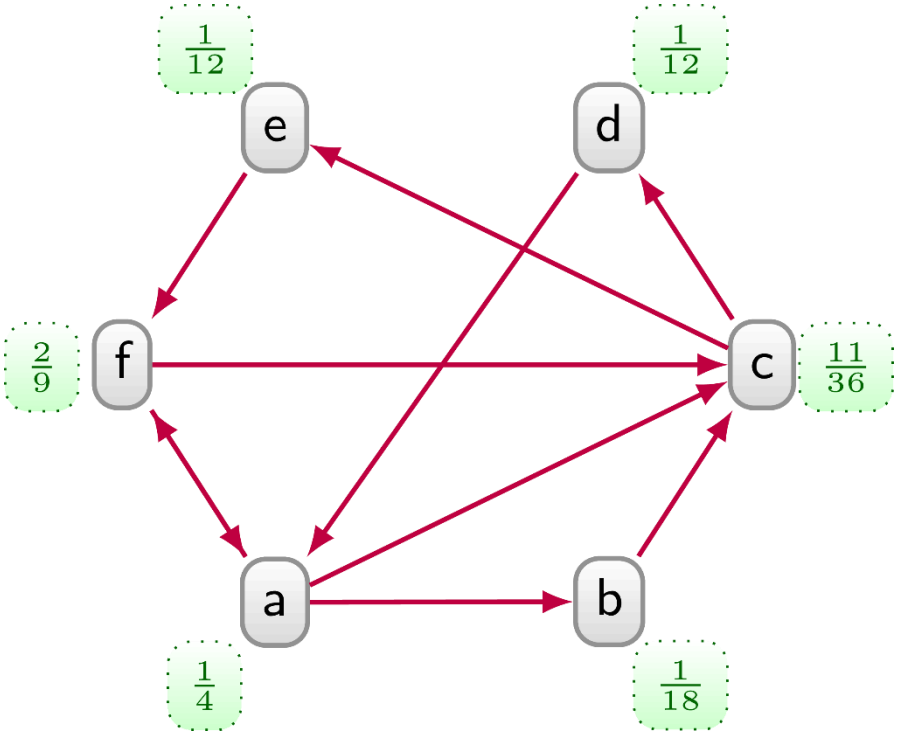
Example: PageRank (simplified)

$\mathrm{PageRank}(G)$:



Example: PageRank (simplified)

$\text{PageRank}(G)$:



...

Other algorithms ...

- Shortest paths / path queries
- Clustering (k-means, label prop.)
- Inferencing (class/property hierarchies)
- Conway's game of life
- Centrality (PageRank, ...)
- Neural Networks
- Turing Machine ☺
- ...

What the framework offers ...

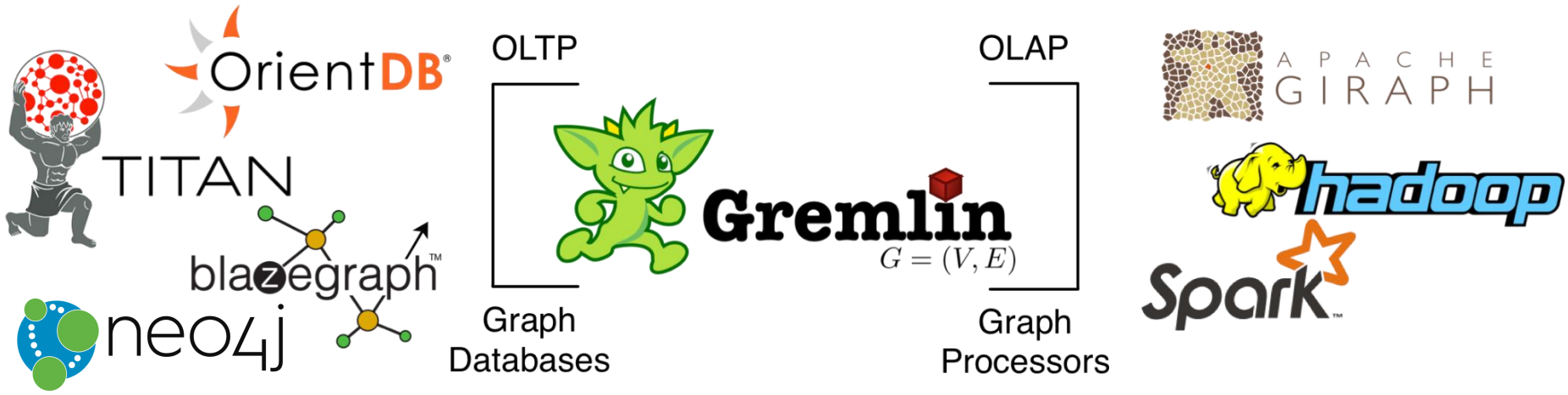
- Communication abstraction
- Distribution/parallel abstraction
- Topological abstraction

Ongoing Research …

- Pregel: A System for Large-Scale Graph Processing
  - Malewicz et al., SIGMOD 2010 [2938 citations]
- Shark: SQL and Rich Analytics at Scale
  - Xin et al. , SIGMOD 2013 [442 citations]
- GraphLab: A New Framework for Parallel Machine Learning
  - Low et al., arXiV 2014 [843 citations]
- GraphX: graph processing in a distributed dataflow framework
  - Gonzalez et al., OSDI 2014 [538 citations]
- GraphFrames: An Integrated API for Mixing Graph and Relational Queries
  - Dave et al., GRADES 2016 [25 citations]
- Signal/Collect
  - Stutz et al., SWJ 2018 [0 citations]

/GRAPHS

Gremlin: Graph Queries + Processing



Questions?