A Little Graph Theory for the Busy Developer

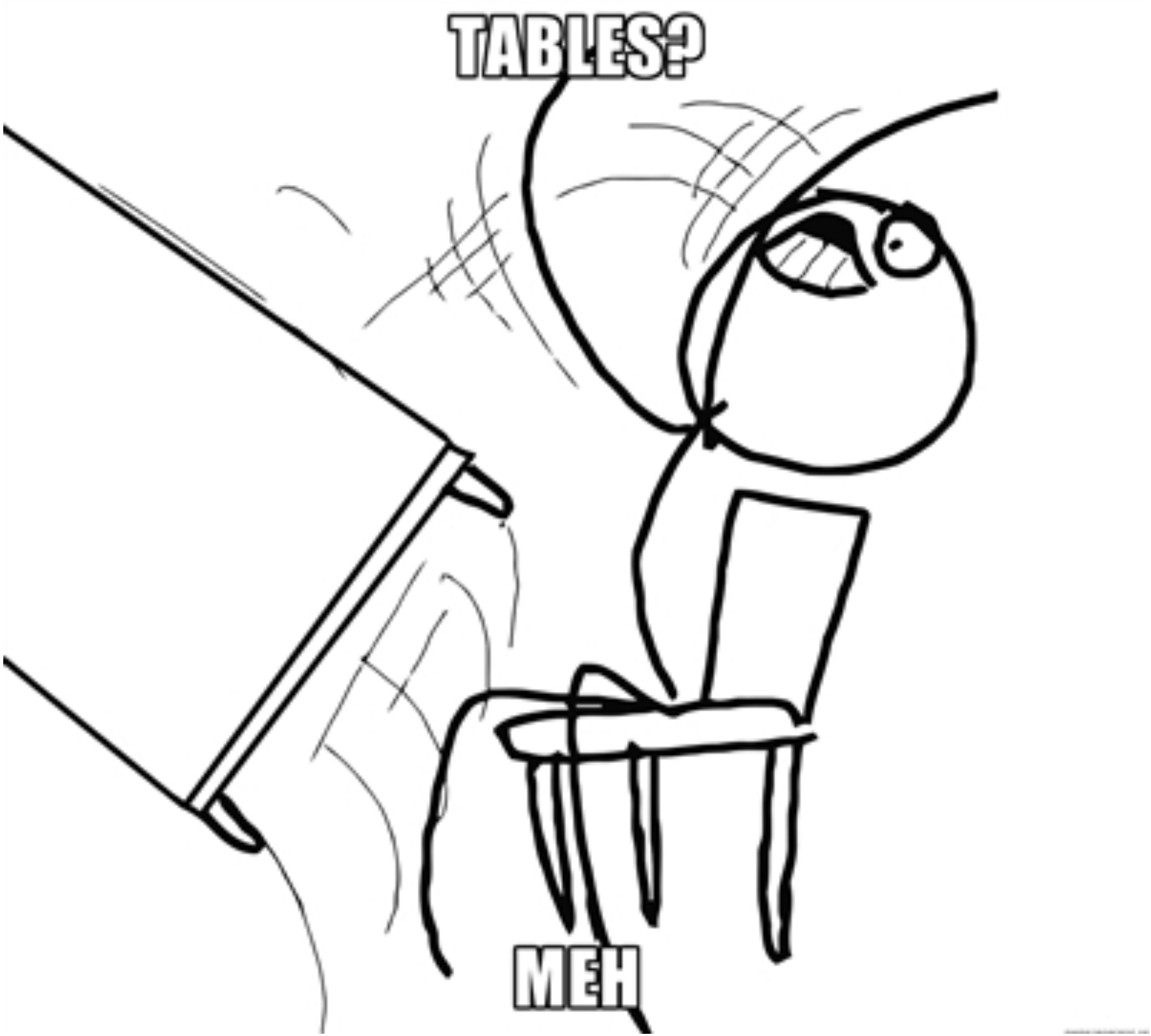
Dr. Jim Webber

Chief Scientist, Neo Technology

@jimwebber

Roadmap

- Imprisoned data
- Labeled Property Graph Model
  - And some cultural imperialism
- Graph theory
  - South East London
  - World War I
- Graph matching
  - Beer, nappies and Xbox
- End


TABLES?
MEH

http://www.flickr.com/photos/crazyneighborlady/355232758/

http://gallery.nen.gov.uk/image82582-.html

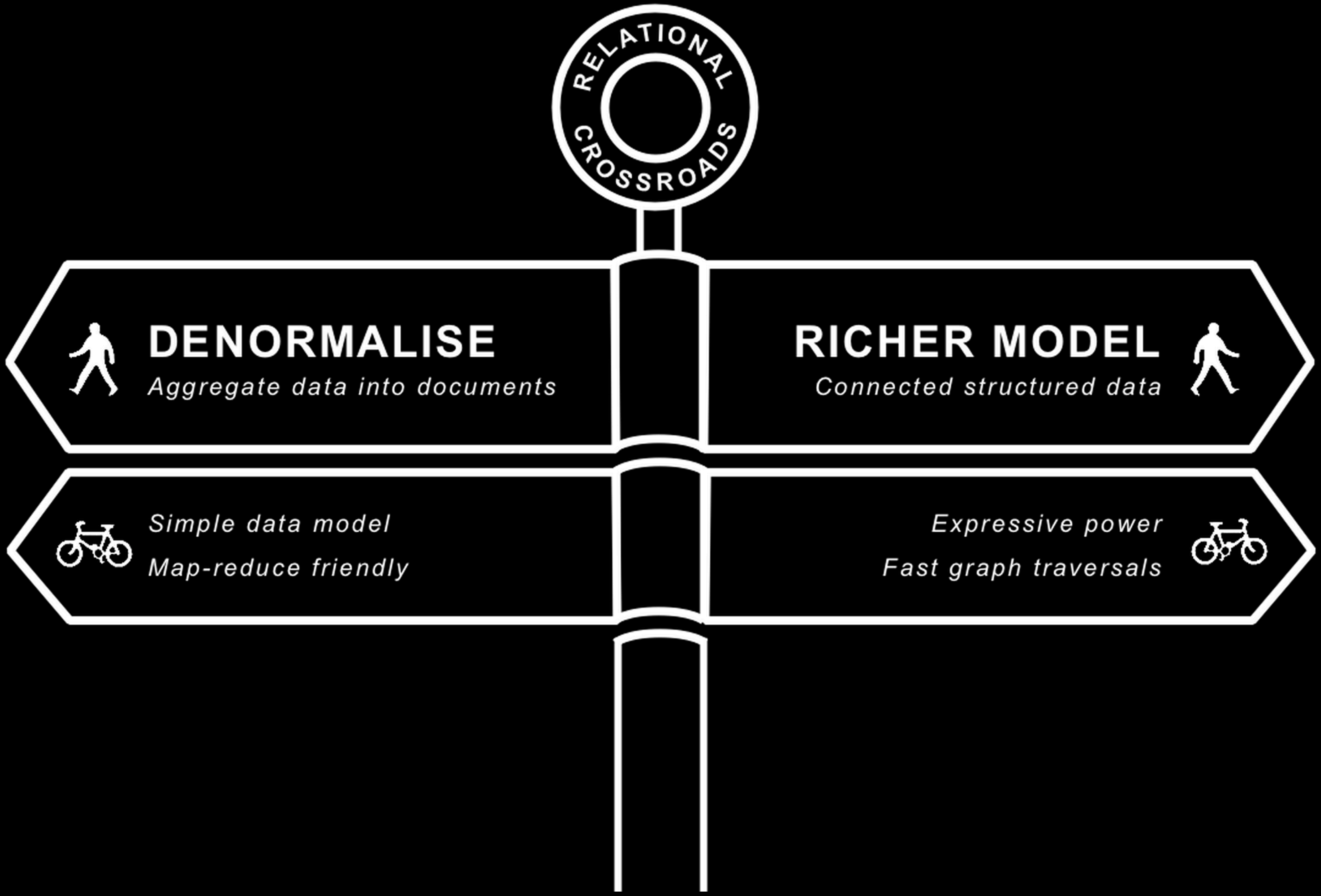
Aggregate-Oriented Data

http://martinfowler.com/bliki/AggregateOrientedDatabase.html

"There is a significant downside - the whole approach works really well when data access is aligned with the aggregates, but what if you want to look at the data in a different way? Order entry naturally stores orders as aggregates, but analyzing product sales cuts across the aggregate structure. The advantage of not using an aggregate structure in the database is that it allows you to slice and dice your data different ways for different audiences.

This is why aggregate-oriented stores talk so much about map-reduce."

complexity = f(size, connectedness, uniformity)



RELATIONAL CROSSROADS

DENORMALISE

Aggregate data into documents

Simple data model

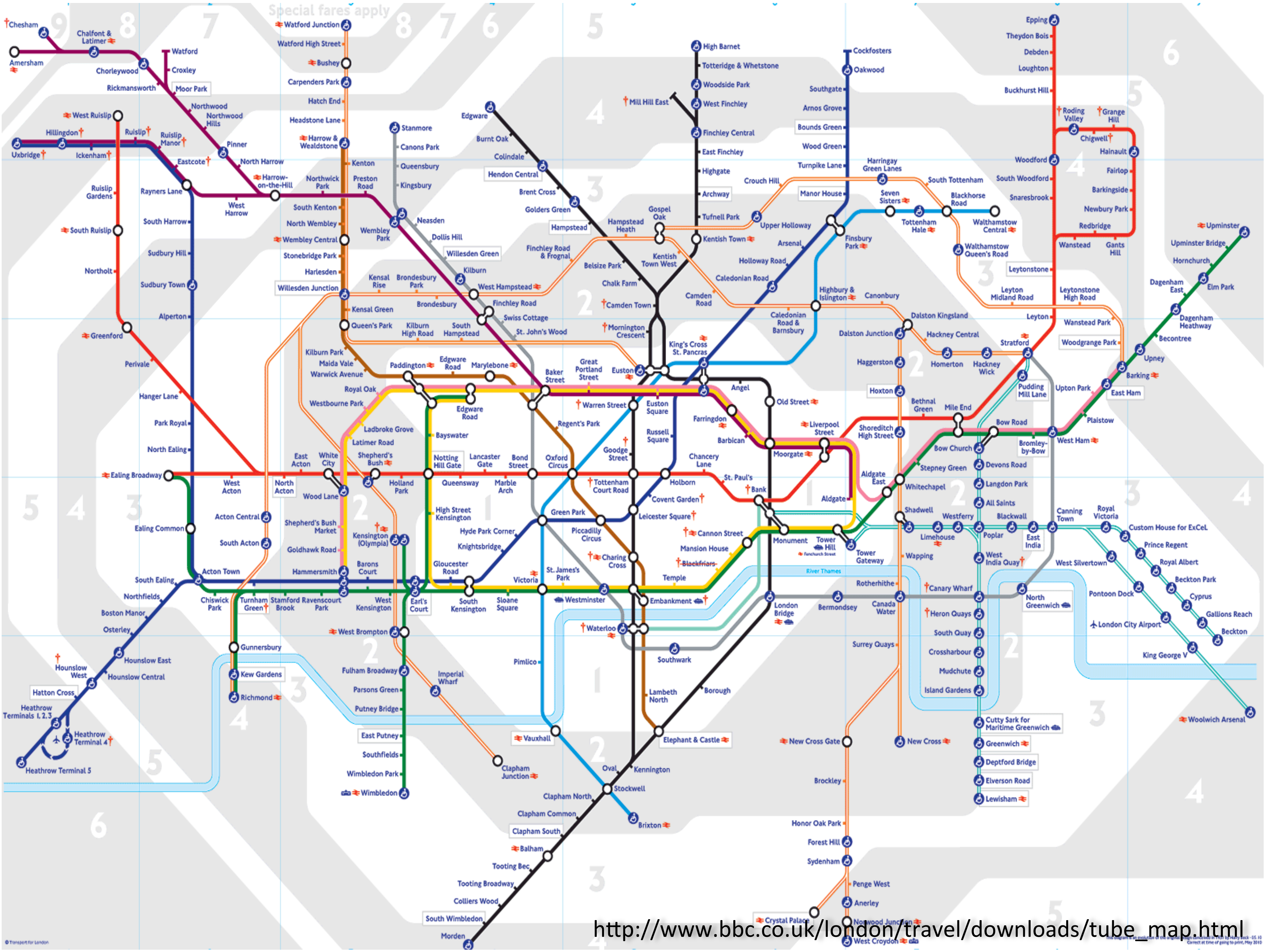
Map-reduce friendly

RICHER MODEL

Connected structured data

Expressive power

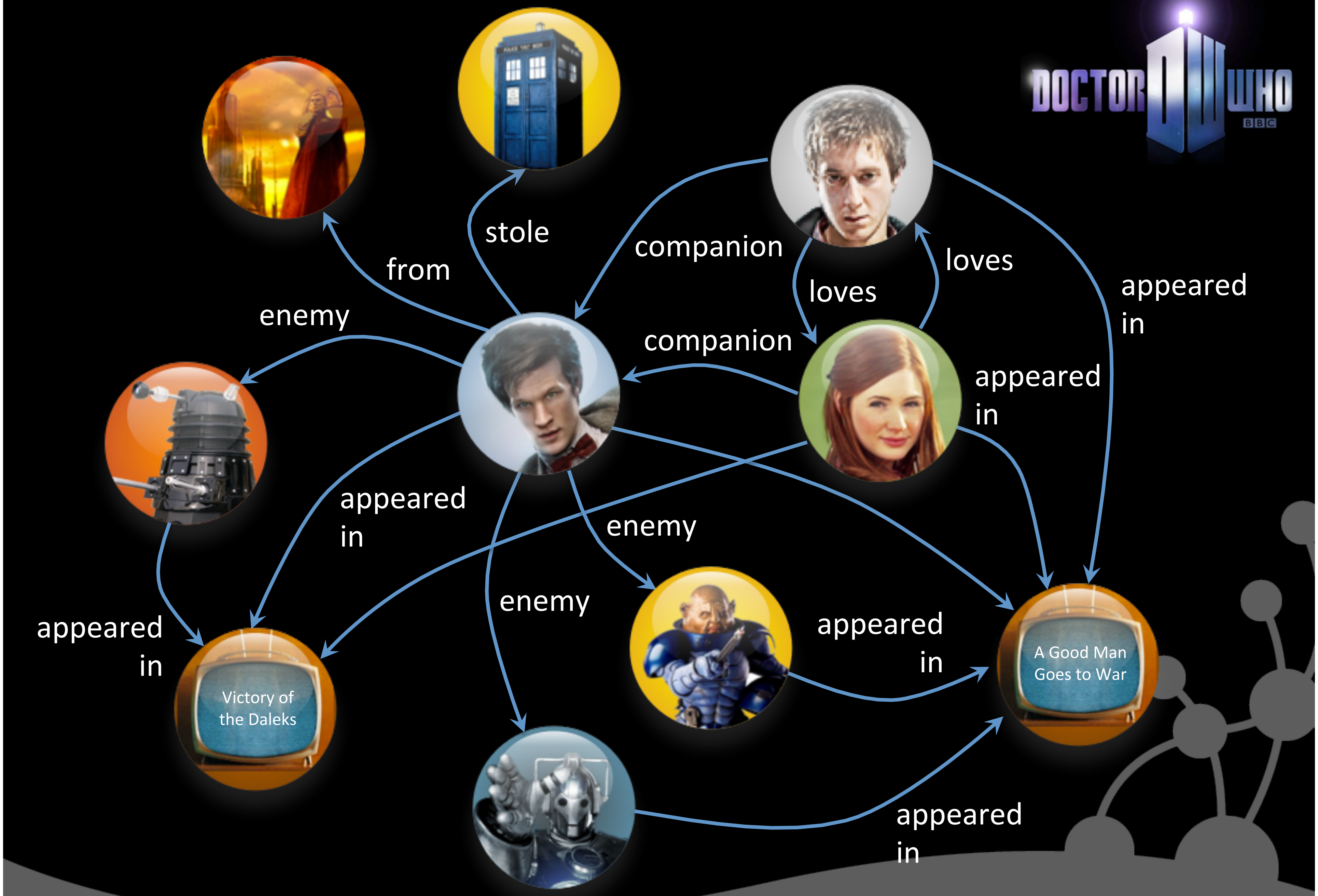
Fast graph traversals



http://www.bbc.co.uk/london/travel/downloads/tube_map.html

Labeled Property graphs

- Property graph model:
  - – Nodes with properties and labels
  - – Named, directed relationships with properties
  - – Relationships have exactly one start and end node
    - Which may be the same node



stole

from

companion

loves

loves

enemy

companion

appeared in

appeared in

appeared in

enemy

appeared in

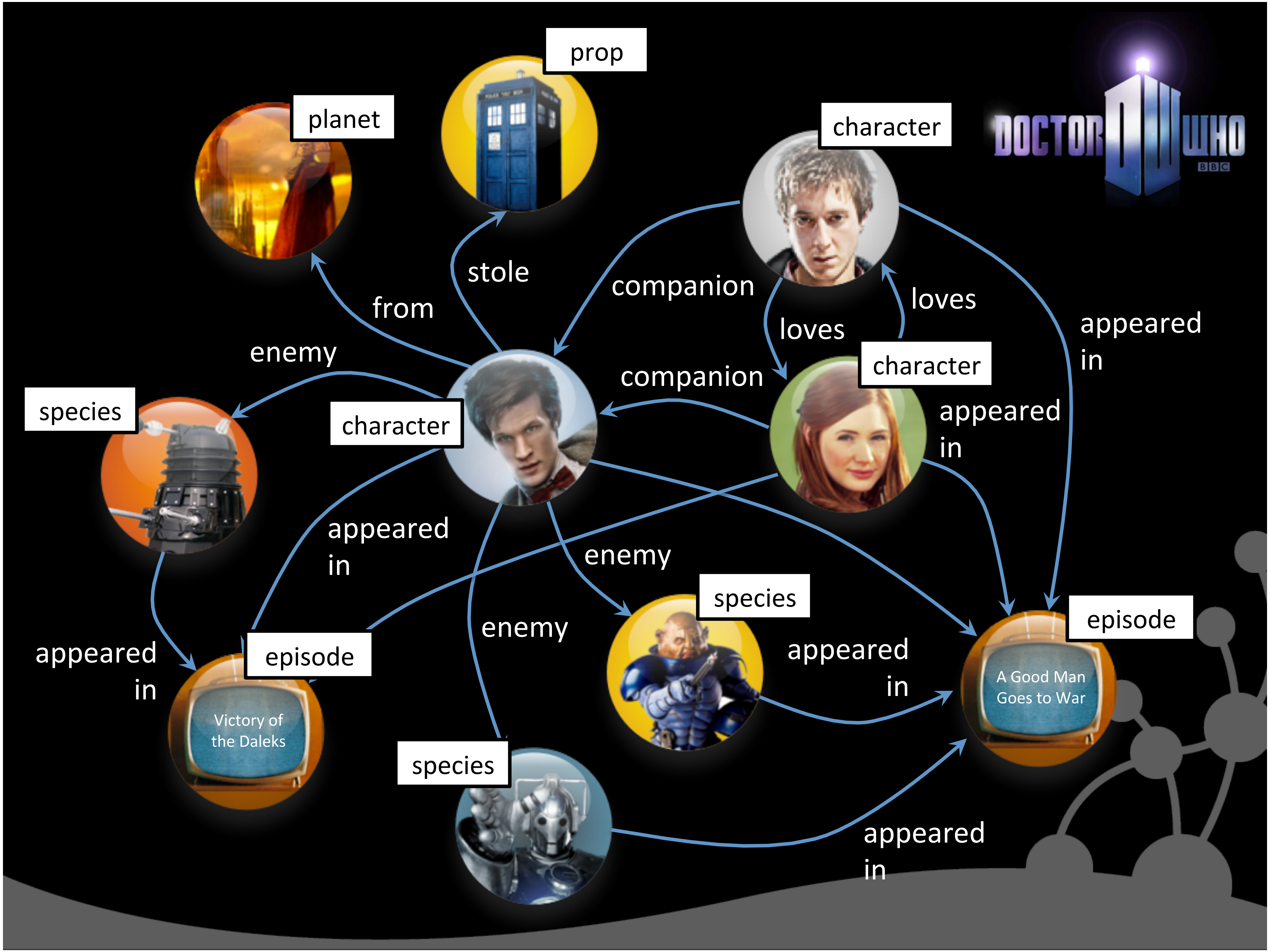
enemy

appeared in

Victory of the Daleks

A Good Man Goes to War

appeared in



planet

prop

character

stole

from

companion

loves

loves

appeared in

enemy

species

character

companion

character

appeared in

appeared in

enemy

species

enemy

appeared in

appeared in

episode

Victory of the Daleks

episode

A Good Man Goes to War

species

appeared in

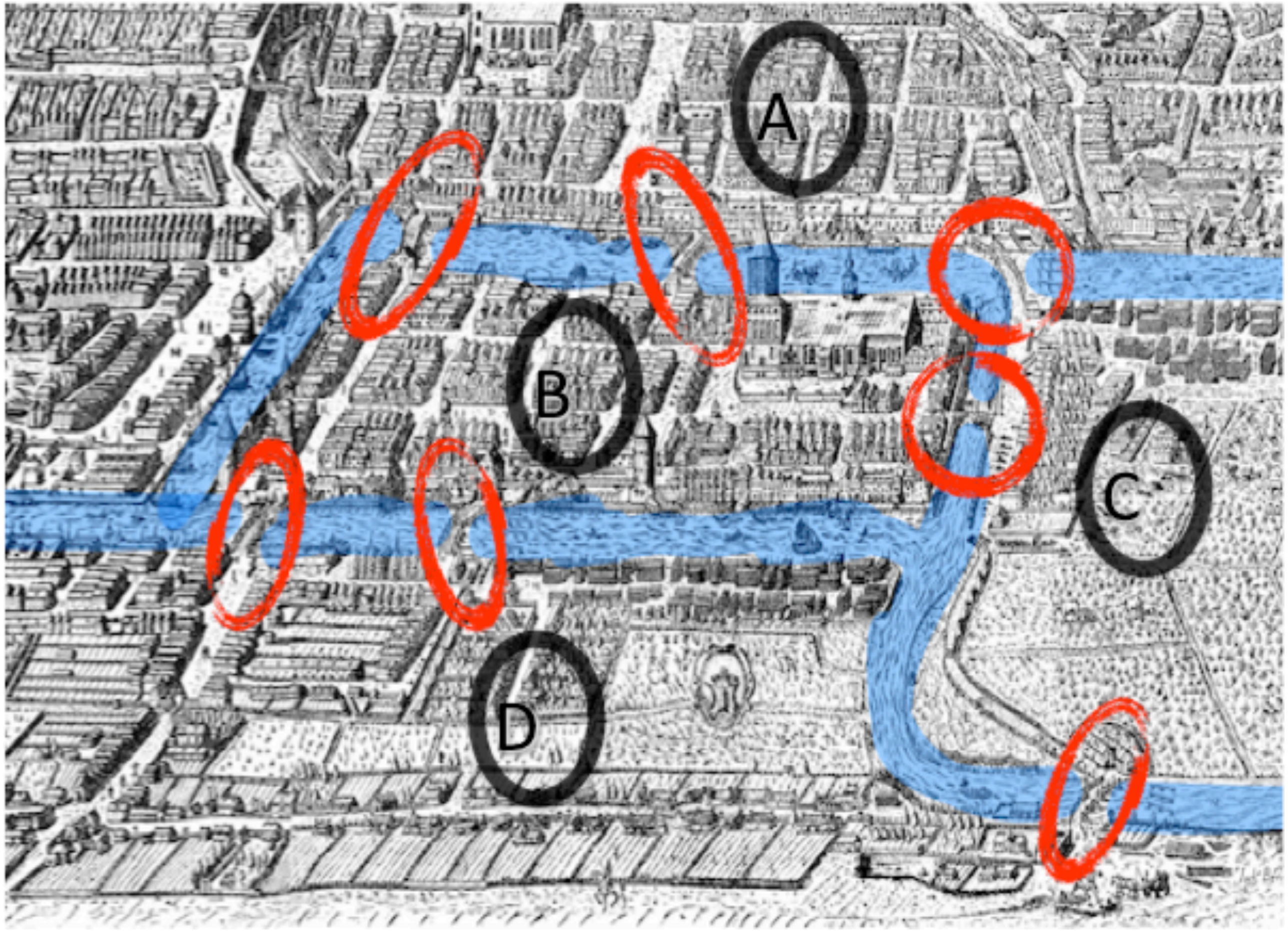
http://blogs.adobe.com/digitalmarketing/analytics/predictive-analytics/predictive-analytics-and-the-digital-marketer/

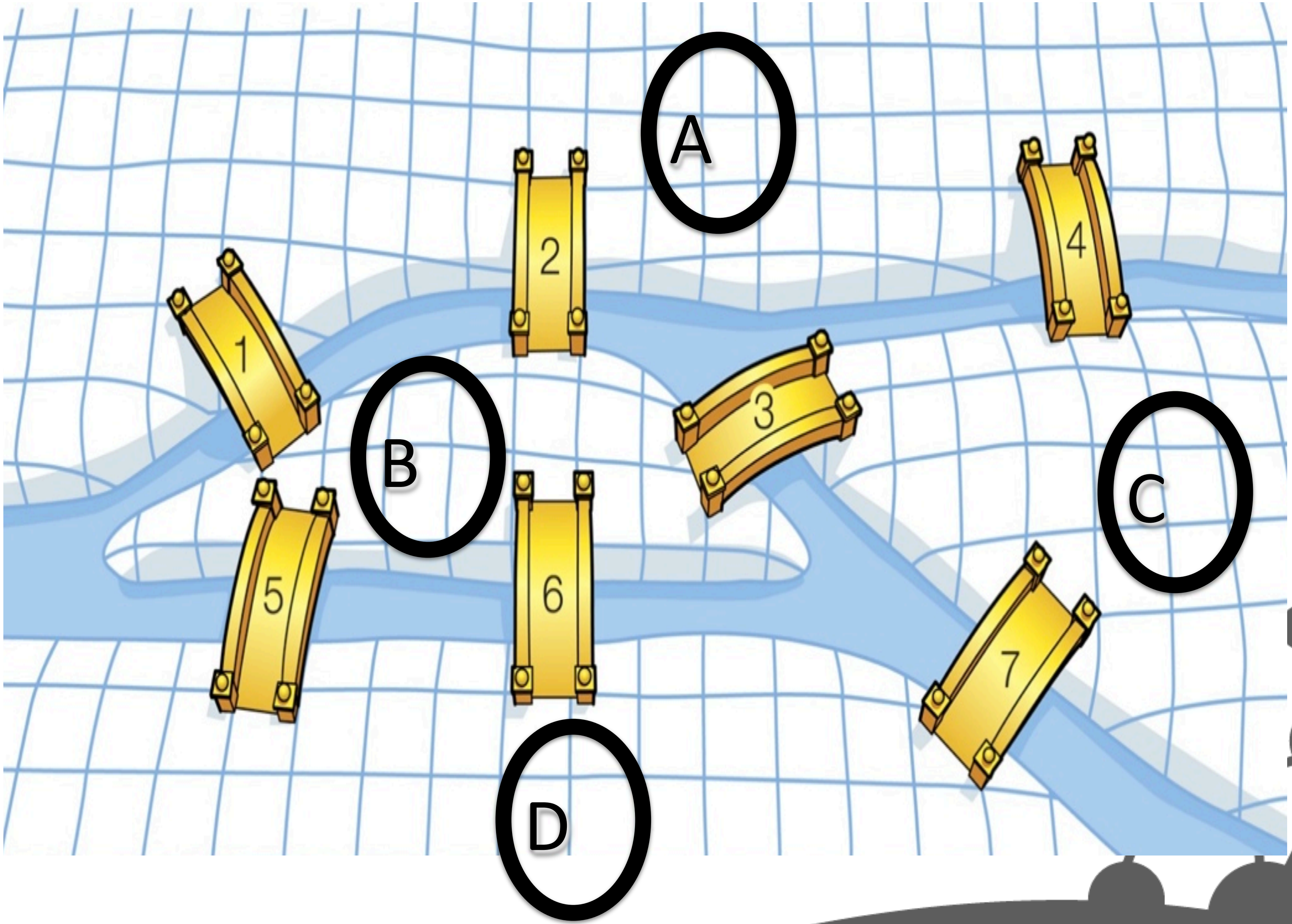
Meet Leonhard Euler

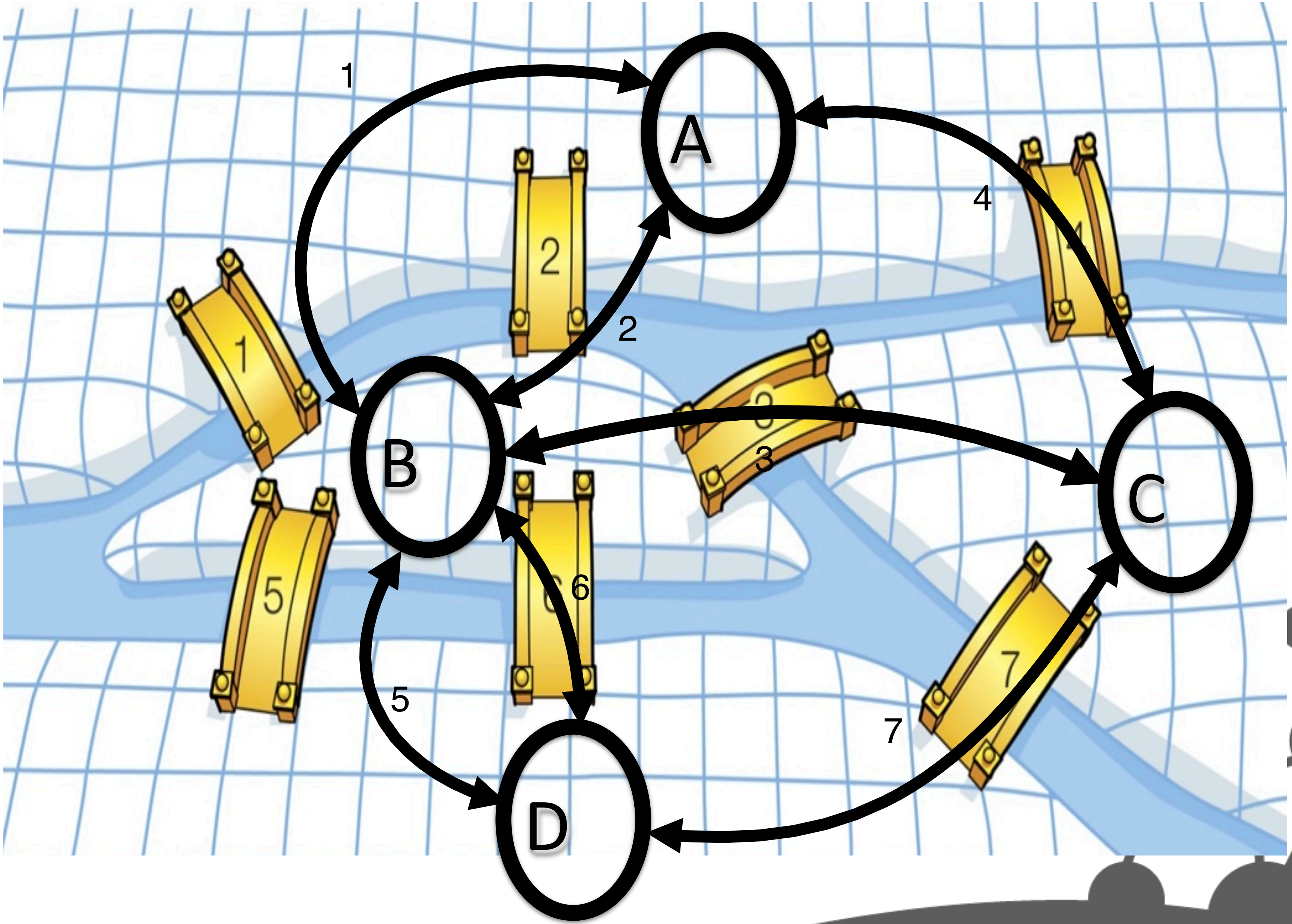
- Swiss mathematician
- Inventor of Graph Theory (1736)

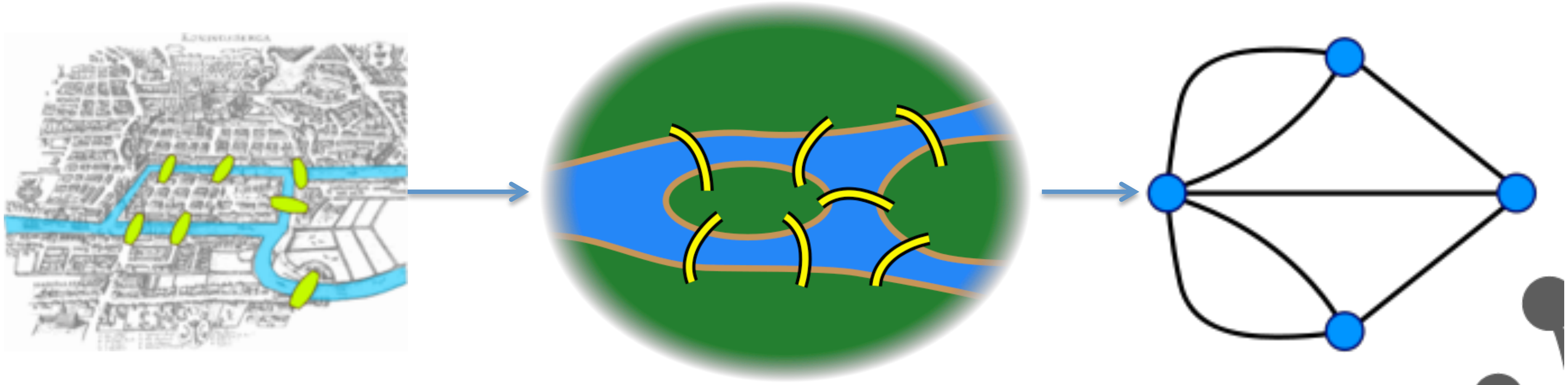
http://en.wikipedia.org/wiki/File:Leonhard_Euler_2.jpg

Königsberg (Prussia) - 1736

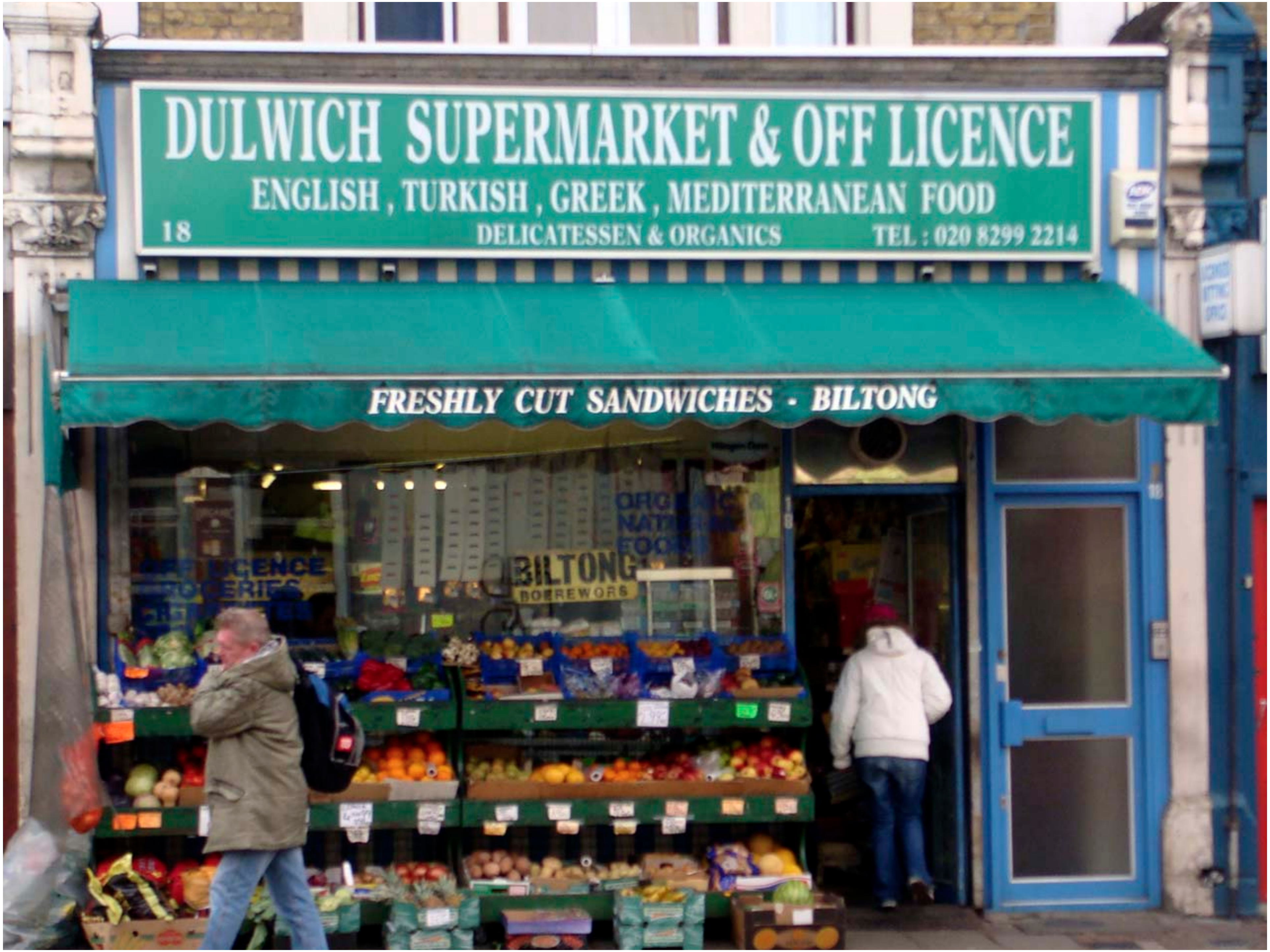



A
2
4
1
3
B
C
5
6
7
D

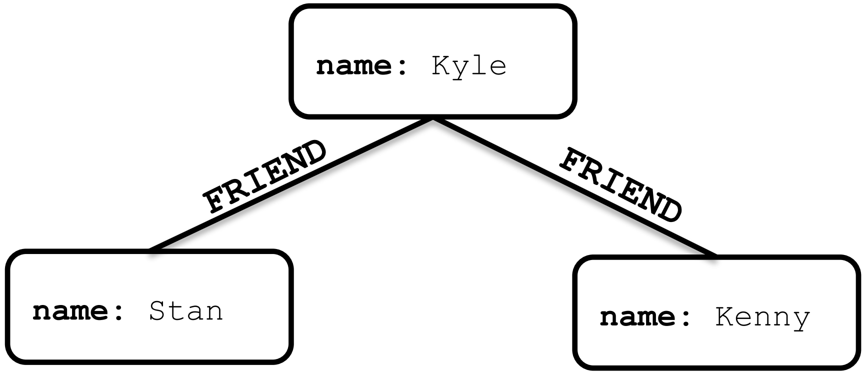

1
A
2
4
1
2
3
B
C
3
5
6
6
4
5
7
7
D



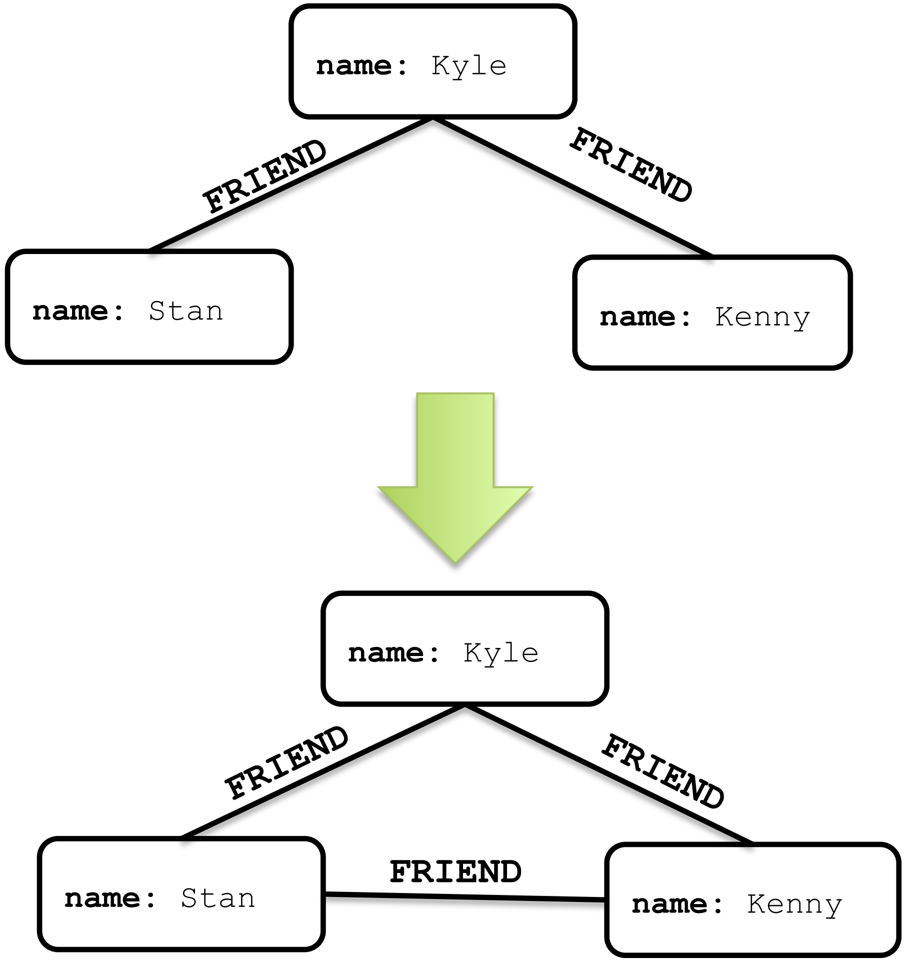
http://en.wikipedia.org/wiki/Seven_Bridges_of_Königsberg


DULWICH SUPERMARKET & OFF LICENCE
ENGLISH , TURKISH , GREEK , MEDITERRANEAN FOOD
18
DELICATESSEN & ORGANICS
TEL : 020 8299 2214
FRESHLY CUT SANDWICHES - BILTONG
BILTONG
BOEREWORS

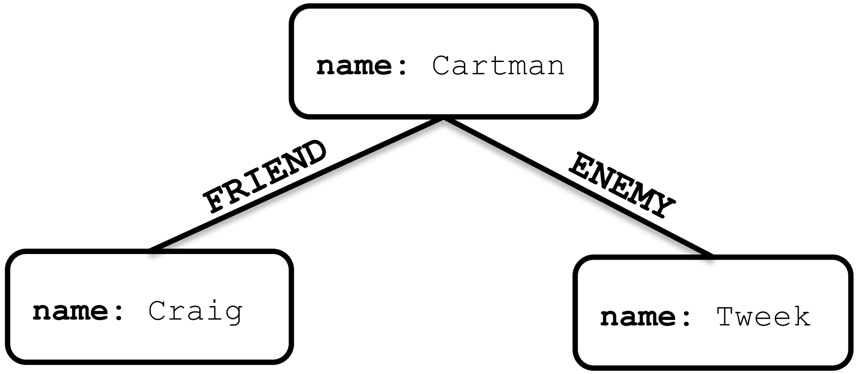
Triadic Closure



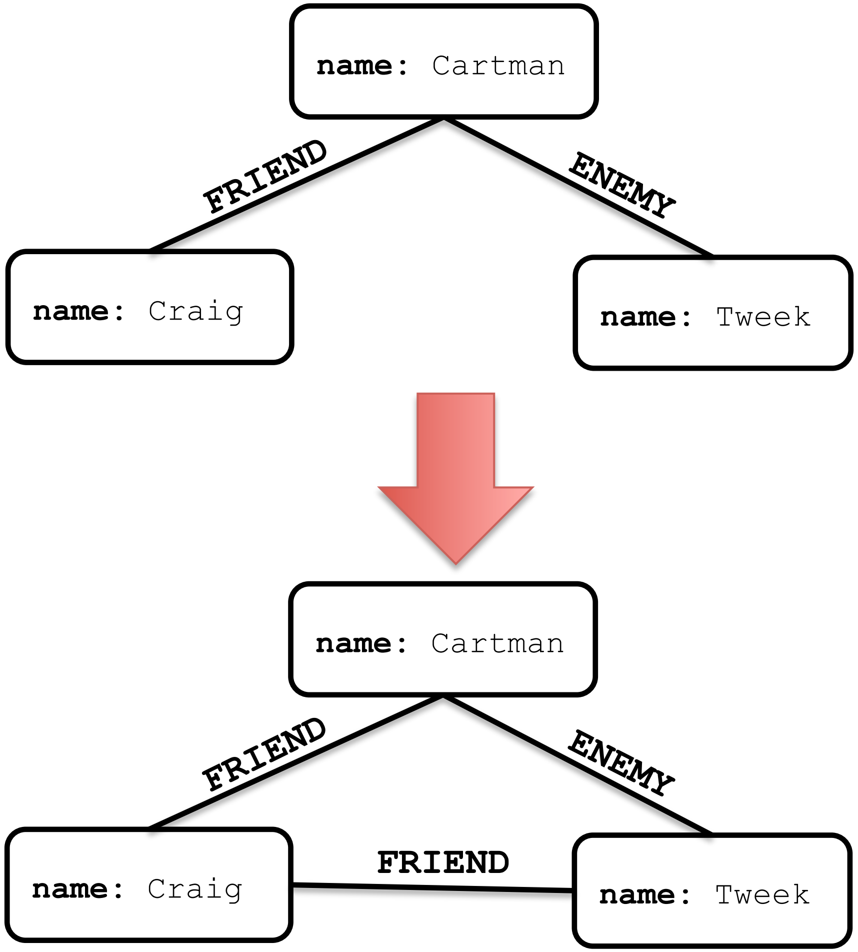
Triadic Closure



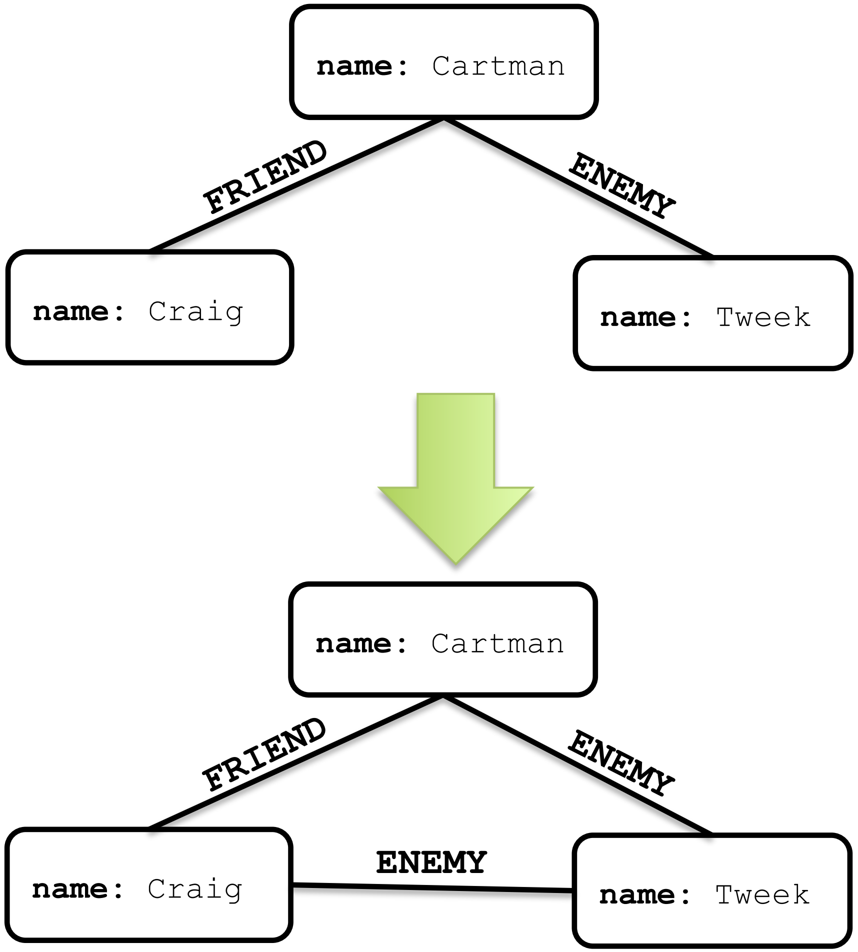
Structural Balance



Structural Balance



Structural Balance



name: Cartman

FRIEND

ENEMY

name: Craig

name: Tweek

name: Cartman

FRIEND

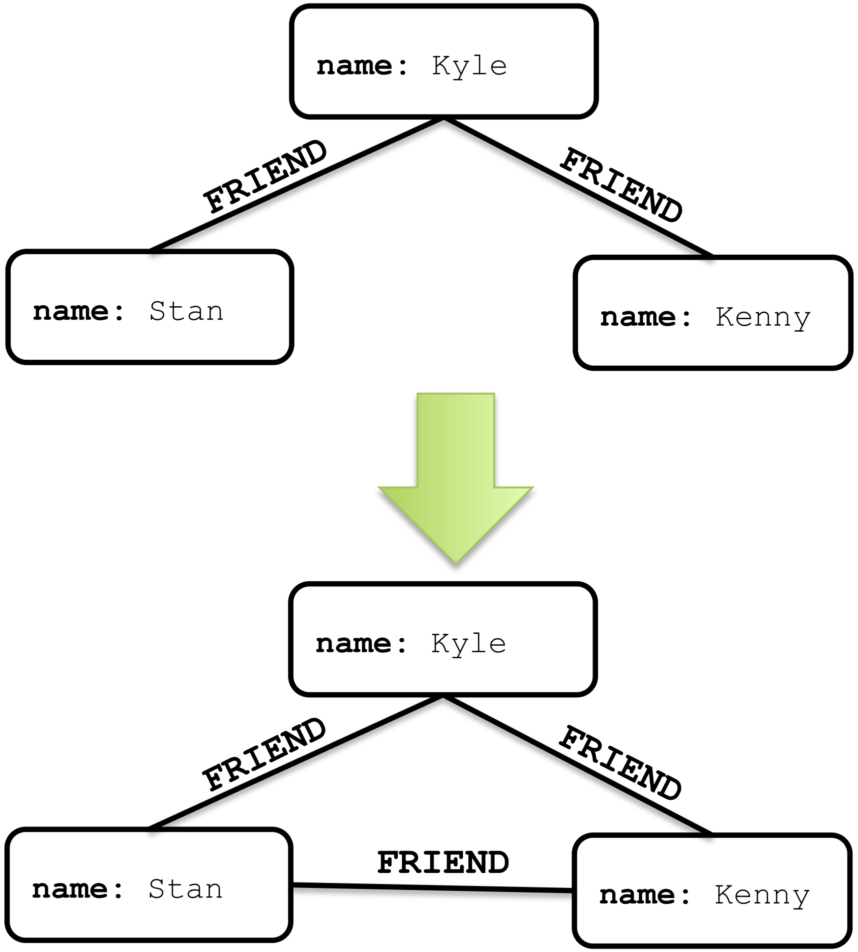
ENEMY

name: Craig

ENEMY

name: Tweek

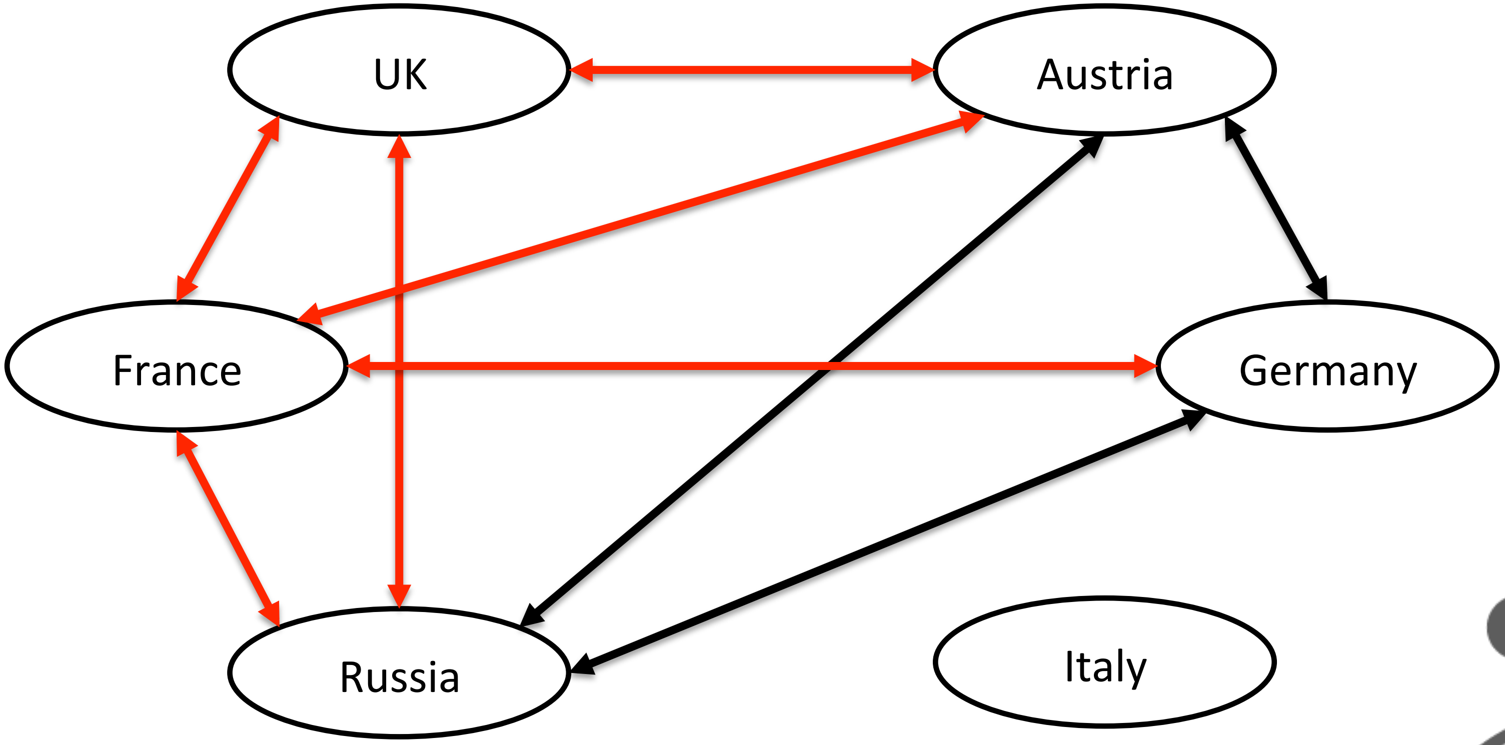
Structural Balance



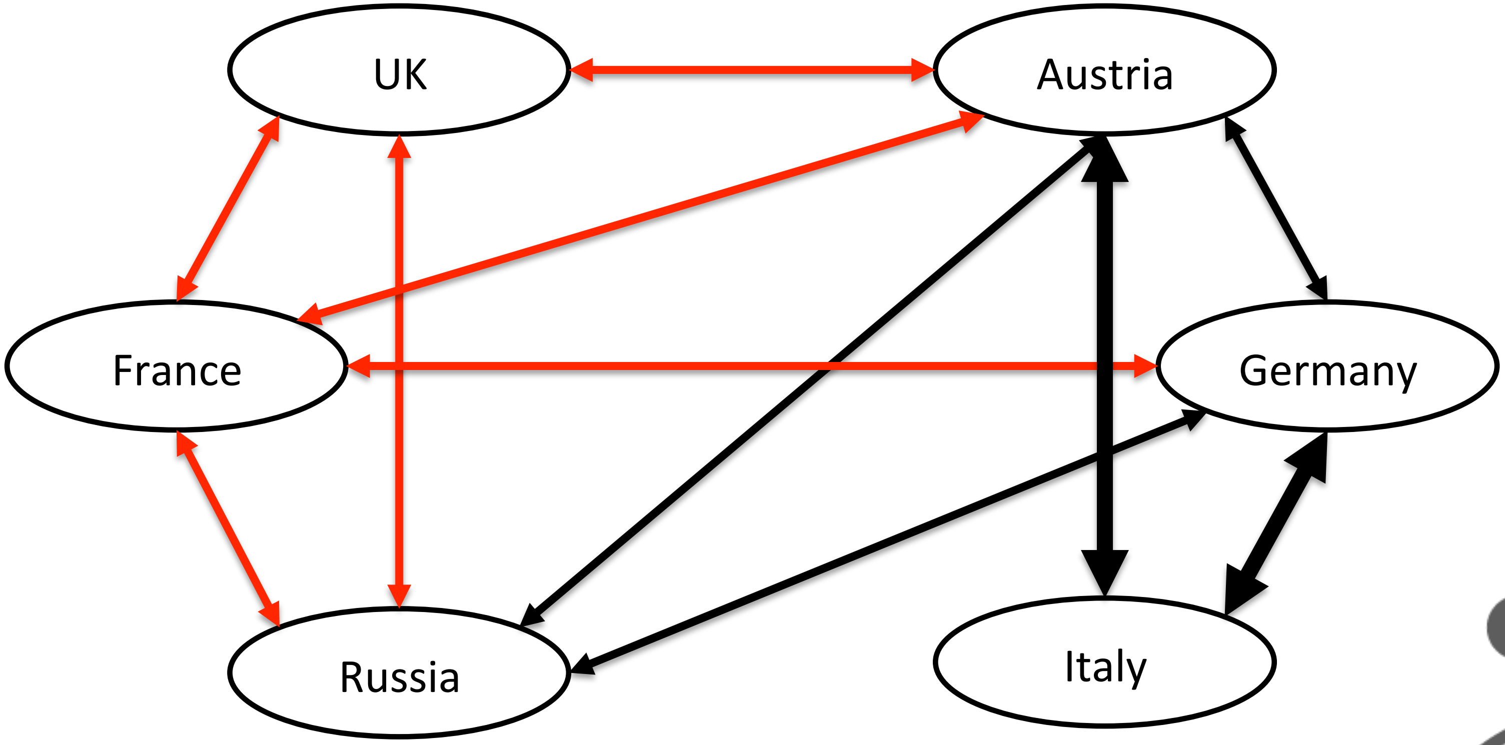
Structural Balance is a *key* predictive technique

And it's domain-agnostic

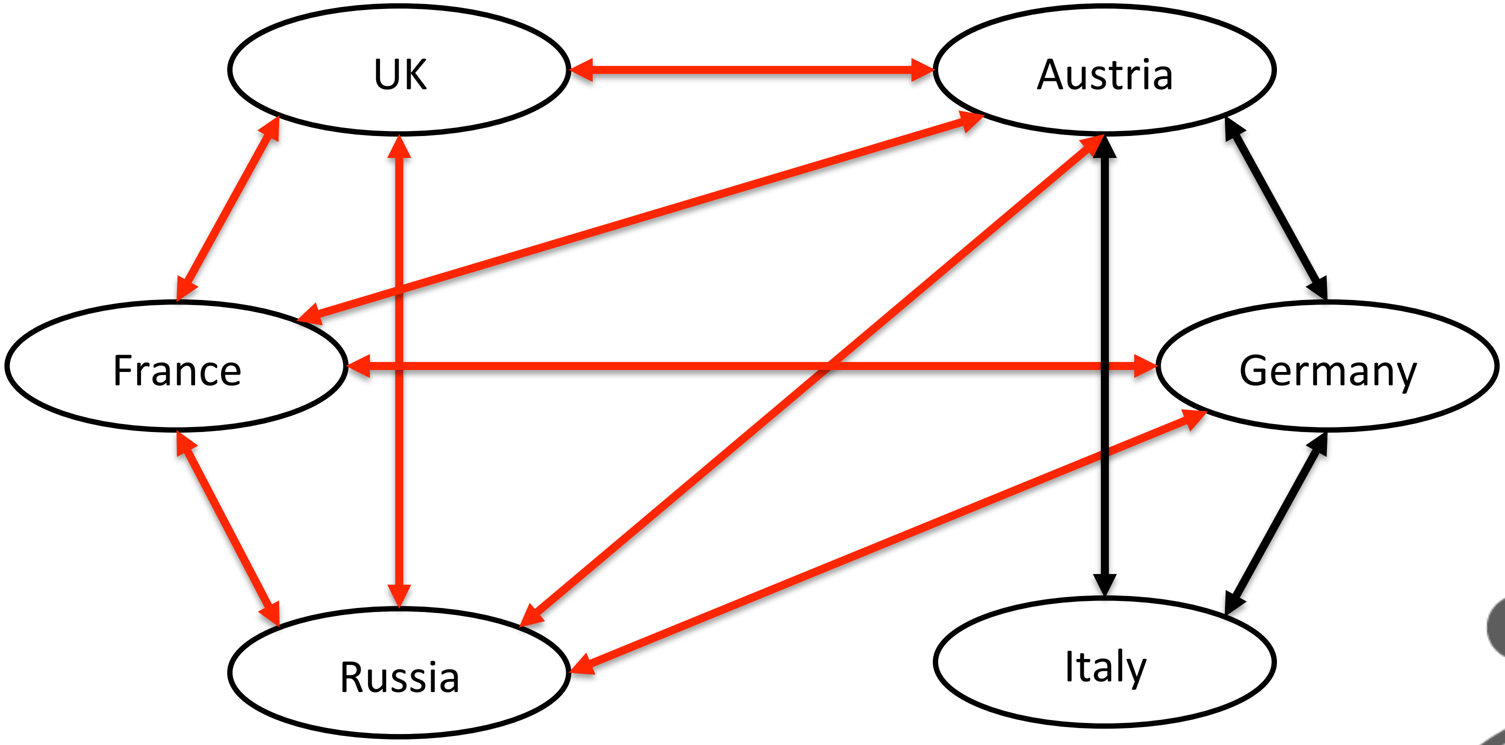
Allies and Enemies



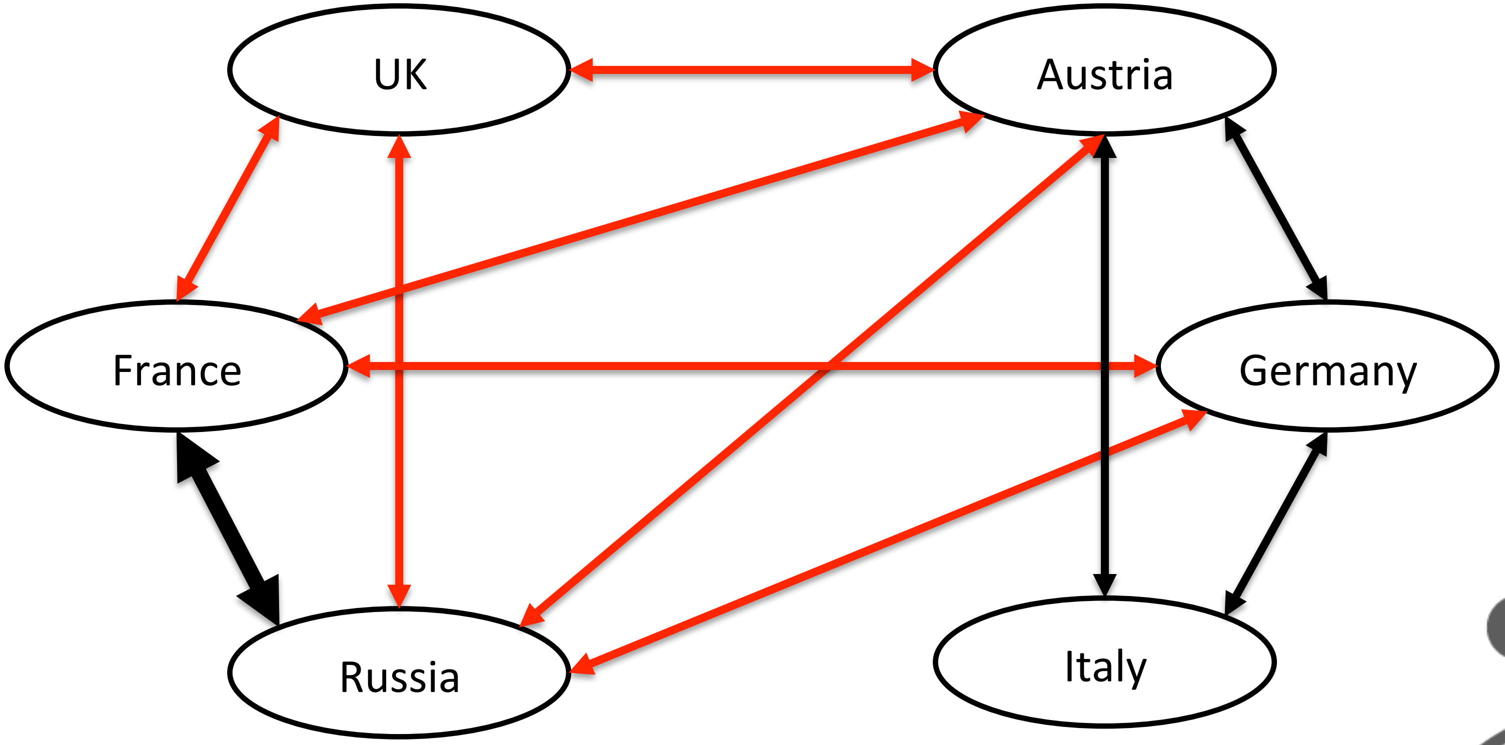
Allies and Enemies



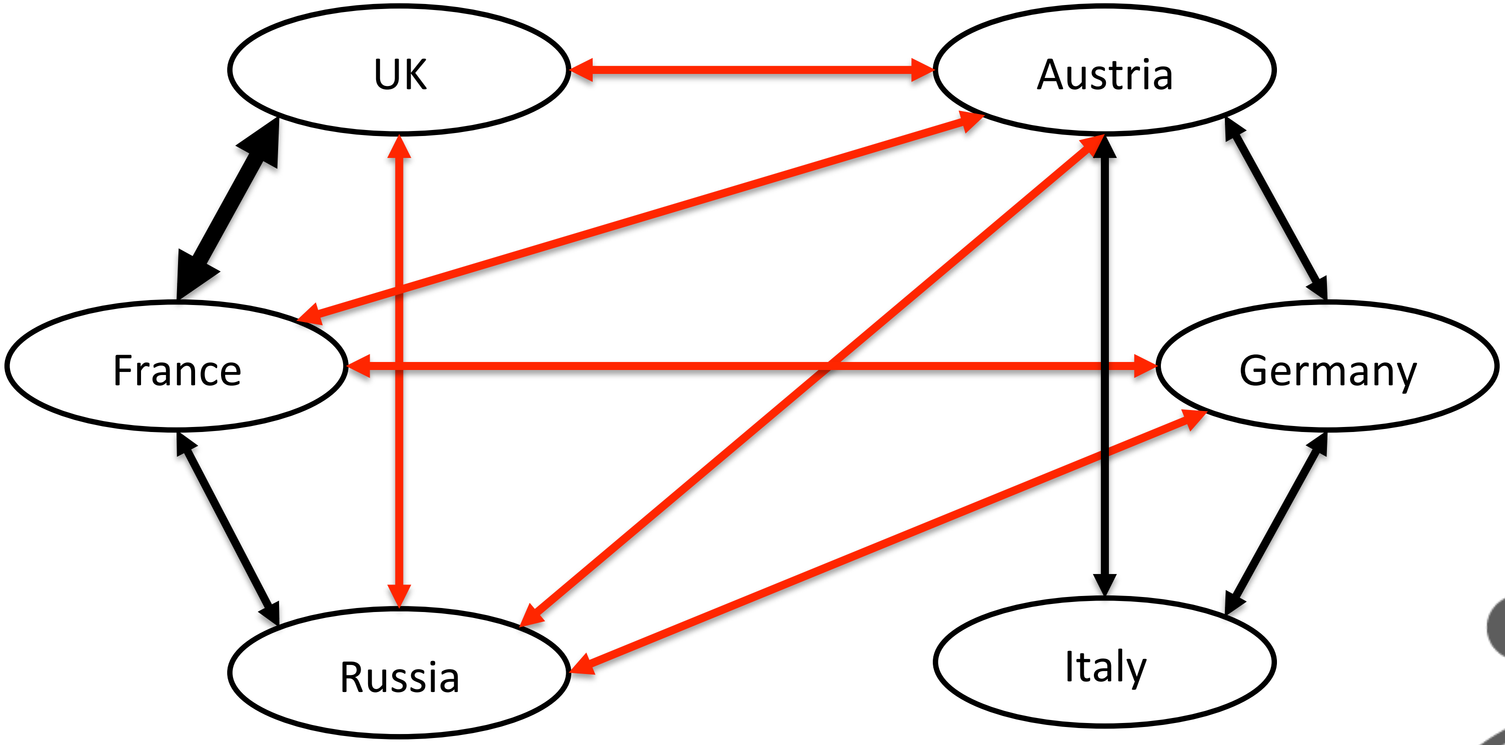
Allies and Enemies



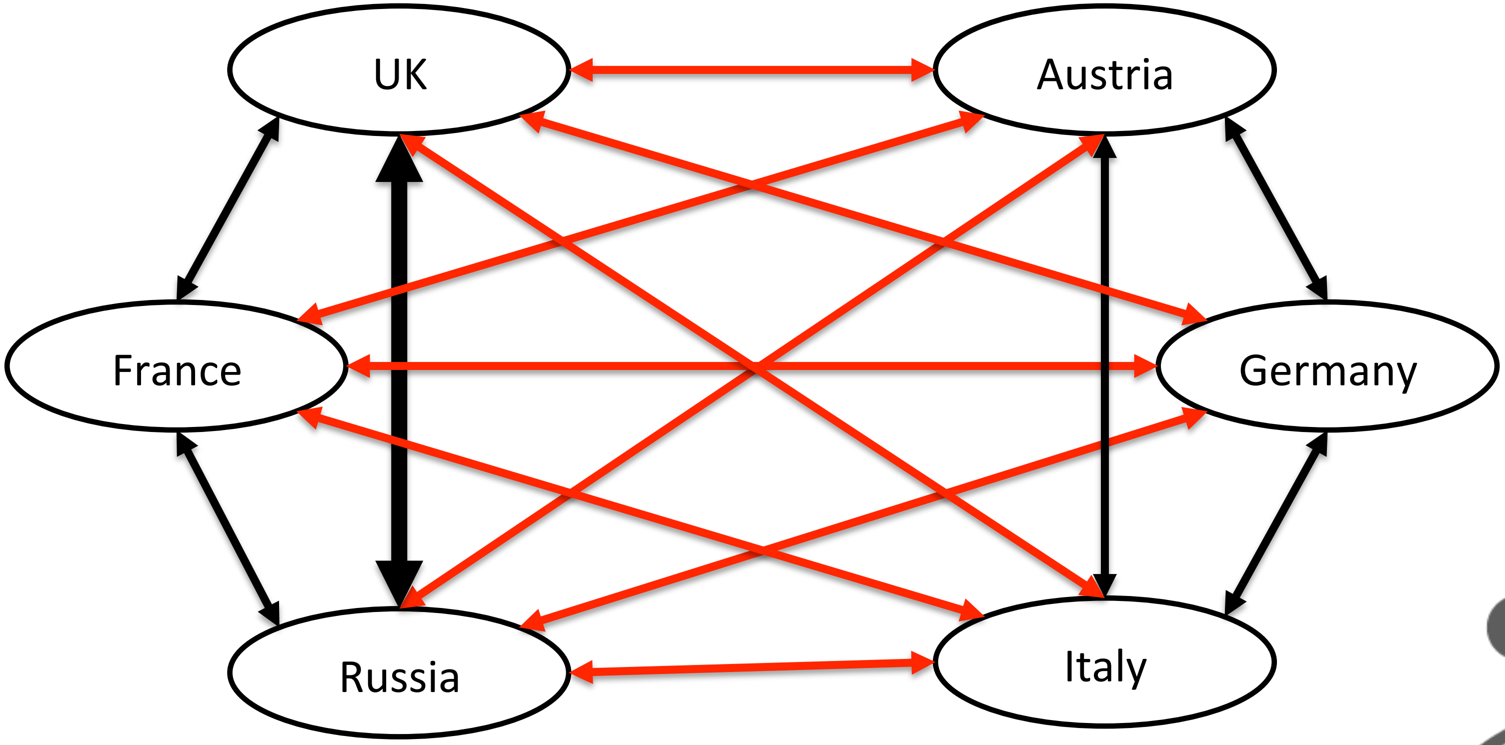
Allies and Enemies



Allies and Enemies

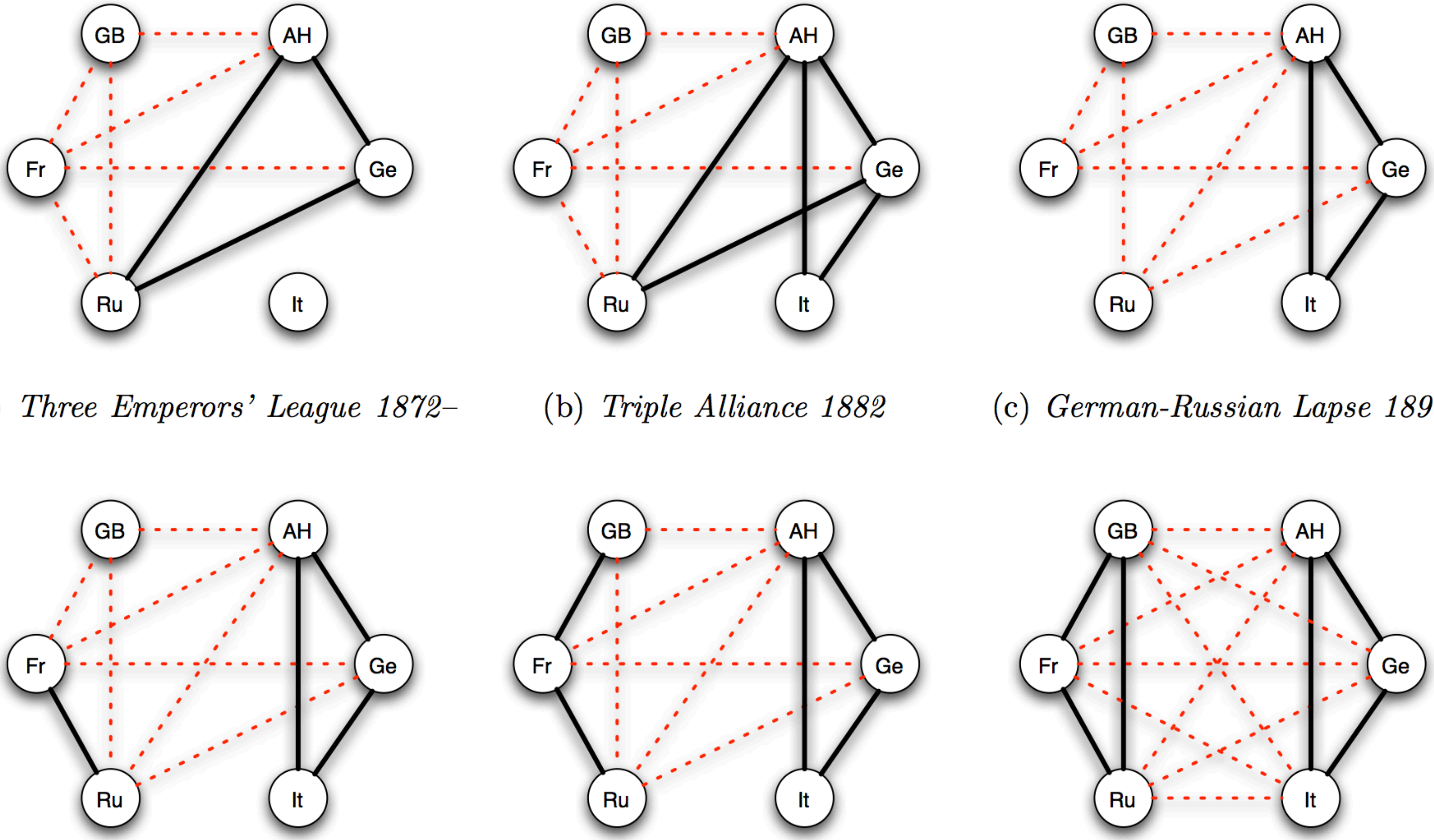


Allies and Enemies



Predicting WWI

[Easley and Kleinberg]



(a) *Three Emperors' League 1872–81*

(b) *Triple Alliance 1882*

(c) *German-Russian Lapse 1890*

(d) *French-Russian Alliance 1891–94*

(e) *Entente Cordiale 1904*

(f) *British Russian Alliance 1907*

Strength

- Relationships can have strength as well as sentiment
- This gives us another dimension to consider in our triangles
  - I **love** you versus I **hate** you
  - I **like** you versus I **dislike** you
- Gives rise to another interesting property

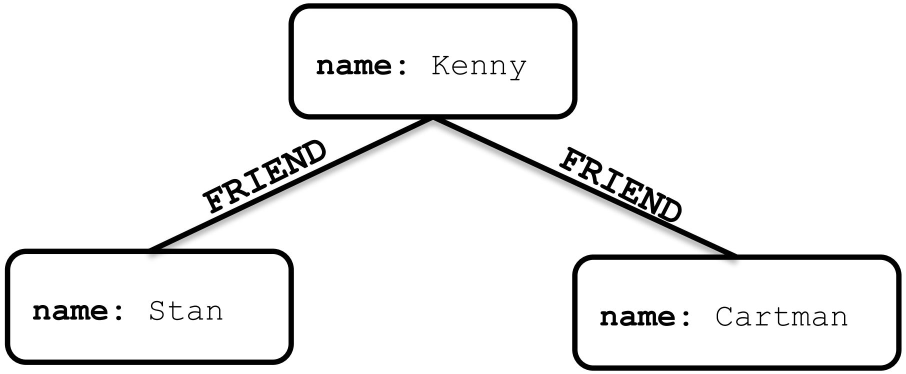
Strong Triadic Closure Property

It if a node has strong relationships to two neighbours, then these neighbours must have at least a weak relationship between them.

[Wikipedia]

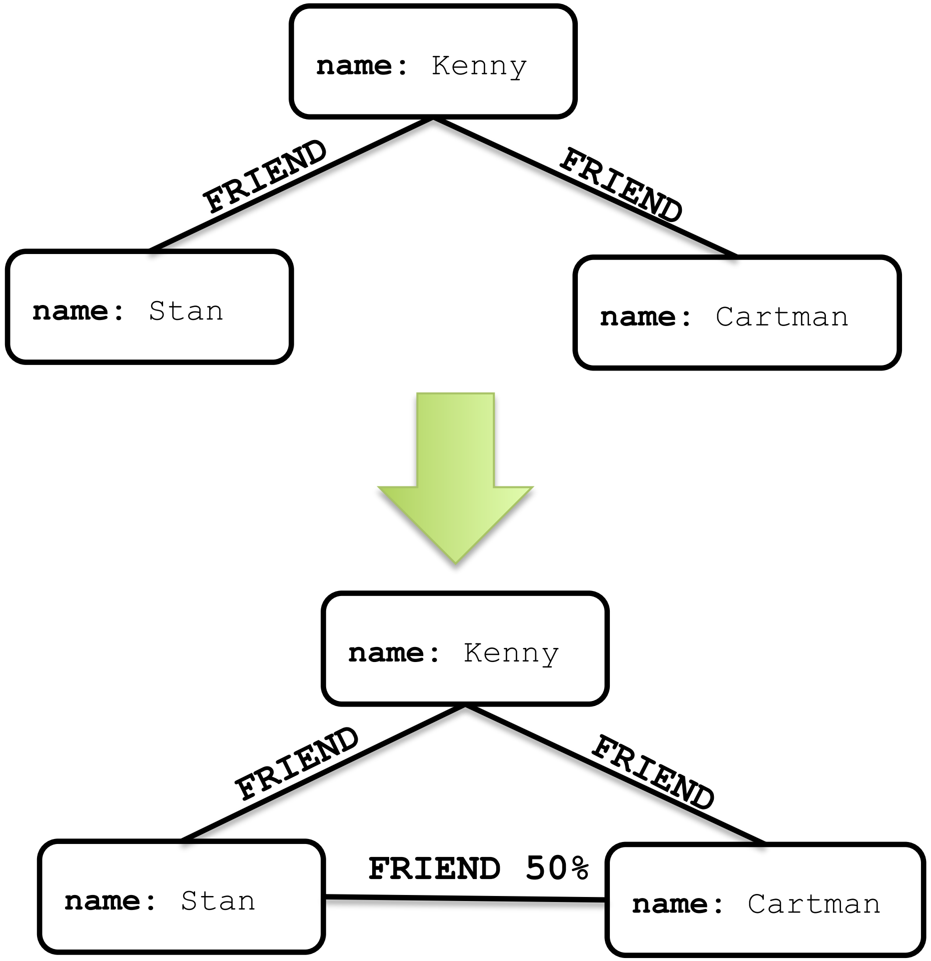
Triadic Closure

(weak relationship)



Triadic Closure

(weak relationship)



name: Kenny

FRIEND

FRIEND

name: Stan

name: Cartman

name: Kenny

FRIEND

FRIEND

name: Stan

FRIEND 50%

name: Cartman

Weak relationships

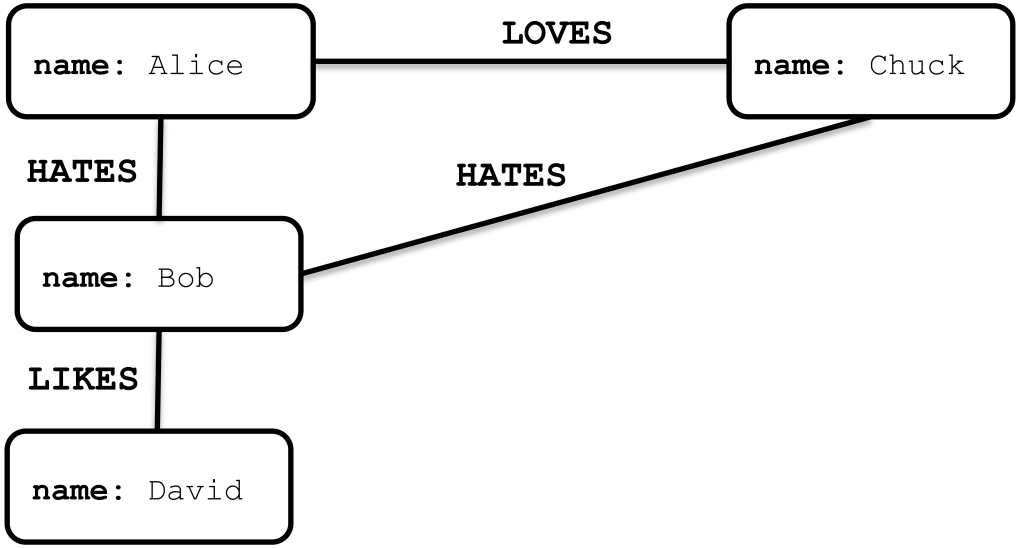
- Weak links play another **super-important** structural role in graph theory
- They bridge neighbourhoods
- Which is allows us to partition graphs

Local Bridge Property

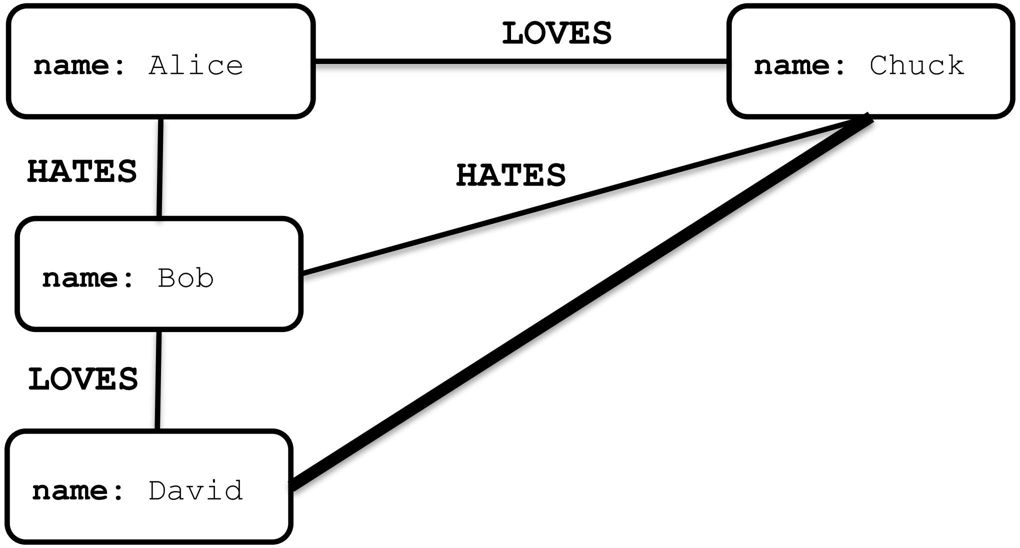
*“If a node **A** in a network satisfies the Strong Triadic Closure Property and is involved in at least two strong relationships, then any local bridge it is involved in must be a weak relationship.”*

[Easley and Kleinberg]

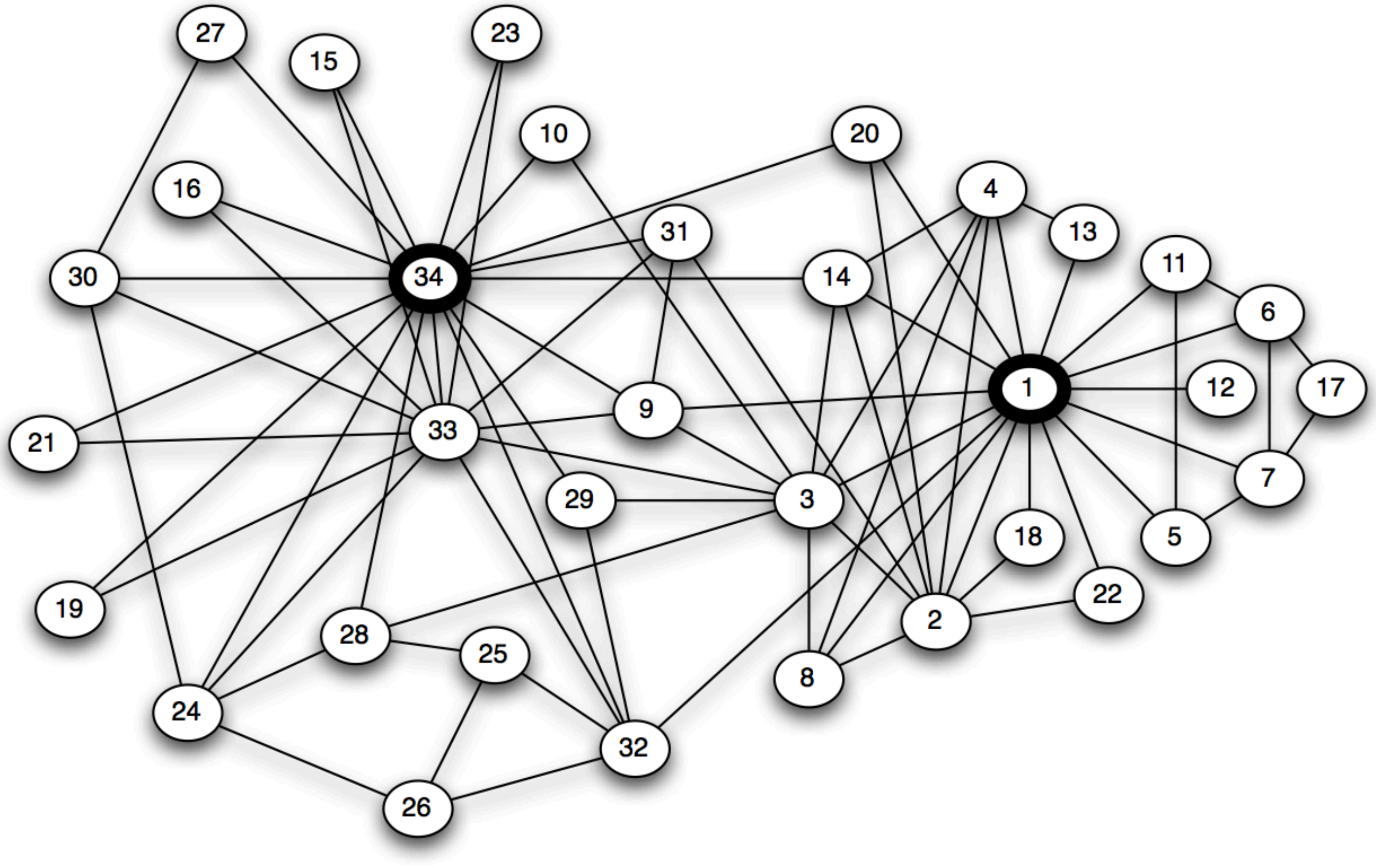
LIKES is a local bridge



LOVES cannot be a local bridge



University Karate Club

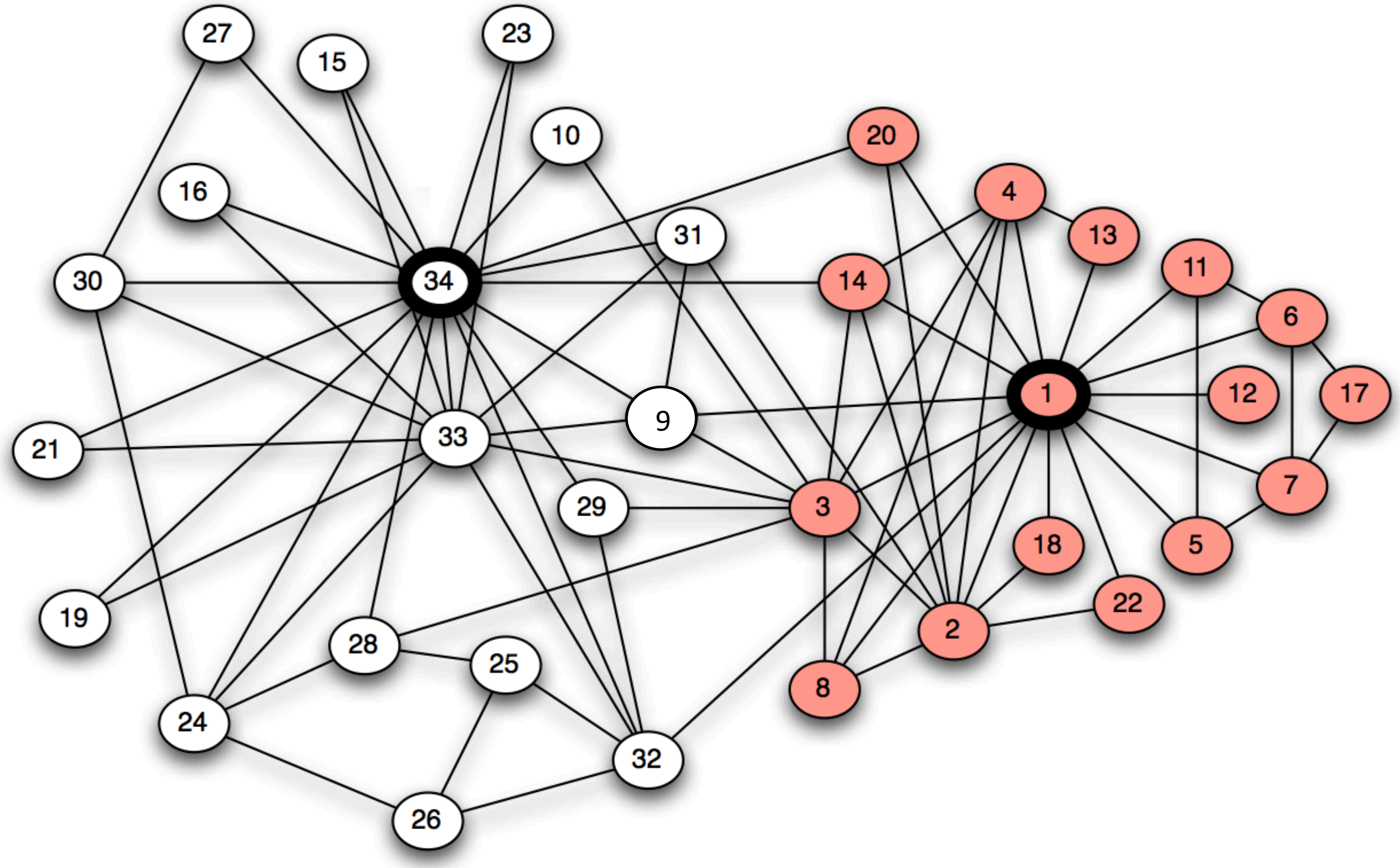


Graph Partitioning

- (NP) Hard problem
  - Recursively remove bridges between dense regions
  - Choose your algorithm carefully – some are better than others for a given domain
- Probabilistic methods are cheaper
- Can use to (almost exactly) predict the break up of the karate club!

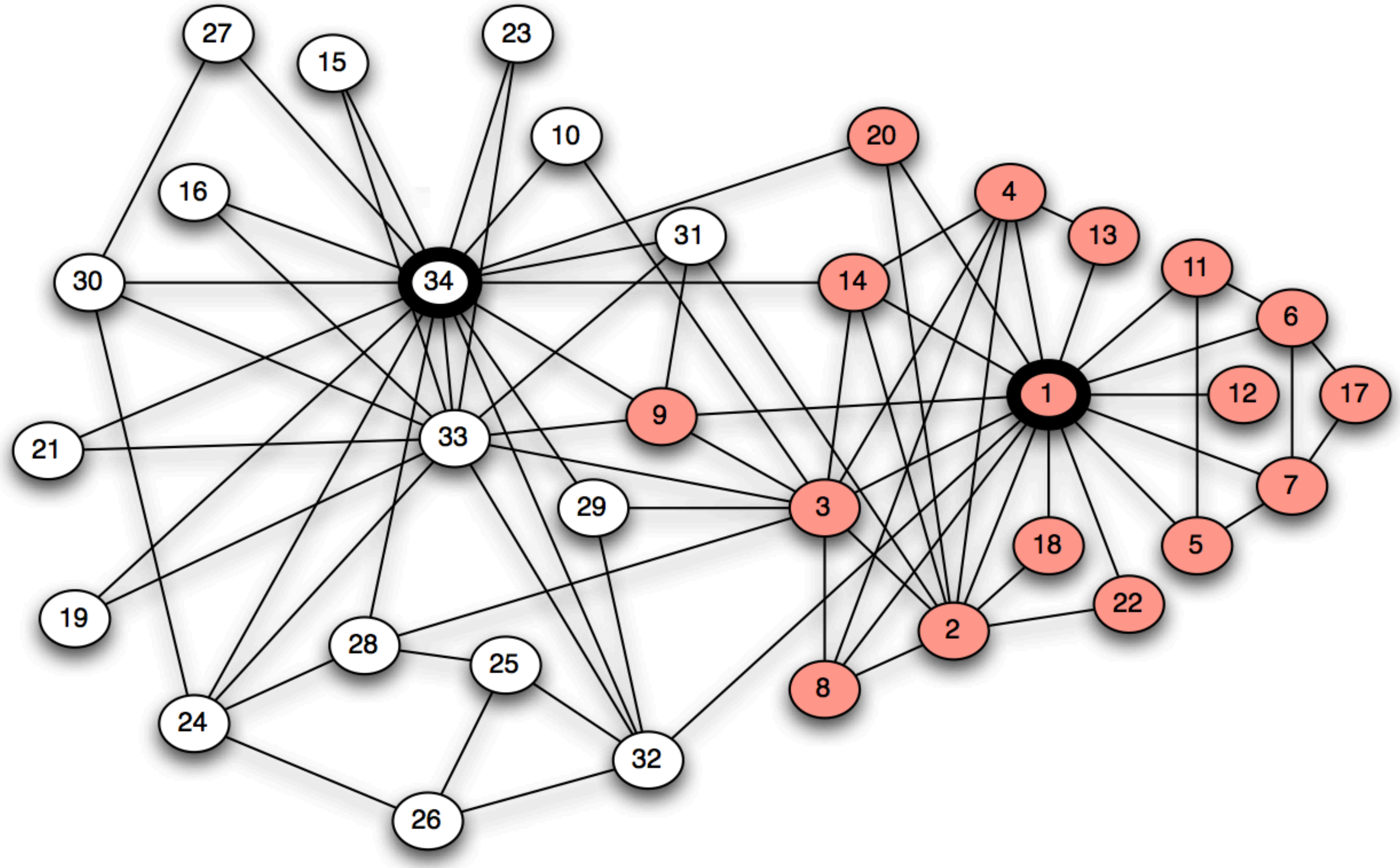
University Karate Clubs

(predicted by Graph Theory)



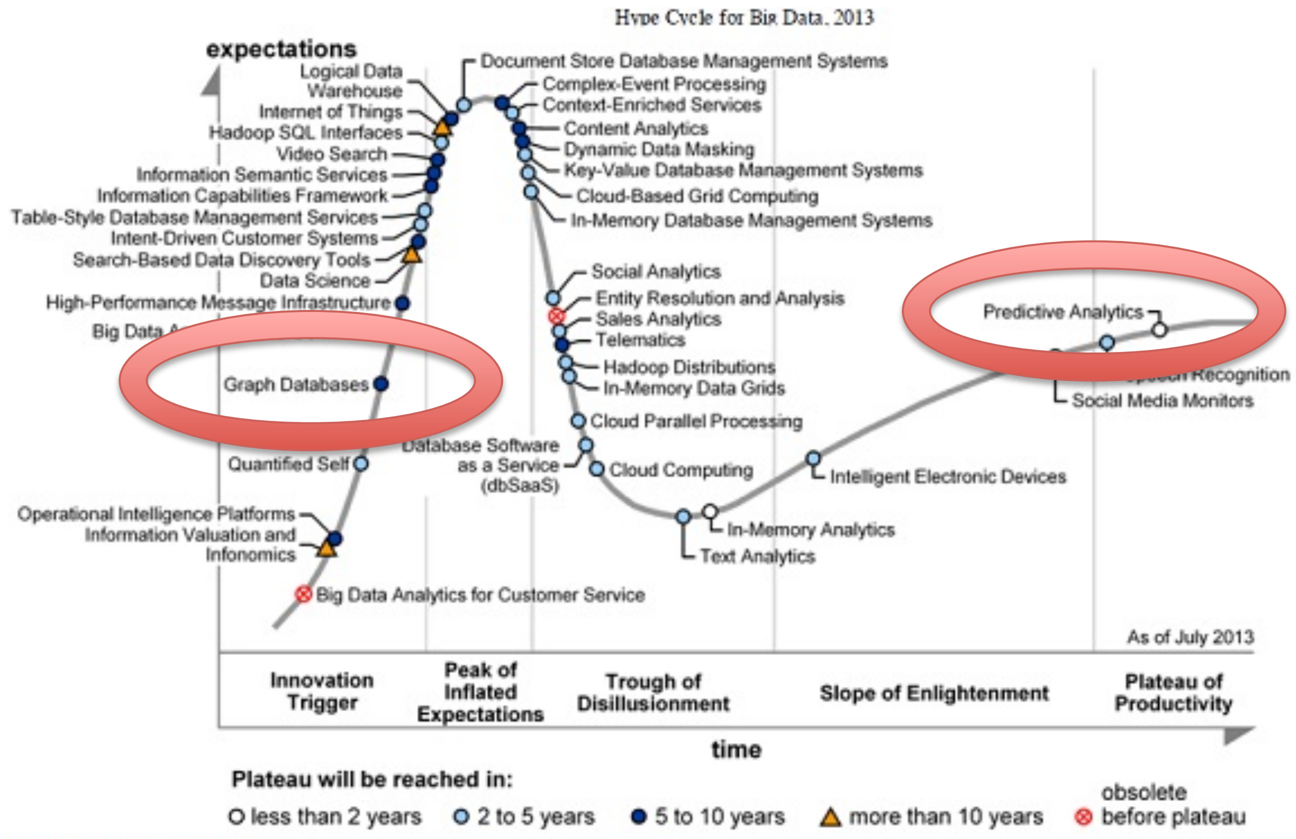
University Karate Clubs

(what actually happened!)

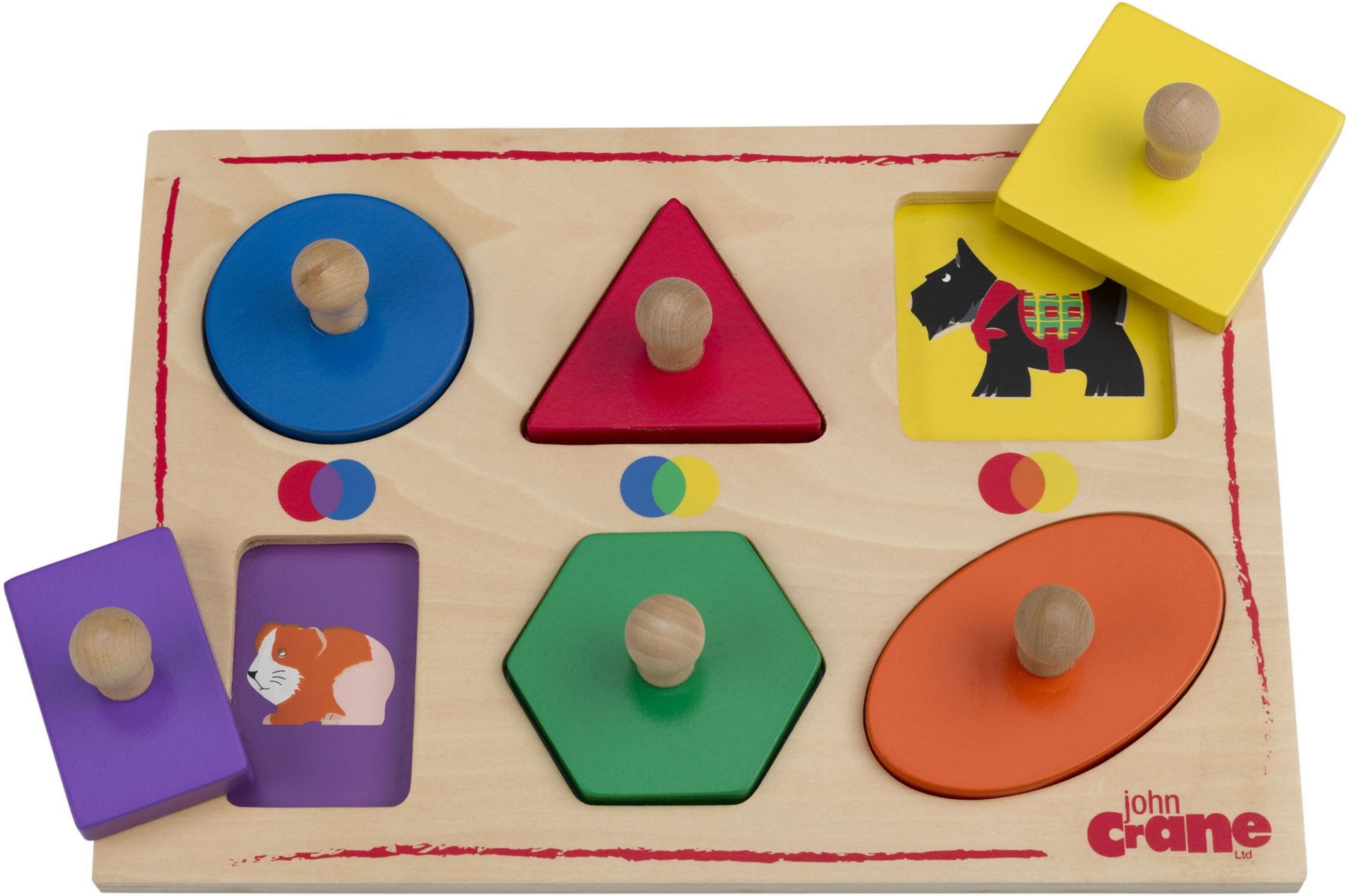


Productive Predictive Analytics in 2 Years

(says Gartner prediction)

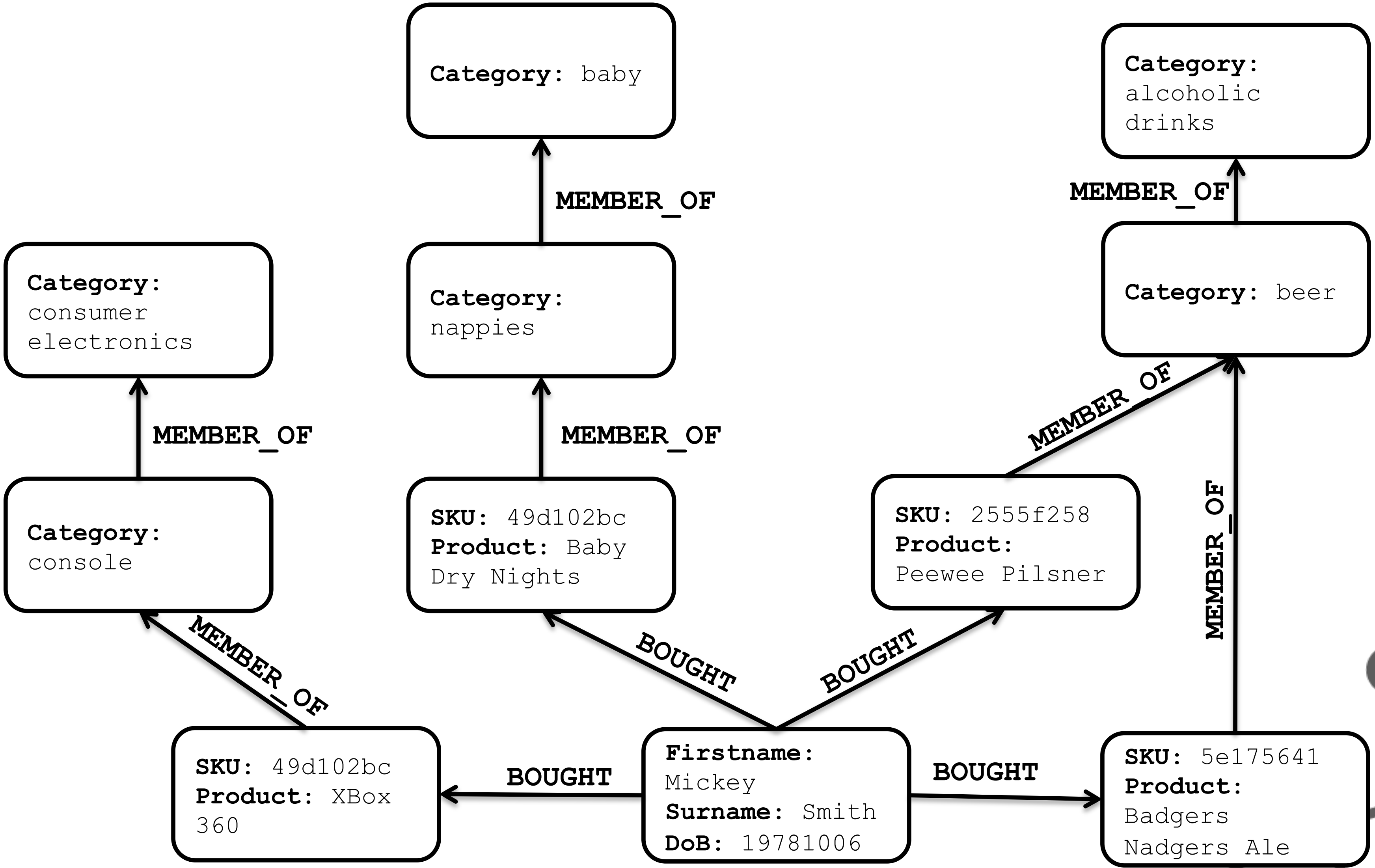


Source: Gartner (July 2013)


john
Crane
Ltd

Cypher

- Declarative graph pattern matching language
  - "SQL for graphs"
  - A humane tool pioneered by a tamed SQL DBA
- A pattern graph matching language
  - Find me stuff like…



Category: baby

Category: alcoholic drinks

Category: consumer electronics

Category: nappies

Category: beer

Category: console

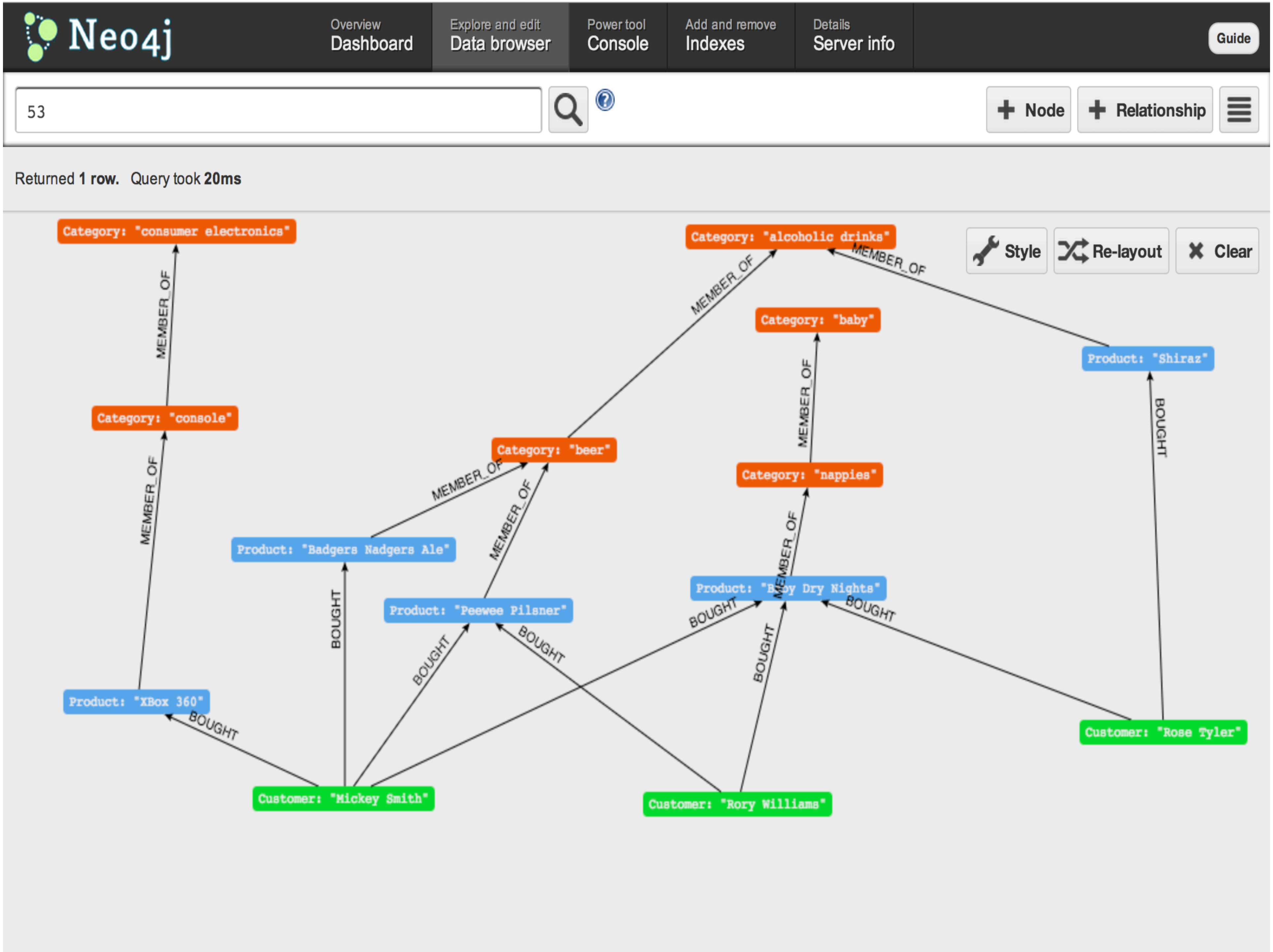
SKU: 49d102bc
Product: Baby Dry Nights

SKU: 2555f258
Product: Peewee Pilsner

SKU: 49d102bc
Product: XBox 360

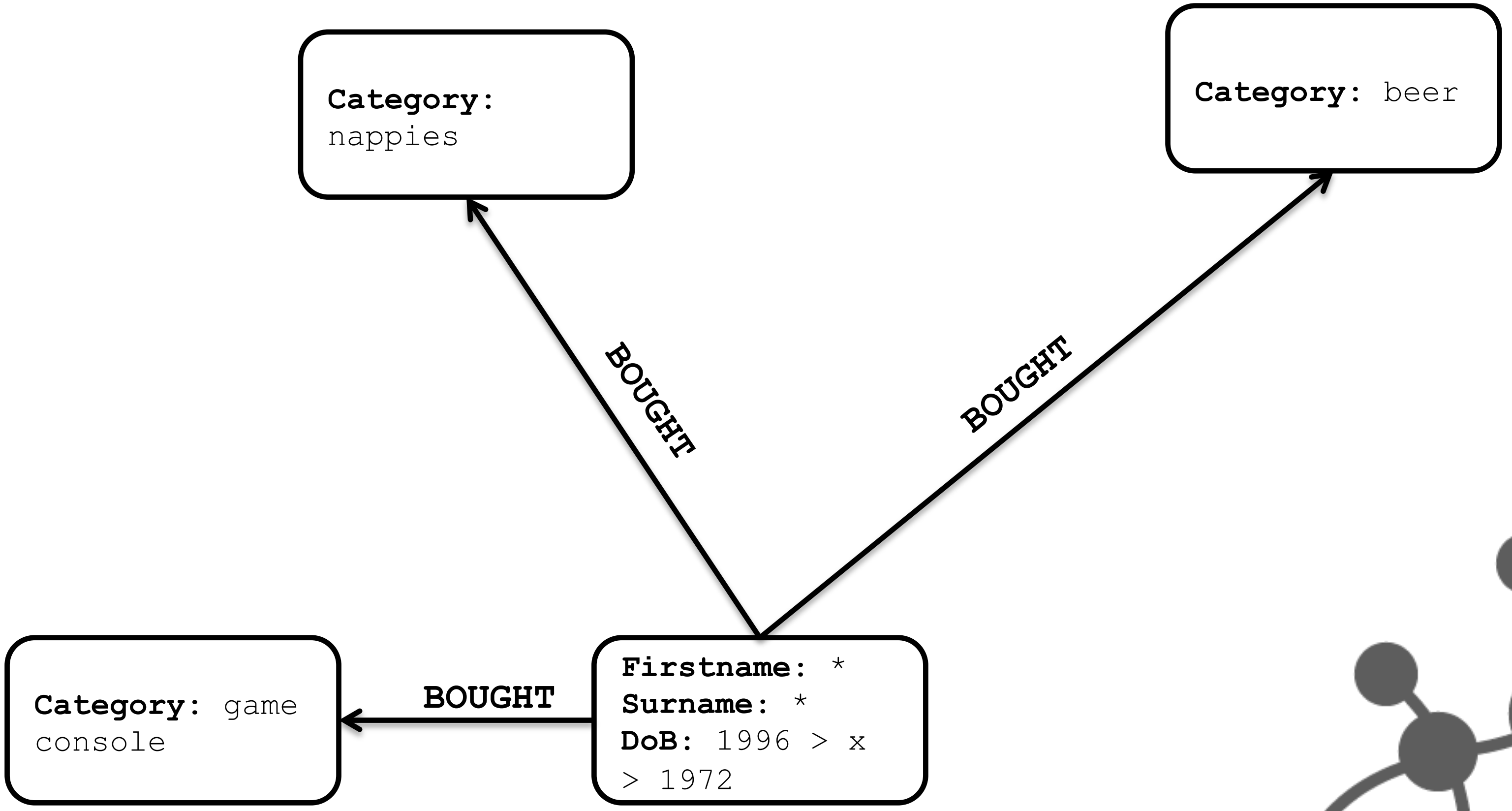
Firstname: Mickey
Surname: Smith
DoB: 19781006

SKU: 5e175641
Product: Badgers Nadgers Ale



53

Returned **1 row.** Query took **20ms**



Category: nappies

Category: beer

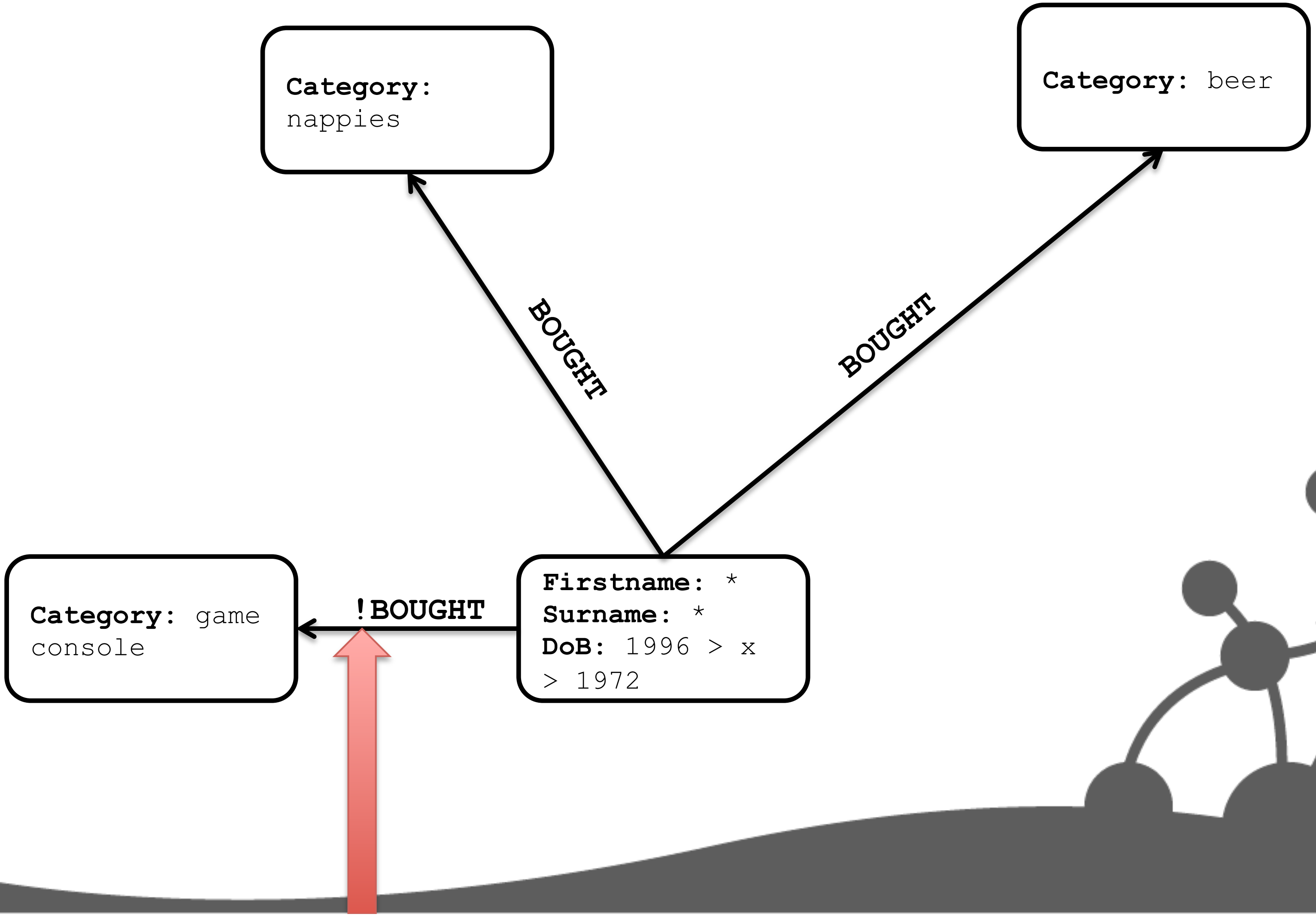
BOUGHT

BOUGHT

Category: game console

BOUGHT

Firstname: *
Surname: *
DoB: 1996 > x > 1972



Category: nappies

Category: beer

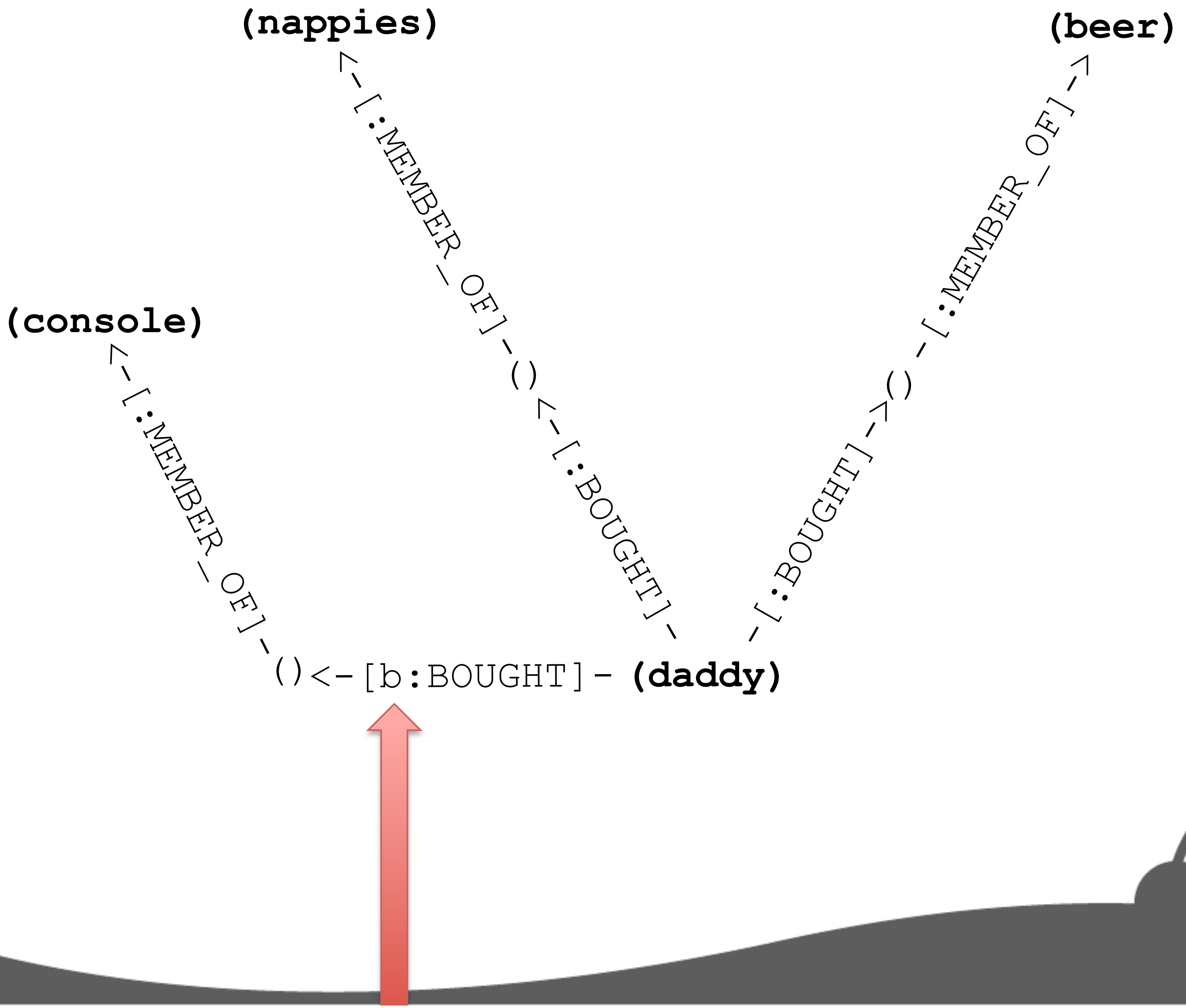
BOUGHT

BOUGHT

Category: game console

!BOUGHT

Firstname: *
Surname: *
DoB: 1996 > x > 1972



(nappies)

<-[:MEMBER_OF]-()<-[:BOUGHT]-

(beer)

-[:BOUGHT]->()-[:MEMBER_OF]->

(console)

<-[:MEMBER_OF]-()<-[b:BOUGHT]- **(daddy)**

Flatten the graph

```
(daddy)-[:BOUGHT]->()-[:MEMBER_OF]->(nappies)
(daddy)-[:BOUGHT]->()-[:MEMBER_OF]->(beer)
(daddy)-[b:BOUGHT]->()-[:MEMBER_OF]->(console)
```

Wrap in a Cypher MATCH clause

```
MATCH (daddy)-[:BOUGHT]->()-[:MEMBER_OF]->(nappies),
(daddy)-[:BOUGHT]->()-[:MEMBER_OF]->(beer),
(daddy)-[b:BOUGHT]->()-[:MEMBER_OF]->(console)
```

Cypher WHERE clause

```
MATCH (daddy)-[:BOUGHT]->()-[:MEMBER_OF]->(nappies),
(daddy)-[:BOUGHT]->()-[:MEMBER_OF]->(beer),
(daddy)-[b:BOUGHT]->()-[:MEMBER_OF]->(console)
WHERE b is null
```

Cypher 1.x recommendations

```
START beer=node:categories(category=‘beer’),
  nappies=node:categories(category=‘nappies’),
  xbox=node:products(product=‘xbox 360’)

MATCH (daddy)-[:BOUGHT]->()-[:MEMBER_OF]->(beer),
  (daddy)-[:BOUGHT]->()-[:MEMBER_OF]->(nappies),
  (daddy)-[b?:BOUGHT]->(xbox)

WHERE b is null

RETURN distinct daddy
```

Cypher 2.0 recommendations

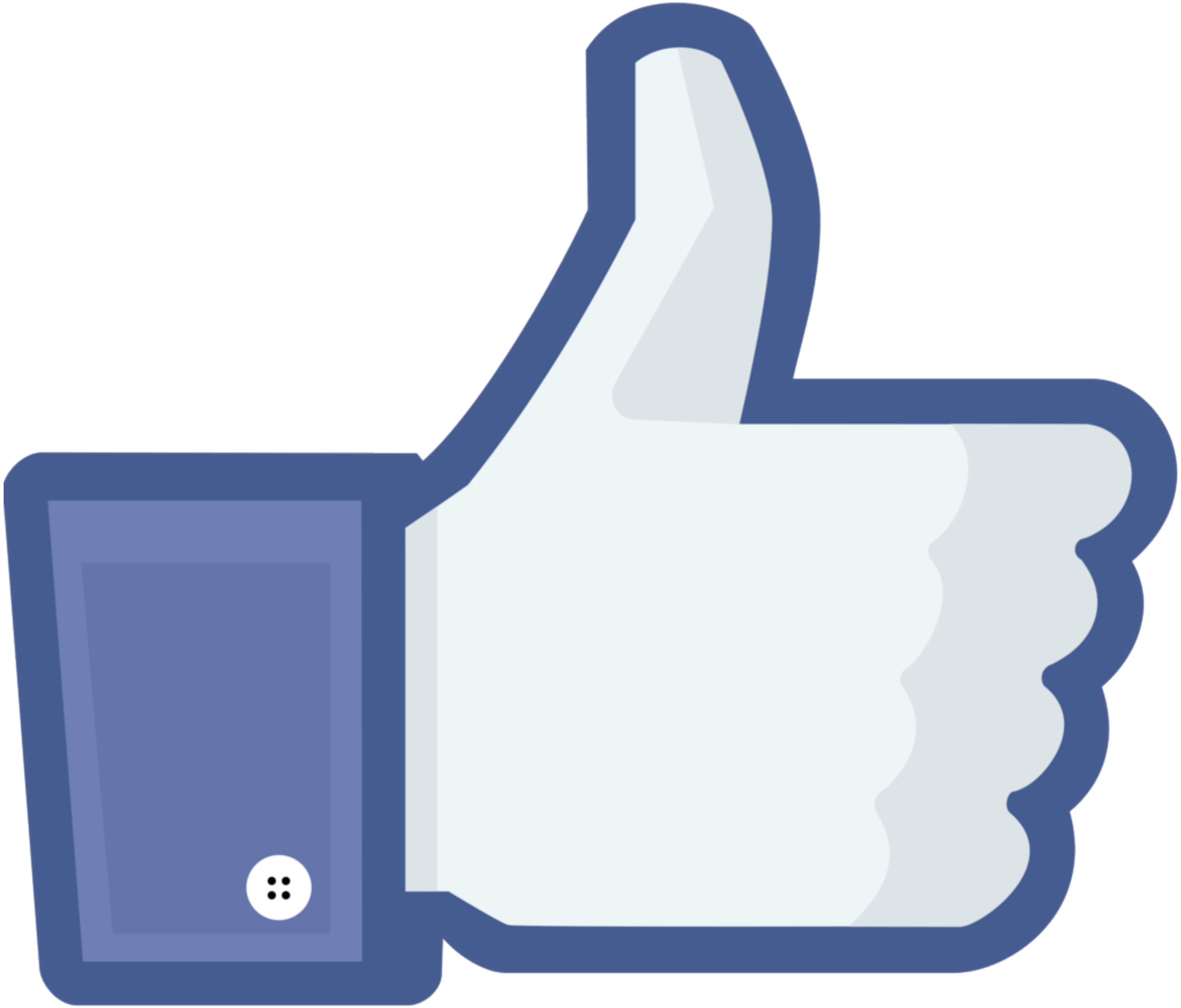
```
MATCH (d:Person)-[:BOUGHT]->()-[:MEMBER_OF]->(n:Category),
      (d:Person)-[:BOUGHT]->()-[:MEMBER_OF]->(b:Category),
      (c:Category)

WHERE n.category = "nappies" AND
      b.category = "beer" AND
      c.category = "console" AND
      NOT((d)-[:BOUGHT]->()-[:MEMBER_OF]->(c))

RETURN DISTINCT d AS daddy
```

Results

```
==> +---------------------------------------------+
==> | daddy                                       |
==> +---------------------------------------------+
==> | Node[15]{name:"Rory Williams",dob:19880121} |
==> +---------------------------------------------+
==> 1 row
==> 0 ms
==>
neo4j-sh (0)$
```

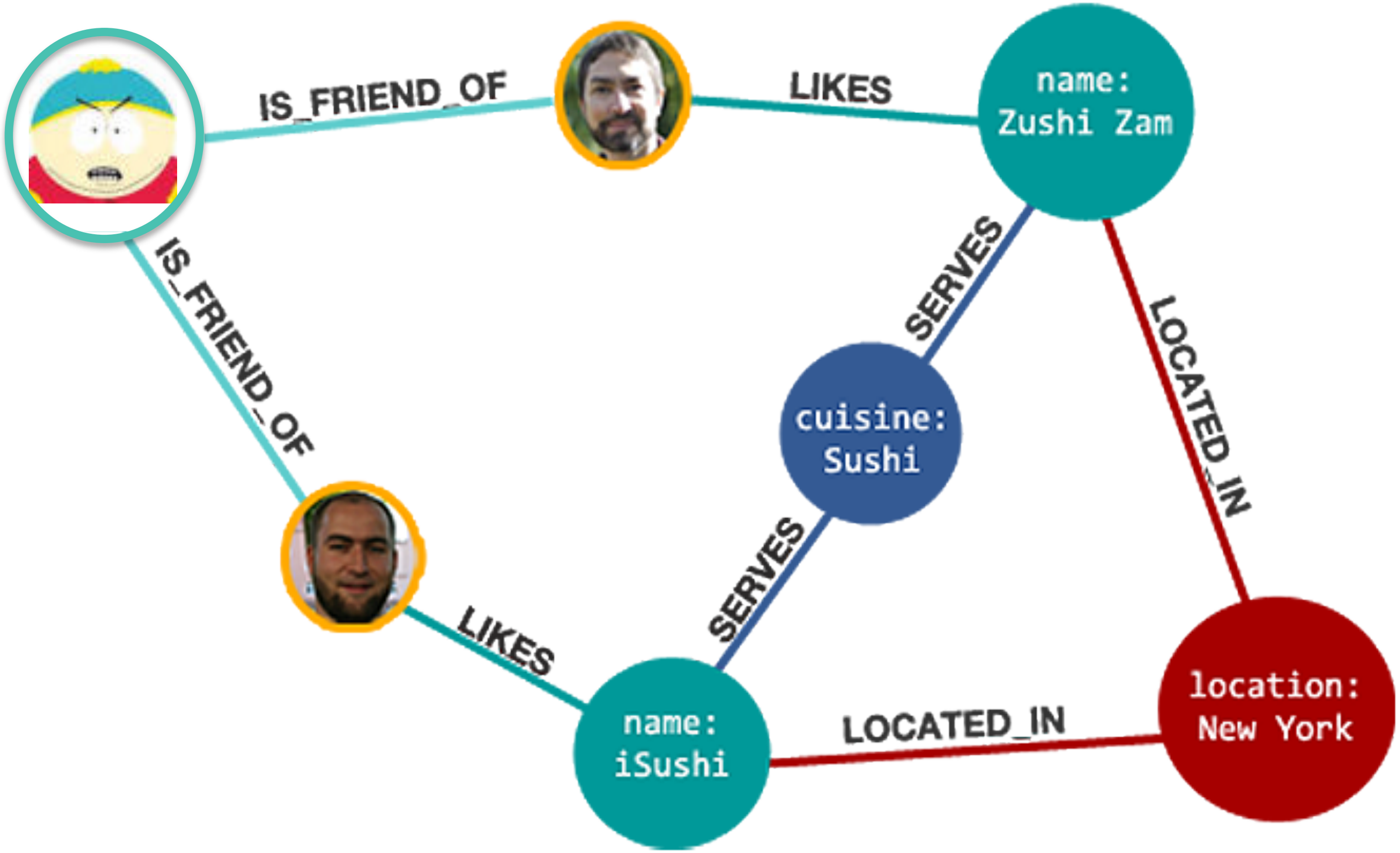


Facebook Graph Search

Which sushi restaurants in NYC do my friends like?

See http://maxdemarzi.com/

Graph Structure



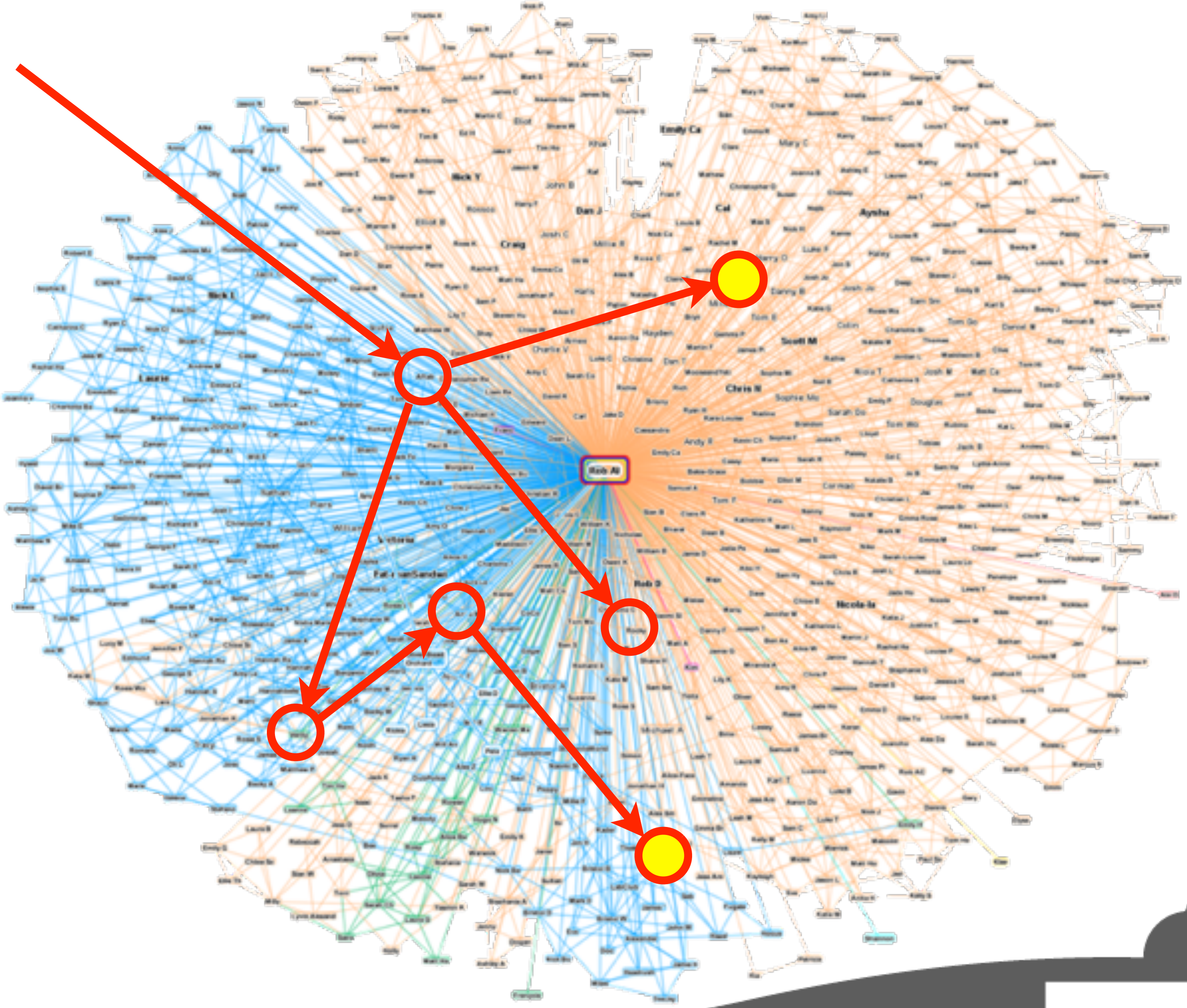
Cypher query

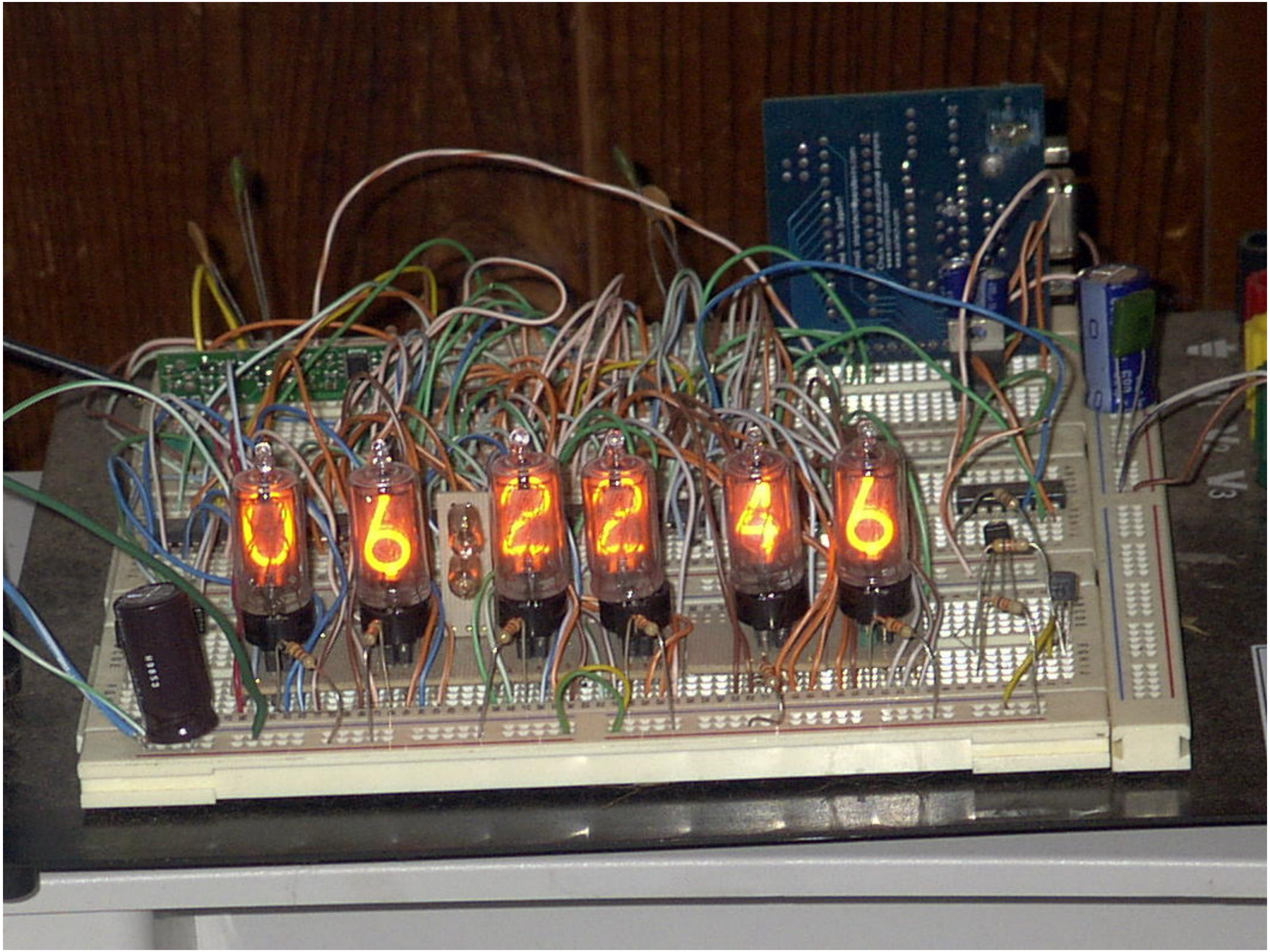
```
START me=node:person(name = 'Jim'),
      location=node:location(location='New York'),
      cuisine=node:cuisine(cuisine='Sushi')

MATCH (me)-[:IS_FRIEND_OF]->(friend)-[:LIKES]->(restaurant)
      -[:LOCATED_IN]->(location),(restaurant)-[:SERVES]->(cuisine)

RETURN restaurant
```

Search structure




062246

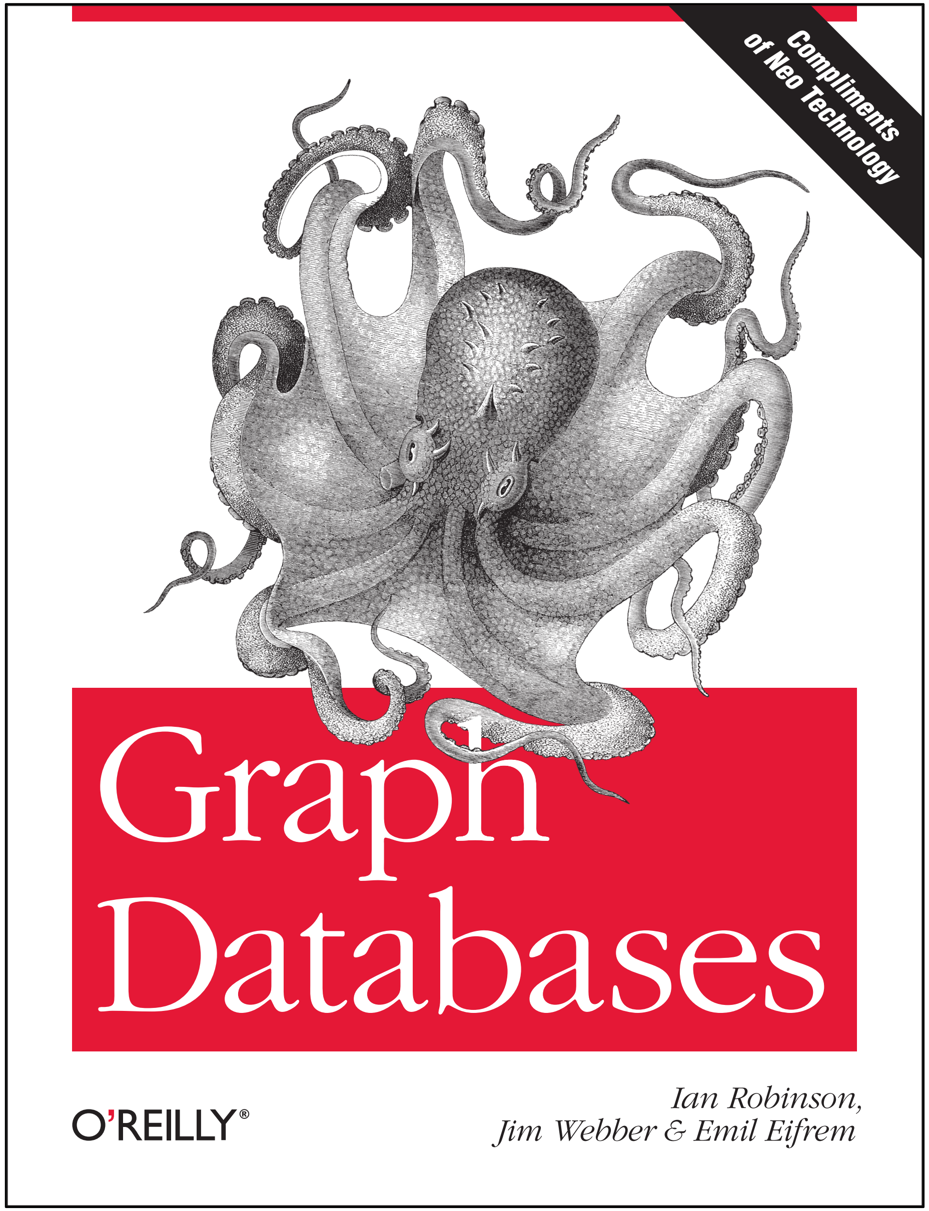
What's Neo4j good for?

- Data centre management
- Supply chain/provenance
- Recommendations
- Business intelligence
- Social computing
- MDM
- Web of things
- Time series/event data
- Product/engineering catalogue
- Web analytics, user journeys
- Scientific computing
- Spatial
- Geo/Seismic/Meteorological
- Bio/Pharma
- And many, many more…

Can Neo4j compute

42?

```
MATCH (you)-[:HAPPINESS]->(answer)
RETURN answer
```



Compliments of Neo Technology

Graph Databases

O'REILLY®

Ian Robinson, Jim Webber & Emil Eifrem

Free O'Reilly book!

Come to the
Neo Technology Stand
or visit
http://graphdatabases.com

Thanks for listening

@jimwebber

http://neo4j.org