Session code: 6191

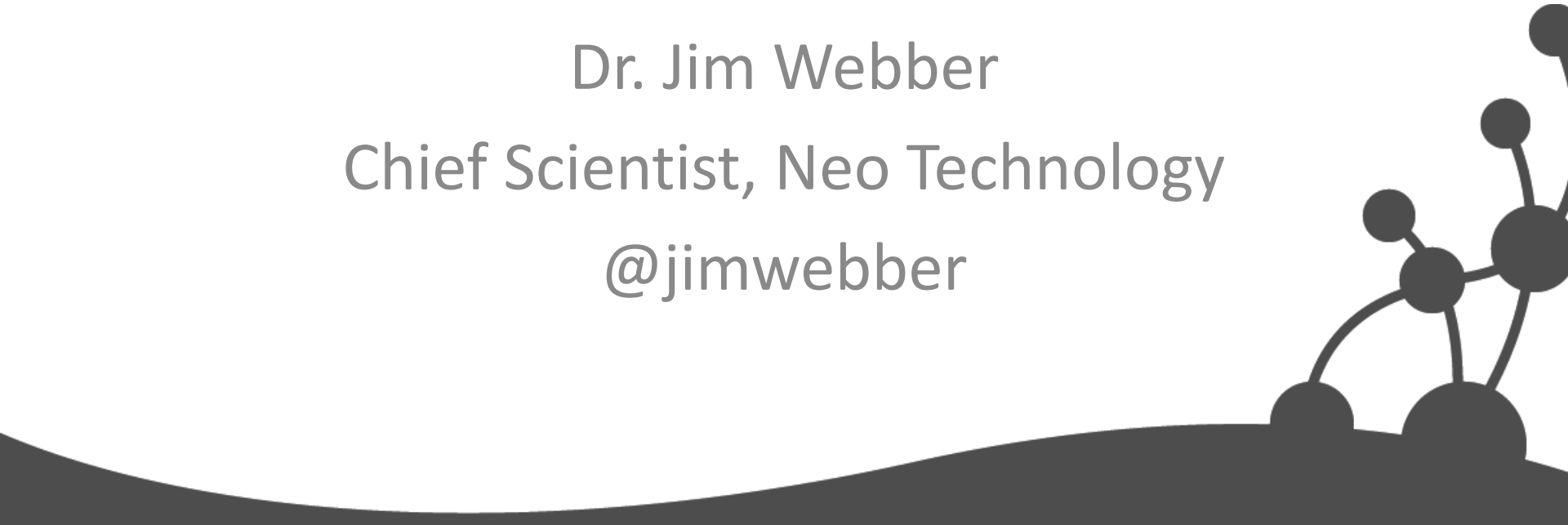
neotechnology
value in relationships

A Little Graph Theory for the Busy Developer

Dr. Jim Webber

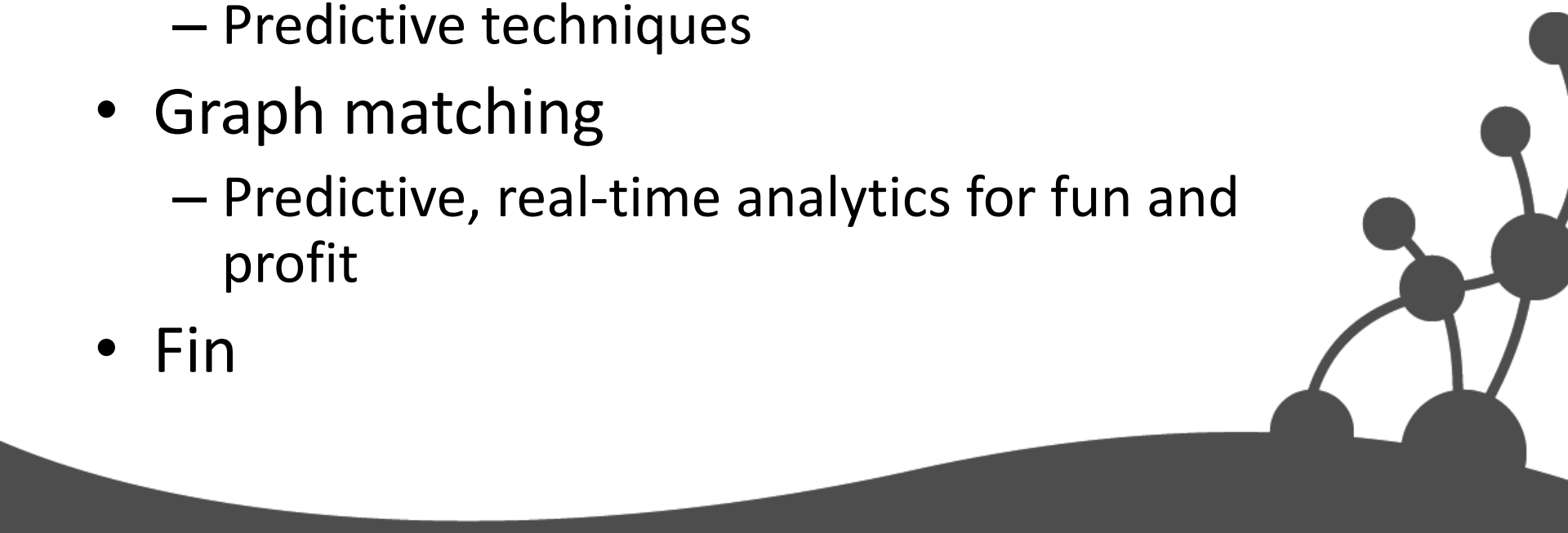
Chief Scientist, Neo Technology

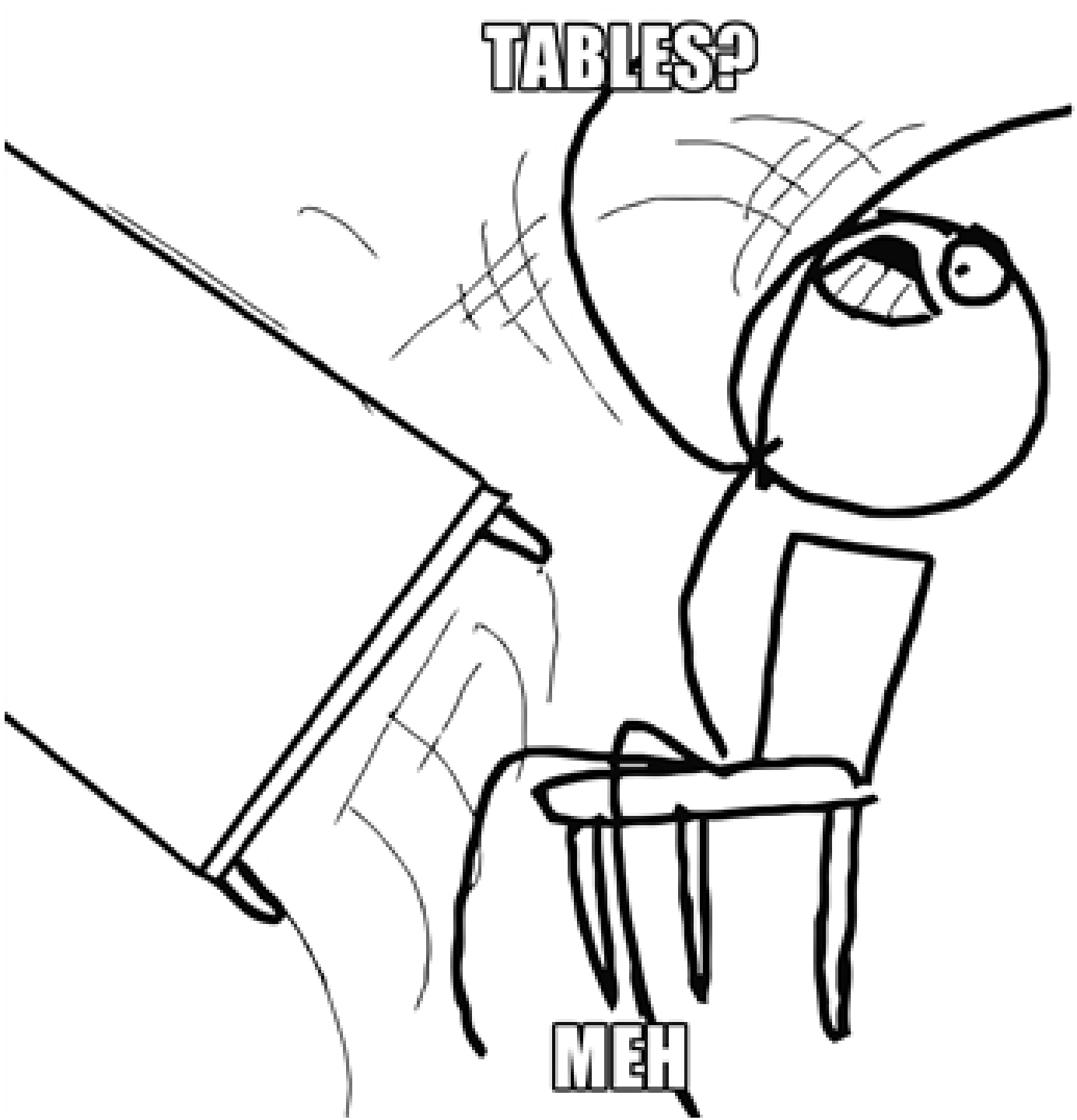
@jimwebber



Roadmap

- Imprisoned data
- Graph models
- Graph theory
  - Local properties, global behaviours
  - Predictive techniques
- Graph matching
  - Predictive, real-time analytics for fun and profit
- Fin




TABLES?
MEH

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


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

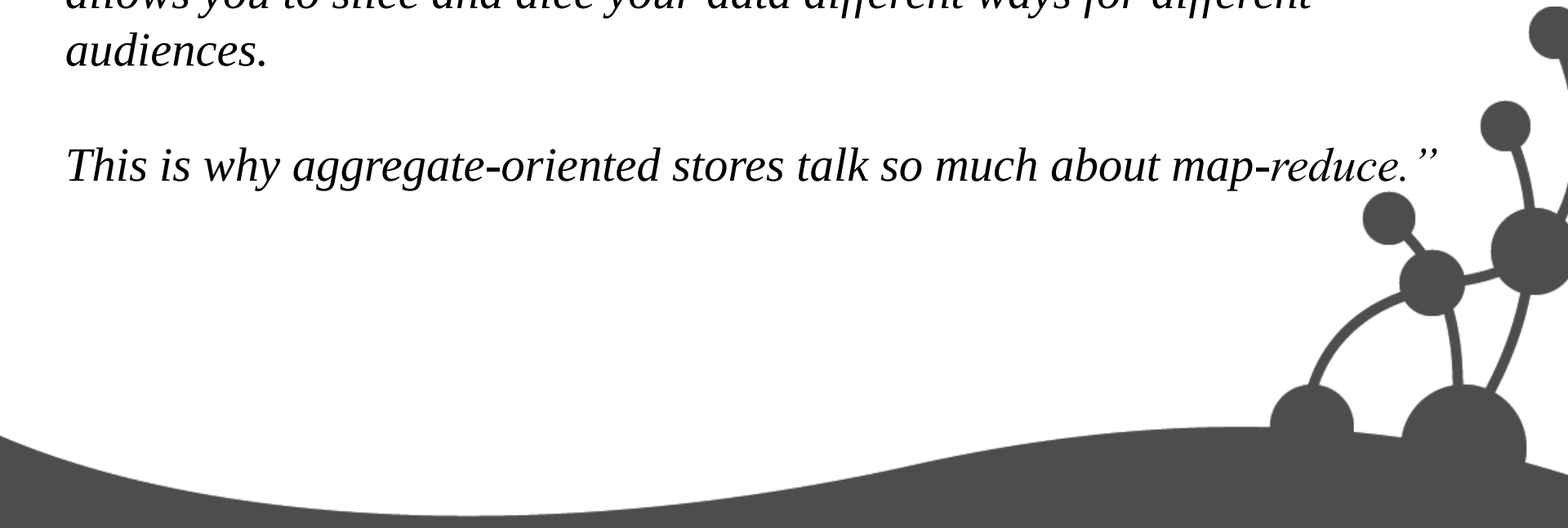

http://www.xtranormal.com/watch/6995033/mongo-db-is-web-scale

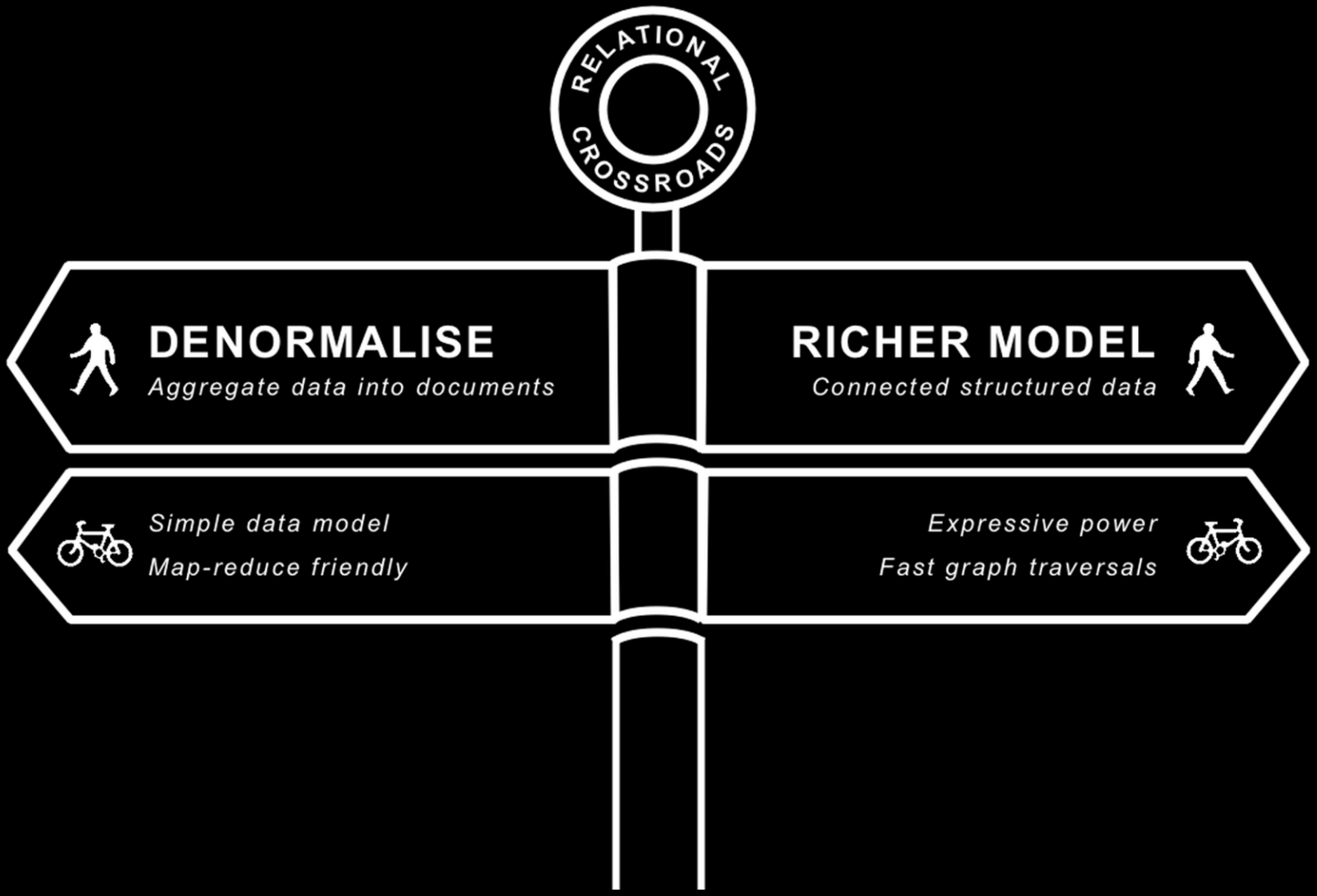
Aggregate-Oriented Data

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

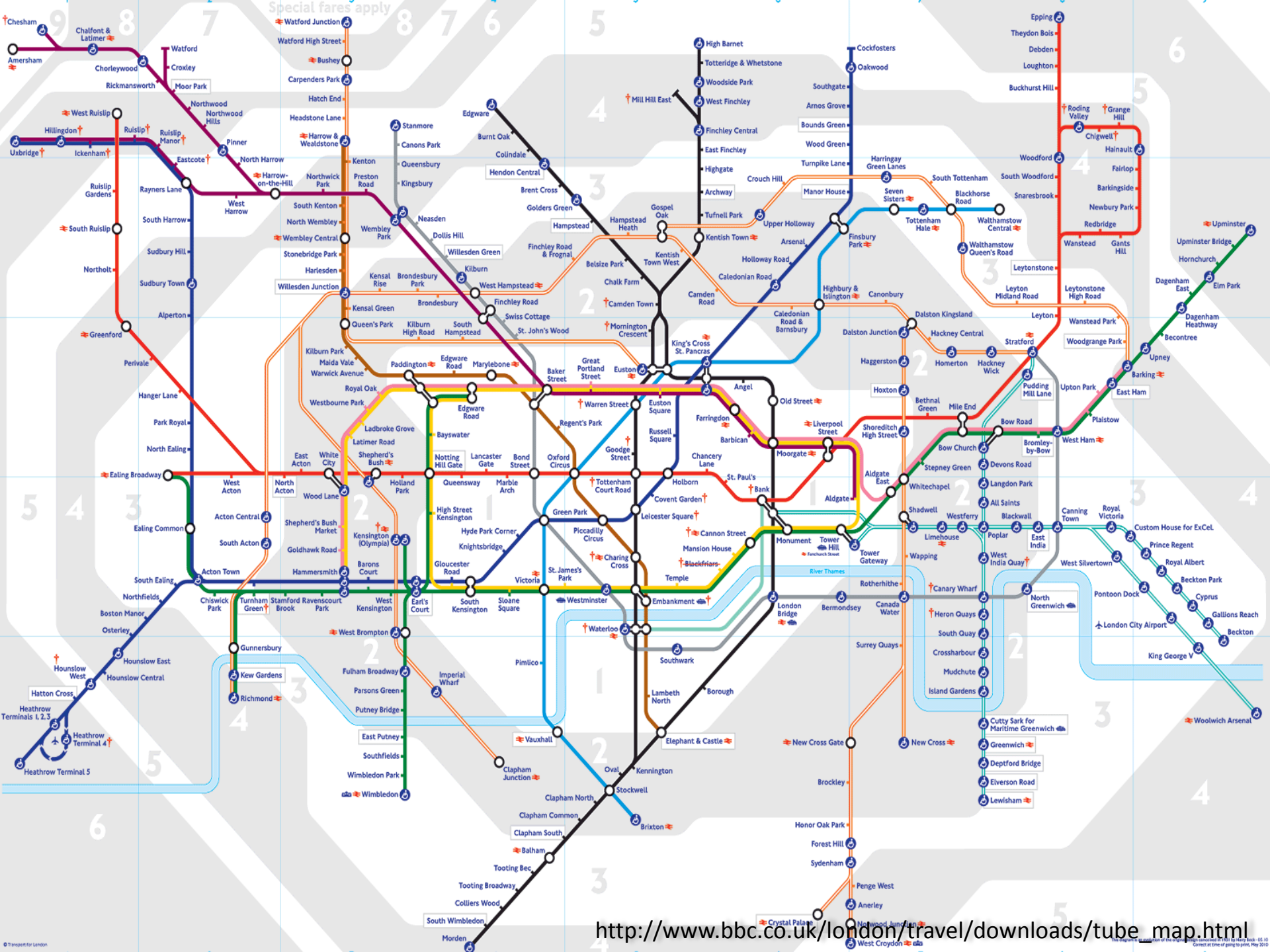
"There is a significant downside - the whole approach works really well when data access is aligned with the aggregates, but what if you want to look at the data in a different way? Order entry naturally stores orders as aggregates, but analyzing product sales cuts across the aggregate structure. The 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."



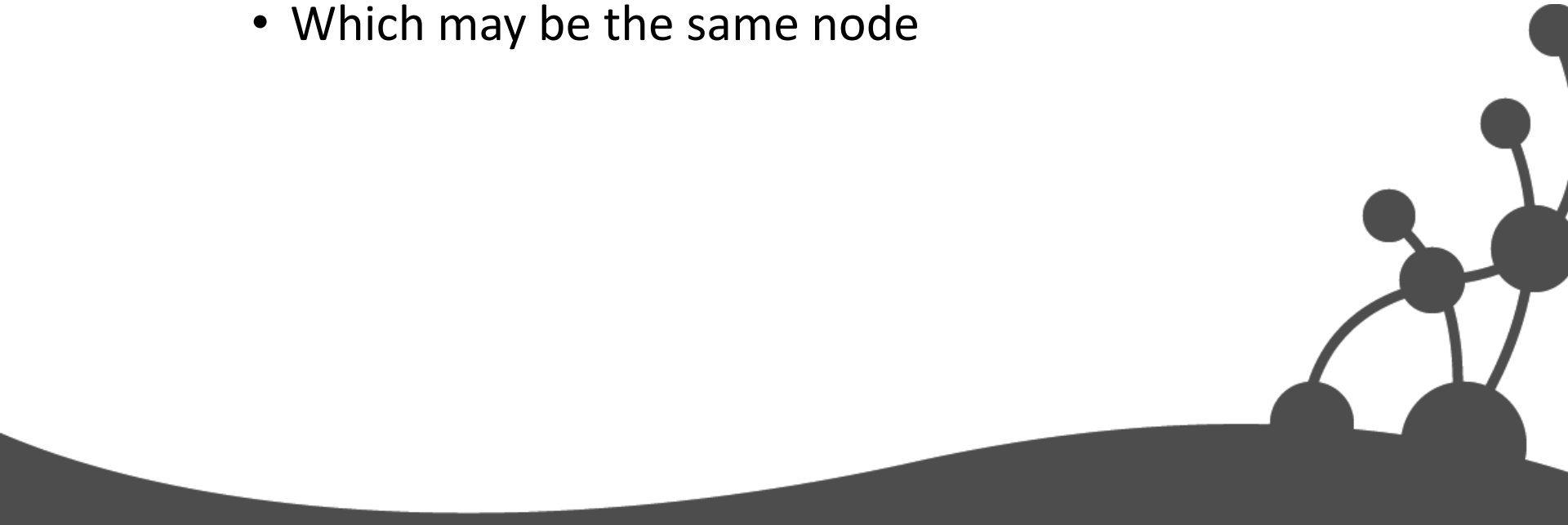

RELATIONAL
CROSSROADS
DENORMALISE
Aggregate data into documents
RICHER MODEL
Connected structured data
Simple data model
Map-reduce friendly
Expressive power
Fast graph traversals

$$complexity = f(size, connectedness, uniformity)$$

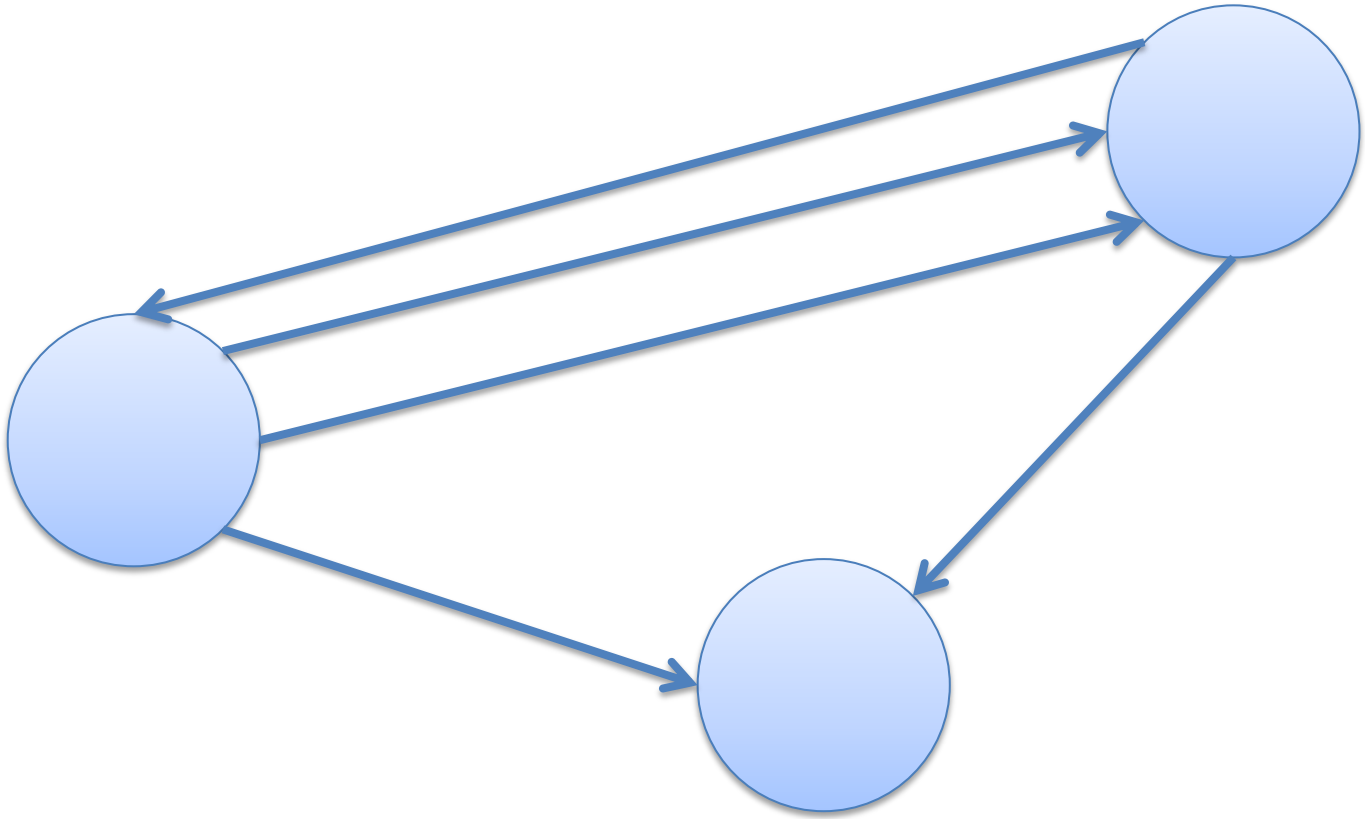

Special fares apply
Chesham
Amersham
Chalfont & Latimer
Chorleywood
Rickmansworth
Watford
Croxley
Moor Park
Northwood
Northwood Hills
Pinner
North Harrow
Harrow-on-the-Hill
West Ruislip
Hillingdon
Uxbridge
Ruislip
Ruislip Manor
Ickenham
Eastcote
Ruislip Gardens
South Ruislip
Northolt
Greenford
Perivale
Hanger Lane
Park Royal
North Ealing
Ealing Broadway
Ealing Common
Acton Central
South Acton
Acton Town
South Ealing
Northfields
Boston Manor
Osterley
Hounslow West
Hounslow East
Hounslow Central
Hatton Cross
Heathrow Terminals 1, 2, 3
Heathrow Terminal 4
Heathrow Terminal 5
Rayners Lane
West Harrow
South Harrow
Sudbury Hill
Sudbury Town
Alperton
Watford Junction
Watford High Street
Bushey
Carpenders Park
Hatch End
Headstone Lane
Harrow & Wealdstone
Kenton
Northwick Park
Preston Road
South Kenton
North Wembley
Wembley Central
Stonebridge Park
Harlesden
Willesden Junction
Kensal Green
Queen's Park
Kensal Rise
Brondesbury Park
Brondesbury
Kilburn High Road
Kilburn
South Hampstead
West Hampstead
Finchley Road
Swiss Cottage
St. John's Wood
Wembley Park
Stanmore
Canons Park
Queensbury
Kingsbury
Neasden
Dollis Hill
Willesden Green
Kilburn Park
Maida Vale
Warwick Avenue
Paddington
Edgware Road
Marylebone
Baker Street
Great Portland Street
Euston
Royal Oak
Westbourne Park
Ladbroke Grove
Latimer Road
Shepherd's Bush
Bayswater
Notting Hill Gate
Lancaster Gate
Bond Street
Oxford Circus
Warren Street
Regent's Park
Goodge Street
Euston Square
Russell Square
East Acton
White City
West Acton
North Acton
Wood Lane
Holland Park
Queensway
Marble Arch
Tottenham Court Road
Holborn
Covent Garden
Leicester Square
Chancery Lane
St. Paul's
Bank
High Street Kensington
Green Park
Piccadilly Circus
Hyde Park Corner
Knightsbridge
Shepherd's Bush Market
Goldhawk Road
Kensington (Olympia)
Barons Court
Hammersmith
Gloucester Road
Victoria
St. James's Park
Westminster
Embankment
Charing Cross
Cannon Street
Mansion House
Blackfriars
Temple
Monument
Tower Hill
Fenchurch Street
Tower Gateway
Aldgate
Aldgate East
Whitechapel
Shadwell
Limehouse
Westferry
Poplar
Blackwall
East India
Canning Town
Royal Victoria
Custom House for ExCeL
Prince Regent
Royal Albert
Beckton Park
Cyprus
Gallions Reach
Beckton
King George V
Woolwich Arsenal
London City Airport
Pontoon Dock
West Silvertown
River Thames
Wapping
Rotherhithe
Canary Wharf
Canada Water
Bermondsey
London Bridge
Borough
Southwark
Waterloo
Lambeth North
Elephant & Castle
Pimlico
Vauxhall
Oval
Kennington
Stockwell
Clapham North
Clapham Common
Clapham South
Balham
Tooting Bec
Tooting Broadway
Colliers Wood
South Wimbledon
Morden
Clapham Junction
Brixton
Wimbledon
Wimbledon Park
Southfields
East Putney
Putney Bridge
Parsons Green
Fulham Broadway
West Brompton
Imperial Wharf
Gunnersbury
Kew Gardens
Richmond
Turnham Green
Stamford Brook
Ravenscourt Park
Chiswick Park
Earl's Court
West Kensington
South Kensington
Sloane Square
High Barnet
Totteridge & Whetstone
Woodside Park
West Finchley
Finchley Central
East Finchley
Highgate
Archway
Tufnell Park
Kentish Town
Kentish Town West
Mill Hill East
Edgware
Burnt Oak
Colindale
Hendon Central
Brent Cross
Golders Green
Hampstead
Hampstead Heath
Gospel Oak
Belsize Park
Chalk Farm
Camden Town
Mornington Crescent
Finchley Road & Frognal
Camden Road
Caledonian Road
Caledonian Road & Barnsbury
Holloway Road
Arsenal
Upper Holloway
Crouch Hill
Harringay Green Lanes
Turnpike Lane
Manor House
Wood Green
Bounds Green
Arnos Grove
Southgate
Cockfosters
Oakwood
King's Cross St. Pancras
Angel
Old Street
Farringdon
Barbican
Moorgate
Liverpool Street
Highbury & Islington
Canonbury
Dalston Kingsland
Dalston Junction
Hackney Central
Homerton
Hackney Wick
Haggerston
Hoxton
Shoreditch High Street
Pudding Mill Lane
Bethnal Green
Mile End
Bow Road
Bromley-by-Bow
West Ham
Bow Church
Devons Road
Stepney Green
Langdon Park
All Saints
Seven Sisters
South Tottenham
Tottenham Hale
Blackhorse Road
Walthamstow Central
Walthamstow Queen's Road
Finsbury Park
Stratford
Leyton
Leyton Midland Road
Leytonstone
Leytonstone High Road
Wanstead Park
Woodgrange Park
Upton Park
East Ham
Plaistow
Barking
Upney
Becontree
Dagenham Heathway
Dagenham East
Elm Park
Hornchurch
Upminster Bridge
Upminster
Epping
Theydon Bois
Debden
Loughton
Buckhurst Hill
Roding Valley
Grange Hill
Chigwell
Hainault
Woodford
South Woodford
Snaresbrook
Fairlop
Barkingside
Newbury Park
Redbridge
Wanstead
Gants Hill
Crossharbour
Mudchute
Island Gardens
Cutty Sark for Maritime Greenwich
Greenwich
Deptford Bridge
Elverson Road
Lewisham
North Greenwich
New Cross
New Cross Gate
Brockley
Honor Oak Park
Forest Hill
Sydenham
Penge West
Anerley
Crystal Palace
Norwood Junction
West Croydon
http://www.bbc.co.uk/london/travel/downloads/tube_map.html

Property graphs

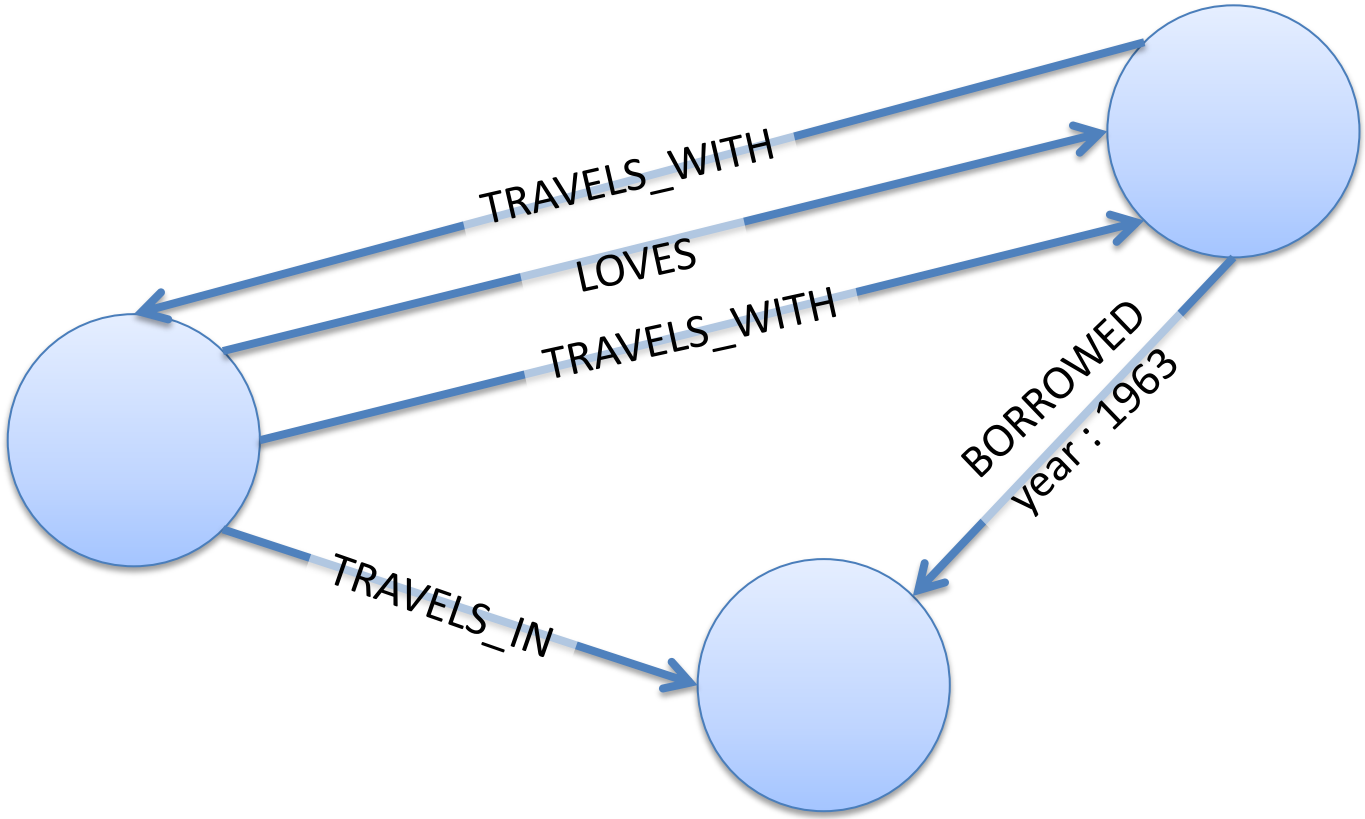
- Property graph model:
  - Nodes with properties
  - Named, directed relationships with properties
  - Relationships have exactly one start and end node
    - Which may be the same node



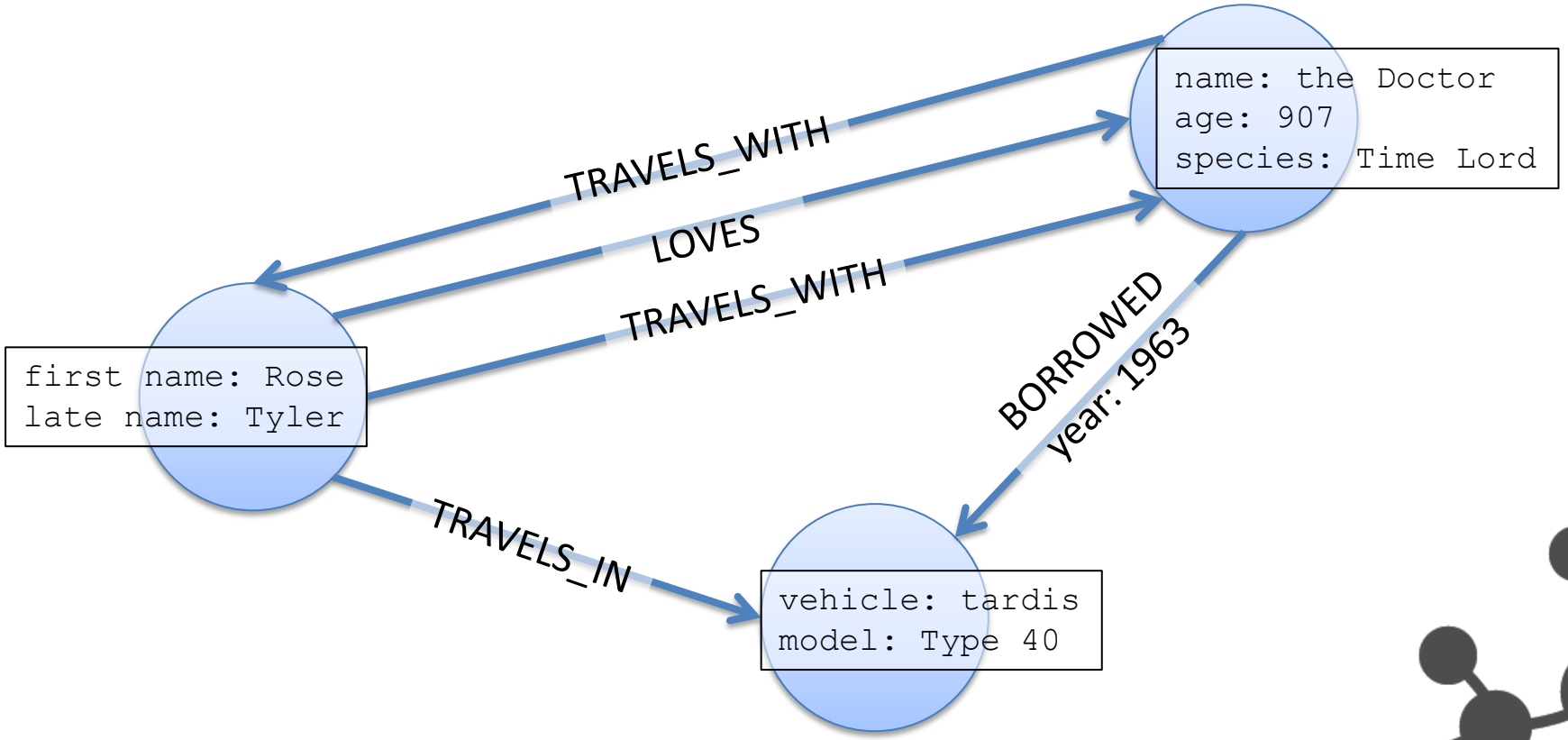
Property Graph Model



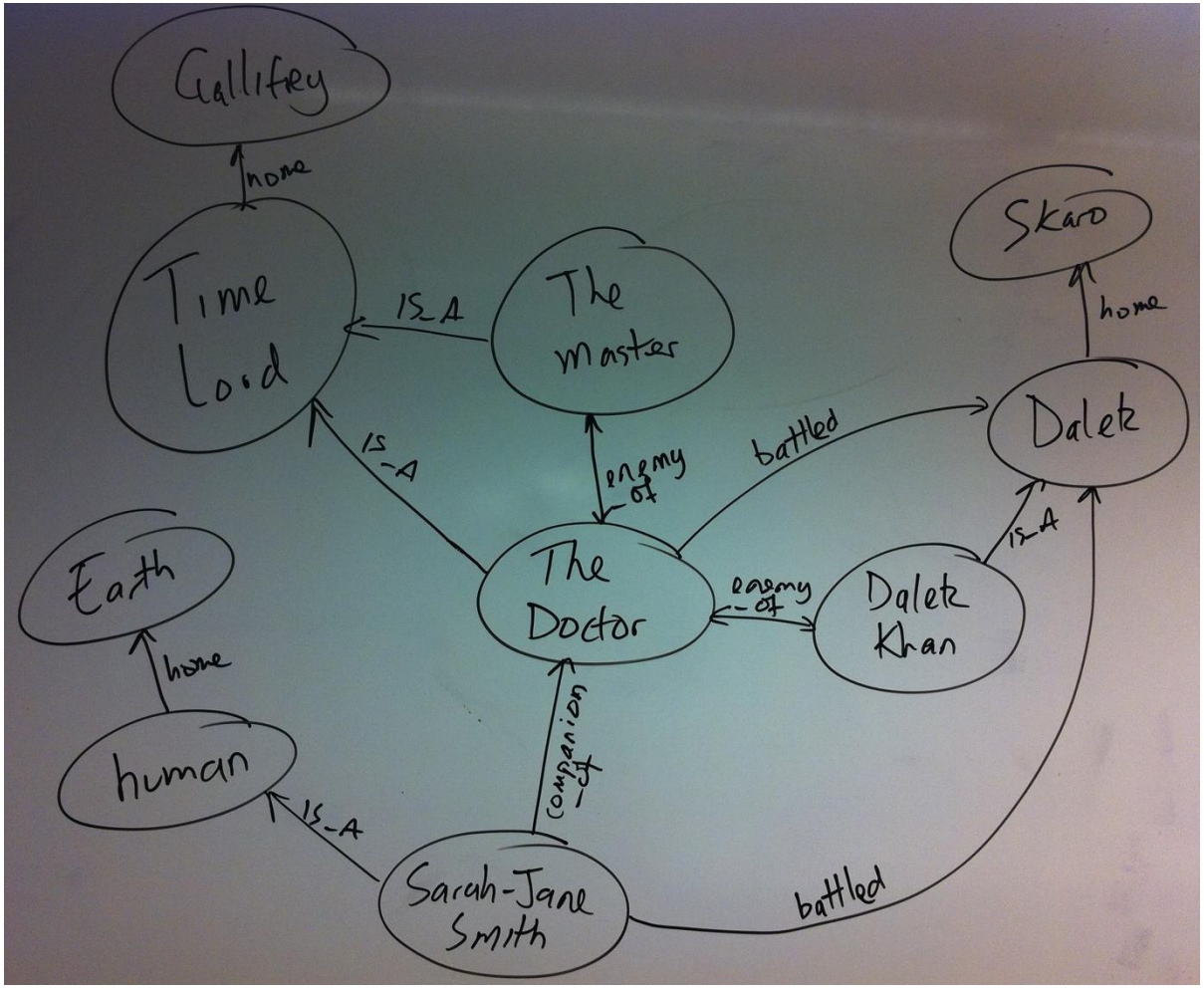
Property Graph Model



Property Graph Model



Property graphs are *very* whiteboard-friendly

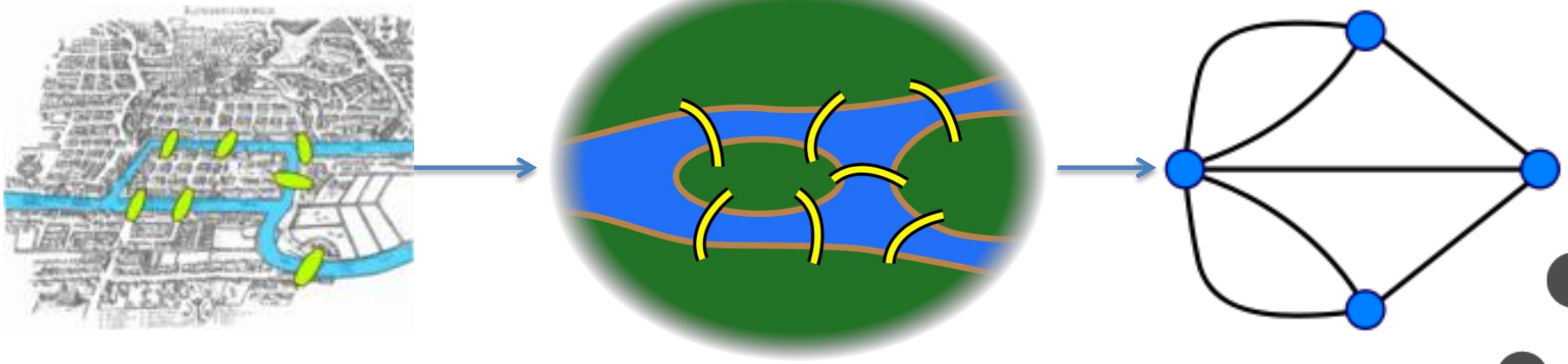


http://blogs.adobe.com/digitalmarketing/analytics/predictive-analytics/predictive-analytics-and-the-digital-marketer/

Meet Leonhard Euler

- Swiss mathematician
- Inventor of Graph Theory (1736)

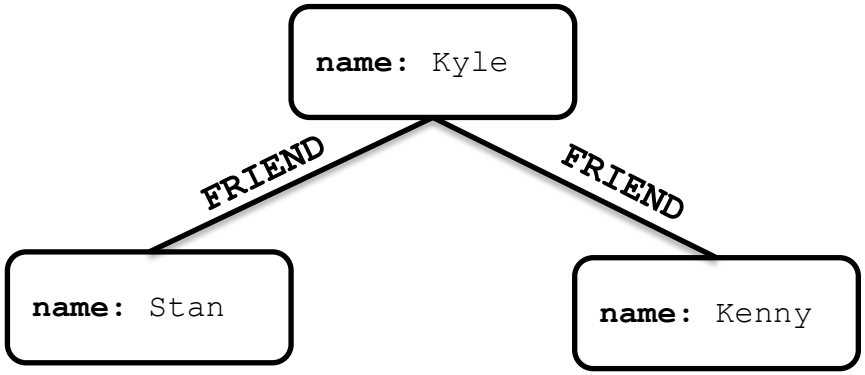
http://en.wikipedia.org/wiki/File:Leonhard_Euler_2.jpg



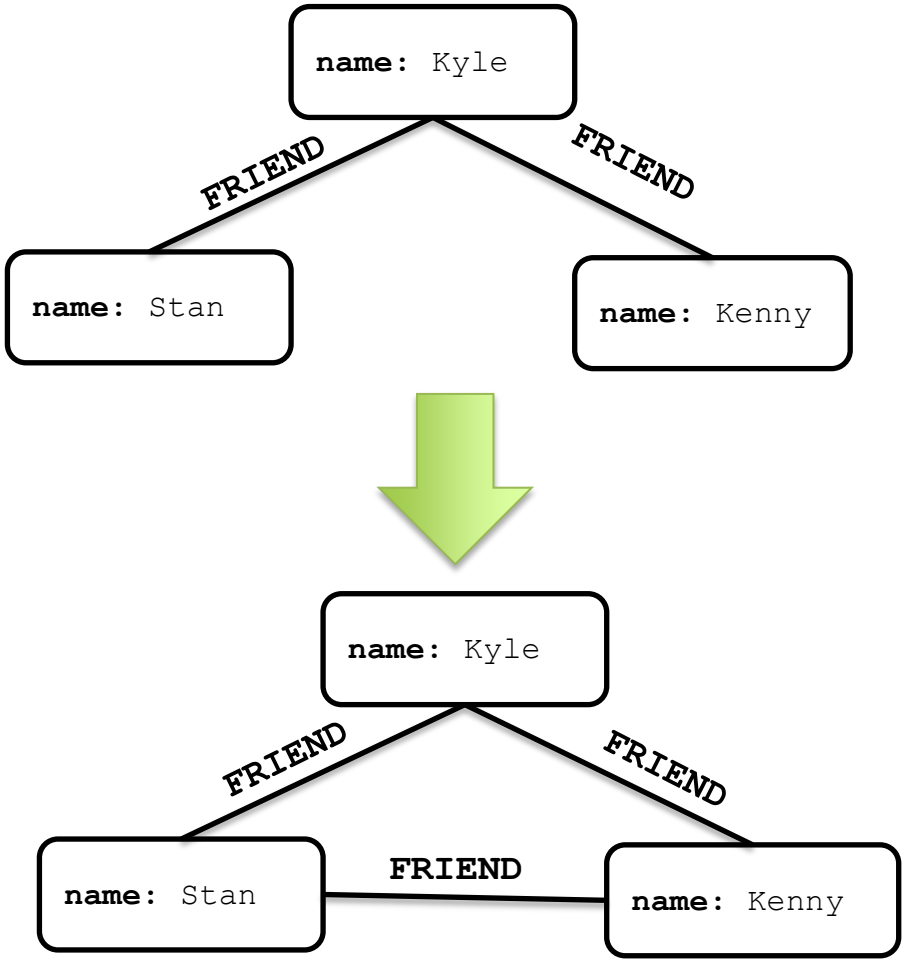
http://en.wikipedia.org/wiki/Seven_Bridges_of_Königsberg


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

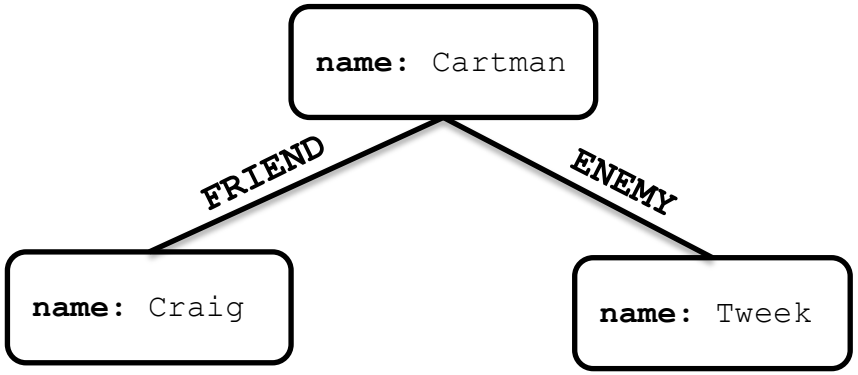
Triadic Closure



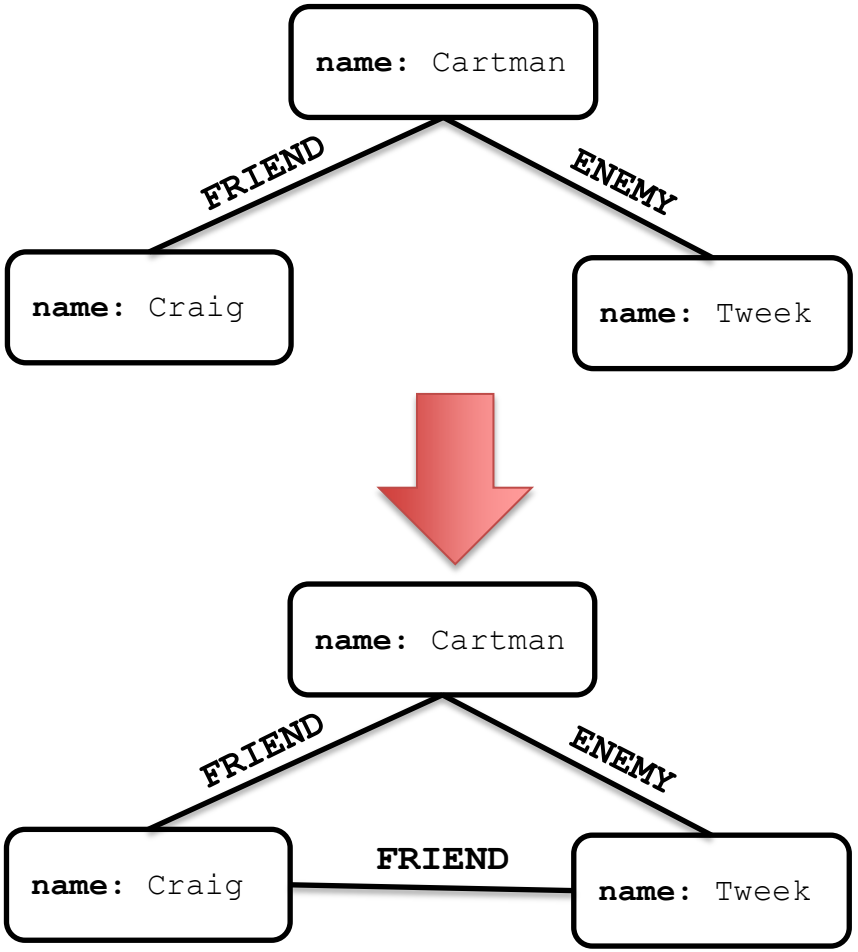
Triadic Closure



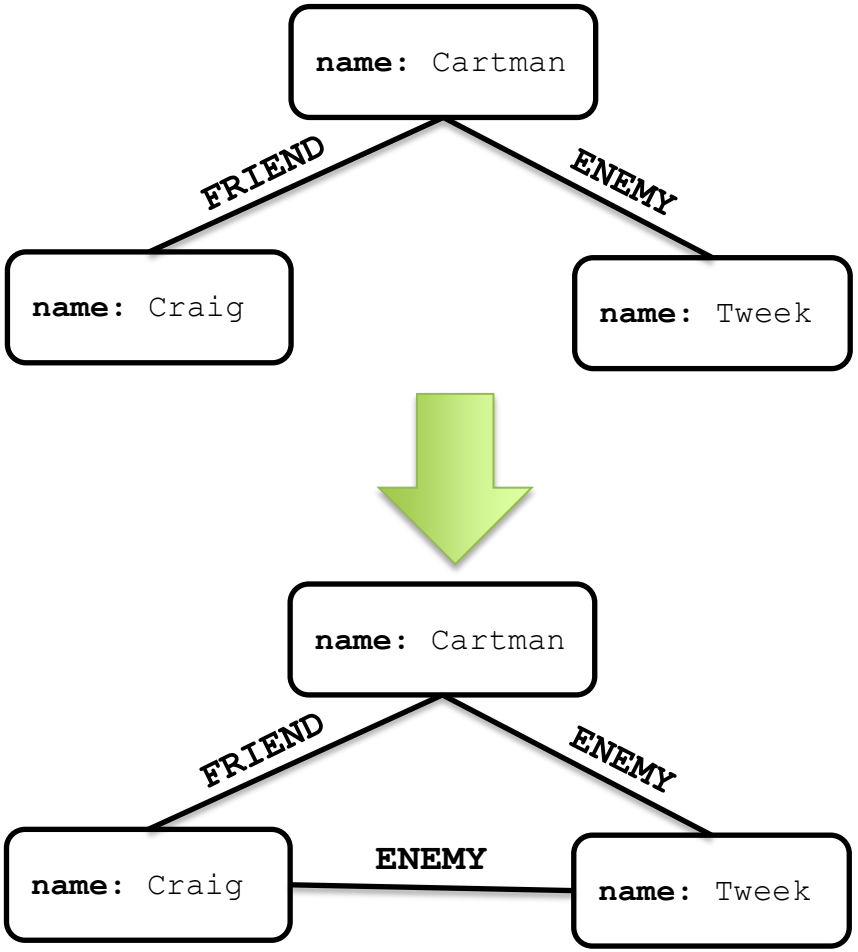
Structural Balance



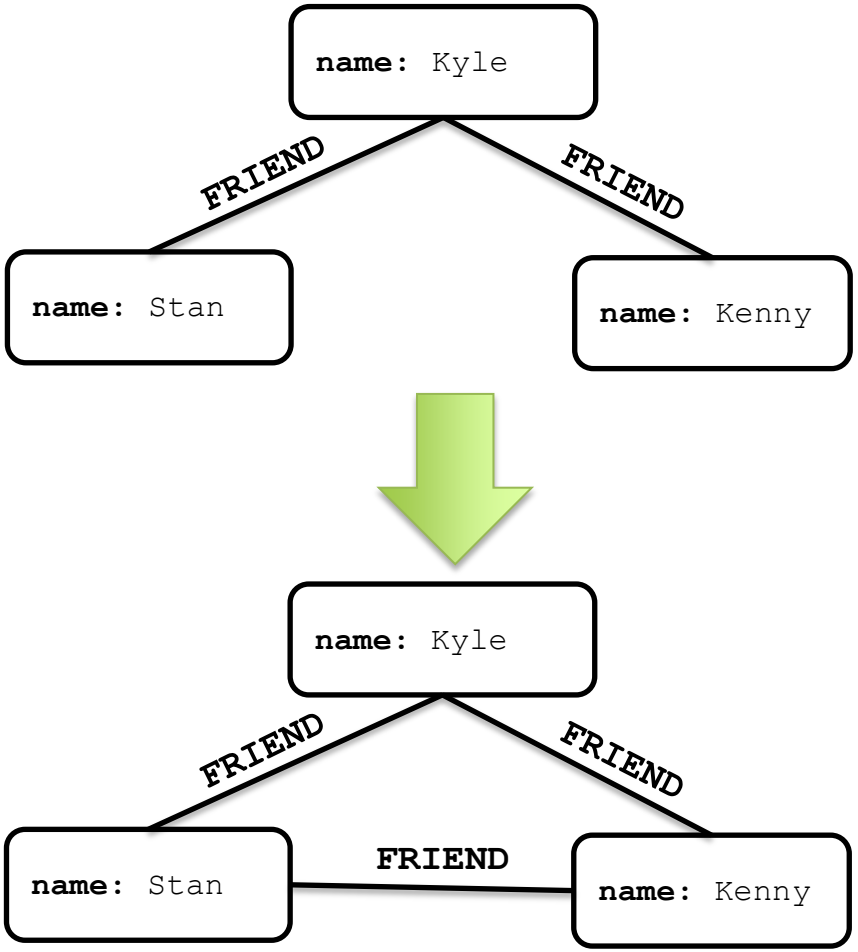
Structural Balance



Structural Balance

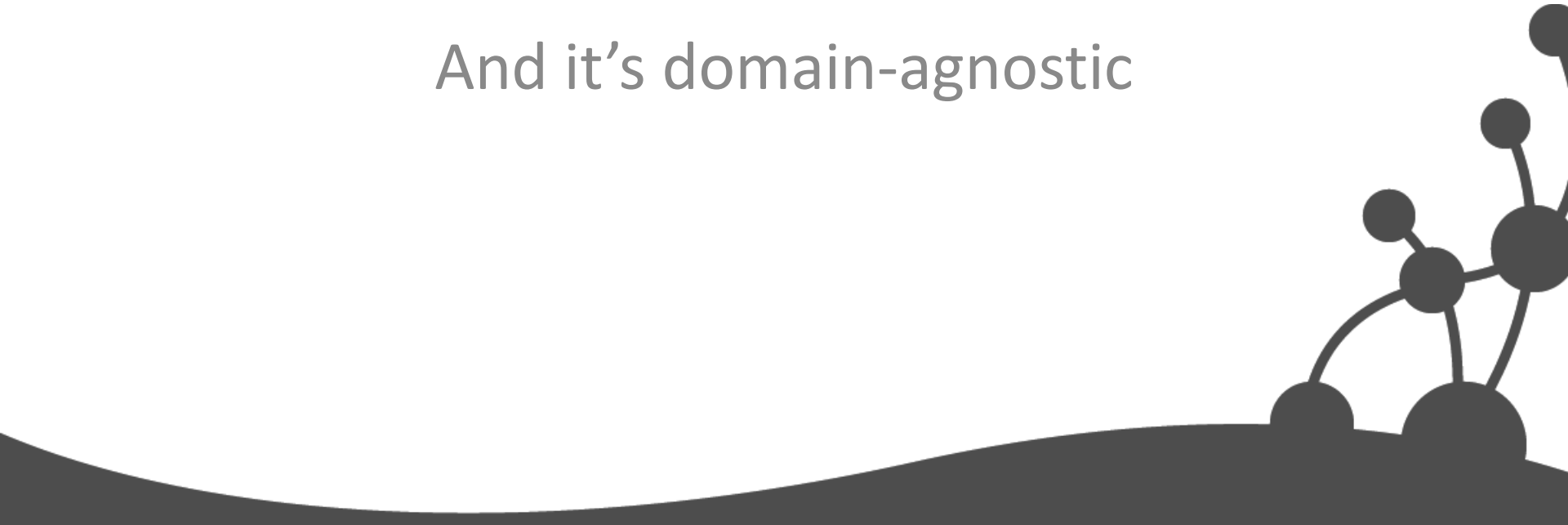


Structural Balance

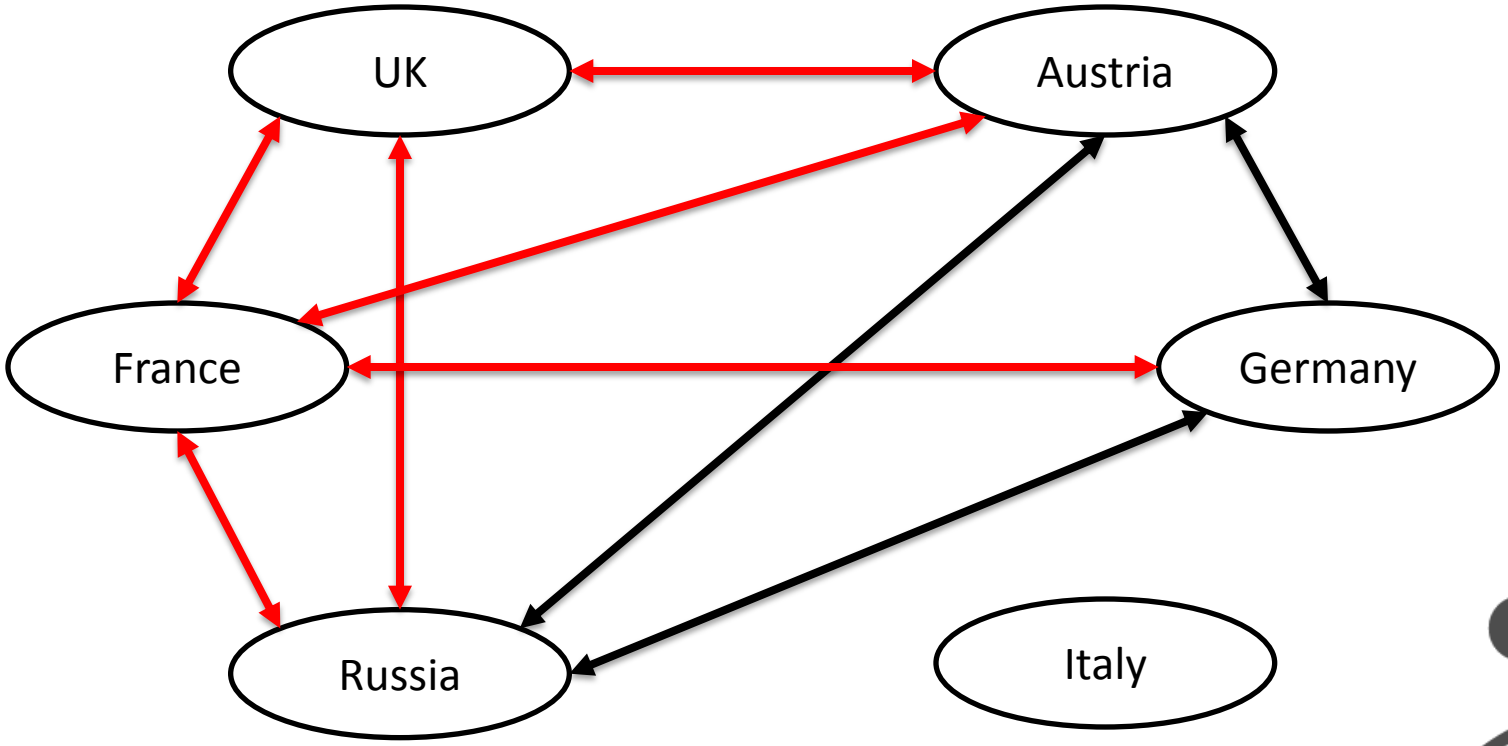


Structural Balance is a *key* predictive technique

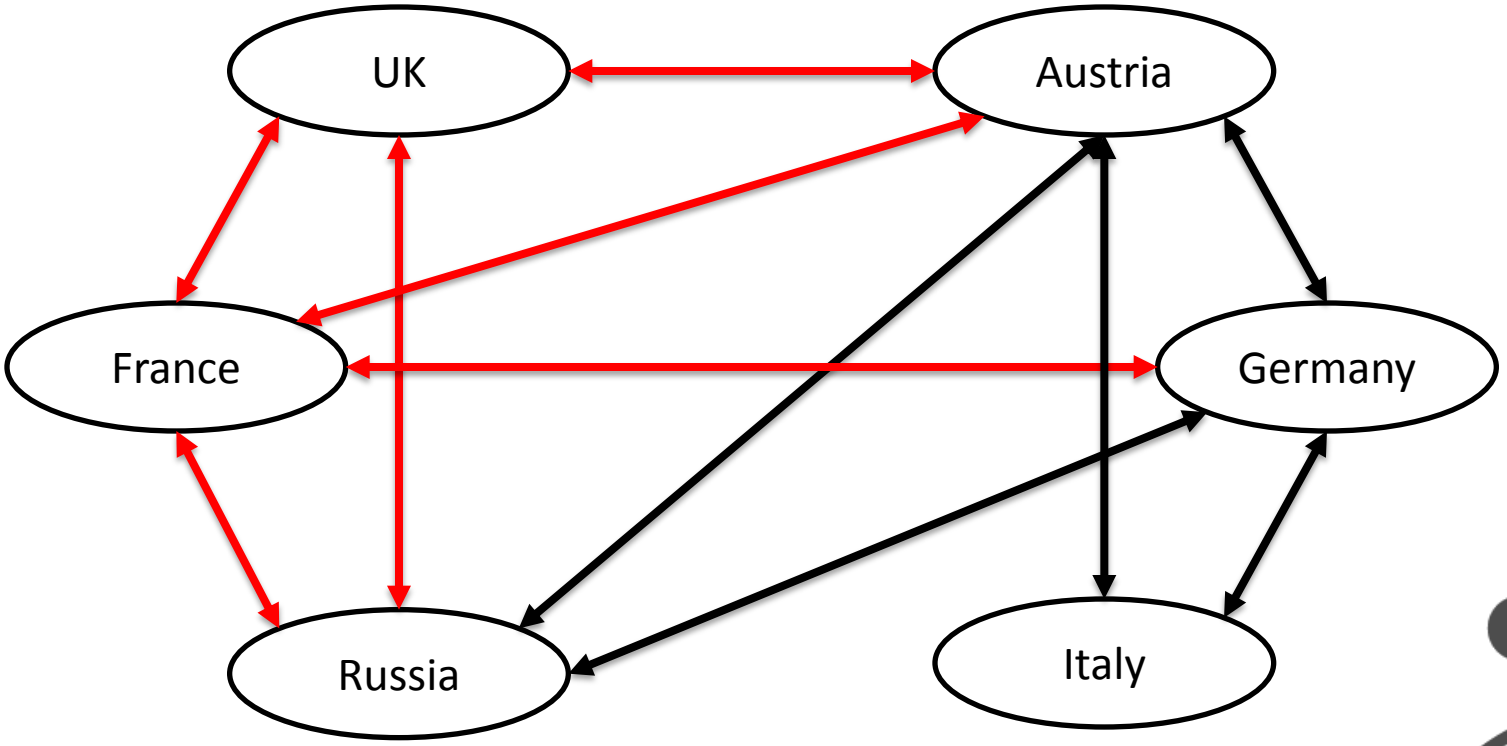
And it’s domain-agnostic



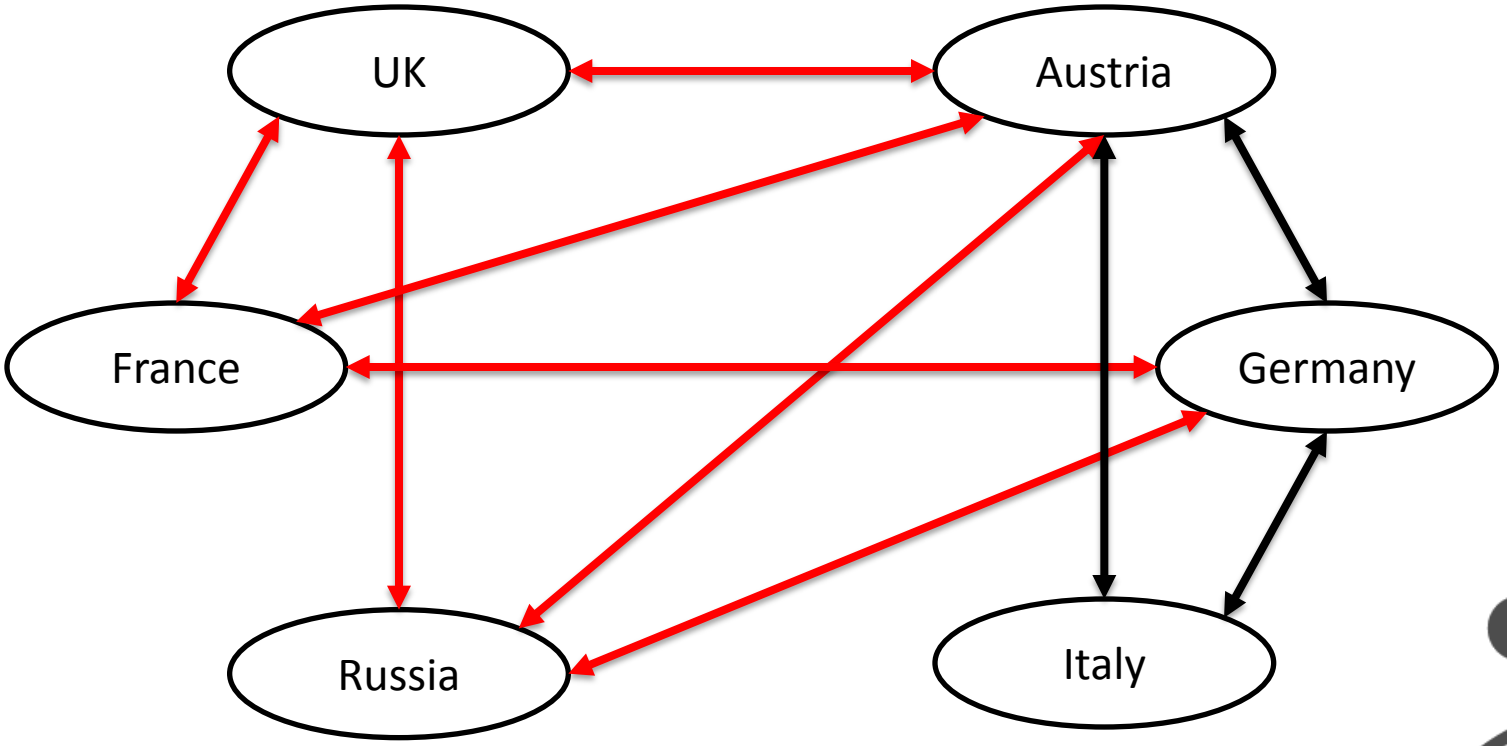
Allies and Enemies



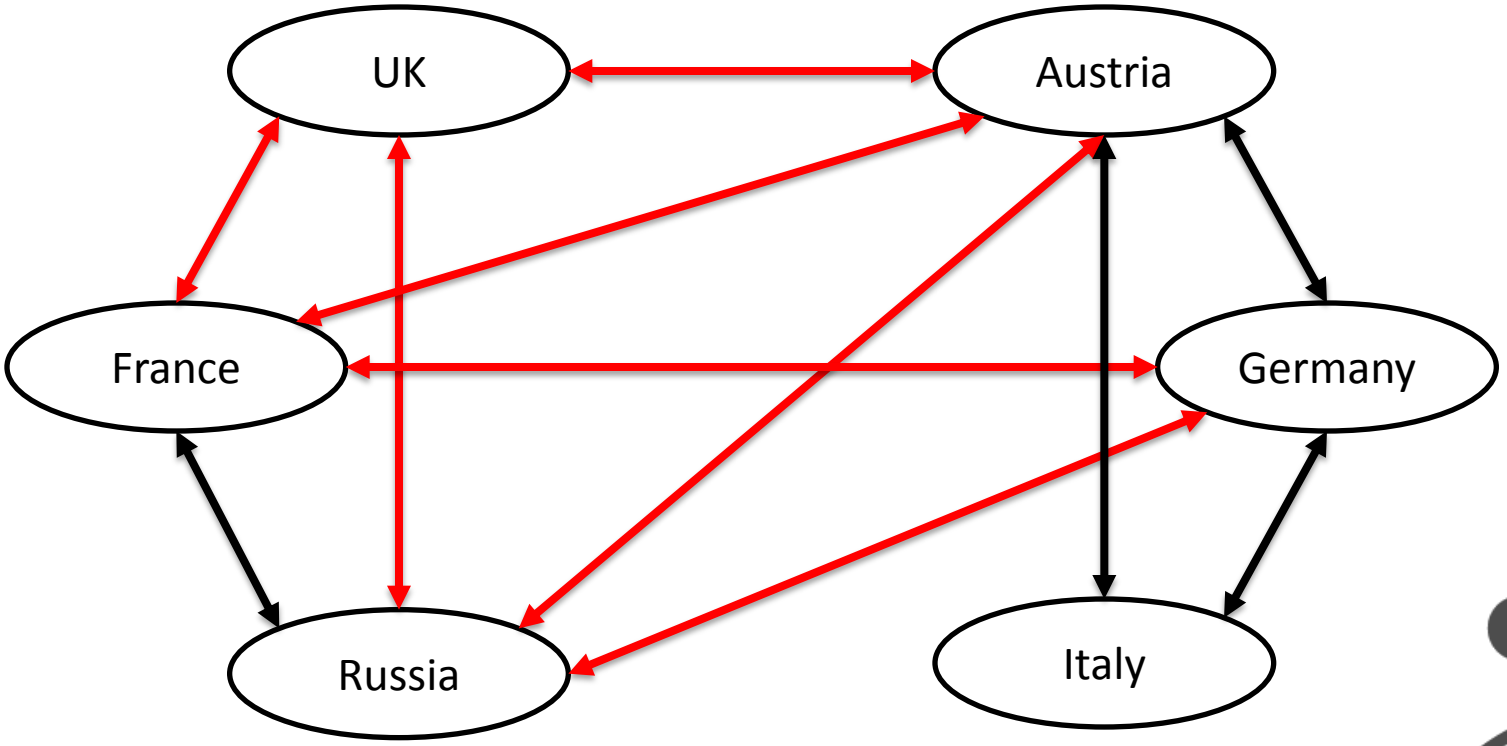
Allies and Enemies



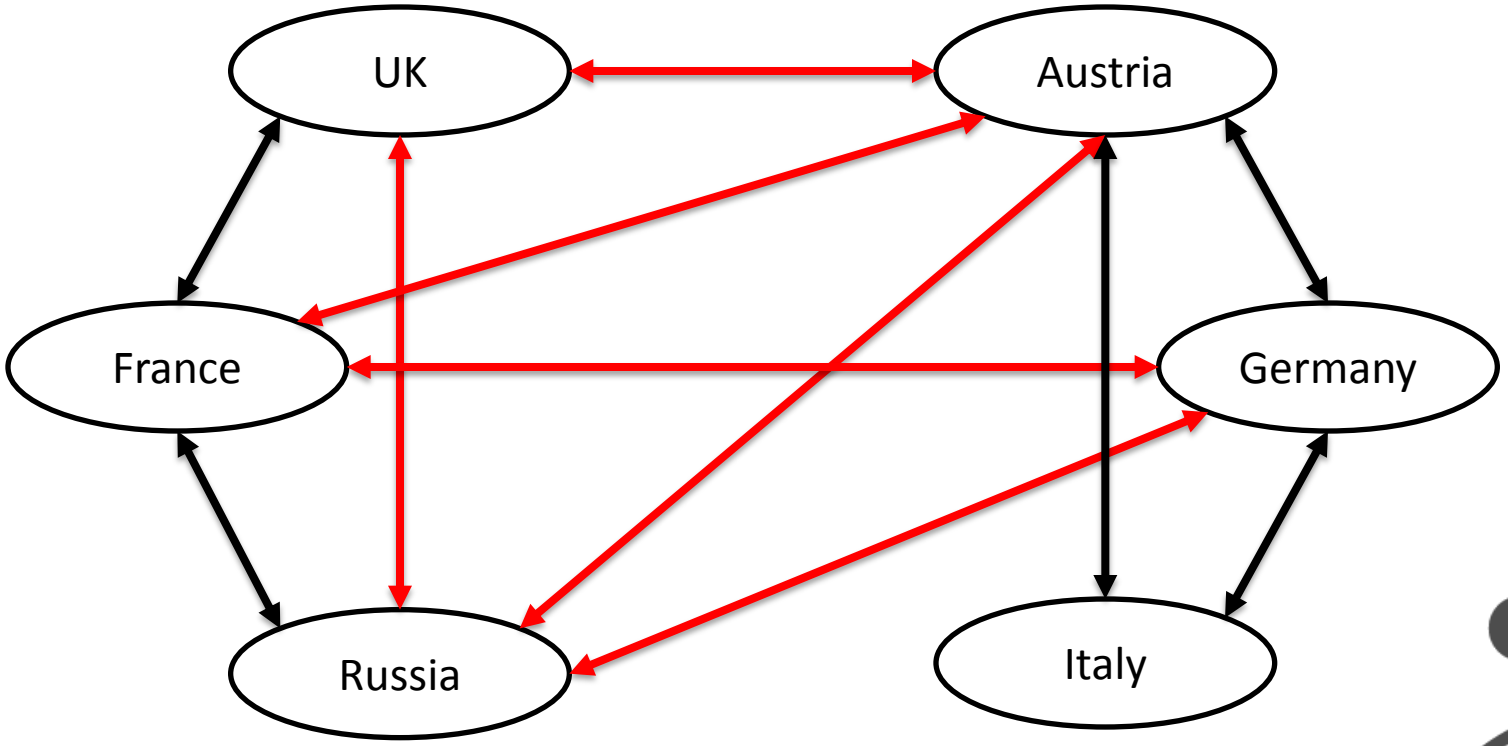
Allies and Enemies



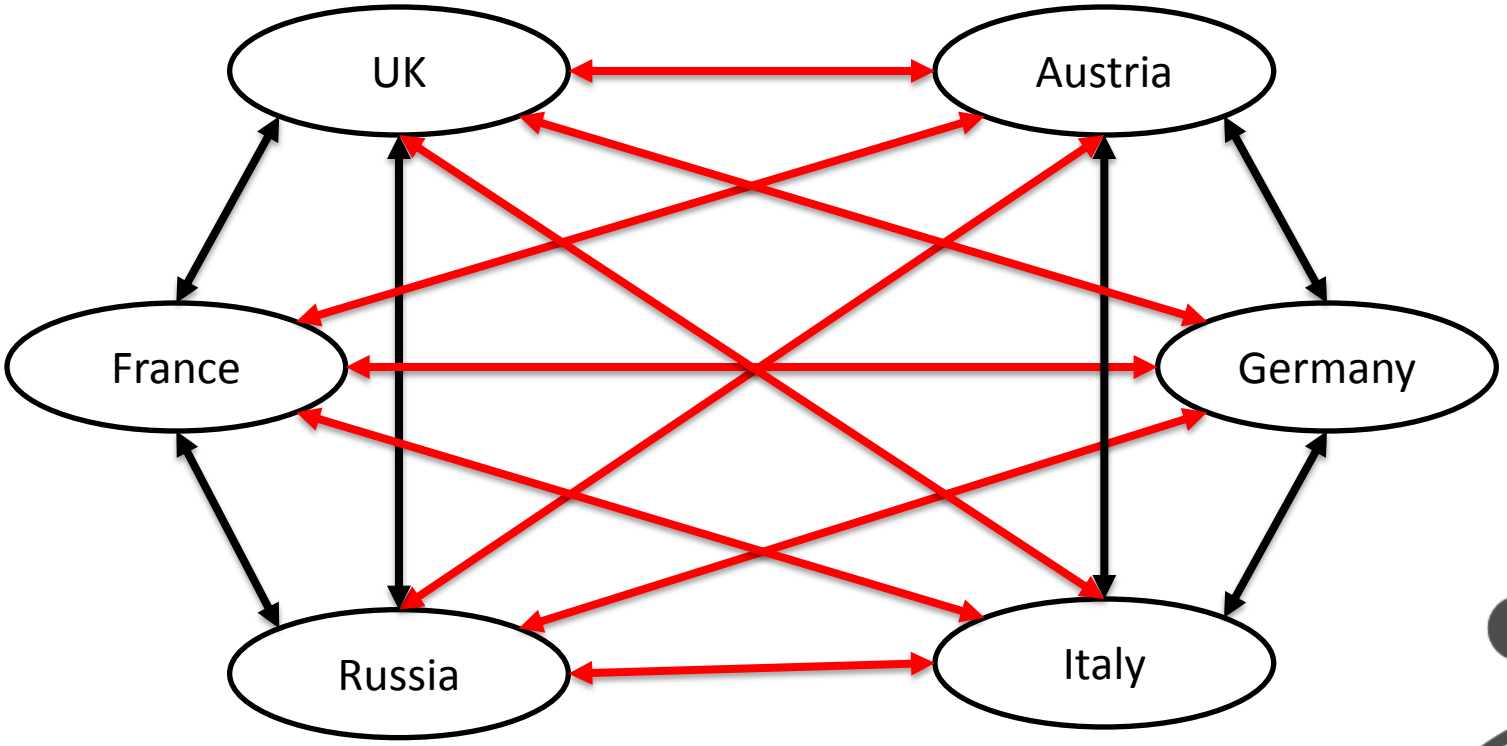
Allies and Enemies



Allies and Enemies

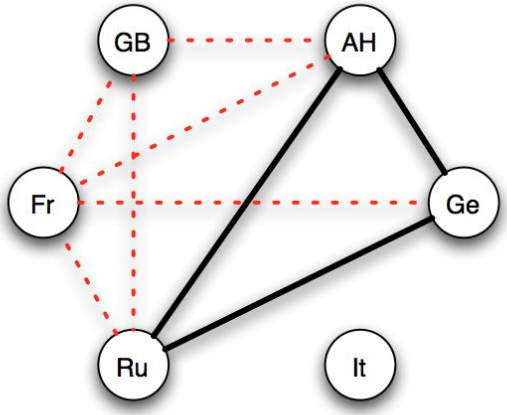


Allies and Enemies

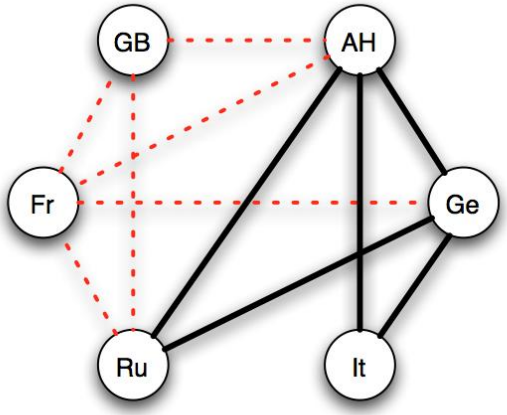


Predicting WWI

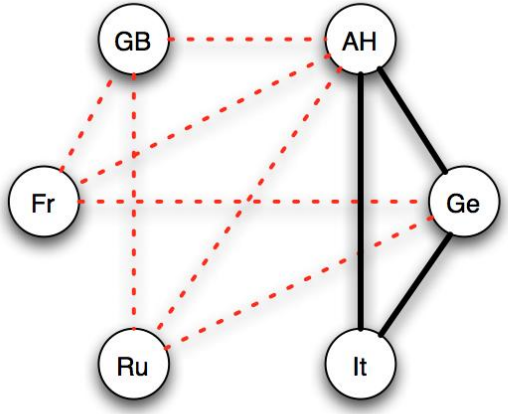
[Easley and Kleinberg]



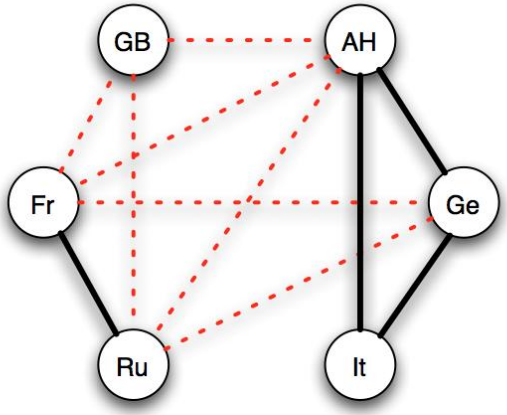
(a) *Three Emperors' League 1872–81*



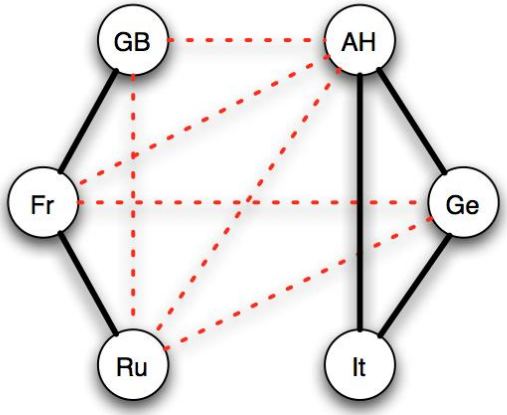
(b) *Triple Alliance 1882*



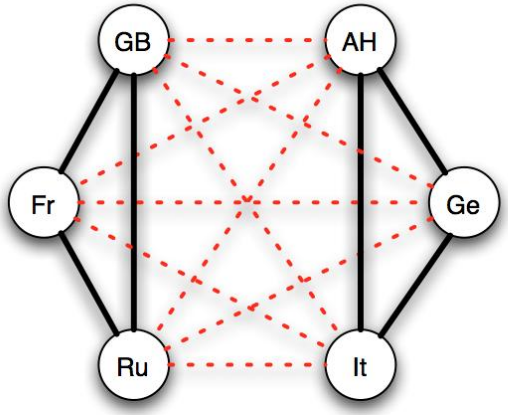
(c) *German-Russian Lapse 1890*



(d) *French-Russian Alliance 1891–94*



(e) *Entente Cordiale 1904*



(f) *British Russian Alliance 1907*

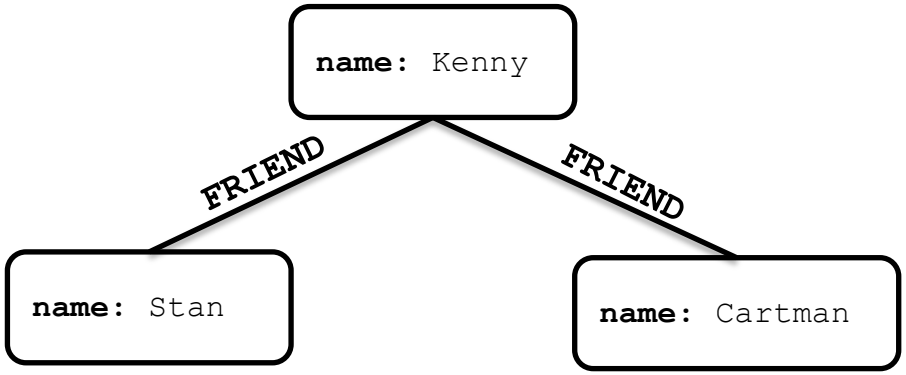
Strong Triadic Closure

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

[Wikipedia]

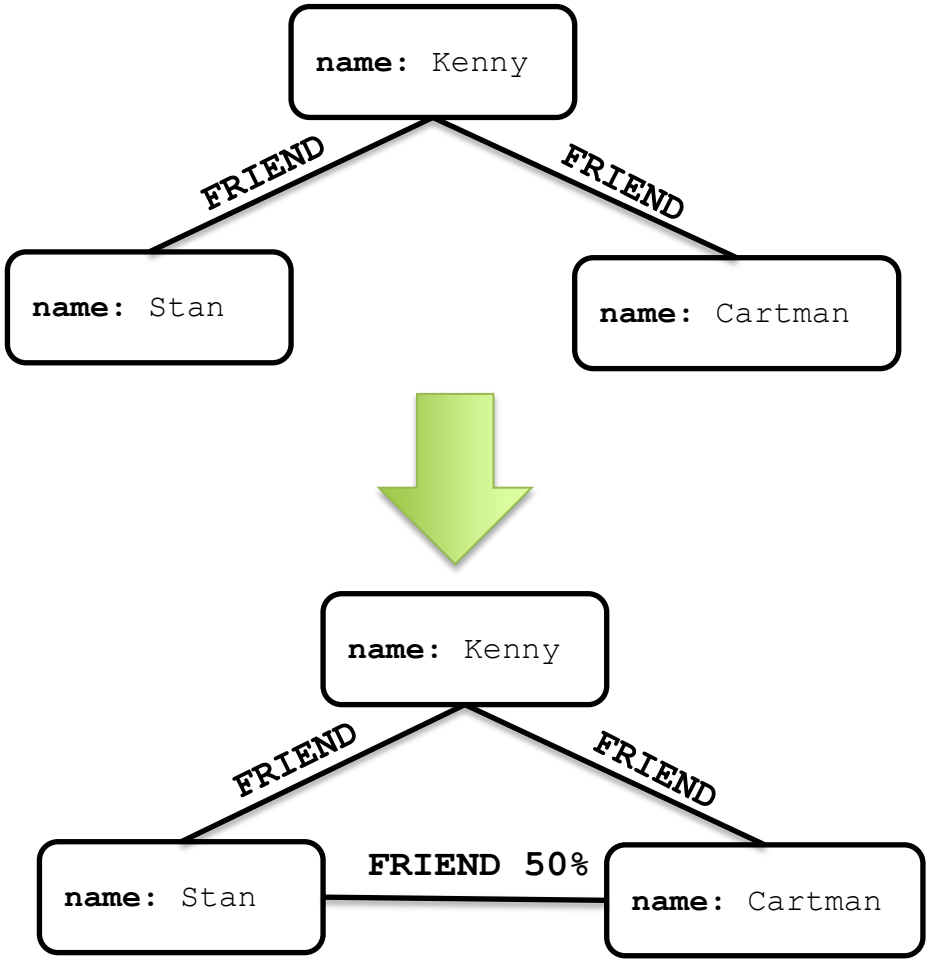
Triadic Closure

(weak relationship)



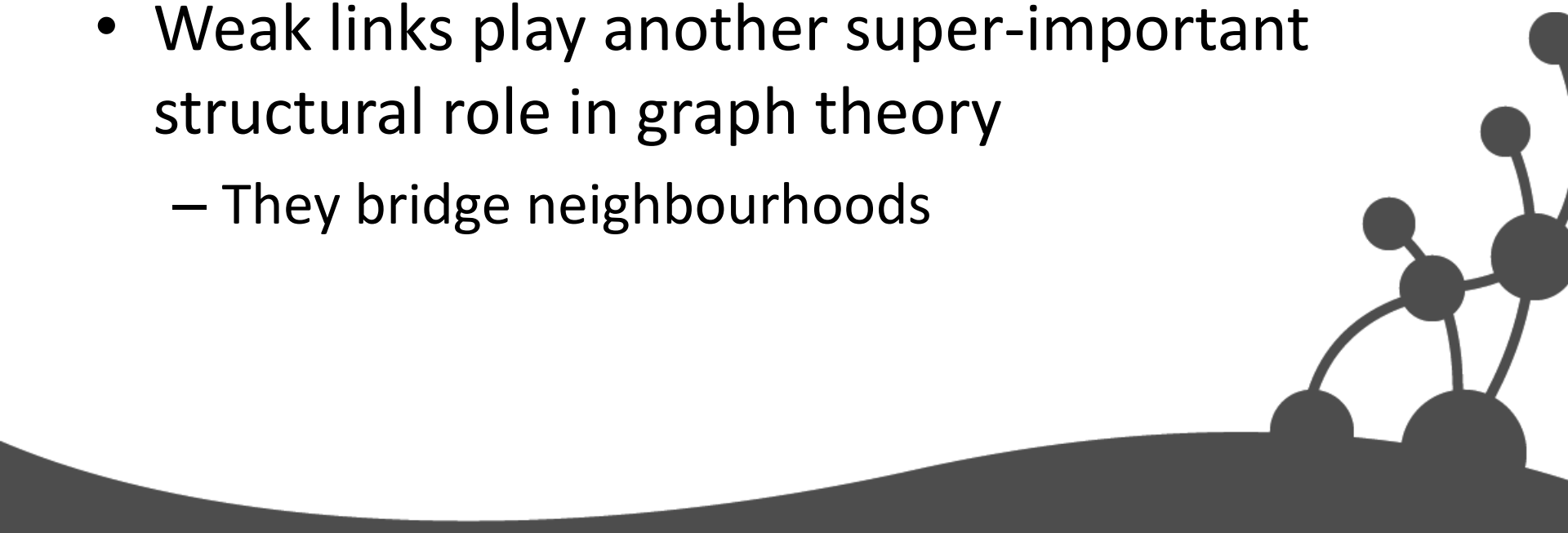
Triadic Closure

(weak relationship)

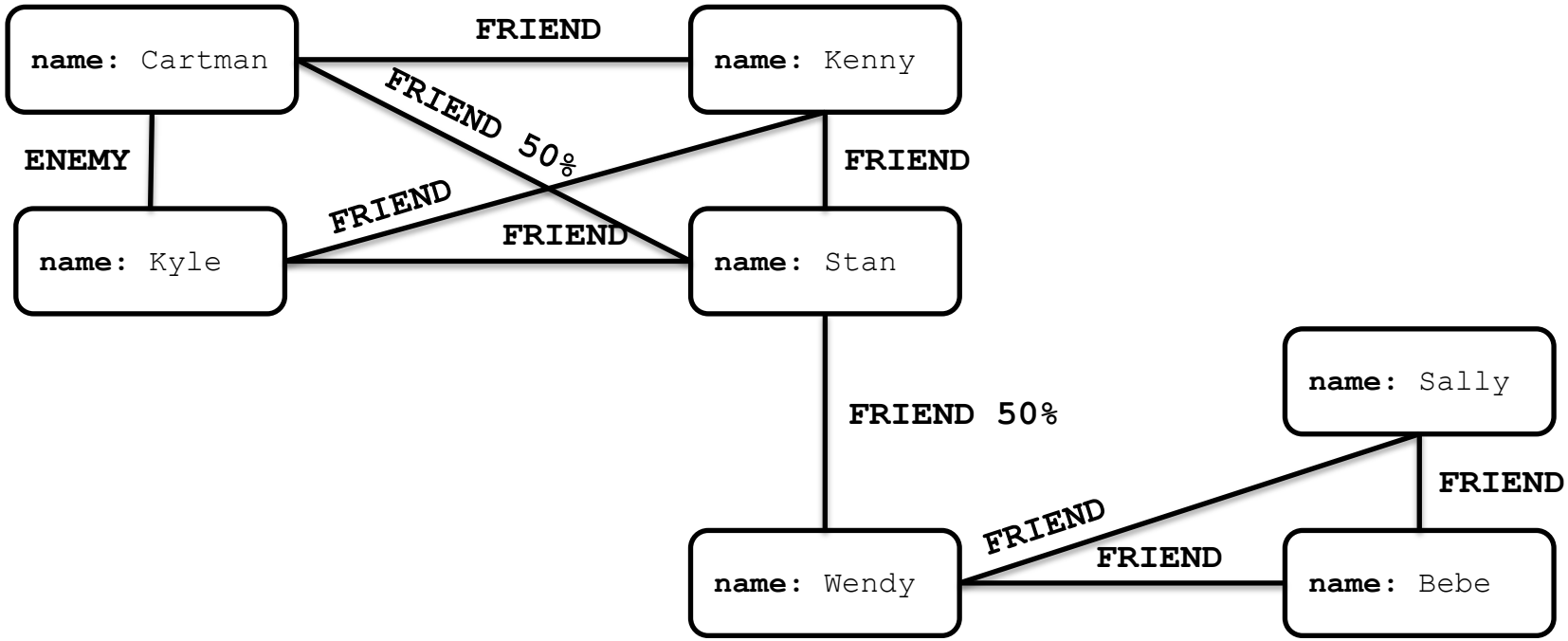


Weak relationships

- Relationships can have “strength” as well as intent
  - Think: weighting on a relationship in a property graph
- Weak links play another super-important structural role in graph theory
  - They bridge neighbourhoods



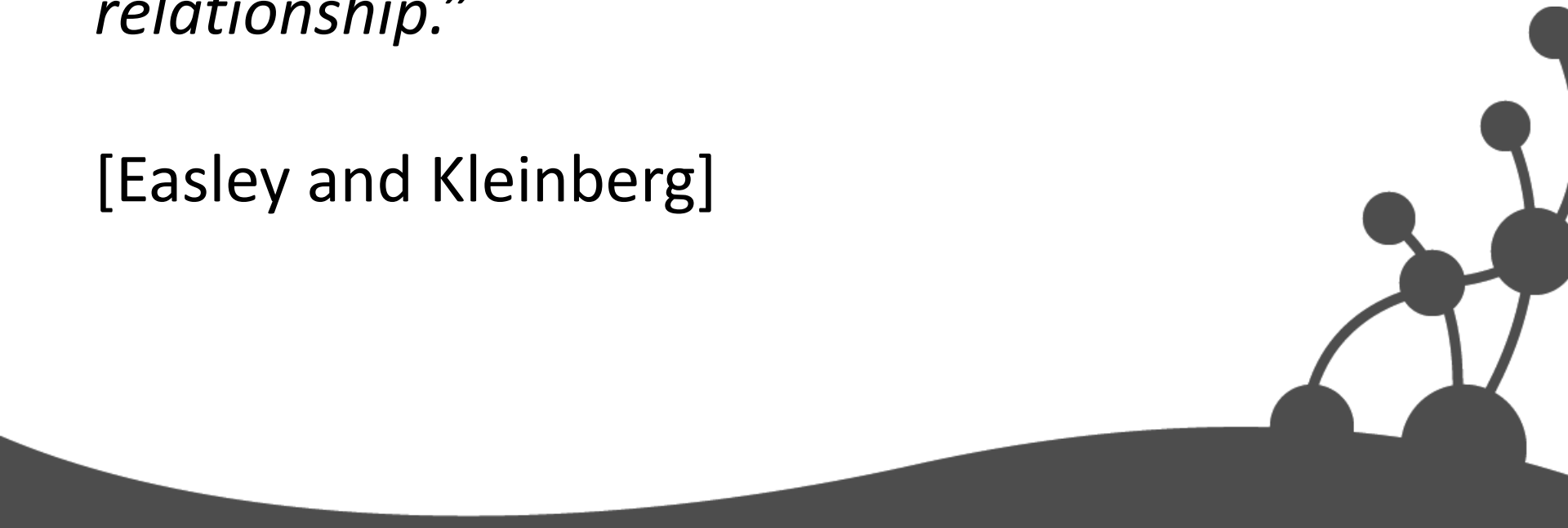
Local Bridges



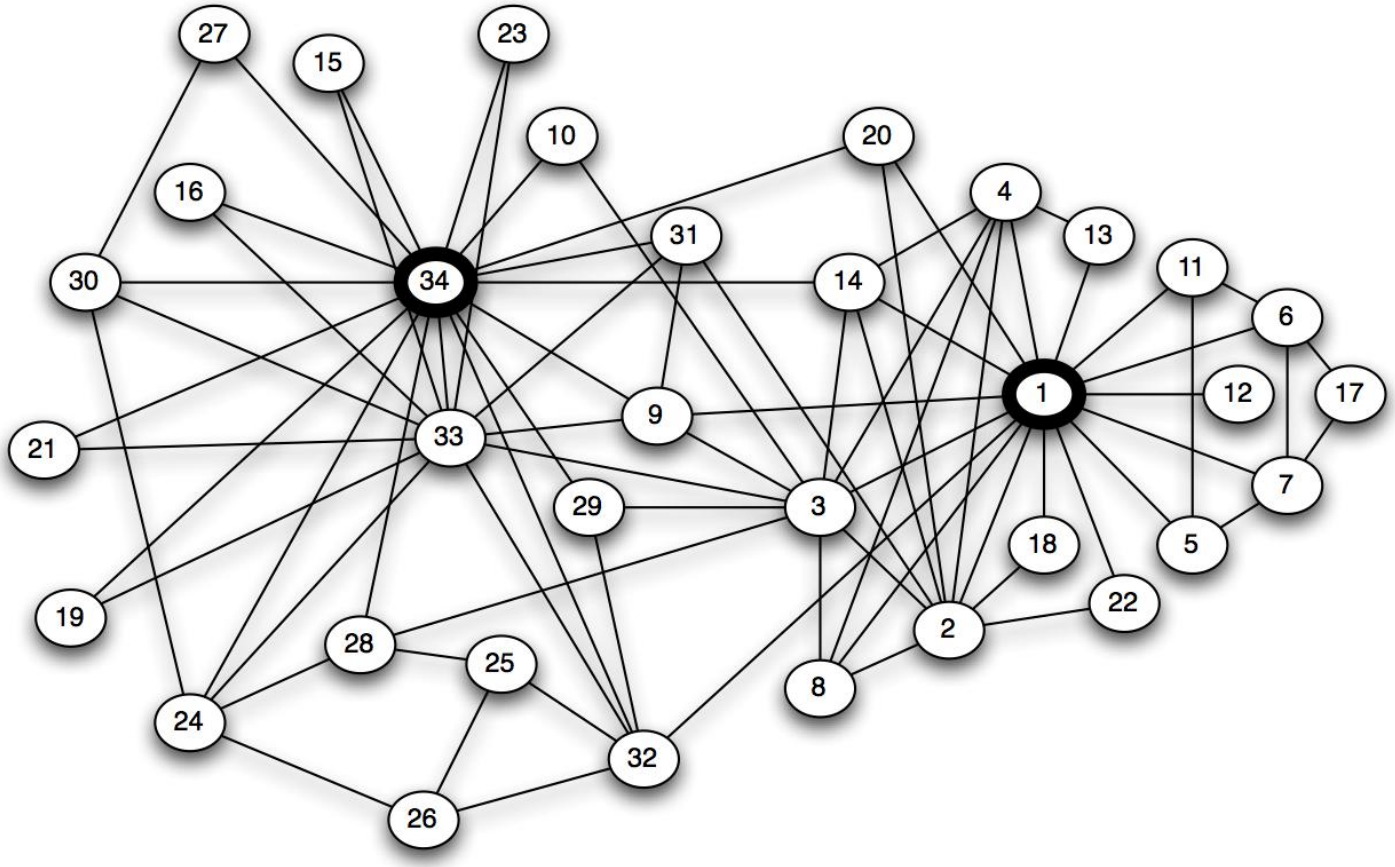
Local Bridge Property

*"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]

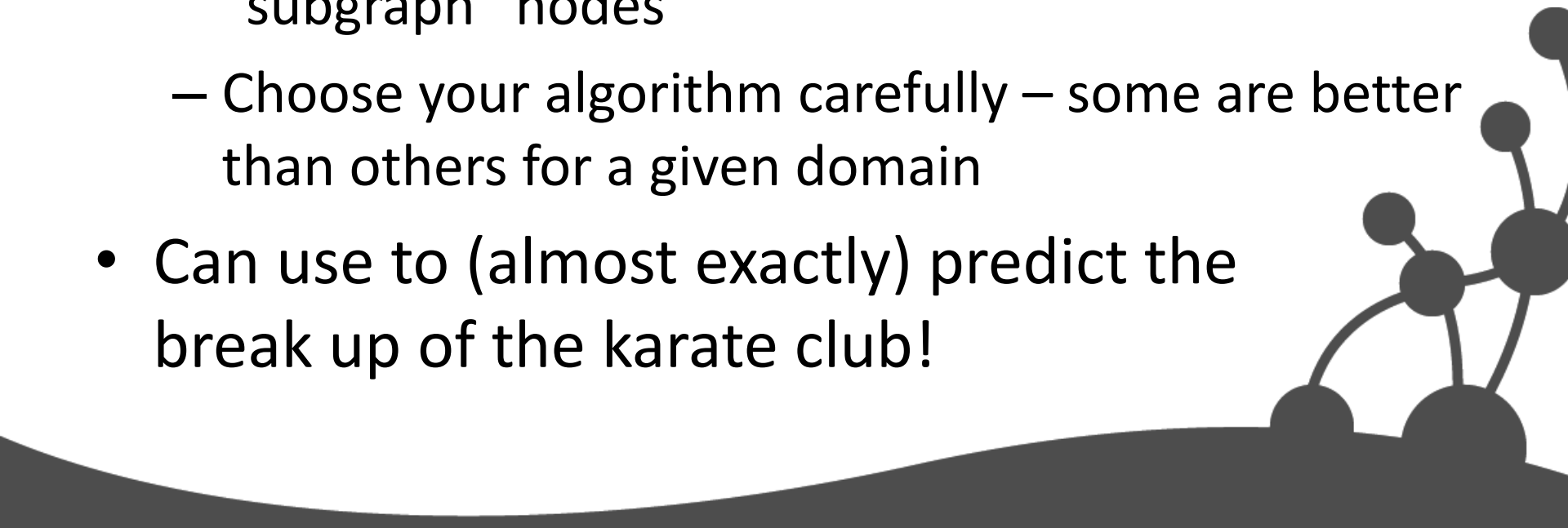


University Karate Club

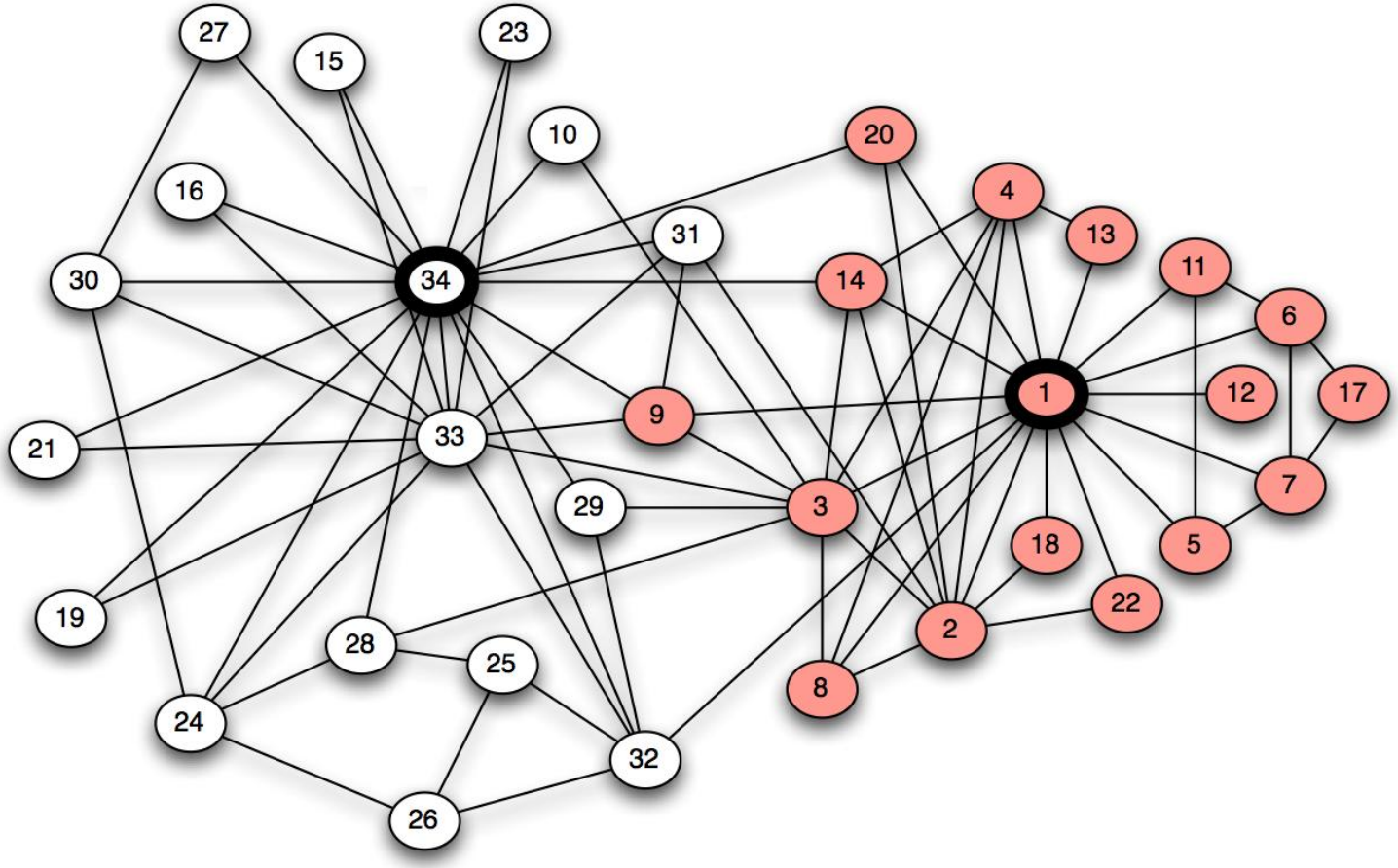


Graph Partitioning

- (NP) Hard problem
  - Recursively remove the spanning links between dense regions
  - Or recursively merge nodes into ever larger “subgraph” nodes
  - Choose your algorithm carefully – some are better than others for a given domain
- Can use to (almost exactly) predict the break up of the karate club!



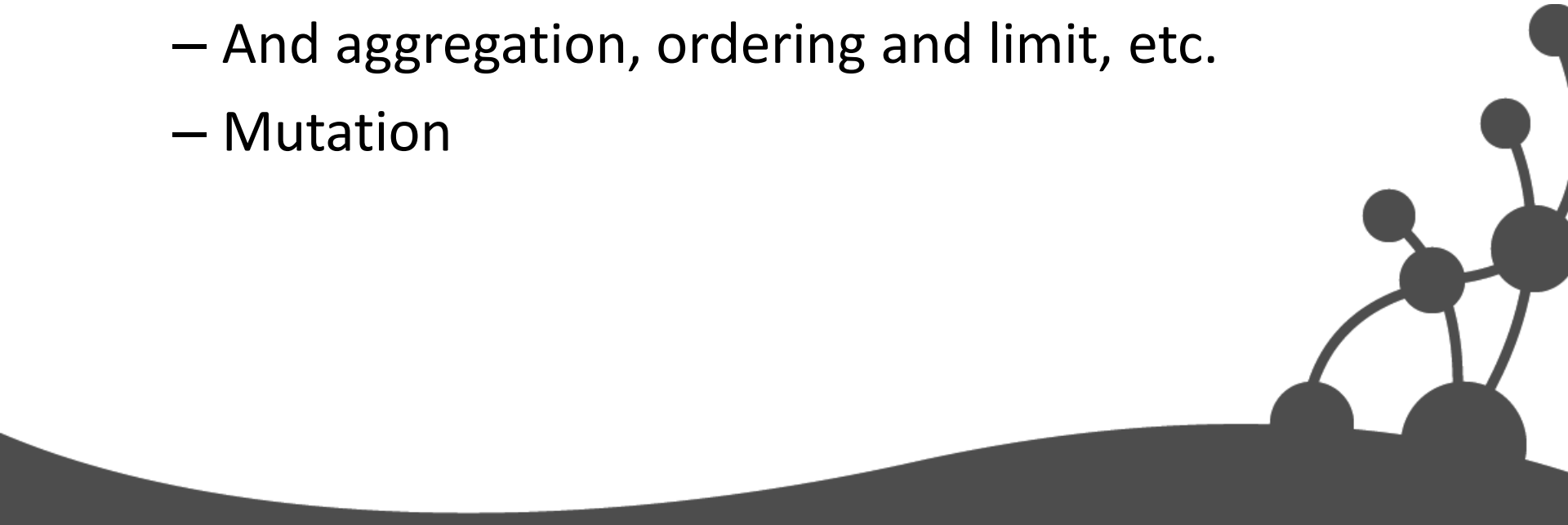
University Karate Clubs




john
crane
Ltd

Cypher

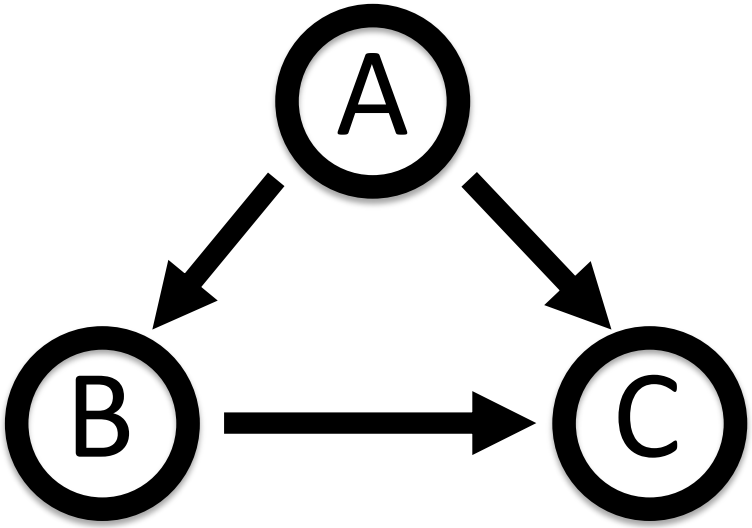
- Declarative graph pattern matching language
  - “SQL for graphs”
  - Columnar results
- Supports graph matching queries
  - And aggregation, ordering and limit, etc.
  - Mutation



Cypher is Declarative

- Imperative
  - follow relationship
  - breadth-first vs depth-first
  - explicit algorithm
- Declarative
  - specify starting point
  - specify desired outcome
  - algorithm adaptable
  - based on query

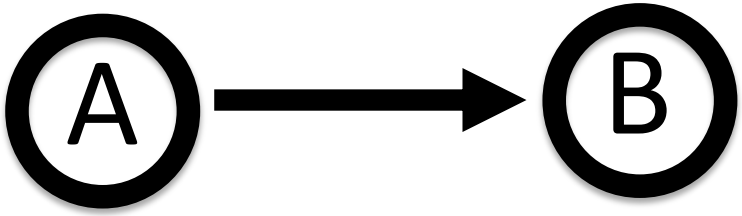
Cypher is a pattern matching language



Un-named Nodes & Rels

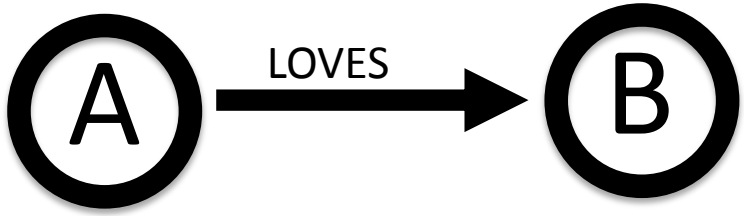
```
() --> ()
```

Un-named Relationship



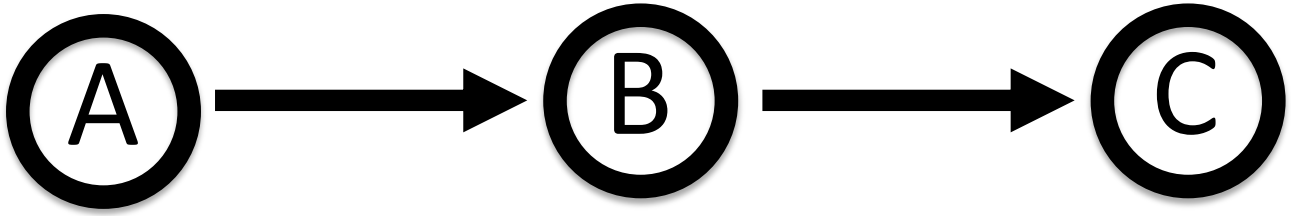
(A) --> (B)

ASCII Art Patterns

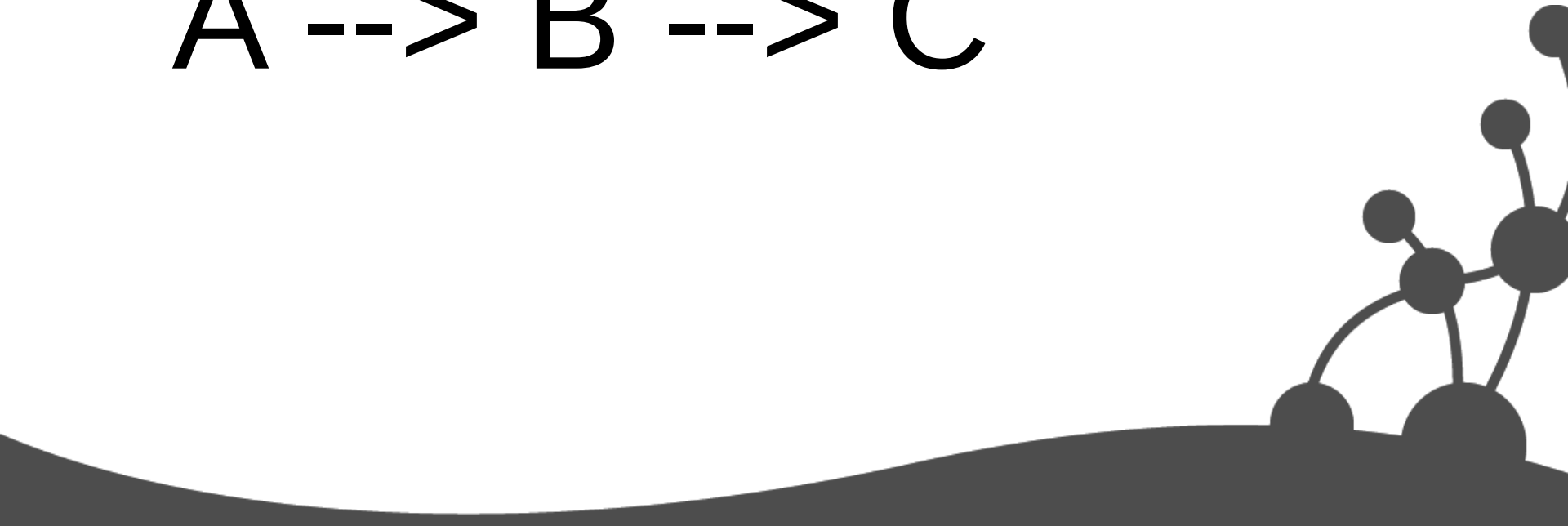



```
A -[:LOVES]-> B
```

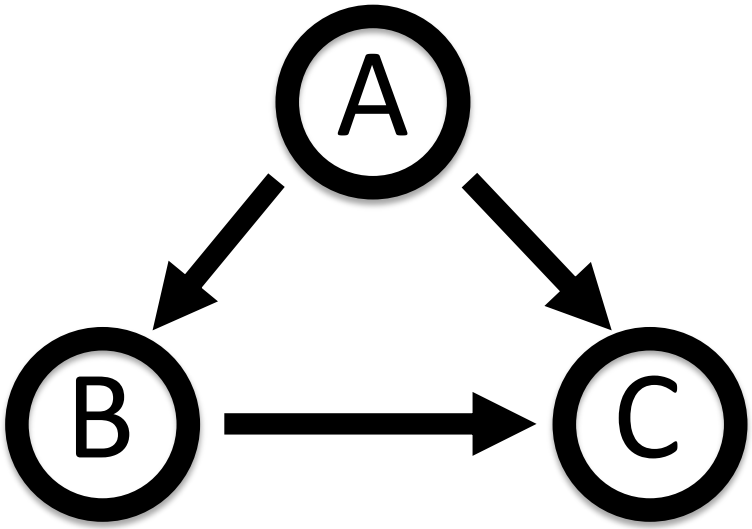
ASCII Art Patterns



A --> B --> C



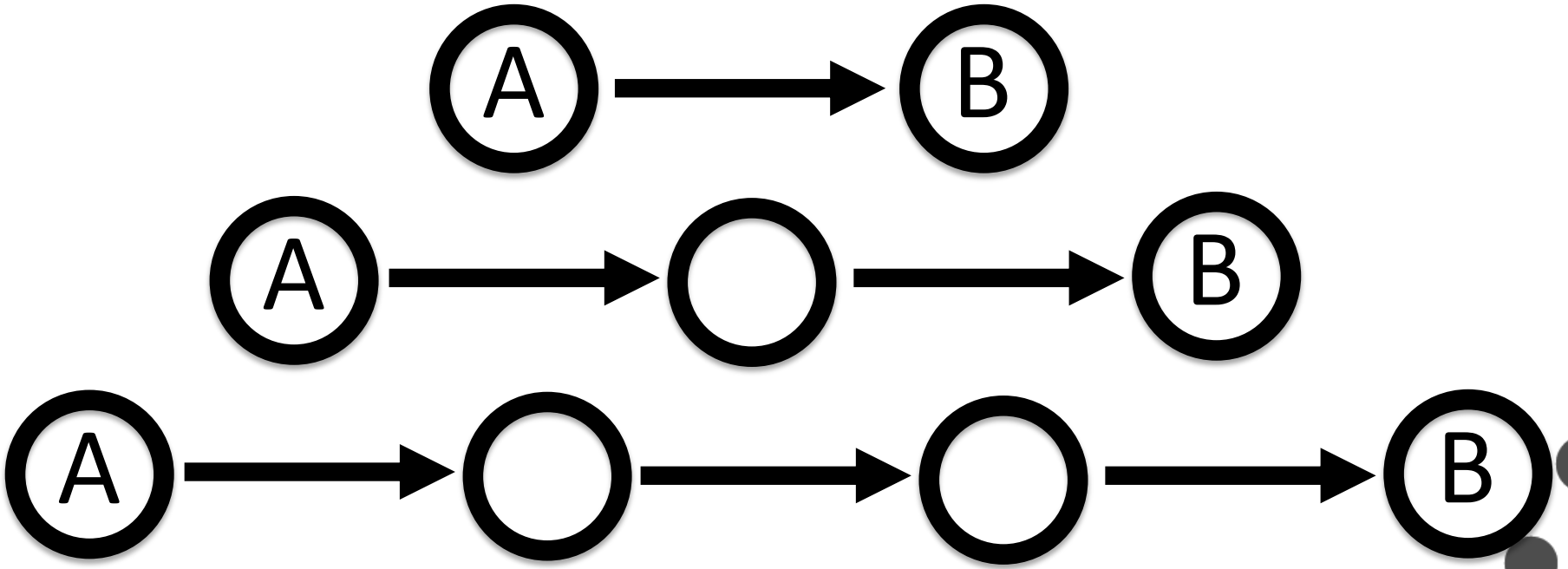
ASCII Art Patterns



A --> B --> C, A --> C

A --> B --> C <-- A

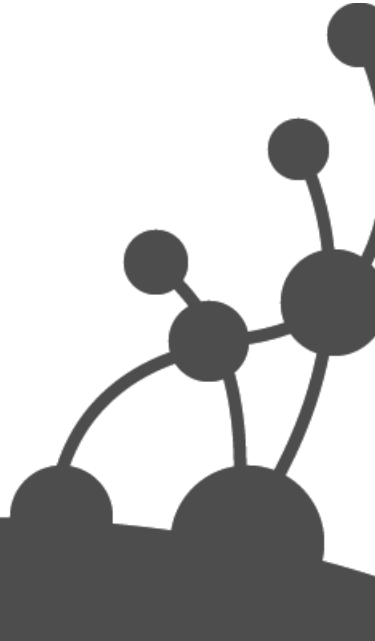
Variable Length Paths



A -[*]-> B

Optional Relationships

A -[?]-> B



Example Query

- The top 5 most frequently appearing companions:

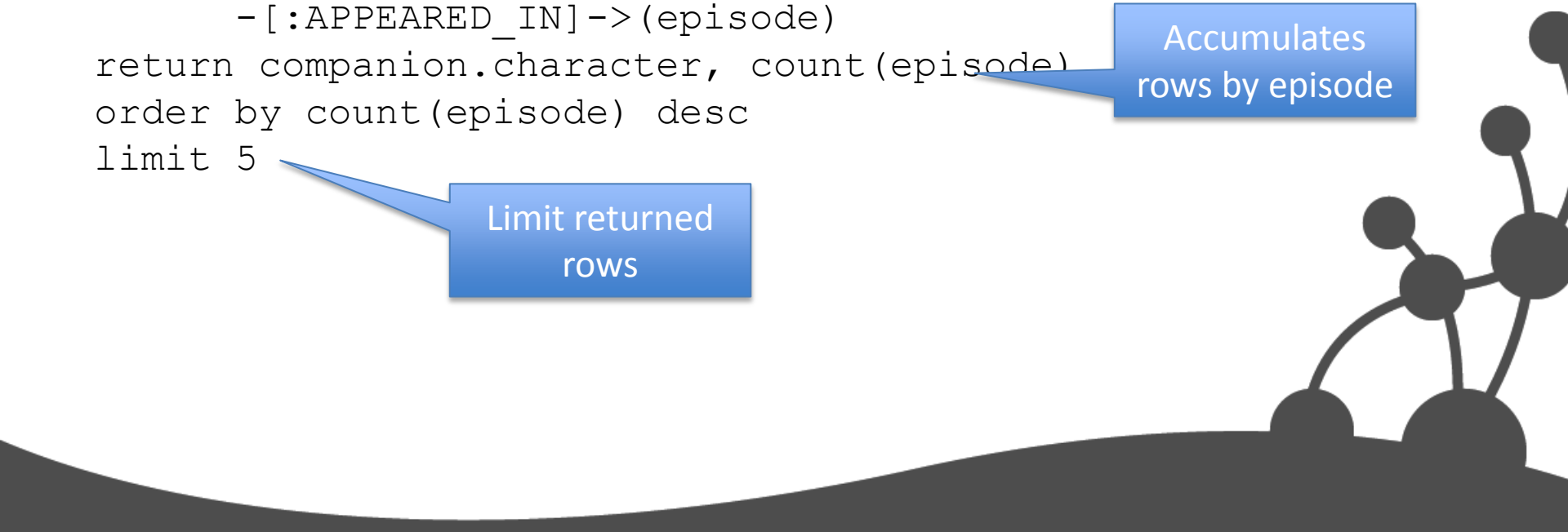
```
start doctor=node:characters(character = 'Doctor')
match (doctor)<-[:COMPANION_OF]-(companion)
      -[:APPEARED_IN]->(episode)
return companion.character, count(episode)
order by count(episode) desc
limit 5
```

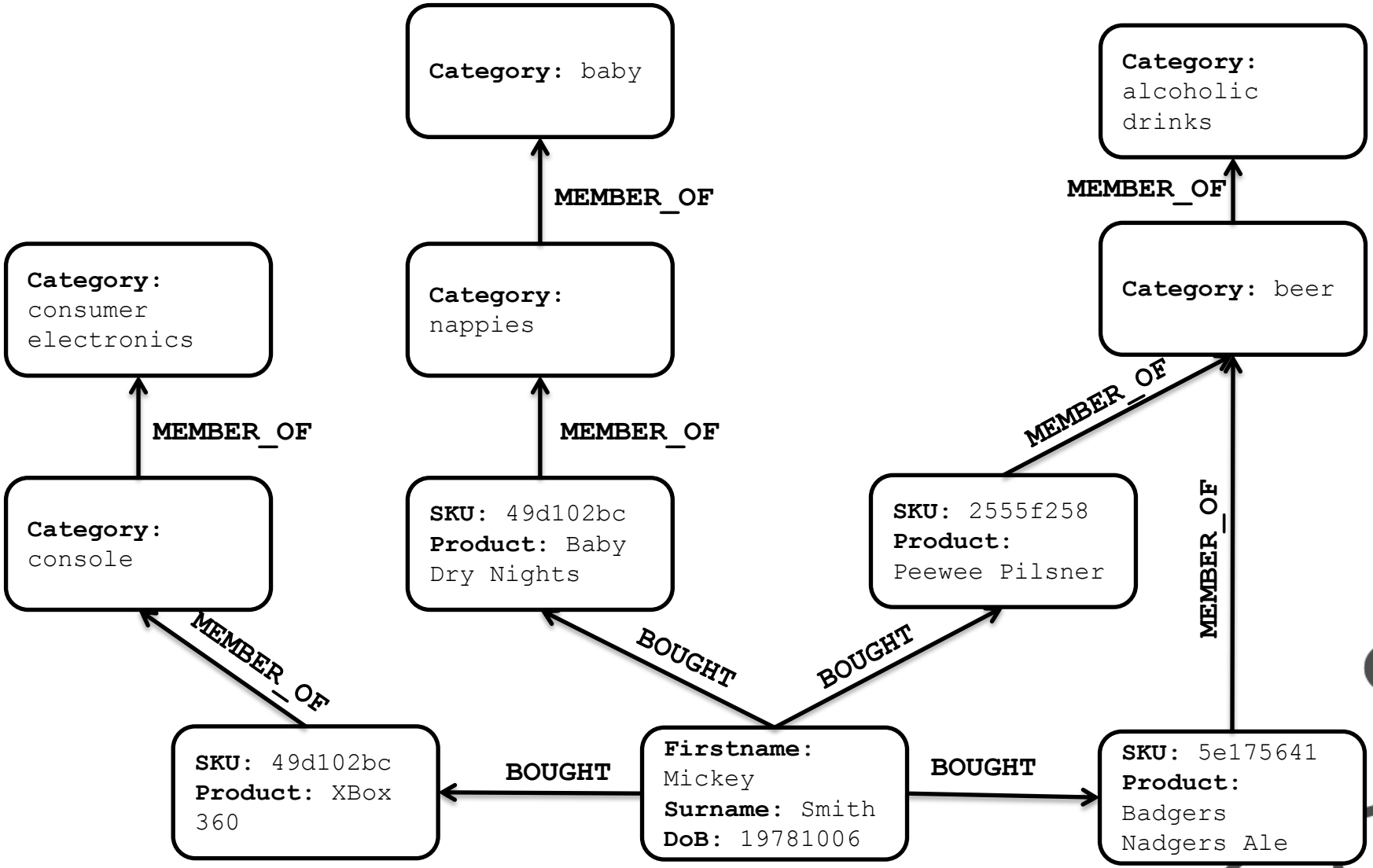
Start node from index

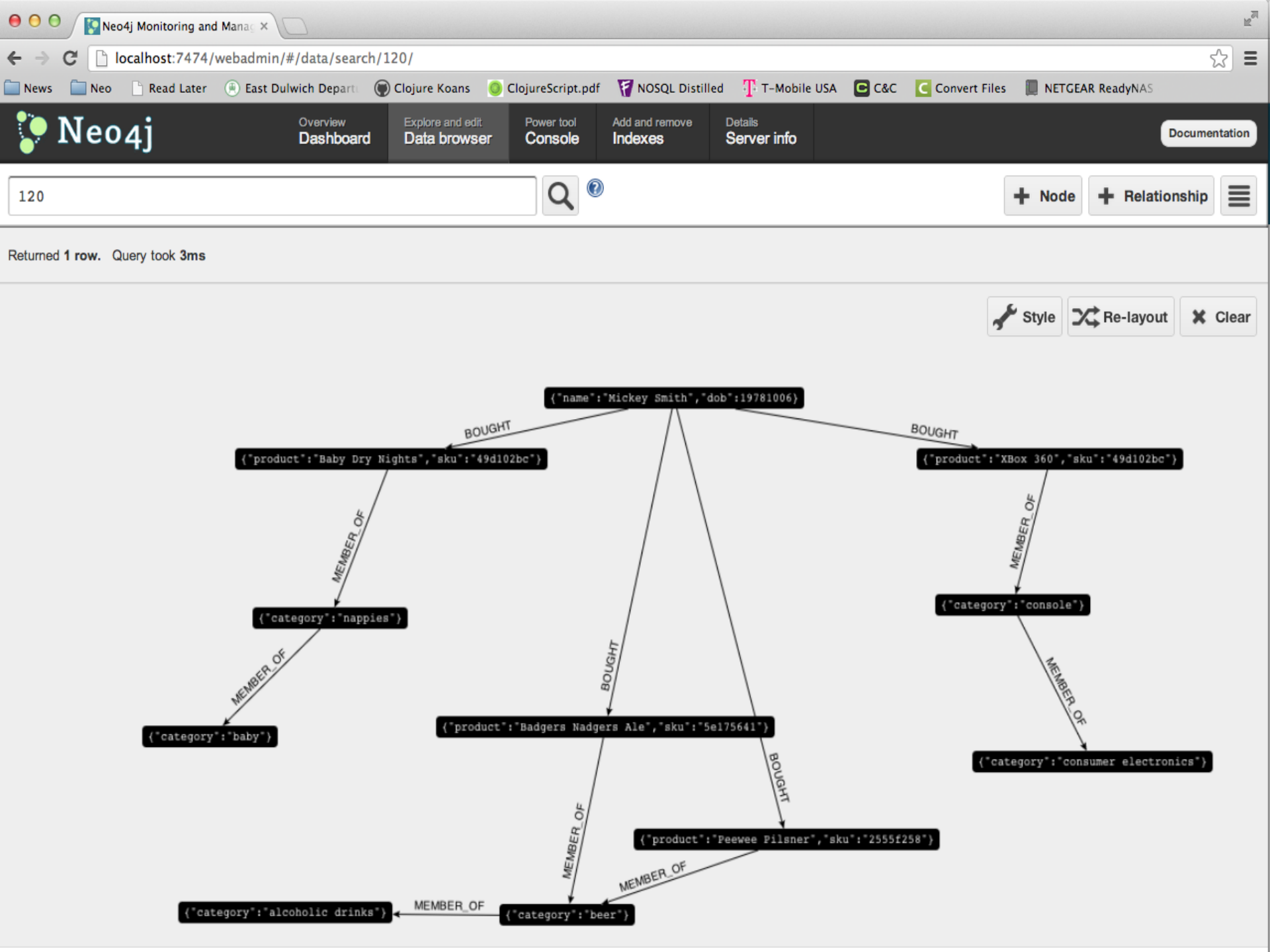
Subgraph pattern

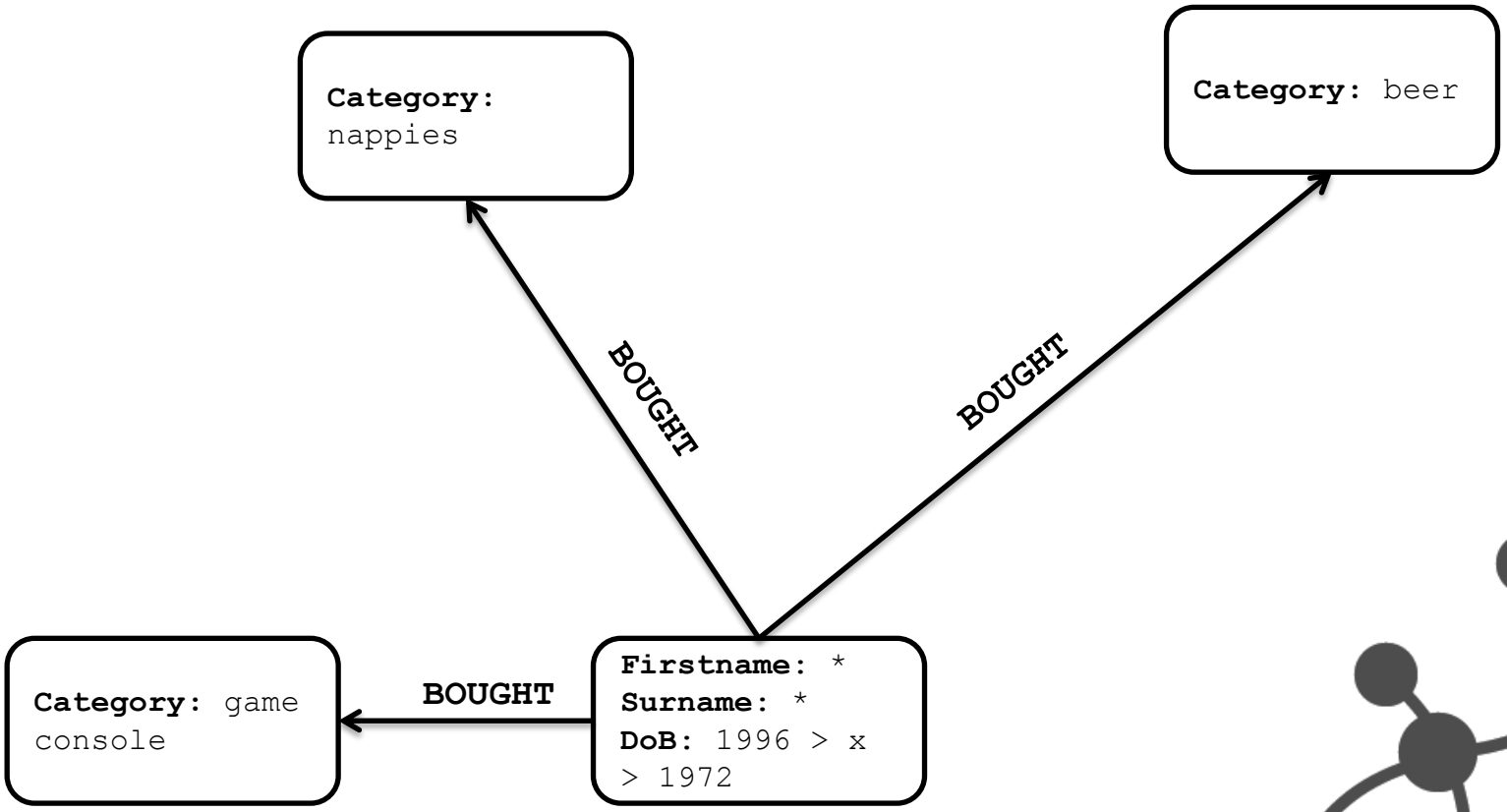
Accumulates rows by episode

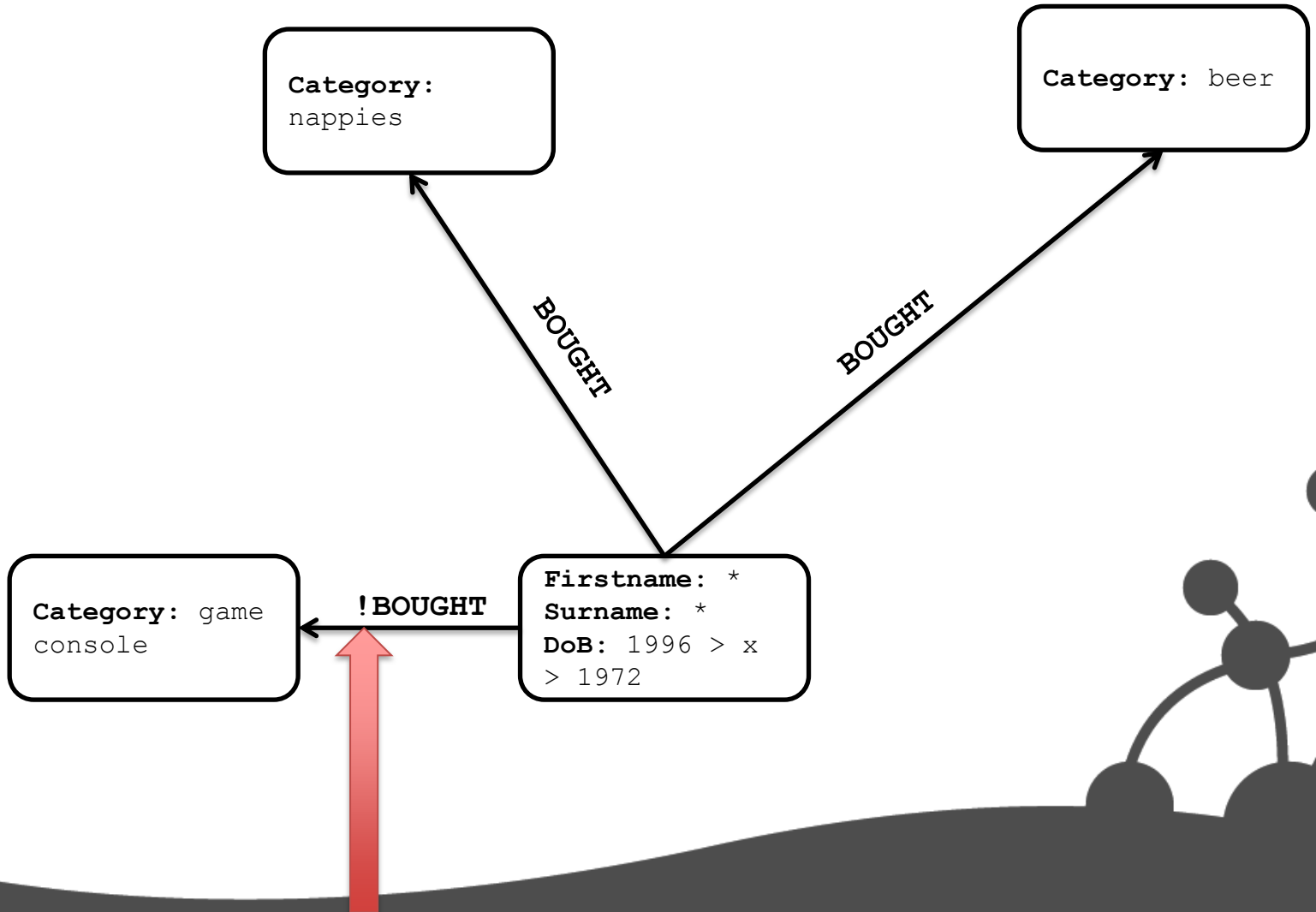
Limit returned rows

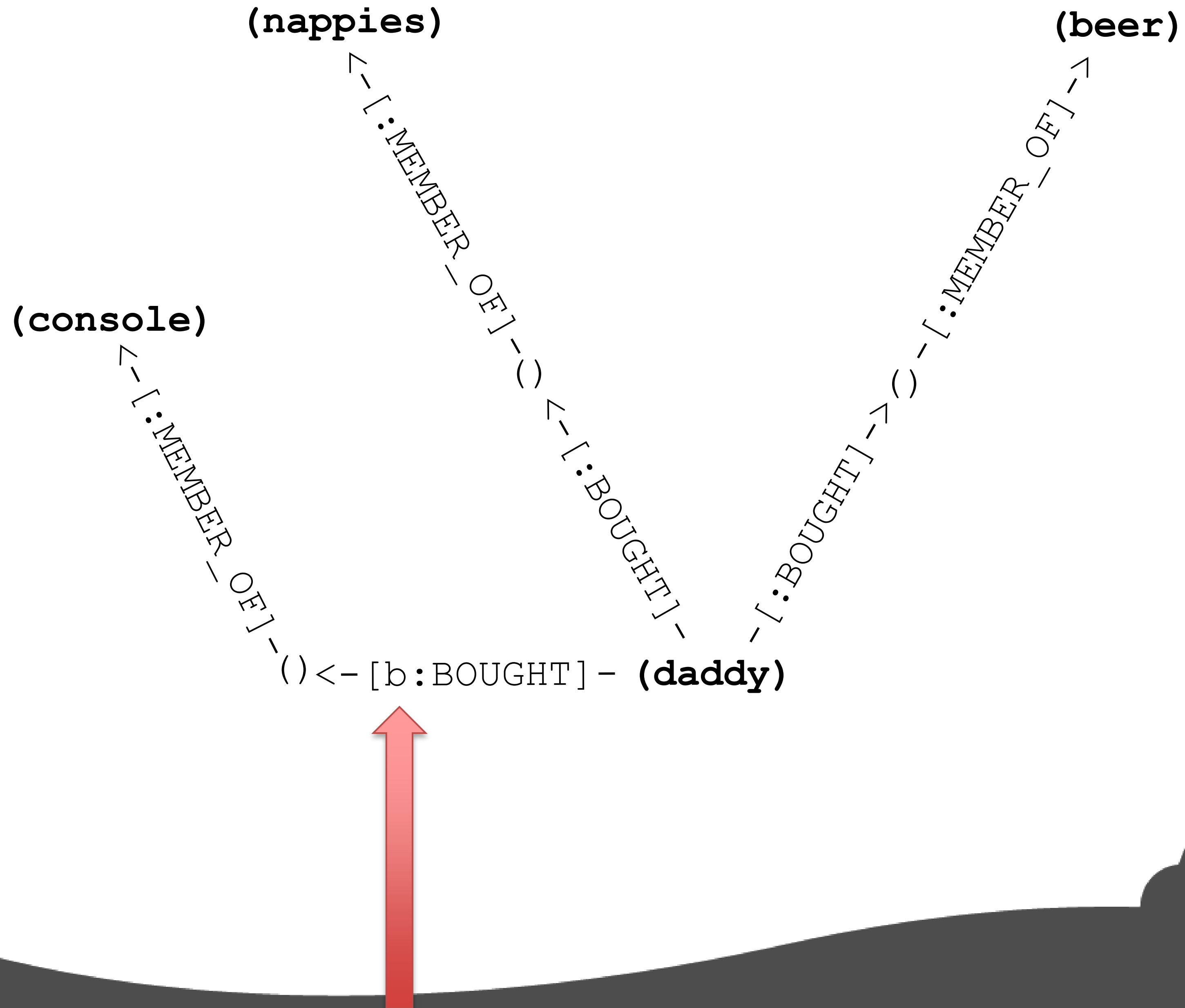



Category: baby
MEMBER_OF
Category: alcoholic drinks
MEMBER_OF
Category: consumer electronics
Category: nappies
Category: beer
MEMBER_OF
MEMBER_OF
MEMBER_OF
Category: console
SKU: 49d102bc Product: Baby Dry Nights
SKU: 2555f258 Product: Peewee Pilsner
MEMBER_OF
BOUGHT
BOUGHT
MEMBER_OF
SKU: 49d102bc Product: XBox 360
BOUGHT
Firstname: Mickey Surname: Smith DoB: 19781006
BOUGHT
SKU: 5e175641 Product: Badgers Nadgers Ale


Neo4j Monitoring and Manag
localhost:7474/webadmin/#/data/search/120/
News
Neo
Read Later
East Dulwich Depart
Clojure Koans
ClojureScript.pdf
NOSQL Distilled
T-Mobile USA
C&C
Convert Files
NETGEAR ReadyNAS
Neo4j
Overview
Dashboard
Explore and edit
Data browser
Power tool
Console
Add and remove
Indexes
Details
Server info
Documentation
120
Node
Relationship
Returned 1 row. Query took 3ms
Style
Re-layout
Clear
{"name":"Mickey Smith","dob":19781006}
BOUGHT
BOUGHT
BOUGHT
{"product":"Baby Dry Nights","sku":"49d102bc"}
{"product":"XBox 360","sku":"49d102bc"}
{"product":"Badgers Nadgers Ale","sku":"5e175641"}
{"product":"Peewee Pilsner","sku":"2555f258"}
MEMBER_OF
MEMBER_OF
MEMBER_OF
MEMBER_OF
MEMBER_OF
MEMBER_OF
{"category":"nappies"}
{"category":"baby"}
{"category":"console"}
{"category":"consumer electronics"}
{"category":"alcoholic drinks"}
{"category":"beer"}


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

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

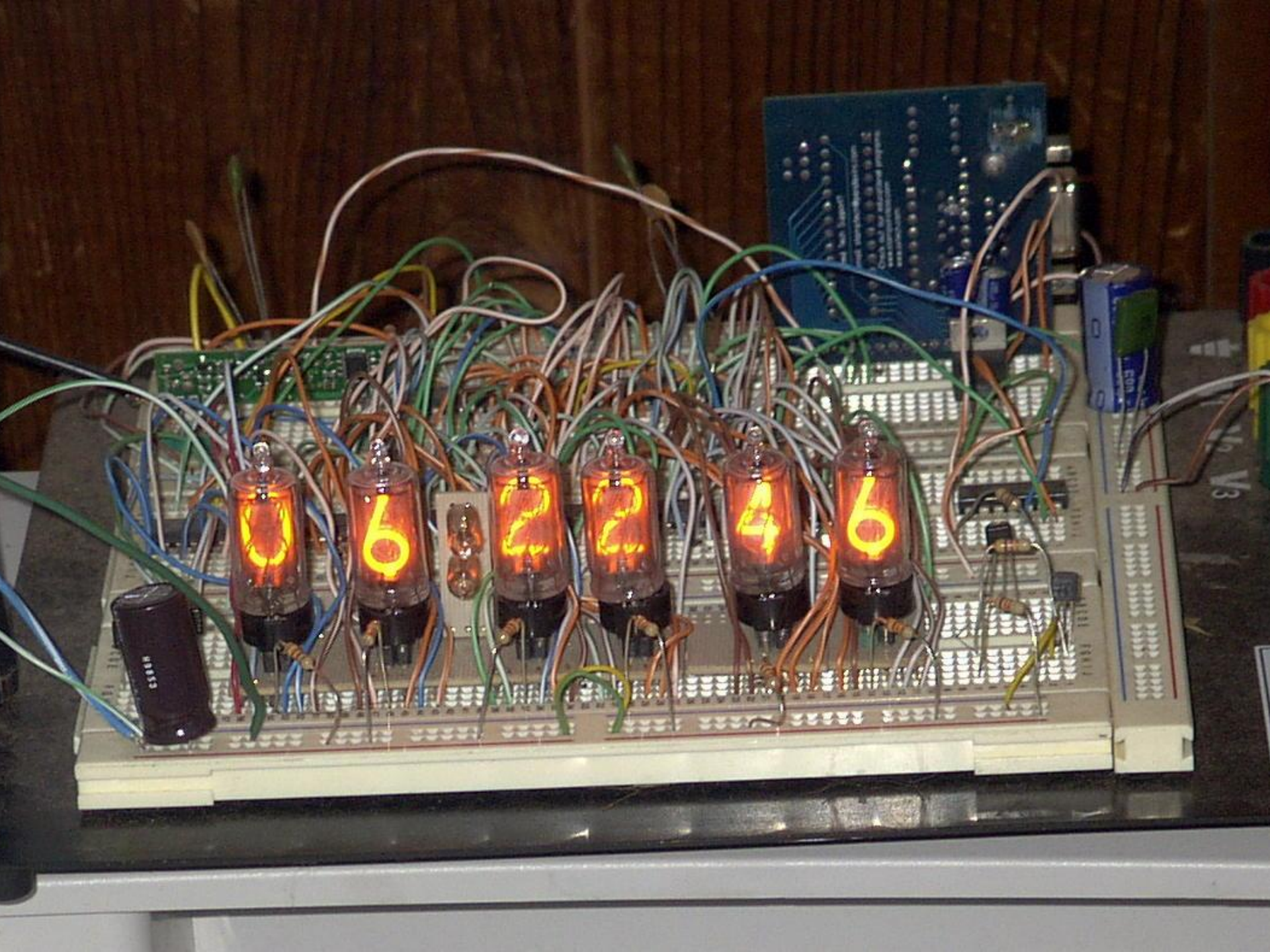
```
MATCH (daddy)-[:BOUGHT]->()-[:MEMBER_OF]->(nappies),
(daddy)-[:BOUGHT]->()-[:MEMBER_OF]->(beer),
(daddy)-[b:BOUGHT]->()-[:MEMBER_OF]->(console)
WHERE b is null
```

Full Cypher query

```
START beer=node:categories(category='beer'),
  nappies=de: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
```

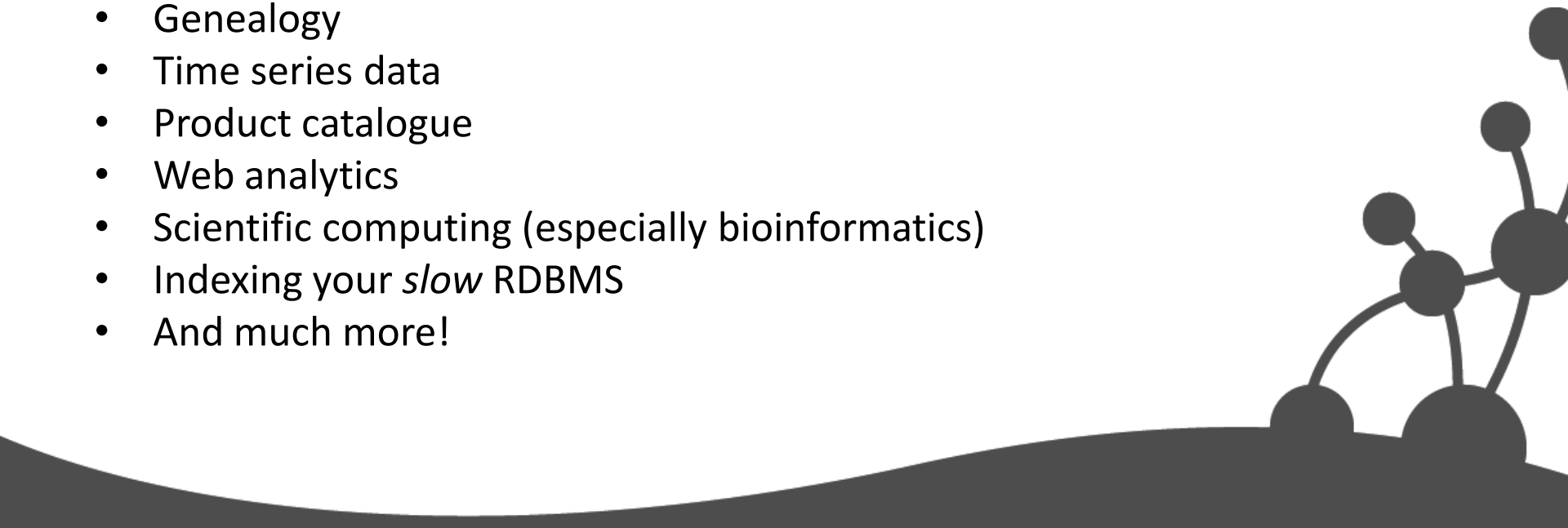
Results

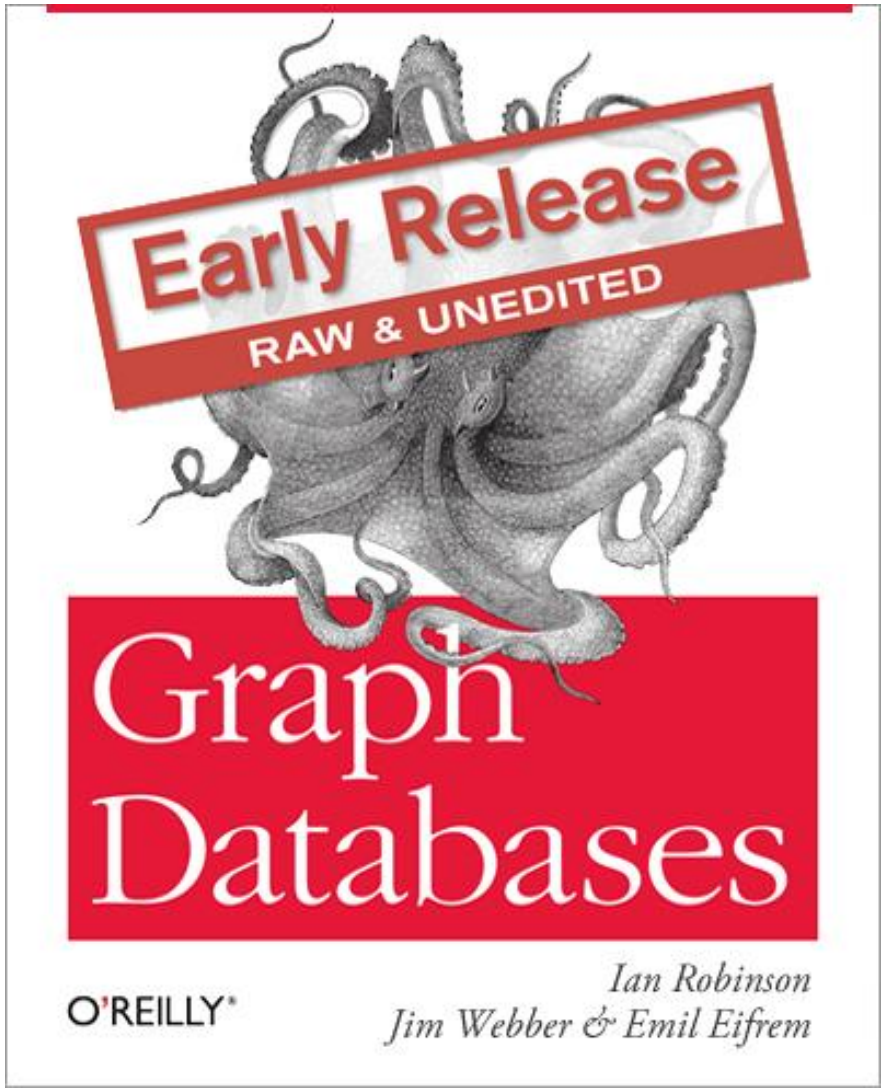
```
==> +-------------------------------------------+
==> | daddy                                     |
==> +-------------------------------------------+
==> | Node[15]{name:"Rory Williams",dob:19880121} |
==> +-------------------------------------------+
==> 1 row
==> 6 ms
==>
neo4j-sh (0)$
```


0 6 2 2 4 6

What are graphs good for?

- Recommendations
- Business intelligence
- Social computing
- Geospatial
- MDM
- Systems management
- Web of things
- Genealogy
- Time series data
- Product catalogue
- Web analytics
- Scientific computing (especially bioinformatics)
- Indexing your *slow* RDBMS
- And much more!





Free O'Reilly eBook!

Visit:
http://GraphDatabases.com

Thanks for listening

Neo4j: **http://neo4j.org**

Neo Technology: **http://neotechnology.com**

Me: **@jimwebber**

Neo4j Meetup in Hilversum Next Week