What <<b u s i n e s s>> can learn from Dating

Rik Van Bruggen
@rvanbruggen

Neo Technology Overview

Company

- Neo Technology, Creator of Neo4j
- 90 employees with HQ in Silicon Valley, London, Munich, Paris and Malmö
- $45M in funding from Fidelity, Sunstone, Conor, Creandum, Dawn Capital

Product

- Neo4j - World's leading graph database
- 1M+ downloads, adding 50k+ per month
- 170+ enterprise subscription customers including over 50 of the Global 2000

How Customers Use Neo4j

Social	Recom–mendations	Search & Discovery	Network & Data Center	Master Data Management	Identity & Access	GEO
eHarmony	careerbuilder	Walmart	IEEE	hp	CISCO	UBS	TOMTOM
classmates	careerbuilder	First Data.	SFR	Pitney Bowes	telenor	ebay now	SNCF
meetic	SNAP	InfoJobs	janssen	ADP ADVENT	LifeWay	classmates
Hinge	viadeo	koobcoodle	NetApp	die Bayerische	KiwiRail
DOWN	NATIONAL GEOGRAPHIC	LIQUID COMMON	e-Spirit	TechCrunch	DingLi.com
maaii	Glowbl	curaspan	codex	Zenoss	veda	Perigee
ZEPHYR HEALTH INC	onefinestay	NOMURA	TRIBAL technologies	gen	ZEPHYR HEALTH INC	DOWN
mallow street	gamesys	research now	Lufthansa	Juni sphere	Didacti	CHRONOTRACK
Care.com	ice	compete	Didacti	VIRTUAL INSTRUMENTS	aikux.com	YELAGO
LIFECHURCH.TV	bwin.party	cobrain	scribestar	EarthLink	Juice PLUS+
doximity	Livestation	shutl	kitedesk	springcm	EQUILAR
noble group

Neo4j Leads the Graph Database Revolution

"Graph analysis is possibly the **single most effective competitive differentiator** for organizations pursuing data-driven operations and decisions after the design of data capture."

"Forrester estimates that ***over 25% of enterprises*** will be using graph databases by 2017"

"Neo4j is the current market **leader in graph databases**."

IT Market Clock for Database Management Systems, 2014
https://www.gartner.com/doc/2852717/it-market-clock-database-management

TechRadar™: Enterprise DBMS, Q1 2014
http://www.forrester.com/TechRadar+Enterprise+DBMS+Q1+2014/fulltext/-/E-RES106801

Graph Databases – and Their Potential to Transform How We Capture Interdependencies (Enterprise Management Associates)
http://blogs.enterprisemanagement.com/dennisdrogseth/2013/11/06/graph-databasesand-potential-transform-capture-interdependencies/

High Business Value in Data Relationships

Data is increasing in volume…

- New digital processes
- More online transactions
- New social networks
- More devices

… and is getting more connected

Customers, products, processes, devices interact and relate to each other

Using Data Relationships unlocks valu

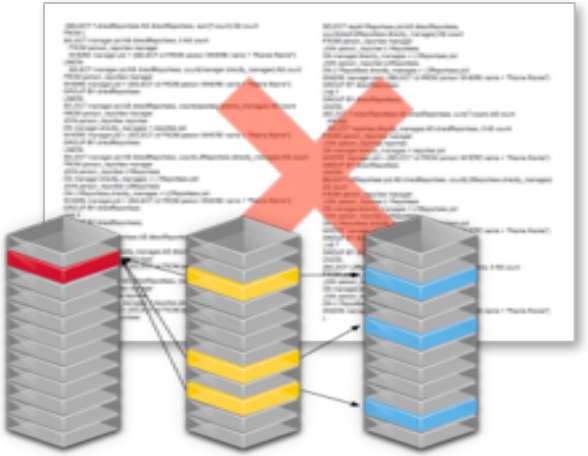
- Real-time recommendations
- Fraud detection
- Master data management
- Network and IT operations
- Identity and access management
- Graph-based search

We need to p
and use the s

Relational DBs Can't Handle Relationships Well

- ***Cannot model or store data and relationships*** without complexity
- ***Performance degrades*** with number and levels of relationships, and database size
- ***Query complexity grows*** with need for JOINs
- ***Adding new types of data and relationships*** requires schema redesign, increasing time to market

… making traditional databases **inappropriate** when data relationships are valuable in **real-time**



Slow development
Poor performance
Low scalability
Hard to maintain

Aggregate Stores *Don't* Handle Relationships

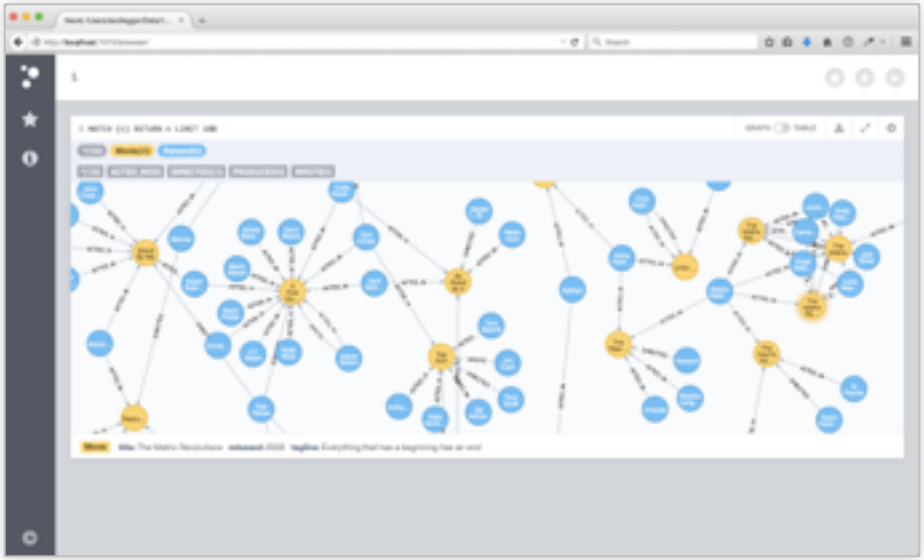
- ***No data structures*** to model or store relationships
- ***No query constructs*** to support data relationships
- ***Relating data requires "JOIN logic"*** in the application
- ***No ACID support*** for transactions

… making NoSQL databases **inappropriate** when data relationships are valuable in **real-time**

Neo4j – Re-Imagine Your Data as a Graph

Neo4j is an ***enterprise-grade graph database*** that enables you to:

- **Model and store** your data as a **graph**
- **Query data relationships** with ease and in real-time
- **Seamlessly evolve applications** to support new requirements by adding new kinds of data and relationships



Agile development
High performance
Vertical and horizontal scale
Seamless evolution

The Whiteboard Model Is the Physical Model

Key Neo4j Product Features

Native Graph Storage
Ensures data consistency and performance

Native Graph Processing
Millions of hops per second, in real time

“Whiteboard Friendly” Data Modeling
Model data as it naturally occurs

High Data Integrity
Fully ACID transactions

Powerful, Expressive Query Language
Requires 10x to 100x less code than SQL

Scalability and High Availability
Vertical and horizontal scaling optimized for graphs

Built-in ETL
Seamless import from other databases

Integration
Drivers and APIs for popular languages

So…
What CAN Business learn from dating?

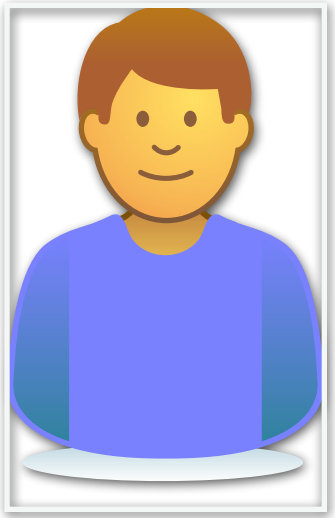
Neo4j
the world's leading graph database
[:AVAILABLE_FOR_FREE_TRIAL]
(NEO4J.ORG)
[:RUNS_ON]
eHarmony
[:USES]
SNAP INTERACTIVE
[:REQUIRES]
[:POWERS]
friends of friends
[:OWNS]
[:NEEDS]
[:USED_BY]
ayi are you interested?
Europe's top dating site
[:IS]
[:REQUIRES]
[:NEEDS]
[:ALLOWS]
[:NEEDS]
Available
meetic
[:USES]
D
BWF
[:FORMERLY_KNOWN_AS]
Hinge
friends of friends of friends...
[:ALSO_RUNS]
[:CREATES]
more CONNECTIONS
ebay
hp
CISCO
INTUIT
[:CREATES]
more INTERACTIONS

neo4j

The 5 graphs of love

- The friends of friends graph
- The passion graph
- The location graph
- The safety graph
- The poser graph

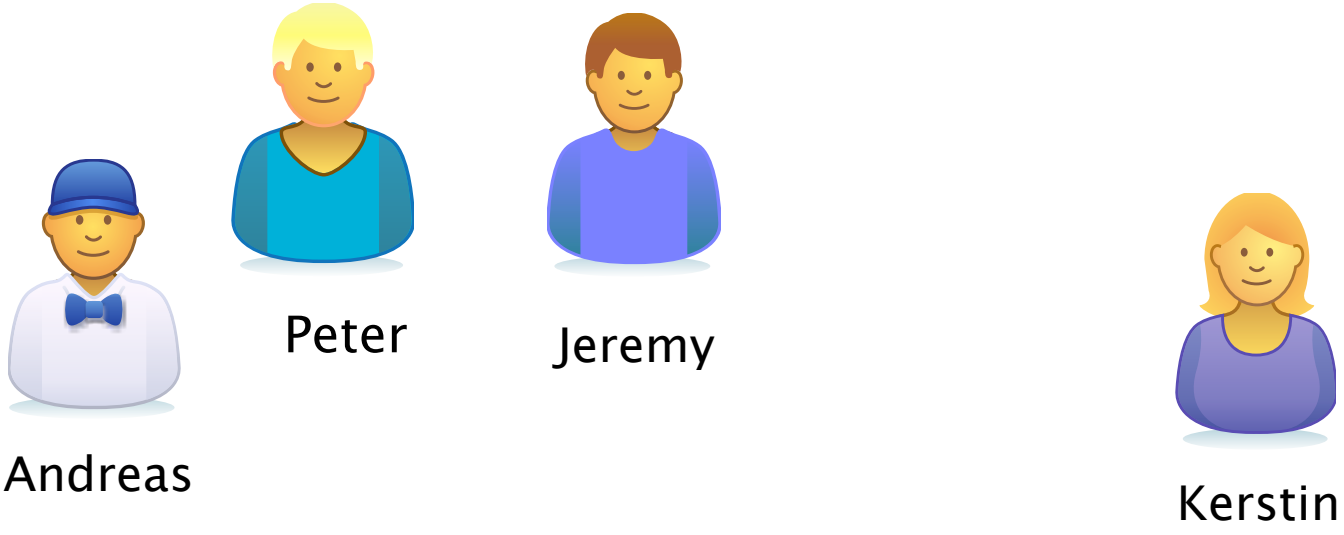
Meet Jeremy



Jeremy

Jeremy has some friends

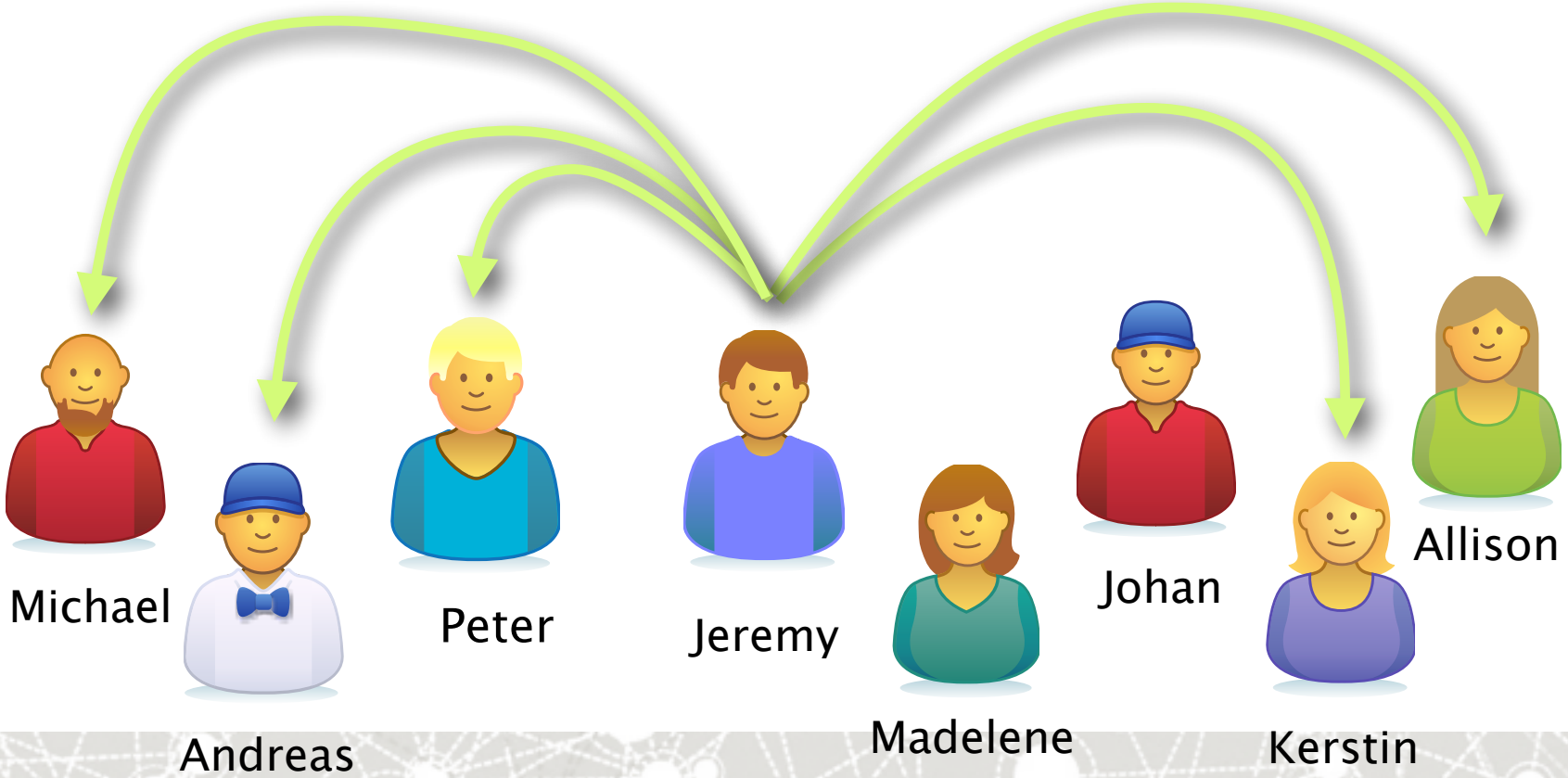
- Kerstin: his sister
- Peter: his friend
- Andreas: his colleague



His friends introduced some more friends

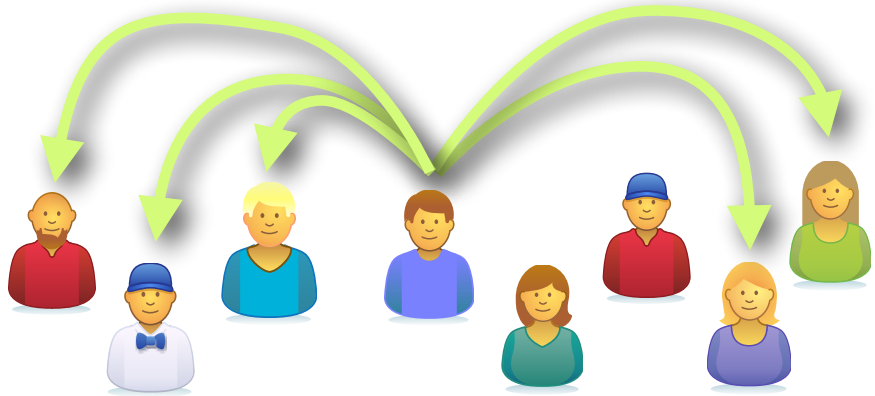
- Michael: master hacker, divorce, 2 children
- Johan: technology sage, likes fast cards
- Madelene: polyglot journalist, loves dogs
- Allison: marketing maven, likes long walks on the beach

So we have a bunch of people



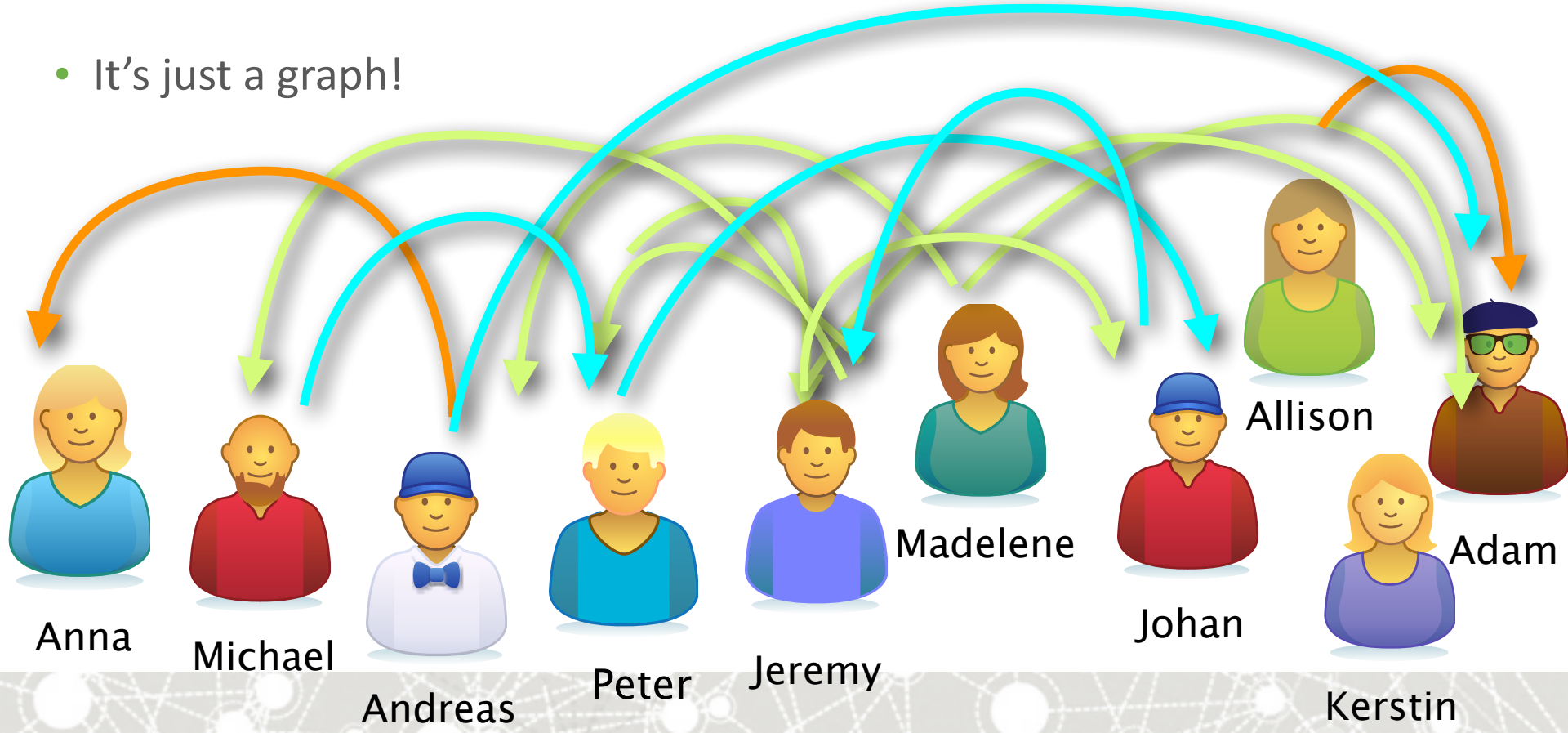
So we have a bunch of people

- How do we know they're friends?
- Either ask each pair: are you friends?
- Or we can add explicit connections
- Twitter, Facebook, LinkedIn, Snapchat etc



So we have a bunch of people

- It's just a graph!



Neo4j: The world's leading graph database

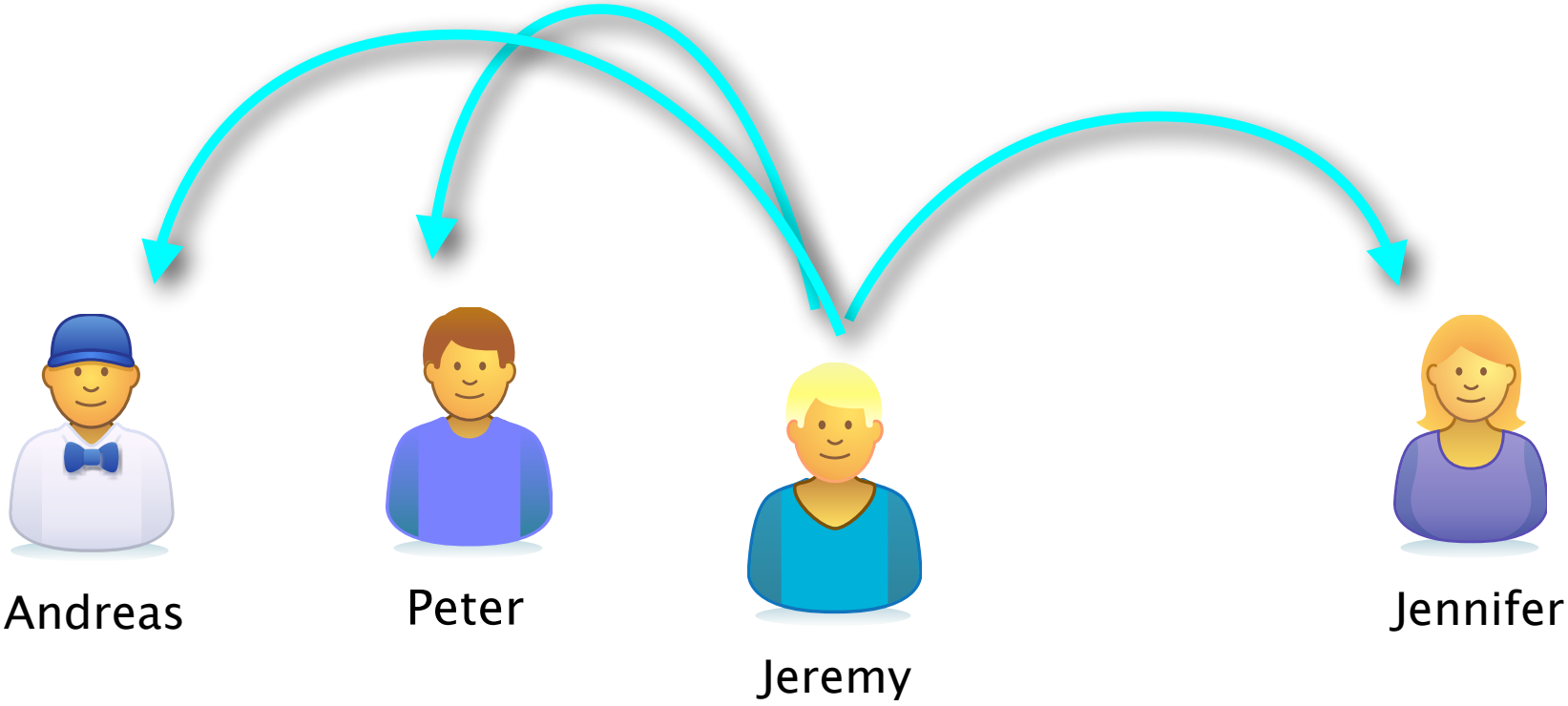
- optimised for the connections between records
- 'pre computed indexes' between records
- really, really fast at querying across records
- A relational database may tell you the average age of everyone here…

…but a graph database will tell you who's most likely to buy you a beer later

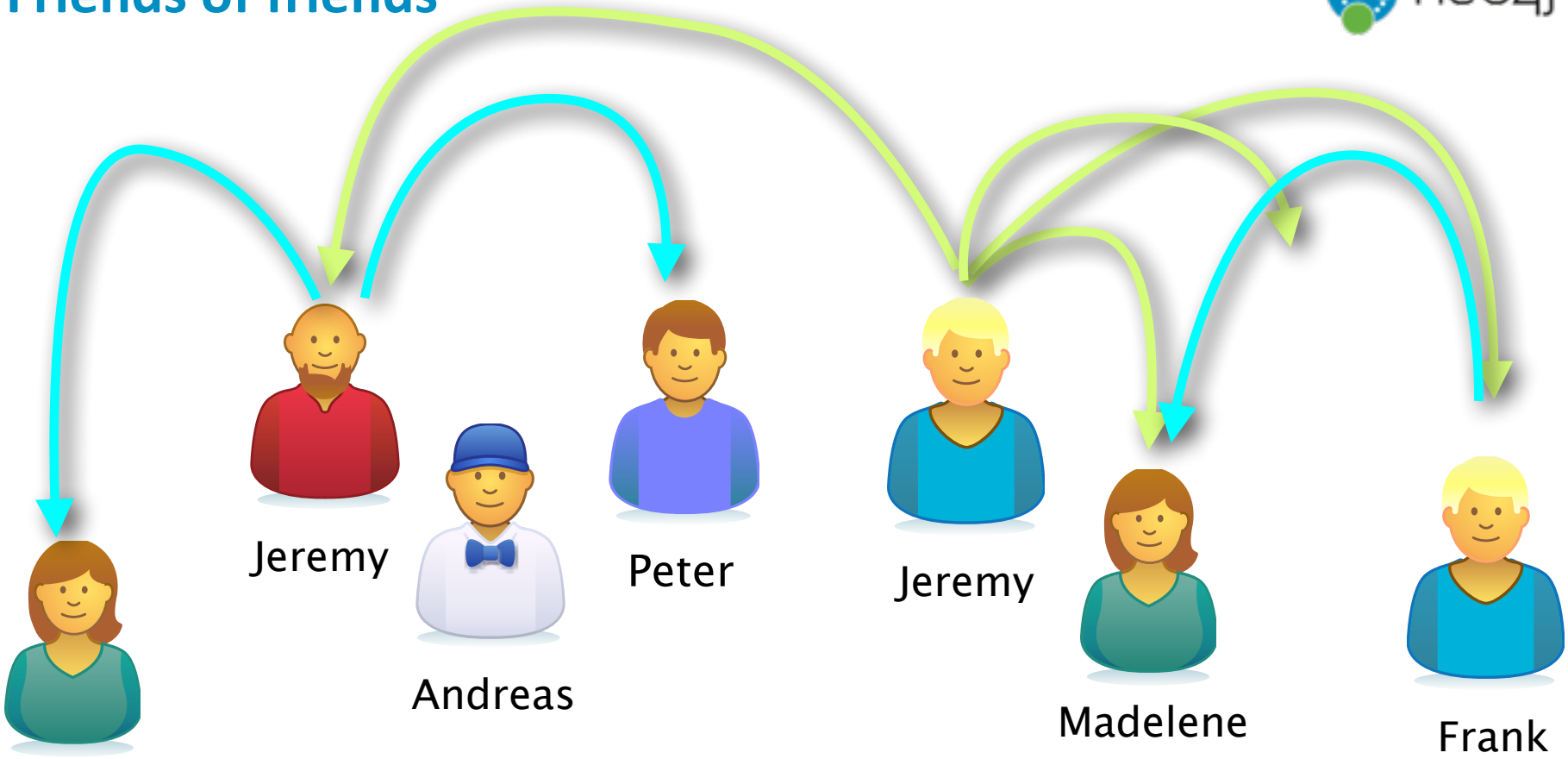
Friends of friends graph

- According to SNAP Interactive if you are a female user, you have a:
  - 4% likelihood of interacting with a stranger
  - 10% likelihood of interacting with a friend of friend
  - 7% likelihood of interacting with a 3rd degree connection (friend of a friend)
  - Connections mean a much larger number of interactions = **#win**

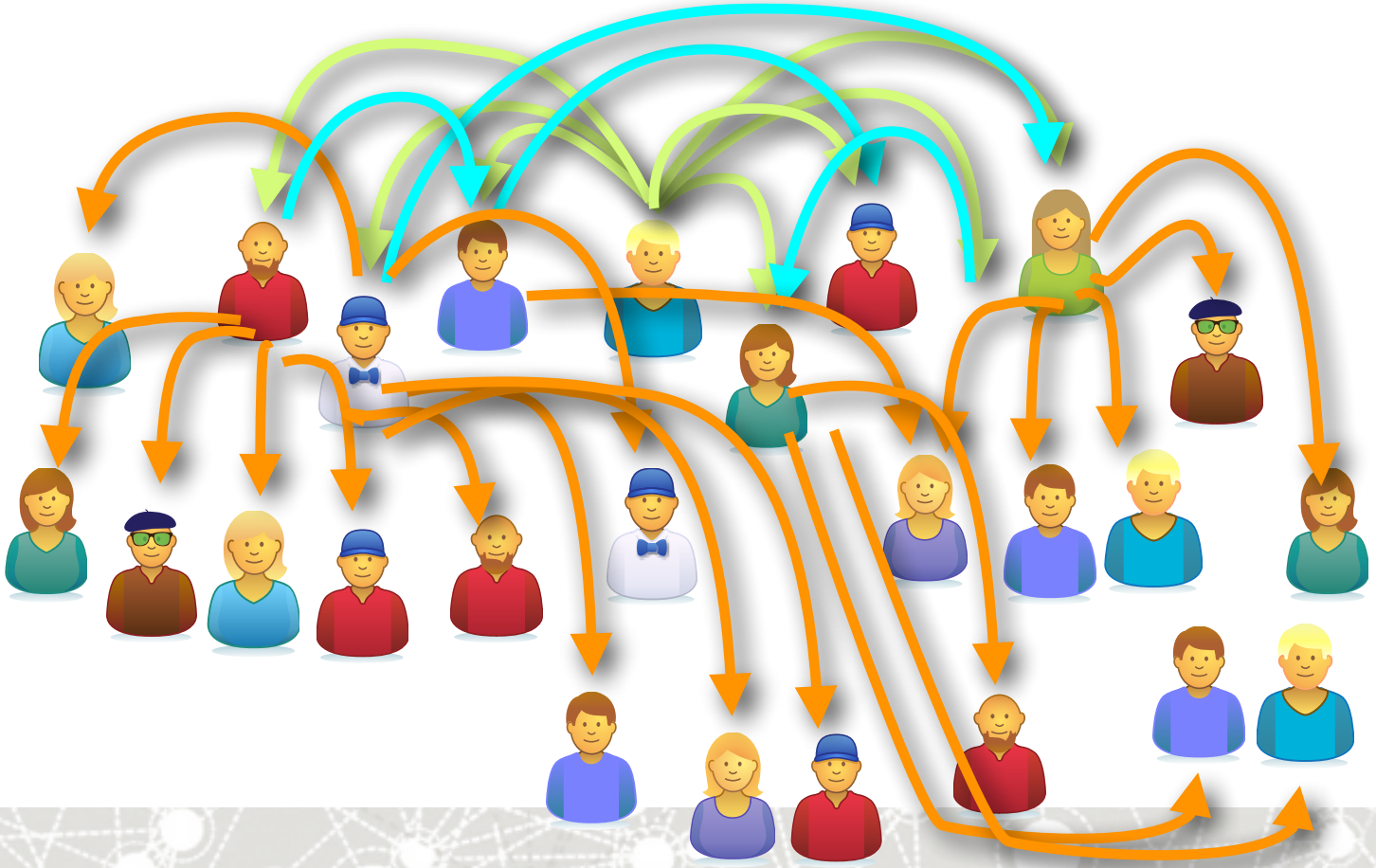
Friends



Friends of friends



Friends of friends of friends



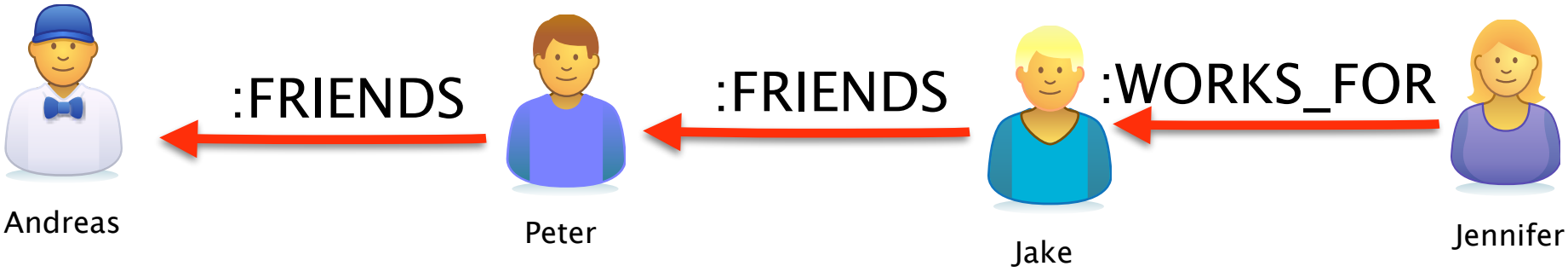
Find Jeremy's FoFs

```
1 //Who are Jeremy's FoFs?
2 MATCH (Jeremy:User{name:"Jeremy"})-[:KNOWS]->
3 (friend)-[:KNOWS]->(fof)
4 RETURN  fof
```

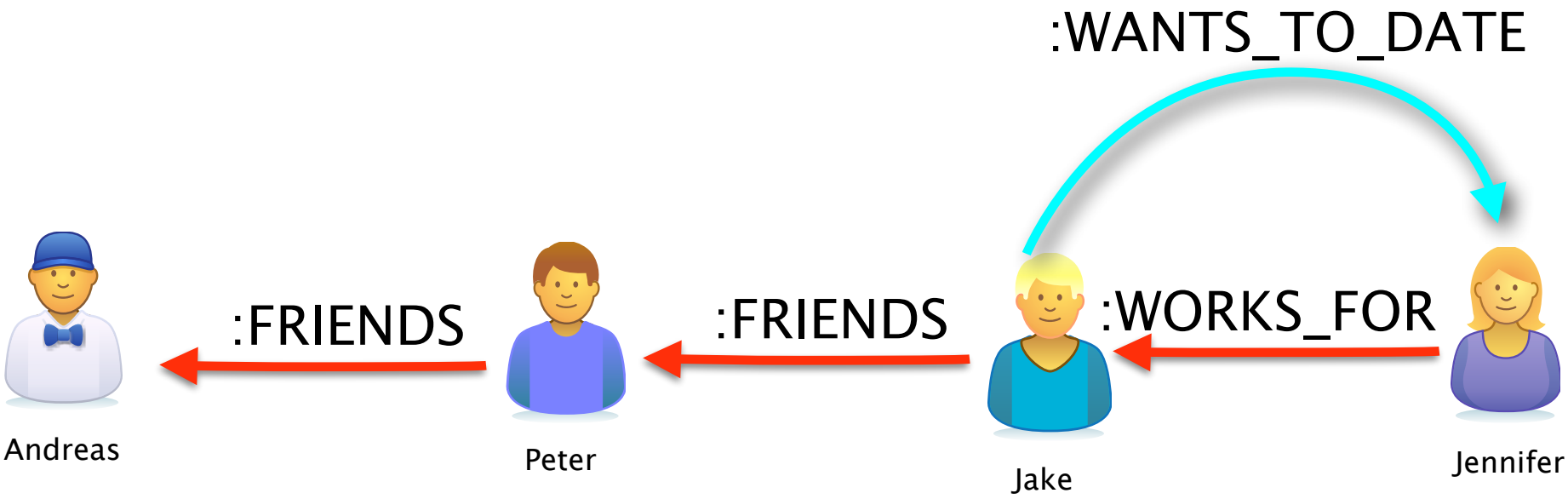
Who does Jeremy shares the most friends with?

```
1 //Who does Jeremy share the most friends with?
2 MATCH (Jeremy:User{name:"Jeremy"})-[:KNOWS]->(friend)
3 -[:KNOWS]->(fof)
4 RETURN distinct fof,length(collect(friend)) as FriendsInCommon
5 ORDER BY FriendsInCommon desc
```

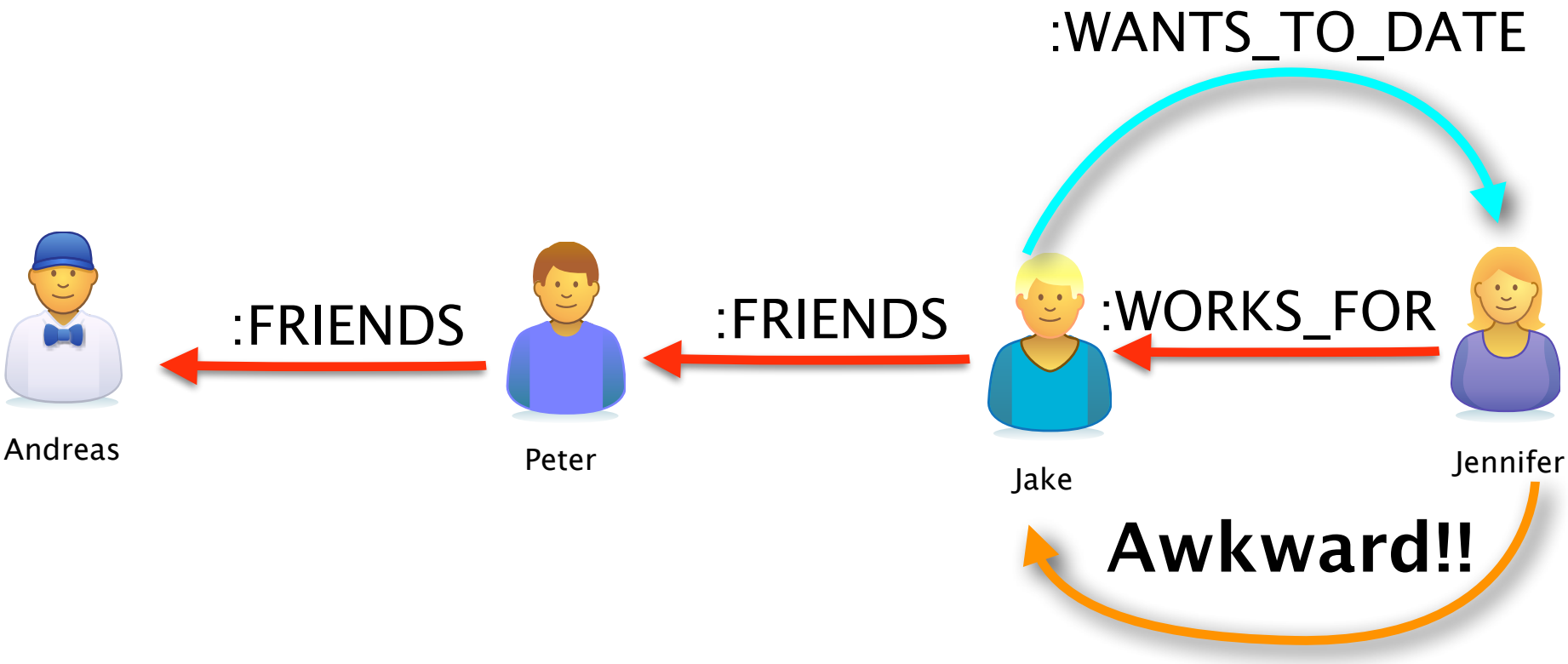
Complicated relationships



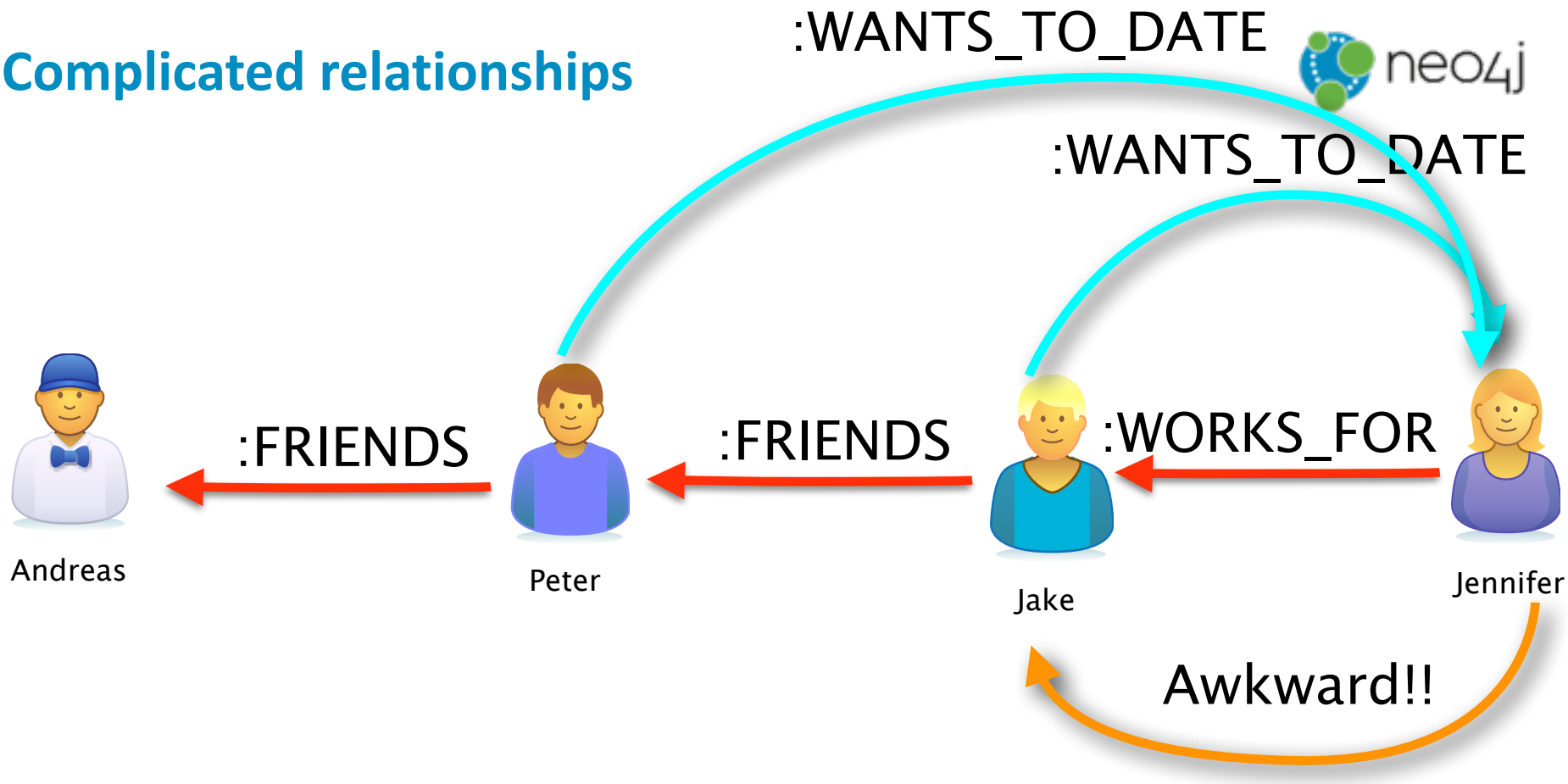
Complicated relationships



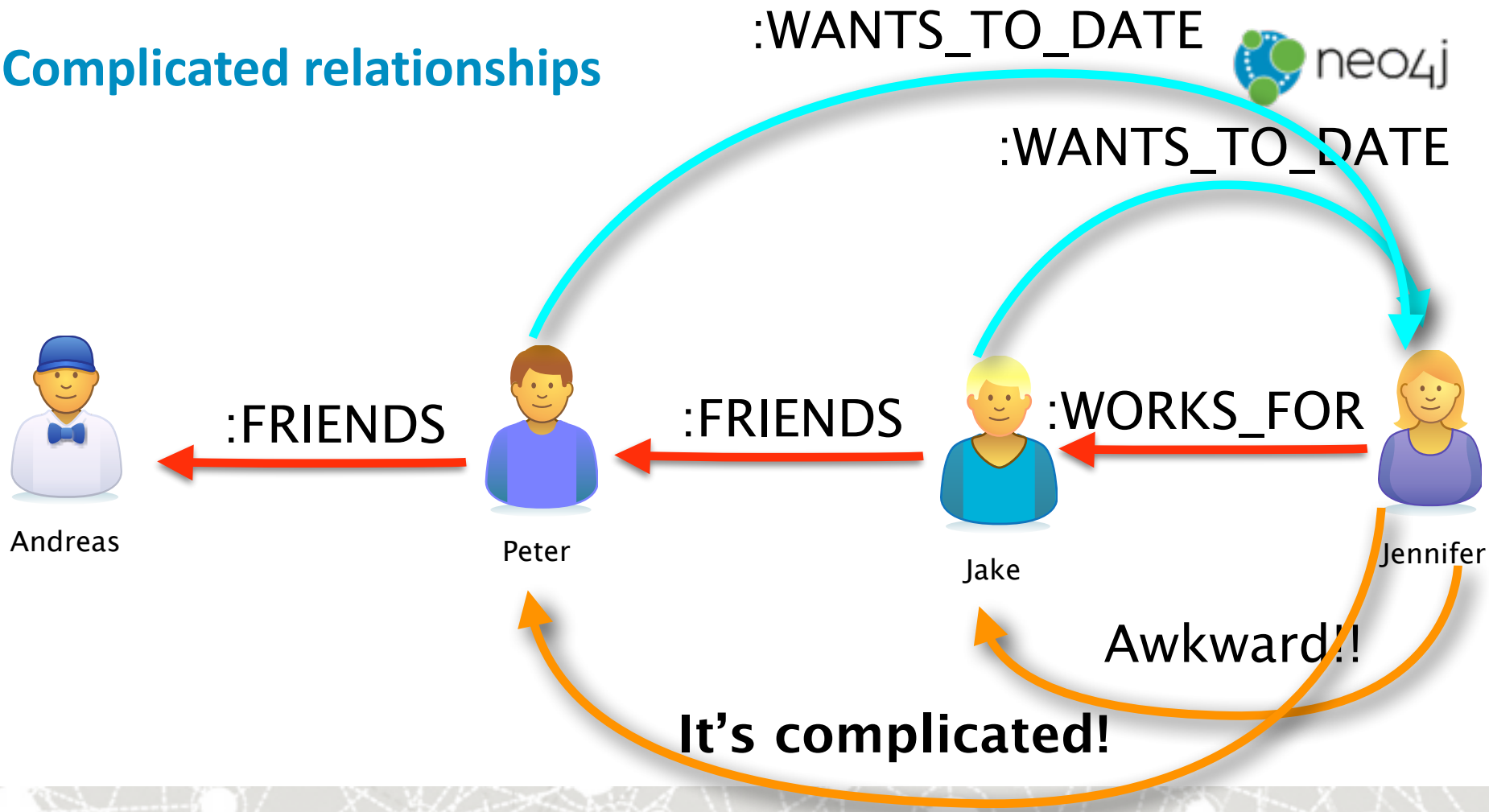
Complicated relationships



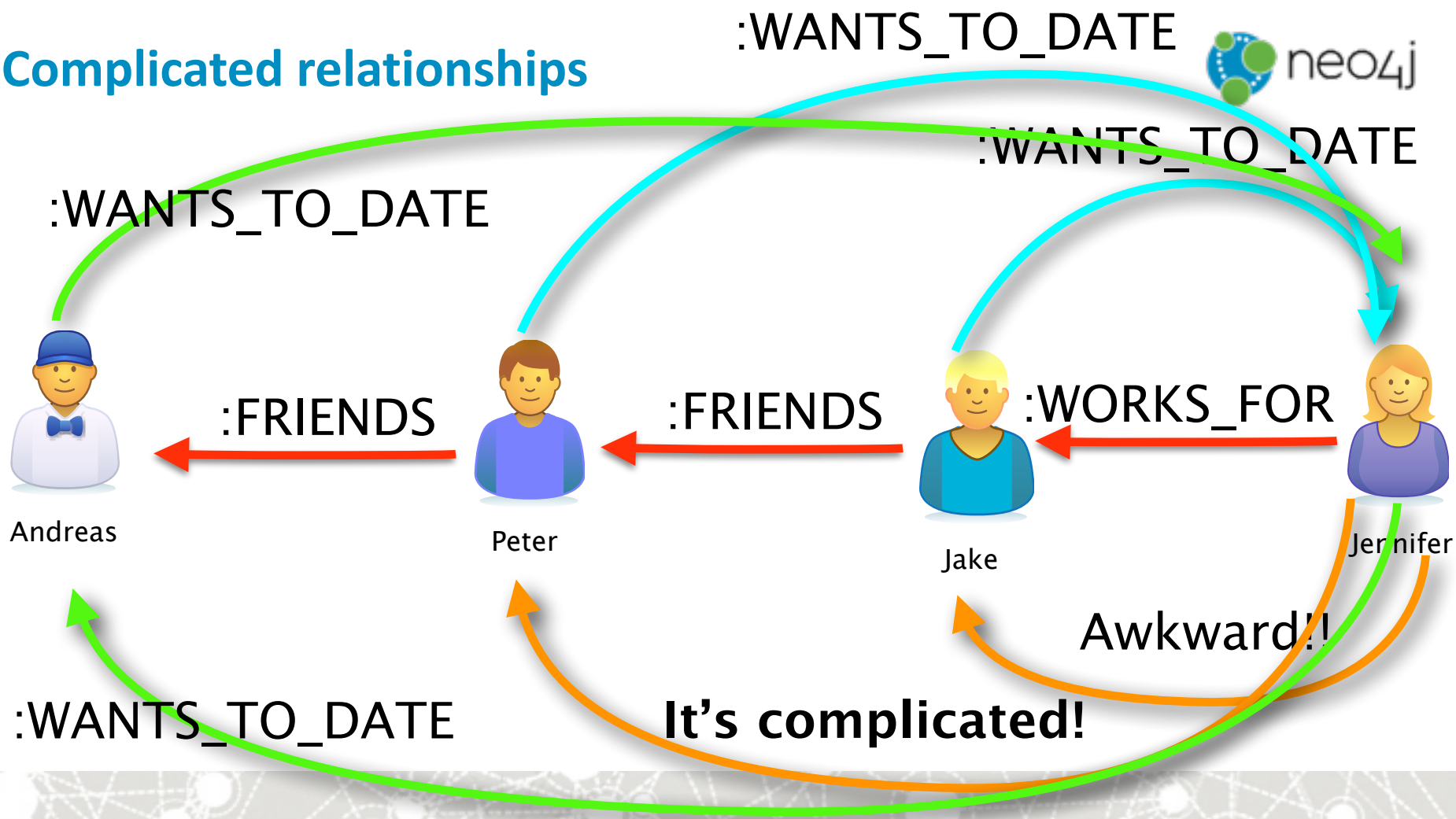
Complicated relationships



Complicated relationships



Complicated relationships



Friends of friends of friends

```
1 //FoFoF Long
2 MATCH (:User)-[:KNOWS]->(friend)-[:KNOWS]->(FoF)-
  [:KNOWS]->(FoFoF)
3 RETURN FoFoF
```

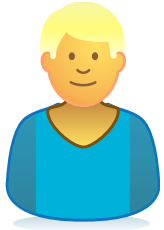
Friends of friends of friends

```
1 //FoFoF Long
2 MATCH (:User)-[:KNOWS]->(friend)-[:KNOWS]->(FoF)-
  [:KNOWS]->(FoFoF)
3 RETURN FoFoF
```

```
1 //FoFoF Short
2 MATCH (:User)-[:KNOWS*3]->(FoFoF)
3 RETURN FoFoF
```

Meet Jon…

- from: UK
- seeking: females
- appearance: hot, hot, hot!
- personality: fun loving, easy going
- interests: cooking, chemistry

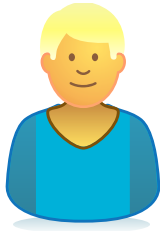


Jon

```
1 MATCH (Jon:User{name:'Jonathan'})
2 RETURN Jon
```

The passion graph

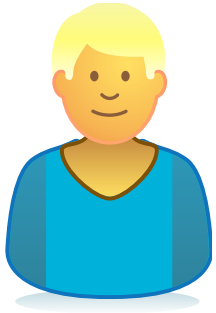
- Jon wants to find someone he can share his passions with



Jon

```
1 MATCH (Jon:User{name:'Jonathan'})
2 RETURN Jon
```

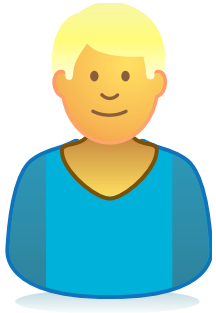
Match specific interests



Jon

:REPORTED_INTEREST

Match specific interests

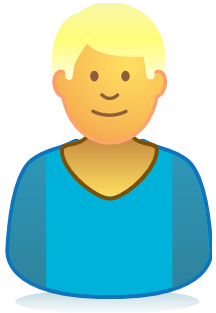


Jon

:REPORTED_INTEREST

```
1 MATCH (date)-[:REPORTED_INTEREST]->(interest)
2 WHERE interest.name = "Foodie" AND date.gender = 'F'
3 RETURN date
```

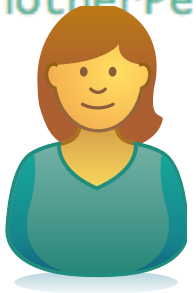
Match specific interests



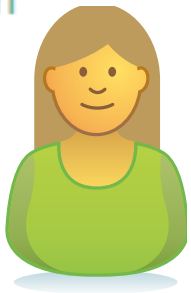
:HAS_INTEREST

Jon

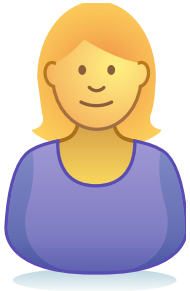
```
MATCH (Jeremy:User {name: "Jonathan"})-[:HAS_INTEREST]->(interest)
MATCH (interest)<-[:HAS_INTEREST]-(anotherPerson)
RETURN interest, anotherPerson
```



Anne



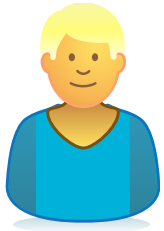
Julia



Jennifer

The location graph

- Jon wants to find a date but refuses to have a long distance relationship



Jon

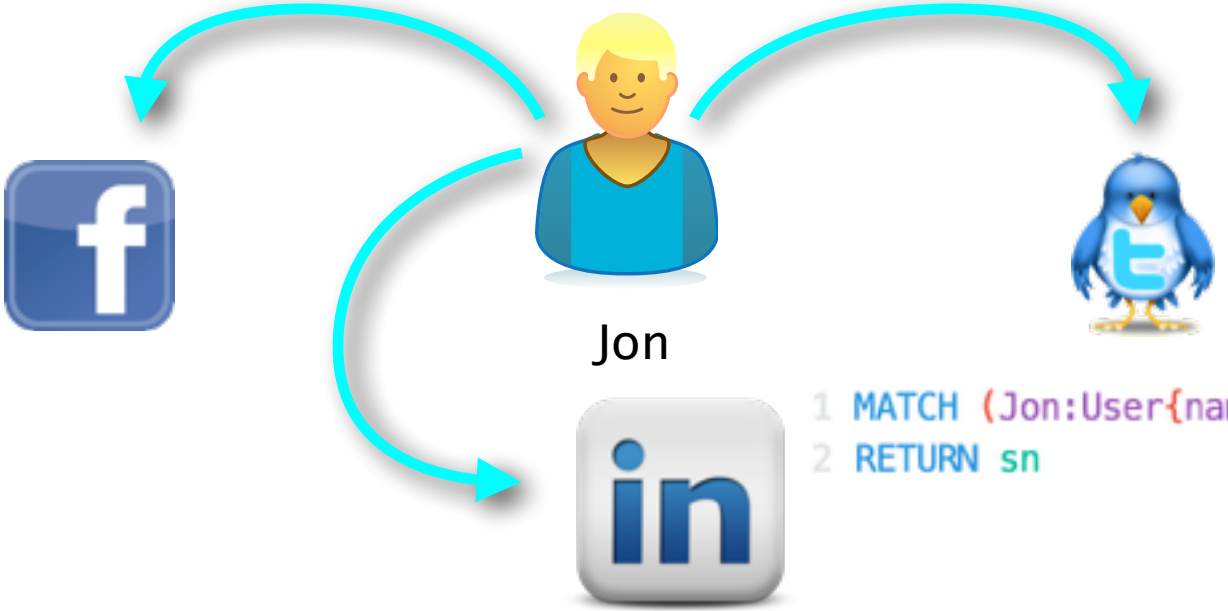
```
1 MATCH (Jon:User{name:'Jonathan'})
2 RETURN Jon
```

The location graph

```
1 START date = node('withinDistance:[51.99,-0.19, 100.0]')
2 WHERE date.gender = 'F'
3 RETURN date
```

The safety graph

- Jon uses social networks




```
1 MATCH (Jon:User{name:'Jonathan'})-[:HAS_SN]-(sn)
2 RETURN sn
```

Let’s dig into his Twitter

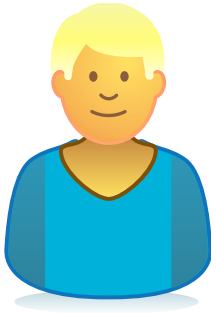
He follows some strange people…

…and tweets about strange things!

Some basic word analysis

Tweets

Let's update based on behaviour



Jon

:DEMONSTRATED_INTEREST

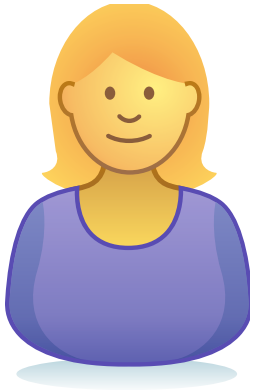
```
CREATE (Jonathan:Person{name:"Jonathan Dreamyman"})
-[:DEMONSTRATED_INTEREST]->
(drugs:Interest{name:"Drugs"})
```

Any ladies ok with this?

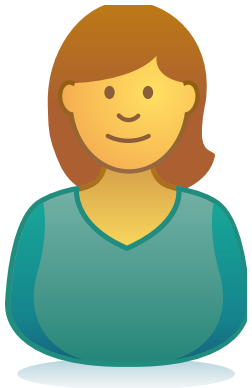
```
1 //Jon's FoFs who like Drugs
2 MATCH (date:User)-[:KNOWS*2]-(Jon:User{name : 'Jonathan'}),
3 (date)-[*..2]-(Drugs:Interest{name:'Drugs'})
4 WHERE date.gender = 'F'
5 RETURN distinct date
```

Any ladies ok with this?

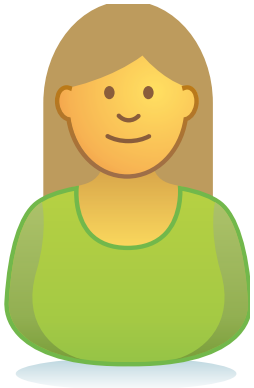
```
1 //Jon's FoFs who like Drugs
2 MATCH (date:User)-[:KNOWS*2]-(Jon:User{name : 'Jonathan'}),
3 (date)-[*..2]-(Drugs:Interest{name:'Drugs'})
4 WHERE date.gender = 'F'
5 RETURN distinct date
```



Jennifer



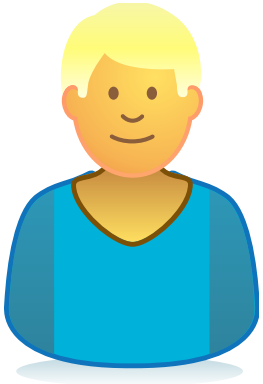
Jane



Maria

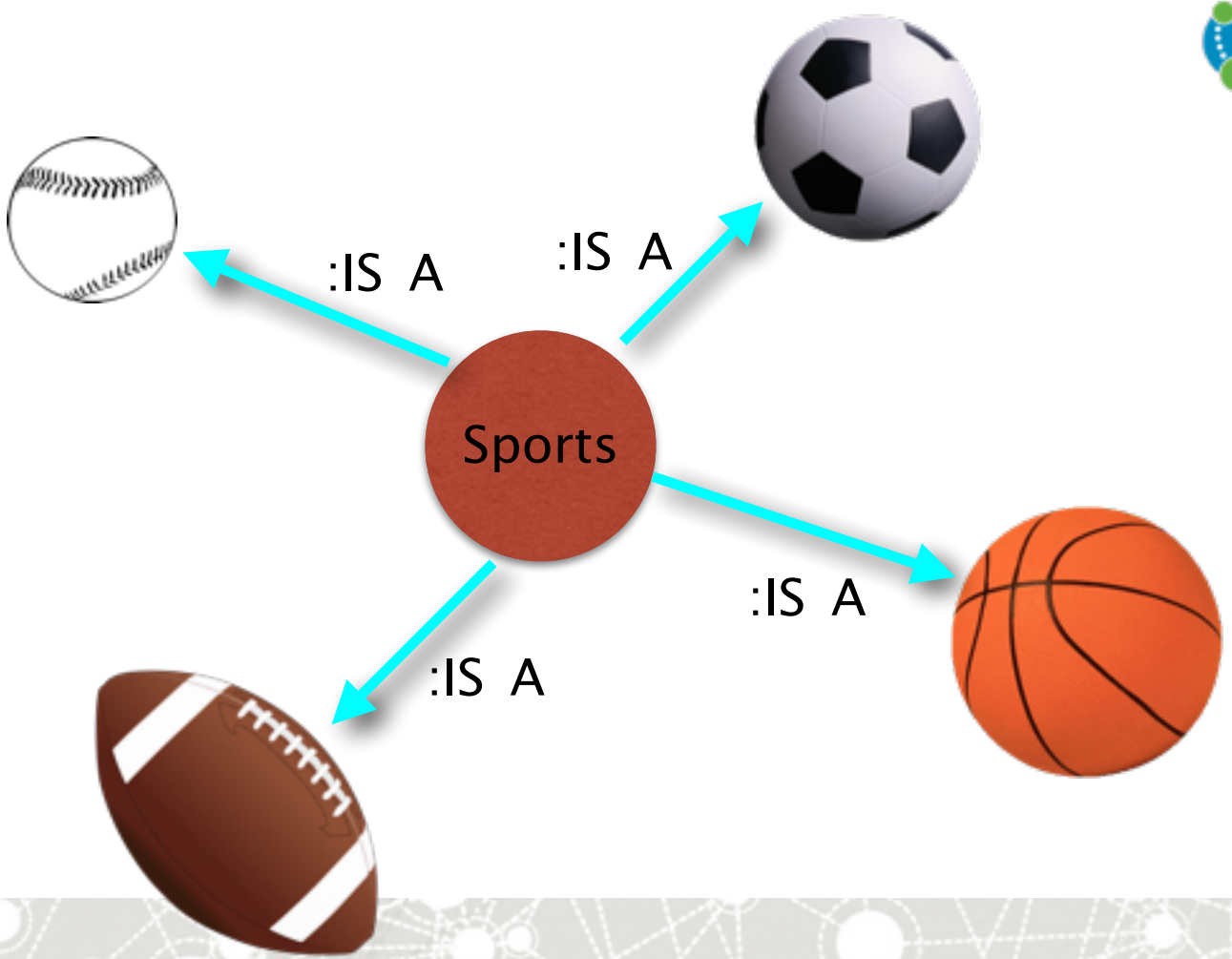
The passion graph

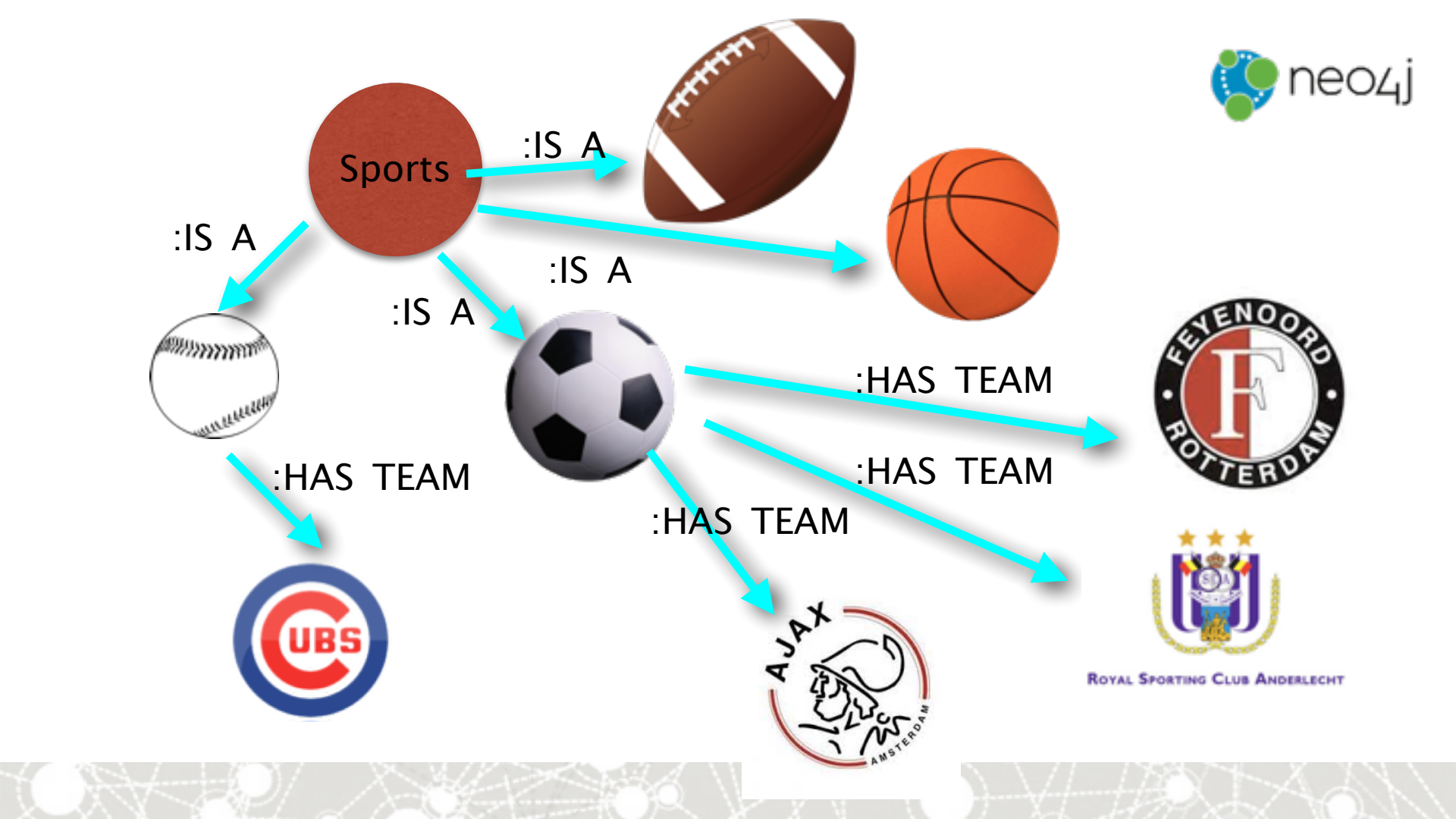
- Jon loves Ajax

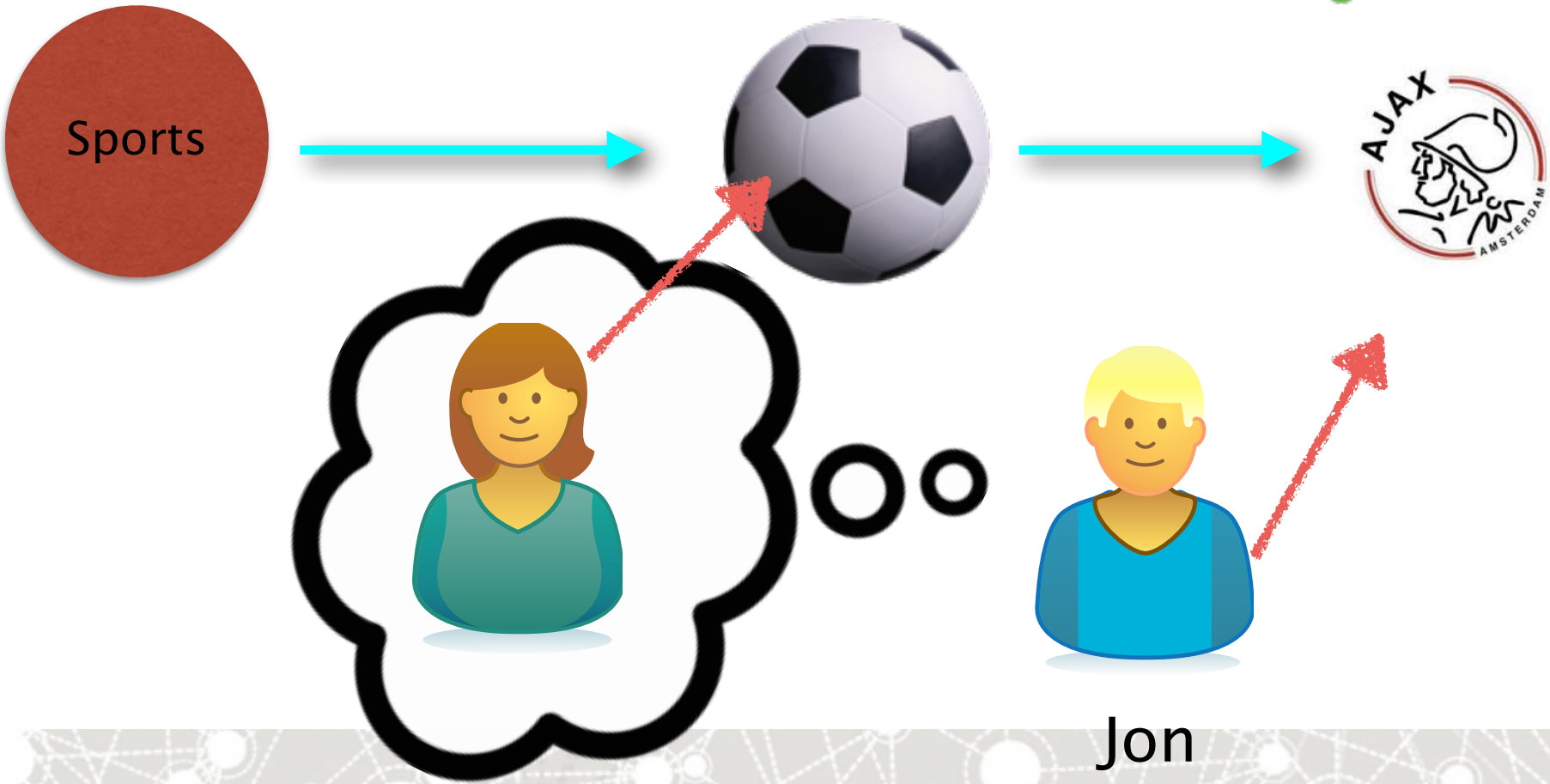


Jon

:HAS_INTEREST


:IS A
:IS A
Sports
:IS A
:IS A


neo4j
Sports
:IS A
:IS A
:IS A
:IS A
:HAS TEAM
:HAS TEAM
:HAS TEAM
:HAS TEAM
CUBS
AJAX
AMSTERDAM
FEYENOORD
ROTTERDAM
ROYAL SPORTING CLUB ANDERLECHT


Sports
AJAX
AMSTERDAM
Jon

The poser graph

- Jon has no luck with online dating. All of his interactions are with spam profiles

The poser graph

- Let’s find real people with at least 1 social network and a minimum of 2 posts

Find ladies who aren't spam bots

```
1 //Find real people
2 MATCH (date:User)-[:KNOWS*2]-(Jon:User{name : 'Jonathan'}),
3 (date)-[r:HAS_SN]->(SN)
4 WHERE date.gender = 'F' AND r.posts > 2
5 RETURN distinct date
```

Find ladies who like football

```
1 // Jon's FoF's who like football
2 MATCH (date:User)-[:KNOWS*2]-(Jon:User{name: 'Jonathan'}),
3       (date)-[*..2]-(a:Interest{name: "Football"})
4 WHERE date.gender = 'F'
5 RETURN distinct date
```

Find ladies who like football

```
1 // Jon's FoF's who like football
2 MATCH (date:User)-[:KNOWS*2]-(Jon:User{name: 'Jonathan'}),
3       (date)-[*..2]-(a:Interest{name: "Football"})
4 WHERE date.gender = 'F'
5 RETURN distinct date
```

Katie Jennifer Greta

Find Jon’s perfect date

```
1 MATCH (date:User)-[:KNOWS*2]-(Jon:User{name: 'Jonathan'}),
2       (date)-[:HAS_INTEREST]->(food:Interest{name:'Foodie'}),
3       (date)-[*..2]-(a:Interest{name:'Football'}),
4       (date)-[*..2]-(Drugs:Interest{name:'Drugs'})
5 WHERE date.gender = 'F'
6 RETURN distinct date
```

Find Jon's perfect date

```
1 MATCH (date:User)-[:KNOWS*2]-(Jon:User{name: 'Jonathan'}),
2       (date)-[:HAS_INTEREST]->(food:Interest{name:'Foodie'}),
3       (date)-[*..2]-(a:Interest{name:'Football'}),
4       (date)-[*..2]-(Drugs:Interest{name:'Drugs'})
5 WHERE date.gender = 'F'
6 RETURN distinct date
```

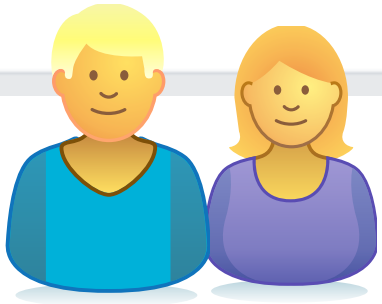
:HAS_DATE_WITH
Jon
Jennifer

Jon and Jennifer delete their profiles and sail off into the sun set

```
1 MATCH (Jon:User{name:'Jonathan'})-[a]-(),(Jen:User{name:'Jennifer'})-[b]-()
2 DELETE a,b,Jon,Jen
```

```
CYPHER MATCH (Jon:User{name:'Jonathan'})-[a]-(),(Jen:User{name:'Jennifer'})-[b]-() DELETE a,b,Jon,Jen
```

✔ Deleted 2 nodes, deleted 16 relationships, returned 0 rows in 165 ms



Jon Jennifer

What's that got to do with Business?

Take a look at "Fraud"

Why Fraud?

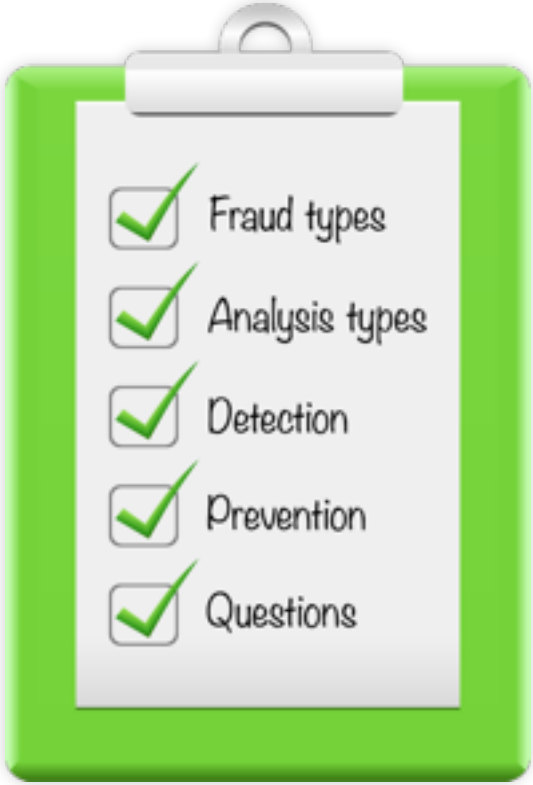
Types of Fraud

- First-Party Fraud
- Synthetic Identities and Fraud Rings
- Insurance Fraud
- Ecommerce Fraud

Types of Analysis

- Traditional Analysis
- Graph-Based Analysis

Fraud Detection and Prevention



First-Party Fraud

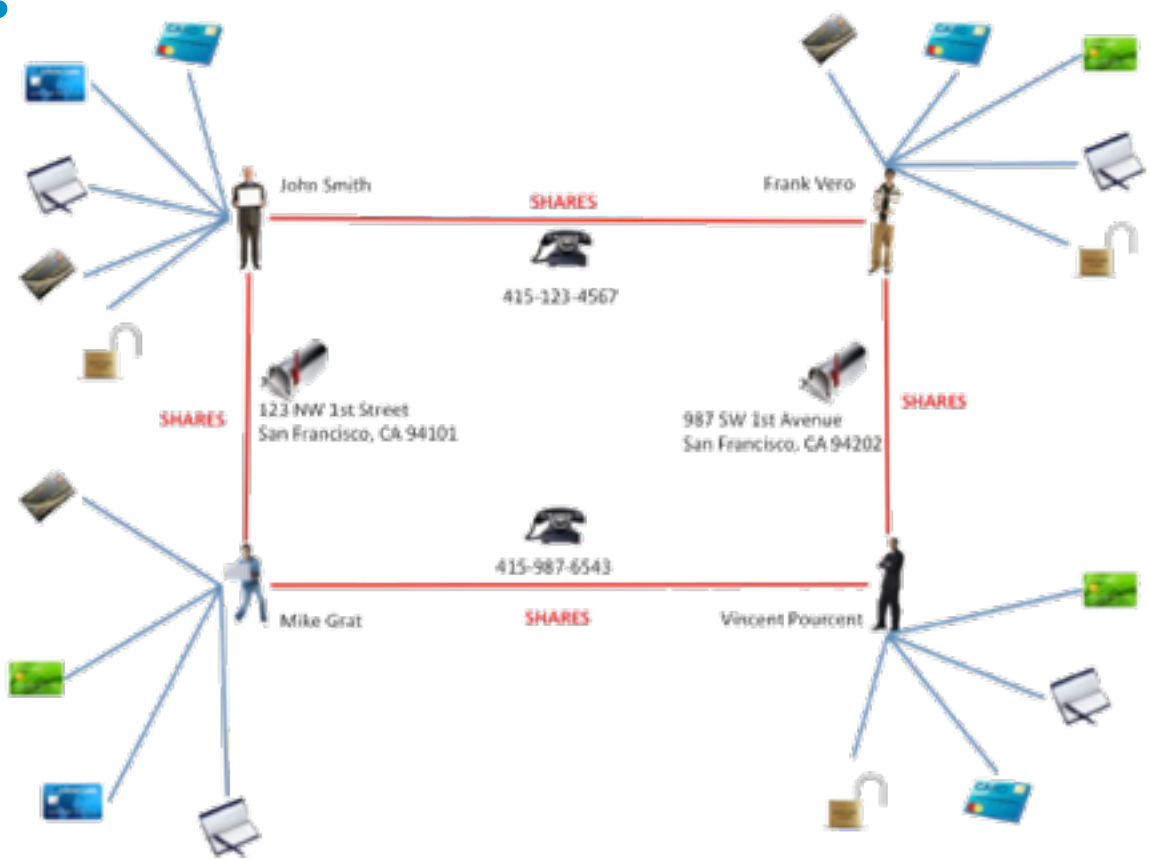
- Fraudster's aim: apply for lines of credit, act normally, extend credit, then…run off with it
- Fabricate a network of synthetic IDs, aggregate smaller lines of credit into substantial value
- Often a hidden problem since only banks are hit
  - Whereas third-party fraud involves customers whose identities are stolen

So what?

- $10's **billions** lost by US banks **every year**
- **25%** of the total consumer credit write-offs in the USA
- Around **20%** of unsecured bad debt in EU and USA is misclassified
  - In reality it is first-party fraud

These are enormous numbers

Fraud Ring


John Smith
SHARES
Frank Vero
415-123-4567
SHARES
123 NW 1st Street
San Francisco, CA 94101
987 SW 1st Avenue
San Francisco, CA 94202
SHARES
415-987-6543
Mike Grat
SHARES
Vincent Pourcent

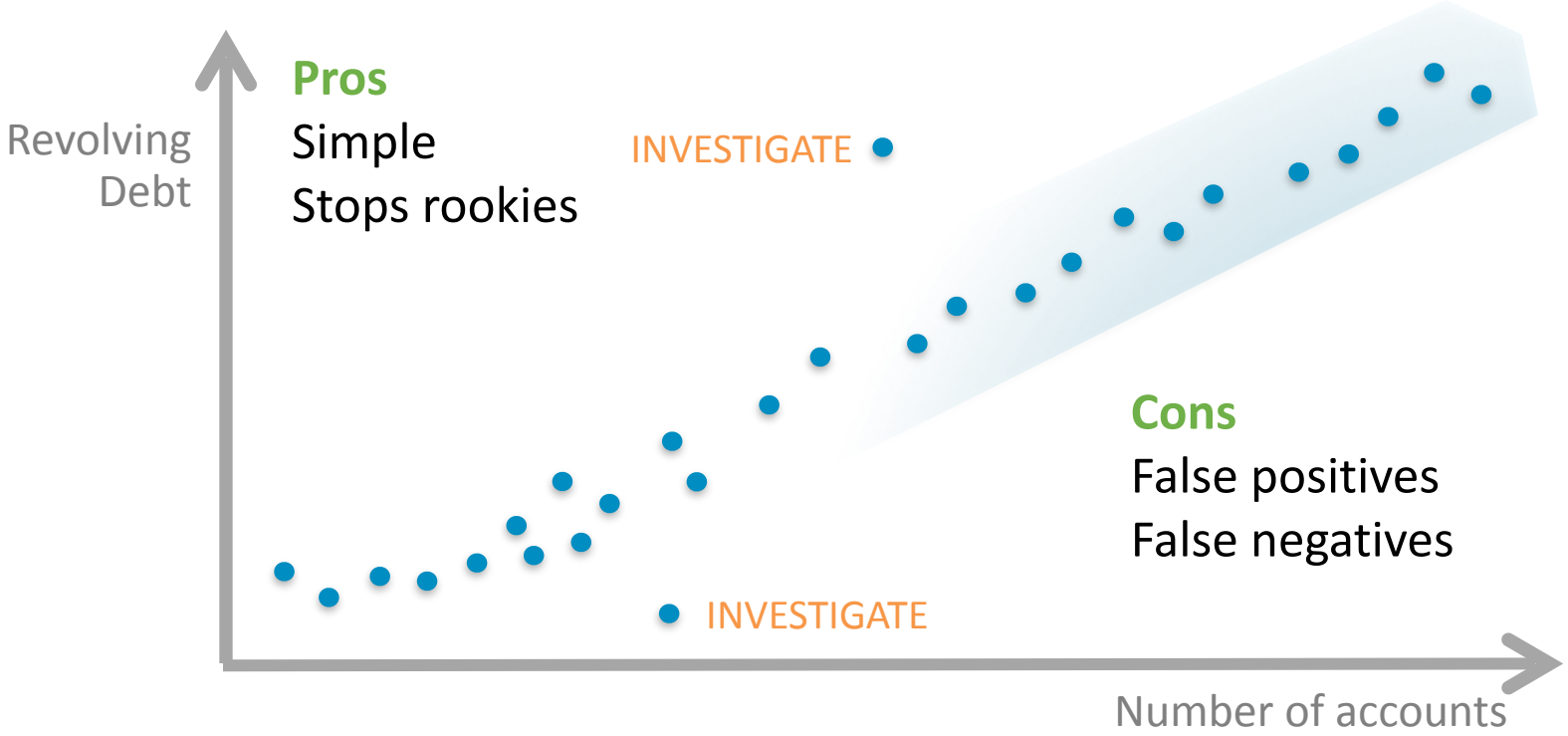
Then the fraud happens…

- Revolving doors strategy
  - Money moves from account to account to provide legitimate transaction history
- Banks duly increase credit lines
  - Observed responsible credit behaviour
- Fraudsters max out all lines of credit and then **bust out**

… and the bank loses

- Collections process ensues
  - Real addresses are visited
  - Fraudsters deny all knowledge of synthetic IDs
  - Bank writes off debt
- Two fraudsters can easily rack up $80k
- Well organised crime rings can rack up many times that

Discrete Data Analysis fails to predict…

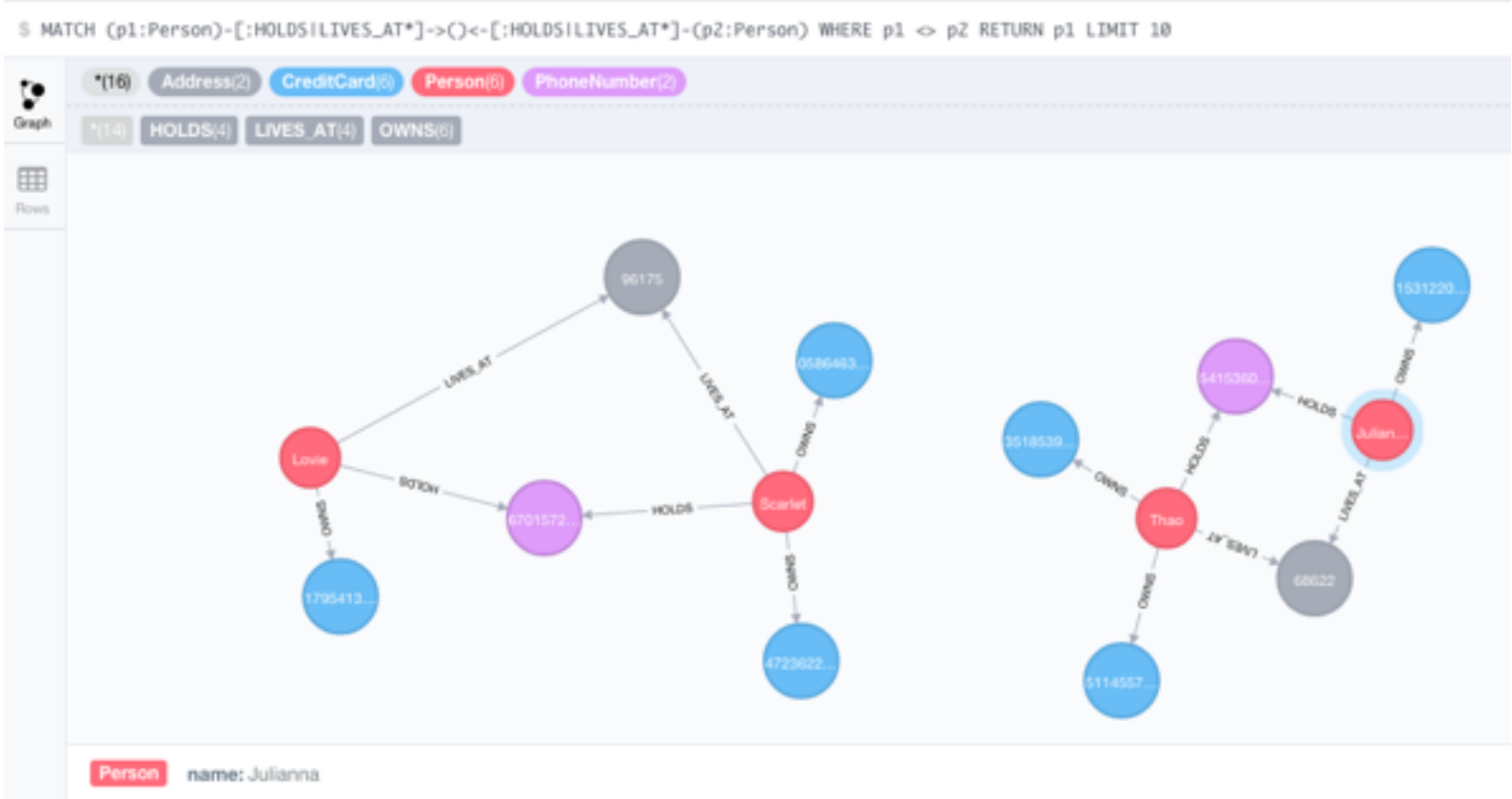


…and Makes it Hard to React

- When the bust out starts to happen, how do you know what to cancel?
- And how do you do it faster then the fraudster to limit your losses?

- A graph, that’s how!

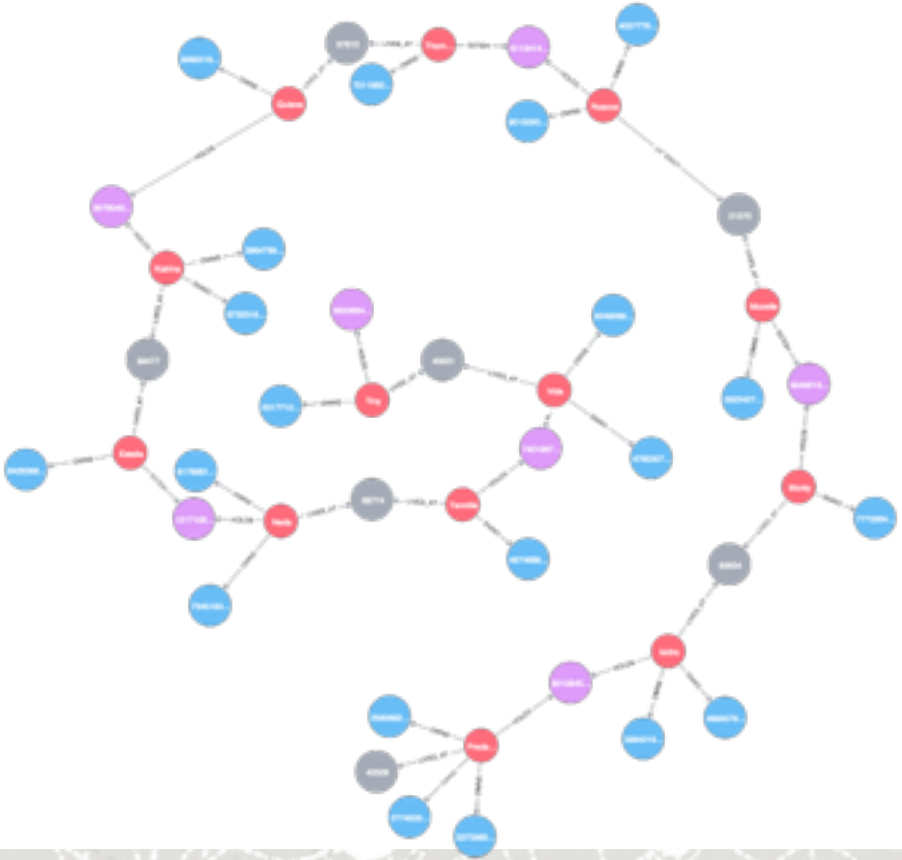
Probably Non-Fraudulent Cohabiters



Probable Cohabiters Query

```
MATCH (p1:Person)-[:HOLDS|LIVES_AT*]->()
    <-[:HOLDS|LIVES_AT*]-(p2:Person)
WHERE p1 <> p2
RETURN DISTINCT p1
```

Dodgy-Looking Chain

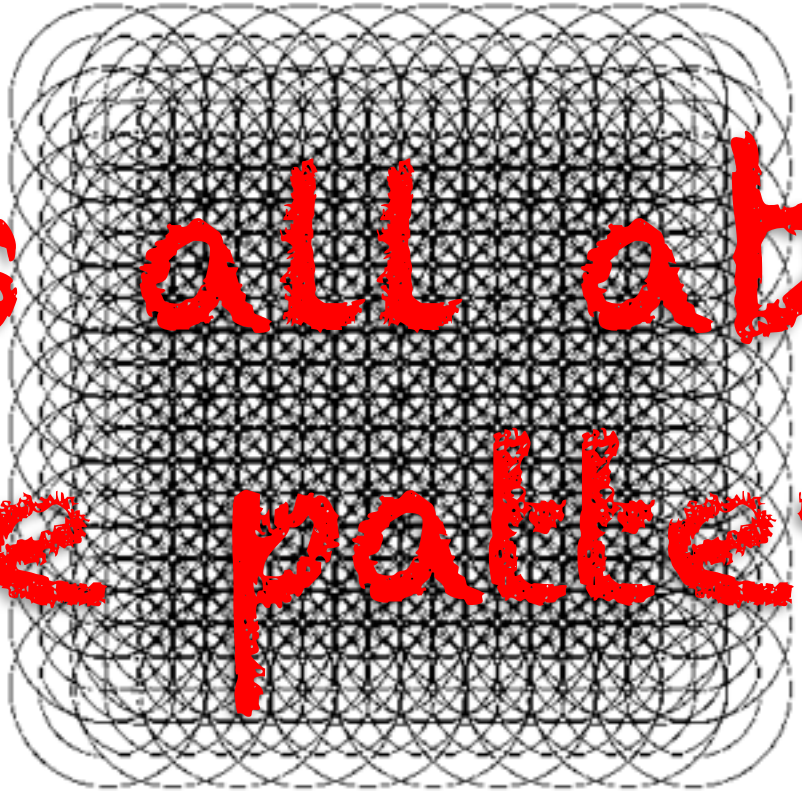


Risky Card Numbers

```
MATCH (p1:Person)-[:HOLDS|LIVES_AT*]->()
  <-[:HOLDS|LIVES_AT*]-(p2:Person)
  -[:HOLDS|LIVES_AT*]->()
  <-[:HOLDS|LIVES_AT*]-(p3:Person)
WHERE p1 <> p2 AND p2 <> p3 AND p3 <> p1
WITH p1, p2, p3
MATCH (p1)-[:OWNS]->(c1:CreditCard),
  (p2)-[:OWNS]->(c2:CreditCard),
  (p3)-[:OWNS]->(c3:CreditCard)
UNWIND [c1, c2, c3] AS creditCardNumbers
RETURN creditCardNumbers
```

neo4j

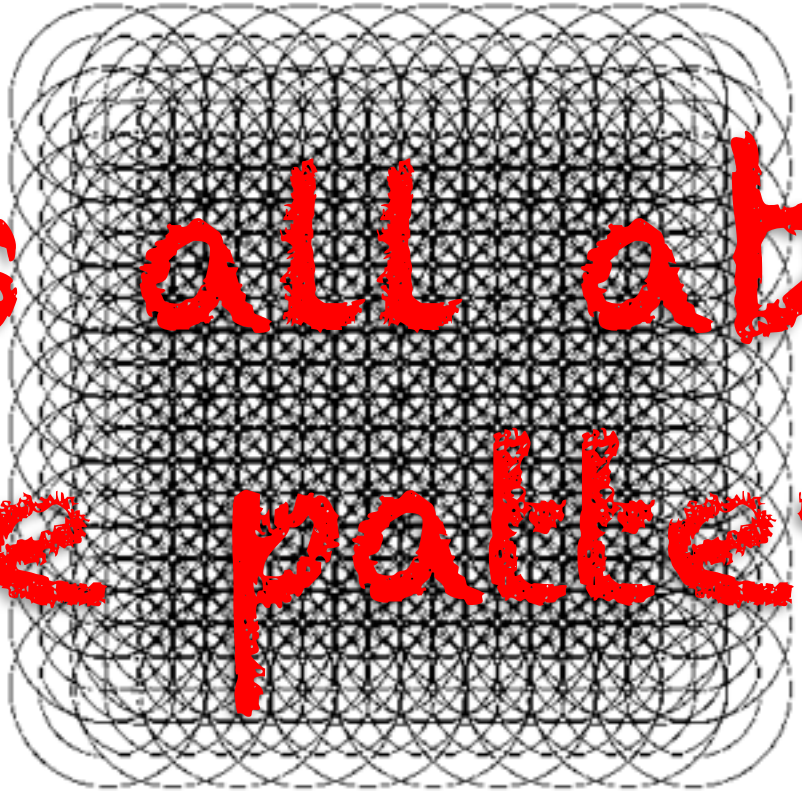
It's all about the patterns



Let me just repeat that

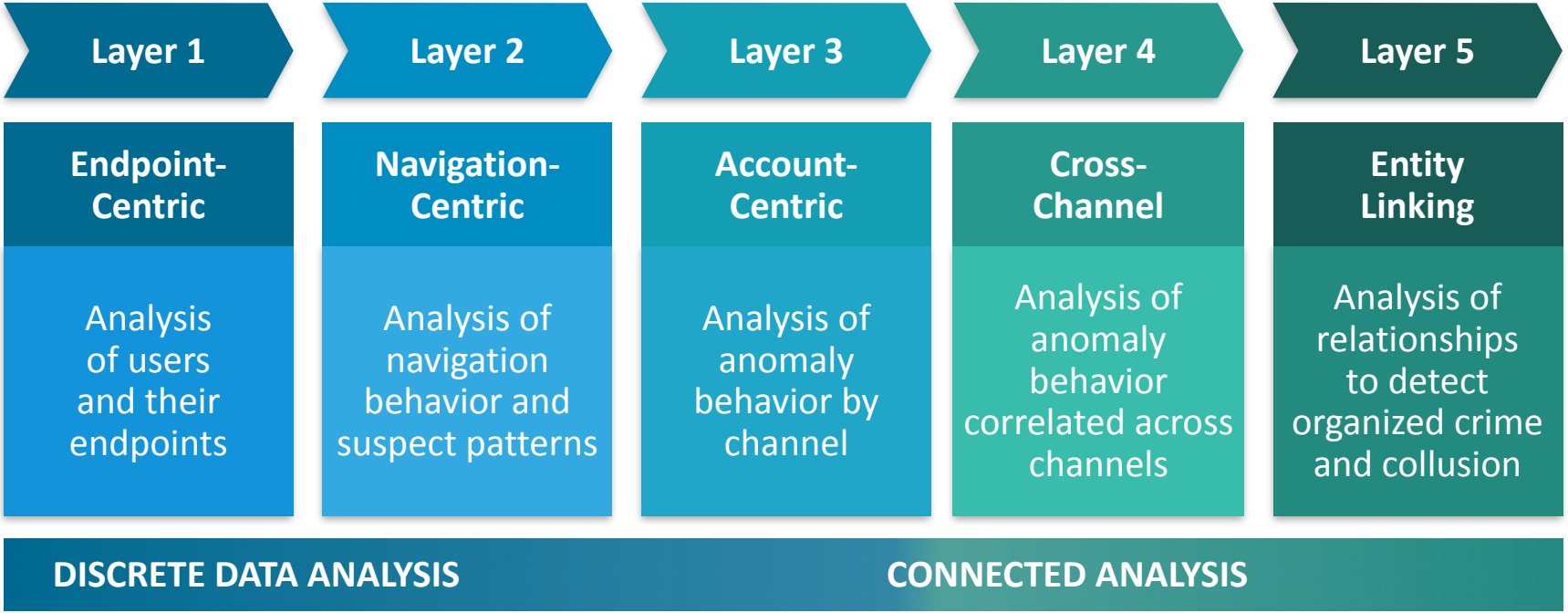
neo4j

It's all about the patterns



How does this fit with traditional fraud prevention?

Gartner's Layered Fraud Prevention Approach (4)



(4) http://www.gartner.com/newsroom/id/1695014

research now
Curaspan HEALTH GROUP
identropy
NexLP
Walmart
teachscape
Glowbl
OPEN Tree of Life
Pitney Bowes
shutl
neo4j
LOCKHEED MARTIN
Livestation
ebay
Pro Integra S.L.
meetic
ConocoPhillips
Wooz world
SevenBridges
NATIONAL GEOGRAPHIC
AXON ACTIVE
mallowstreet
telenor
doximity
DingLi.com
Accenture
Global 500 Finance
Global 500 Finance
intuit
WANDERU
Global 500 Finance
kitedesk
gen
opencorporates
EQUILAR
StarHub
die Bayerische
Global 500 Finance
ice
noble group
SNAP INTERACTIVE
CISCO
<fuseworks/>
UBS
DRW TRADING GROUP
First Data
onefinestay
janssen
spring
KiwiRail
webshots
Mapillary
SMARTMONSTERS
ZEPHYR HEALTH INC
classmates
TOMTOM
tallcat
SharePractice
careerbuilder.com
SHRM
SOCIETY FOR HUMAN RESOURCE MANAGEMENT
superserious
Lufthansa Systems
timeBlend!
Zenoss
CareerArcGroup
splink
megree
wazoku
DRAKER
minettabrook
Beamly
Justdial
Humanvest.co
InfoJobs
maaii
aikux.com
compete
viadeo
ePals
SERENA
gamesys

research now
Curaspan
HEALTH GROUP
entropy
NexLP
Walmart
Global 500
Finance
OPEN
intuit
Bowes
shutl
neo4j
ebay
Pro Integra S.L.
meetic
ConocoPhillips
Global 500
Finance
NATIONAL
GEOGRAPHIC
telenor
doximity
DingUcom
AXON ACTIVE
mallow street
opencorporates
WANDERU
Accenture
die Bayerische
kitedesk
gen
EQUILAR
StarHub
noble
group
Global 500
Finance
SNAP
INTERACTIVE
CISCO
<fuseworks/>
onefinestay
ICE
janssen
spring
KiwiRail
webshots
Mapillary
SmartMonsters
ZEPHYR
classmates
TOMTOM
SharePractice
careerbuilder.com
SHRM
tallcat
superserious
Lufthansa Systems
timeBlend!
First Data.
UBS
megree
wazoku
DRW TRADING GROUP
DRAKER
Beamly
Justdial
Humanvest.co
InfoJobs
aikux.com
compete
viadeo
ePals
SERENA
gamesys

Ask for help if you get stuck

- Online training - http://neo4j.com/graphacademy/
- Videos - http://vimeo.com/neo4j/videos
- Use cases - http://www.neotechnology.com/industries-and-use-cases/
- Meetups every week in Southwark
- Book @ http://www.graphdatabases.com

Or ping me?

rik@neotechnology.com

@rvanbruggen

+32 478 686800

blog.bruggen.com

www.neo4j.com

Thanks for listening