

# Graph Connect Europe 2016



- ▶ 26th April 2016
- ▶ **HERE** QEll Centre, Westminster, London
- ▶ <http://www.graphconnect.com>
- ▶ Use **QCON50** to get 50% off

# **Building a Recommendation Engine with Neo4j**

Michael Hunger @mesirii  
created by

Mark Needham @markhneedham

(Michael)-[:WORKS\_FOR]->(Neo4j)



[michael@neo4j.org](mailto:michael@neo4j.org) | [@mesirii](https://twitter.com/mesirii) | [github.com/jexp](https://github.com/jexp) | [jexp.de/blog](https://jexp.de/blog)

Michael Hunger - Community Caretaker @Neo4j

# Once Upon a Time in Sweden





**Solution**

# History of Neo4j



- 0.x ...  
small embeddable persistent graph library
- 1.x ...  
adding indexes, server, first stab of Cypher
- 2.x ...  
ease of use, data-model, optional schema,  
cost based optimizer, import, Neo4j-Browser
- 3.x ...  
binary protocol, bytecode compiled queries,  
sharding





**facebook**

**(graphs)-[:ARE]->(everywhere)**



# Value from Data Relationships

## Common Use Cases

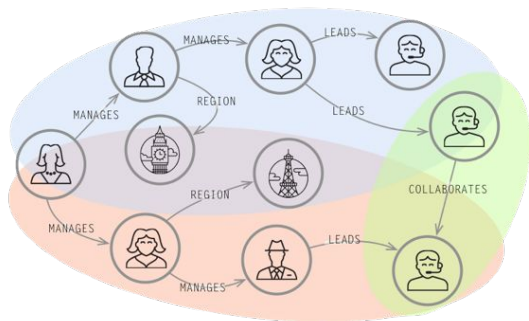


### Internal Applications

Master Data Management

Network and  
IT Operations

Fraud Detection

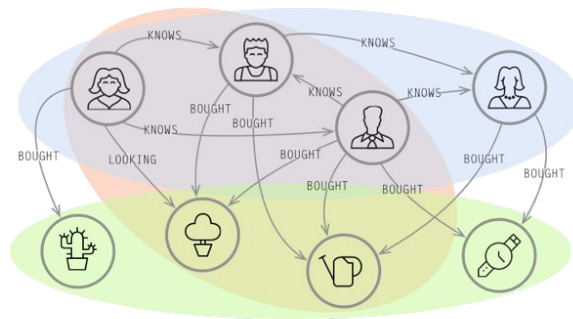


### Customer-Facing Applications

Real-Time Recommendations

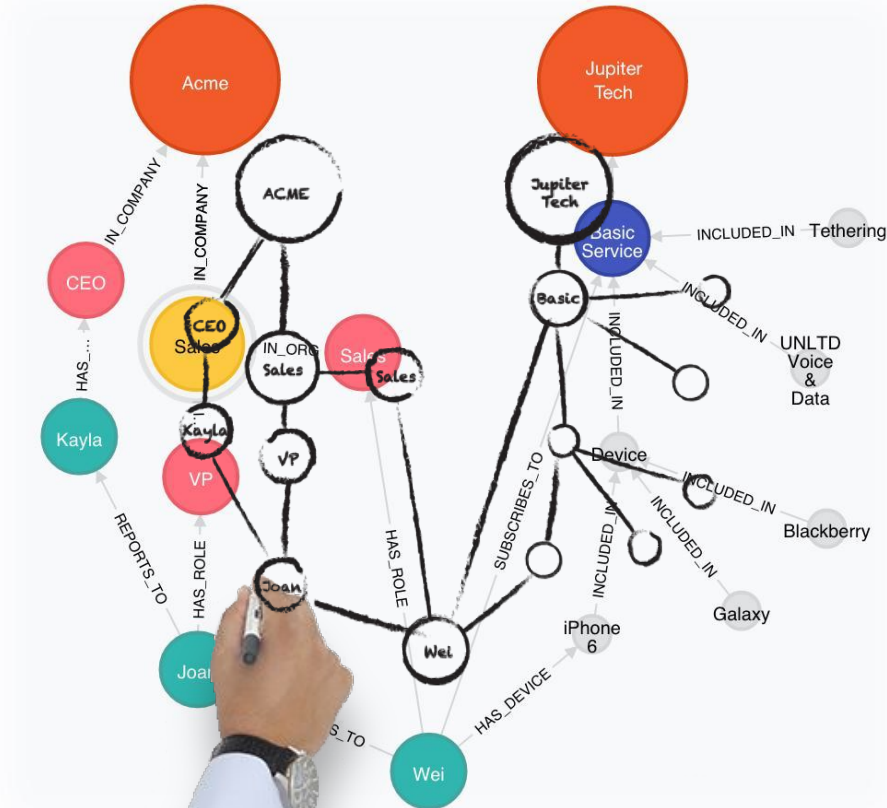
Graph-Based Search

Identity and  
Access Management





# The Whiteboard Model Is the Physical Model



# Property Graph Model

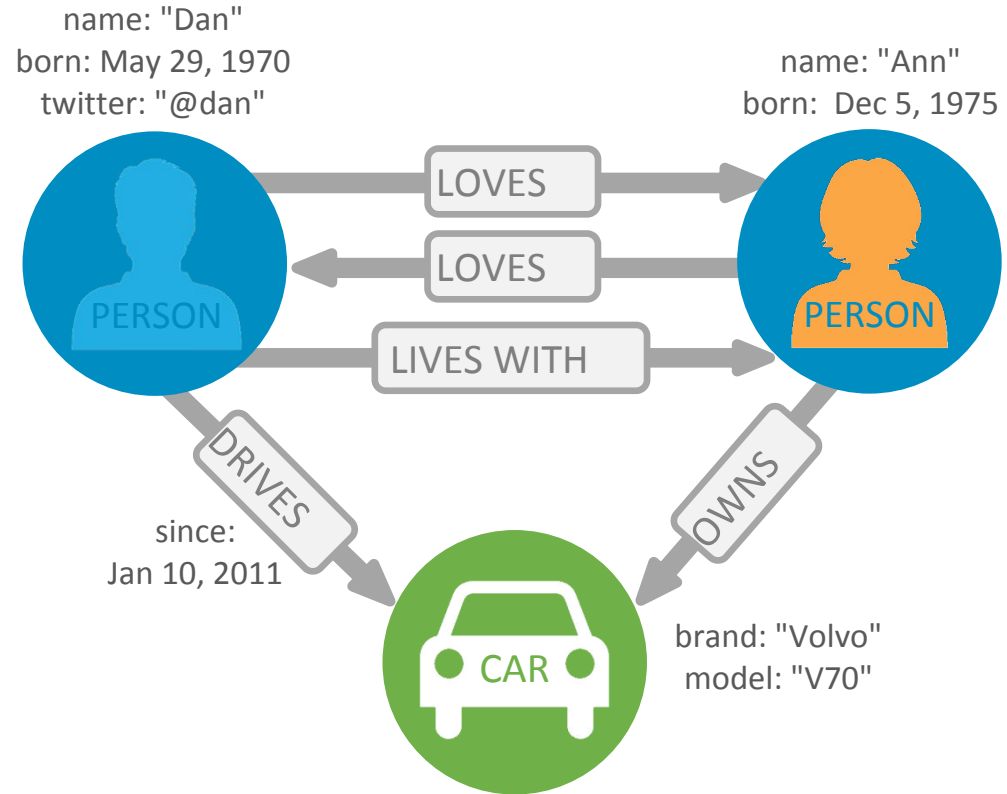


## Nodes

- The objects in the graph
- Can have name-value *properties*
- Can be *labeled*

## Relationships

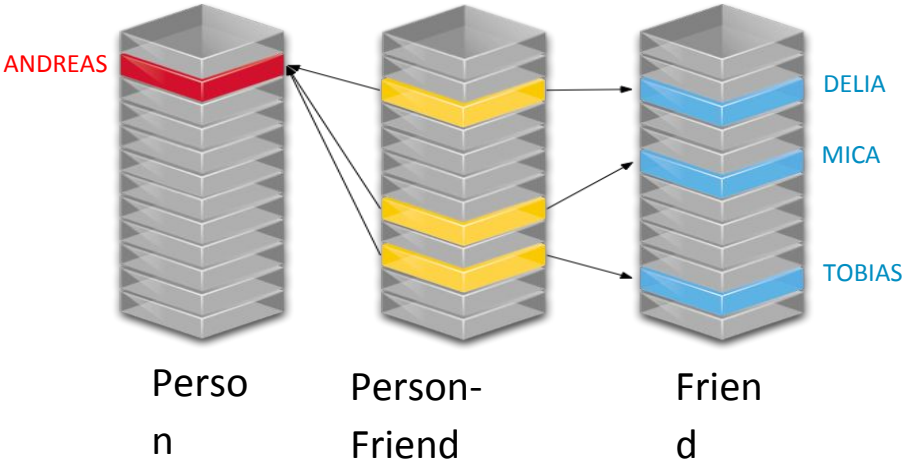
- Relate nodes by type and direction
- Can have name-value *properties*



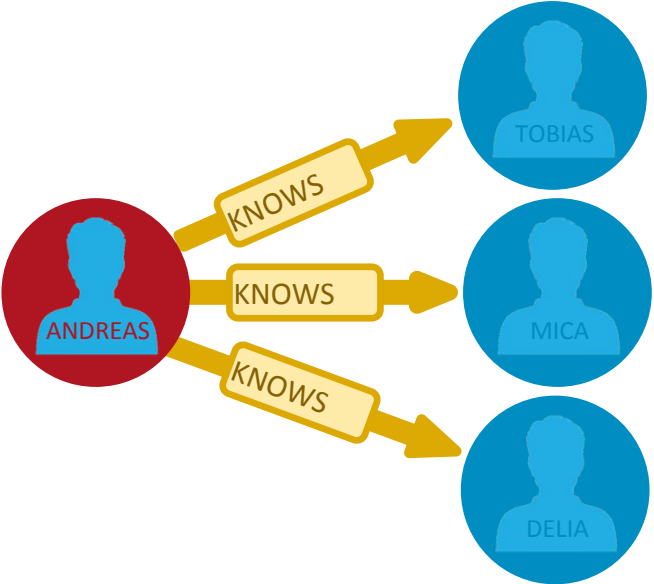
# Relational to Graph



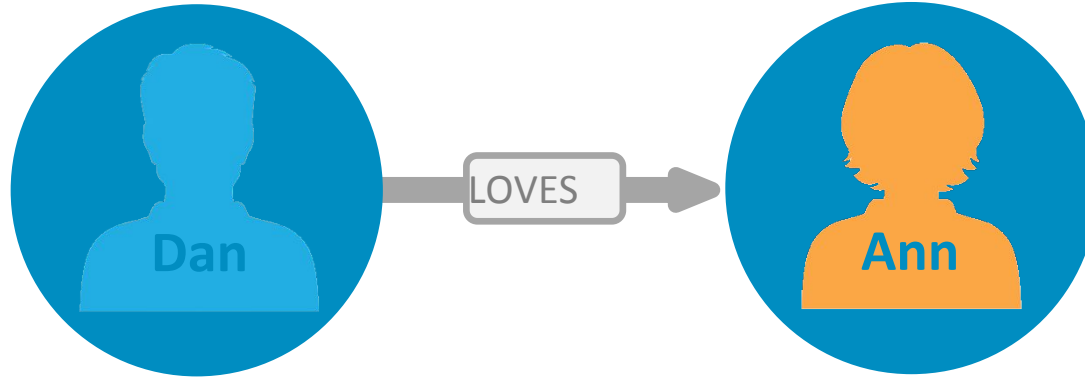
## Relational



## Graph



# Neo4j: All About Patterns



NODE

NODE

```
(:Person { name:"Dan" } ) -[:LOVES]-> (:Person { name:"Ann" } )
```

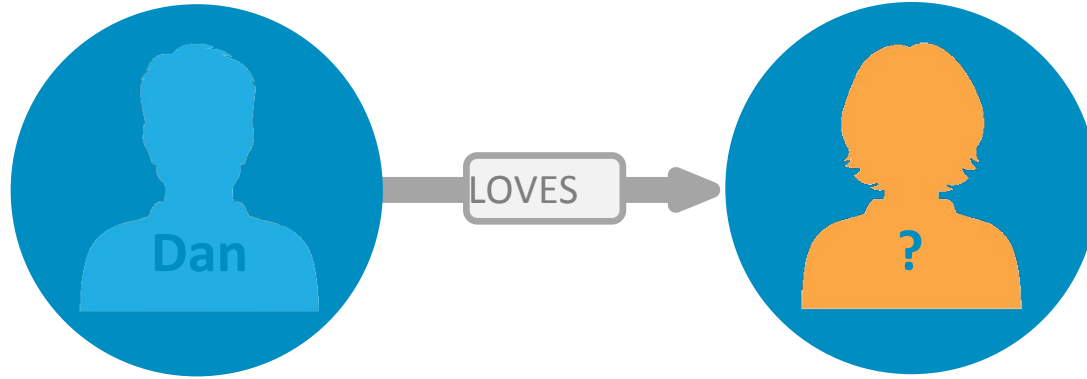
LABEL

PROPERTY

LABEL

PROPERTY

# Cypher: Find Patterns



```
MATCH (:Person { name:"Dan" } ) -[:LOVES]-> (love:Person) RETURN love
```

LABEL

PROPERTY

ALIAS LABEL

ALIAS

# Introducing our data set...



# meetup.com's recommendations



## People in this

### Meetup are also in:



Data & Analytics Innovation & Entrepreneurship

1,167 Members



Society of Data Miners

531 Analytics Practitioners



Cassandra London

1,635 Members



London New Tech

5,558 New technologists



Meteor London

1,596 Meteorites



London Ajax User Group

1,407 Software Engineers

ay: Join 190 R Users at "LondonR Meeting (and Workshop)" Inbox x

LondonR <info@meetup.com> [Unsubscribe](#) 15:40 (15 hours ago)  
to me

#### Events in this message

LondonR Meeting (and Workshop) Mon 30 Nov 2015 14:30 – 22:00 (WET)

[Add to Google Calendar](#)

MONDAY

## LondonR Meeting (and Workshop)

LondonR

Monday, November 30, 2015

2:30 PM

Balls Brothers

Minster Court, Mincing Lane, EC3R 7PP

London

190 R Users going, including:



John Van Praag

"R in my main coding language. Always looking to gain deeper understanding of it."



Lonnie Hamm

New Meetup Group: Agile without Borders Inbox x

Meetup <info@meetup.com> [Unsubscribe](#)  
to me

21 Nov (4 days ago)



## Agile without Borders

Community of Agile practitioners with in interest in developing techniques to scale across geography, time and cultures.

[Join us](#)

[Find out more](#)



Organized by  
Peter Verster

Already joined



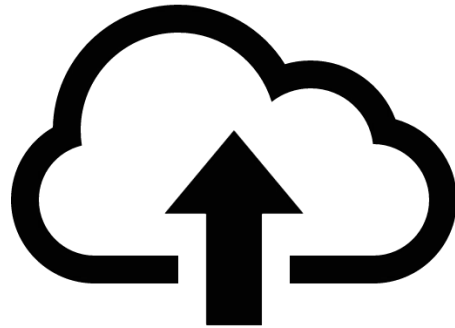
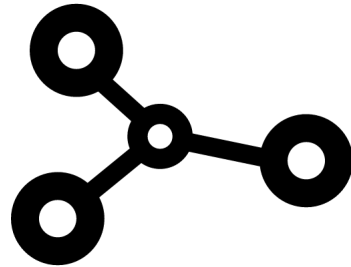
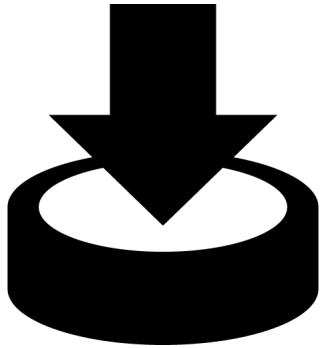
# Recommendation queries



- ▶ Several different types
  - groups to join
  - topics to follow
  - events to attend
- ▶ As a user of [meetup.com](https://www.meetup.com) trying to find groups to join, events to attend and people to meet



# How will this talk be structured?





### London, United Kingdom

Founded Jun 1, 2011

About us...

Invite friends

People 2,706

Group reviews 39

Upcoming Meetups 6

Past Meetups 150

Our calendar



Organizers: Neo4j, Mark Needham

Contact

We're about: Data Mining · New Technology · Web Development · Beta

## Graphs are everywhere. Use them.

+ SUGGEST A NEW MEETUP

Upcoming 6 Past Calendar

### Relational to graph: A worked example

Neo Technology 5 - 11 Lavington Street, London, SE10NZ, London (map)



In this session we'll take an existing relational database and port it into a Neo4j graph. We'll start off by coming up with some 'graphy' questions that we'd be able...

LEARN MORE

Hosted by: Mark Needham (Co-Organizer)

Thu Mar 10 6:30 PM

RSVP

2 days left

44 going

0 comments

### Modelling a recommendation engine: A worked example

Neo Technology 5 - 11 Lavington Street, London, SE10NZ, London (map)



In this session we'll build a recommendation engine from scratch while paying particular attention to the modelling choices made along the way. Our solution will be a...

LEARN MORE

Hosted by: Mark Needham (Co-Organizer)

Tue Mar 15 6:30 PM

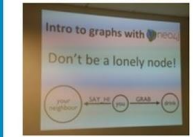
RSVP

7 days left

43 going

0 comments

### What's new



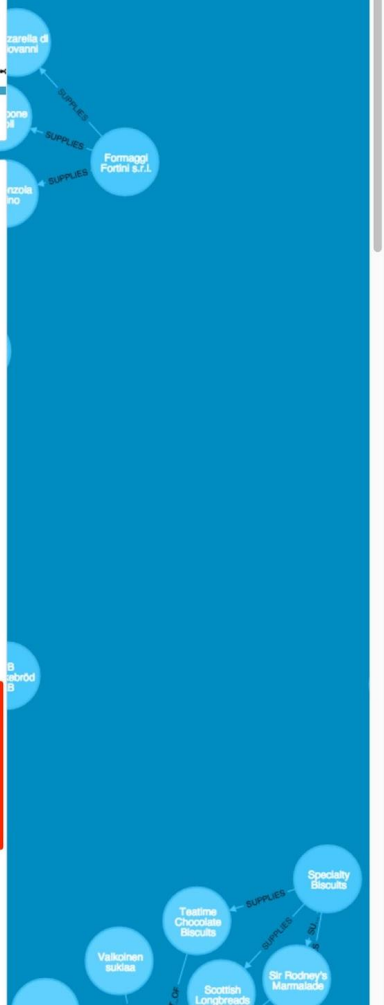
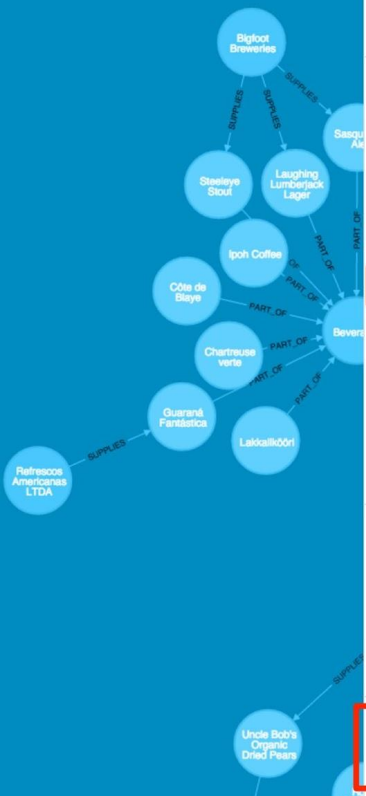
MORE

NEW MEMBER Joy Braun joined Yesterday

NEW MEMBER Arturas Smorgun joined Yesterday

NEW MEMBER Kacper Gunia joined Yesterday

NEW MEMBER Dean Connor joined

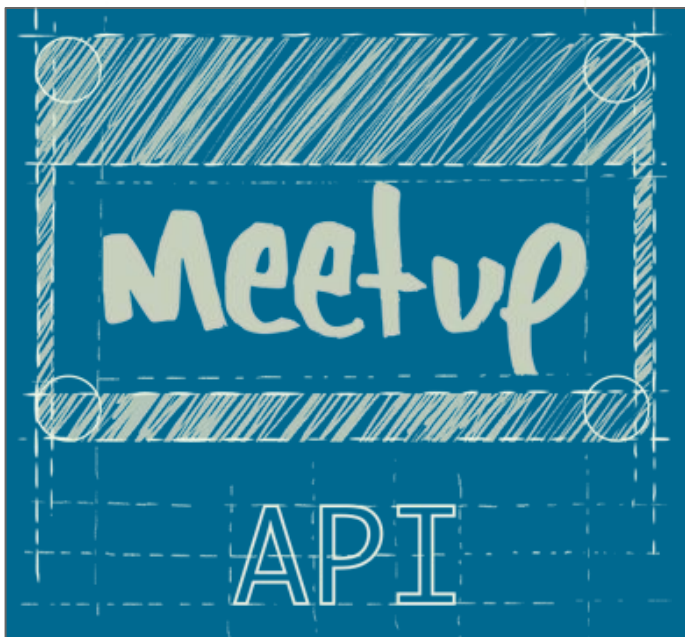


# Data ?



- ▶ Groups
- ▶ Members
- ▶ Events
- ▶ Topics
- ▶ Time & Date
- ▶ Location

# Get Data: Meetup API + jq



./jq



[meetup.com/meetup\\_api/](https://meetup.com/meetup_api/)

[stedolan.github.io/jq/](https://stedolan.github.io/jq/)

# Find similar groups to Neo4j



**As** a member of the Neo4j London group  
**I want** to find other similar meetup groups  
**So that** I can join those groups



# What makes groups similar?



We're about:

Data Mining · New Technology · Web Development · Data Visualization · Data Analytics · Open Source · Cloud Computing · Graph Databases · Big Data · NoSQL · Neo4j · Database Development · Java · Computer programming



We're about:

Open Source · Technology · Web Development · Computer programming · Agile Project Management · Java · Software Development



We're about:

Big Data Analytics · Artificial Intelligence · Computer programming · Big Data · Computer Science · Natural Language Processing · Machine Learning · Data Analytics · Data Visualization · Data Mining · Data Science · Algorithms · Deep Learning · neural networks



We're about:

Data Science · Machine Learning · Predictive Analytics · Data Mining · Big Data · Artificial Intelligence · Statistical Computing · Applied Statistics · Data Analytics · Open Source · Web Analytics · Text Analytics · Natural Language Processing · Hadoop · NoSQL



We're about:

BigData · Intellectual Discussion · Big Data · Debate · New Technology · Data Analytics · Data Visualization · Online Marketing · Database Development · Information Architecture · Information Science · Freedom · Cloud Computing · Business Strategy · Internet Professionals



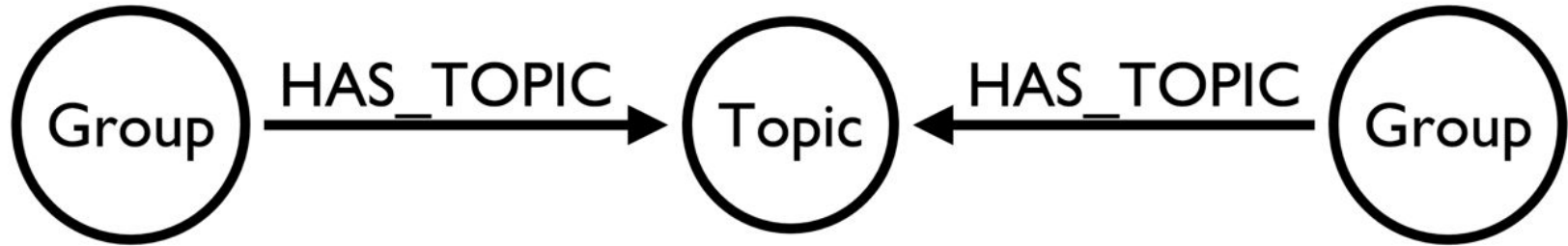
extract

We're about:

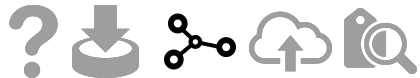
Marketing · Data · Software Development · New Technology · Web Technology · Business Intelligence · Cloud Computing · Business Strategy · Big Data · Machine Learning · Data Analytics · Data Visualization · Data Mining · Data Science · Big Data Analytics



# Find similar groups to Neo4j



**As** a member of the Neo4j London group  
**I want** to find other similar meetup groups  
**So that** I can join those groups



# LOAD CSV



```
LOAD CSV FROM "file:///groups.csv"  
AS row  
RETURN row LIMIT 5;
```

```
LOAD CSV WITH HEADERS FROM "file:///groups.csv"  
AS row WITH row  
WHERE row.rating > 4.5  
RETURN row;
```





# groups.csv



id	name	urlname	rating	created
841735	LJC - London Java Community	Londonjavacommunity	4.54	1196081014000
18313232	Kubernetes London	Kubernetes-London	5	1420729836000
18581527	data+visual London	data-visual-London	4.67	1431021679000
163876	London Web	londonweb	4.11	1034097743000
15734842	Ansible London	Ansible-London	4.42	1405439359000
12963902	Scalability London	Scalability-London	4.95	1392824462000
4062902	Ember London	London-Emberjs-User-Group	4.66	1339522219000



# Create groups



```
LOAD CSV WITH HEADERS FROM "file:///groups.csv"
AS row

CREATE (:Group { id:row.id,
                 name:row.name,
                 urlname:row.urlname,
                 rating:toInt(row.rating),
                 created:toInt(row.created) })
```



# Create groups



```
LOAD CSV WITH HEADERS FROM "file:///groups.csv"
AS row

CREATE (:Group { id:row.id,
                 name:row.name,
                 urlname:row.urlname,
                 rating:toint(row.rating),
                 created:toint(row.created) })
```

We use CREATE because the database is empty.



# groups\_topics.csv



id	name	urlkey
827	.NET	dotnet
2109	System Administration	sysadmin
2260	C#	csharp
10105	Microsoft Windows	mswindows
15167	Cloud Computing	cloud-computing
46810	Configuration Management	configuration-management
52210	PowerShell	powershell
66339	Windows Azure Platform	windows-azure-platform
84706	Scripting	scripting
87614	DevOps	devops
99537	Microsoft Technology	microsoft-technology
189	Java	java
563	Open Source	opensource



# Create topics



```
LOAD CSV WITH HEADERS FROM "file:///groups_topics.csv"  
AS row  
MERGE (topic:Topic {id: row.id})  
ON CREATE SET topic.name = row.name,  
              topic.urlkey = row.urlkey
```



# Create topics



```
LOAD CSV WITH HEADERS FROM "file:///groups_topics.csv"  
AS row  
MERGE (topic:Topic {id: row.id})  
ON CREATE SET topic.name = row.name,  
             topic.urlkey = row.urlkey
```

We use **MERGE** because we want to avoid creating duplicate topics



# Create unique constraints



```
CREATE CONSTRAINT ON (t:Topic)  
ASSERT t.id IS UNIQUE
```

```
CREATE CONSTRAINT ON (g:Group)  
ASSERT g.id IS UNIQUE
```



# Create unique constraints



```
CREATE CONSTRAINT ON (t:Topic)
```

```
ASSERT t.id IS UNIQUE
```

```
CREATE CONSTRAINT ON (g:Group)
```

```
ASSERT g.id IS UNIQUE
```

We create unique constraints to:

- ensure uniqueness across a (label, property) pair
- allow fast lookup of nodes which match these (label, property) pairs.

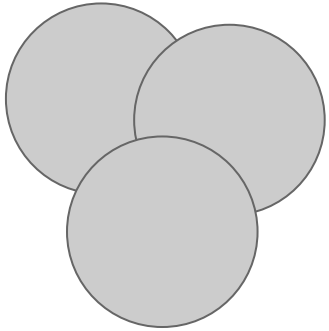




# How does Neo4j use indexes?

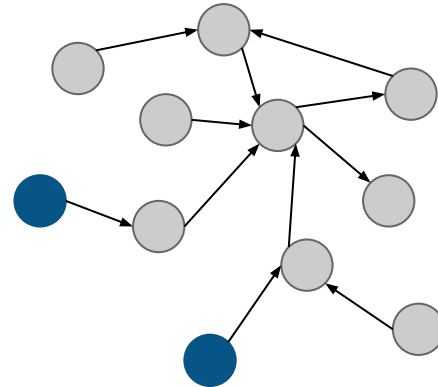


Indexes are **only** used to find the starting point for queries.



## Relational

Use index scans to look up rows in tables and join them with rows from other tables



## Graph

Use indexes to find the starting points for a query.



# Groups and topics



id	groupId
827	18780165
2109	18780165
2260	18780165
10105	18780165
15167	18780165
46810	18780165
52210	18780165



# Connect groups and topics



```
LOAD CSV WITH HEADERS FROM "file:///groups_topics.csv"  
AS row  
MATCH (topic:Topic {id: row.id})  
MATCH (group:Group {id: row.groupId})  
MERGE (group)-[:HAS_TOPIC]->(topic)
```



# Connect groups and topics



```
LOAD CSV WITH HEADERS FROM "file:///groups_topics.csv"  
AS row  
MATCH (topic:Topic {id: row.id})  
MATCH (group:Group {id: row.groupId})  
MERGE (group)-[:HAS_TOPIC]->(topic)
```

We can use MERGE to uniquely create relationships as well



# Create index



```
CREATE INDEX ON :Group(name)
```



# Create index



```
CREATE INDEX ON :Group(name)
```

We create an index on :Group(name) so that we can quickly look up groups by name.



# Find similar groups to Neo4j



```
MATCH (group:Group {name: "Neo4j - London User Group"})
      -[:HAS_TOPIC]->(topic)<-[:HAS_TOPIC]-(otherGroup)
RETURN otherGroup.name,
       COUNT(topic) AS topicsInCommon,
       COLLECT(topic.name) AS topics
ORDER BY topicsInCommon DESC, otherGroup.name
LIMIT 10
```



# Find similar groups to Neo4j



```
$ MATCH (group:Group {name: "Neo4j - London User Group"}) -[:HAS_TOPIC]->(topic)<-[:HAS_TOPIC]-(otherGroup) RETURN otherGroup.name, COUNT(topic) AS to...
```

	otherGroup.name	topicsInCommon	topics
Rows	Python for Quant Finance	8	[New Technology, Cloud Computing, Big Data, Data Analytics, Data Visualization, Computer programming, Open Source, Data Mining]
Code	Closed scaling group	7	[Cloud Computing, Web Development, Big Data, Computer programming, Java, Open Source, NoSQL]
	Couchbase London	7	[Cloud Computing, Big Data, Database Development, Data Analytics, Open Source, NoSQL, Data Mining]
	London PostgreSQL Meetup Group	7	[New Technology, Cloud Computing, Web Development, Big Data, Database Development, Open Source, NoSQL]
	Business, Analytics and Data Science	7	[New Technology, Big Data, Data Analytics, Data Visualization, Computer programming, Open Source, Data Mining]
	Analytics.Club London	7	[Cloud Computing, Big Data, Data Analytics, Data Visualization, Open Source, NoSQL, Data Mining]
	Hadoop Users Group UK	7	[Cloud Computing, Big Data, Data Analytics, Java, Open Source, NoSQL, Data Mining]
	London NoSQL	6	[New Technology, Cloud Computing, Web Development, Big Data, Open Source, NoSQL]
	Data Science & Business Analytics London Meetup	6	[Big Data, Data Analytics, Data Visualization, Open Source, NoSQL, Data Mining]
	Big Data Week London Meetup	6	[Big Data, Data Analytics, Data Visualization, Open Source, NoSQL, Data Mining]

Returned 10 rows in 22 ms.





# I'm already a member of these!



```
$ MATCH (group:Group {name: "Neo4j - London User Group"}) -[:HAS_TOPIC]->(topic)<-[:HAS_TOPIC]-(otherGroup) RETURN otherGroup.name, COUNT(topic) AS to...
```

	otherGroup.name	topicsInCommon	topics
Rows	Python for Quant Finance	8	[New Technology, Cloud Computing, Big Data, Data Analytics, Data Visualization, Computer programming, Open Source, Data Mining]
Code	Closed scaling group	7	[Cloud Computing, Web Development, Big Data, Computer programming, Java, Open Source, NoSQL]
	Couchbase London	7	[Cloud Computing, Big Data, Database Development, Data Analytics, Open Source, NoSQL, Data Mining]
	London PostgreSQL Meetup Group	7	[New Technology, Cloud Computing, Web Development, Big Data, Database Development, Open Source, NoSQL]
	Business, Analytics and Data Science	7	[New Technology, Big Data, Data Analytics, Data Visualization, Computer programming, Open Source, Data Mining]
	Analytics.Club London	7	[Cloud Computing, Big Data, Data Analytics, Data Visualization, Open Source, NoSQL, Data Mining]
	Hadoop Users Group UK	7	[Cloud Computing, Big Data, Data Analytics, Java, Open Source, NoSQL, Data Mining]
	London NoSQL	6	[New Technology, Cloud Computing, Web Development, Big Data, Open Source, NoSQL]
	Data Science & Business Analytics London Meetup	6	[Big Data, Data Analytics, Data Visualization, Open Source, NoSQL, Data Mining]
	Big Data Week London Meetup	6	[Big Data, Data Analytics, Data Visualization, Open Source, NoSQL, Data Mining]

Returned 10 rows in 22 ms.



# Exclude groups I'm a member of



**As** a member of the Neo4j London group  
**I want** to find other similar meetup groups  
that I'm not already a member of  
**So that** I can join those groups

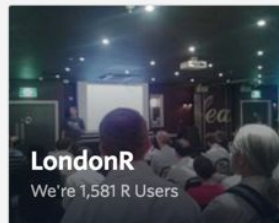
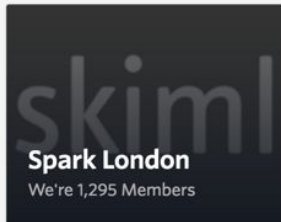
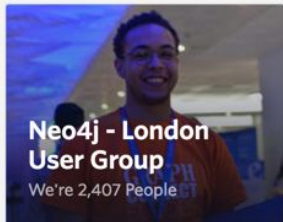


# What other data can we get?



## Your Meetups

Sort by Recommended



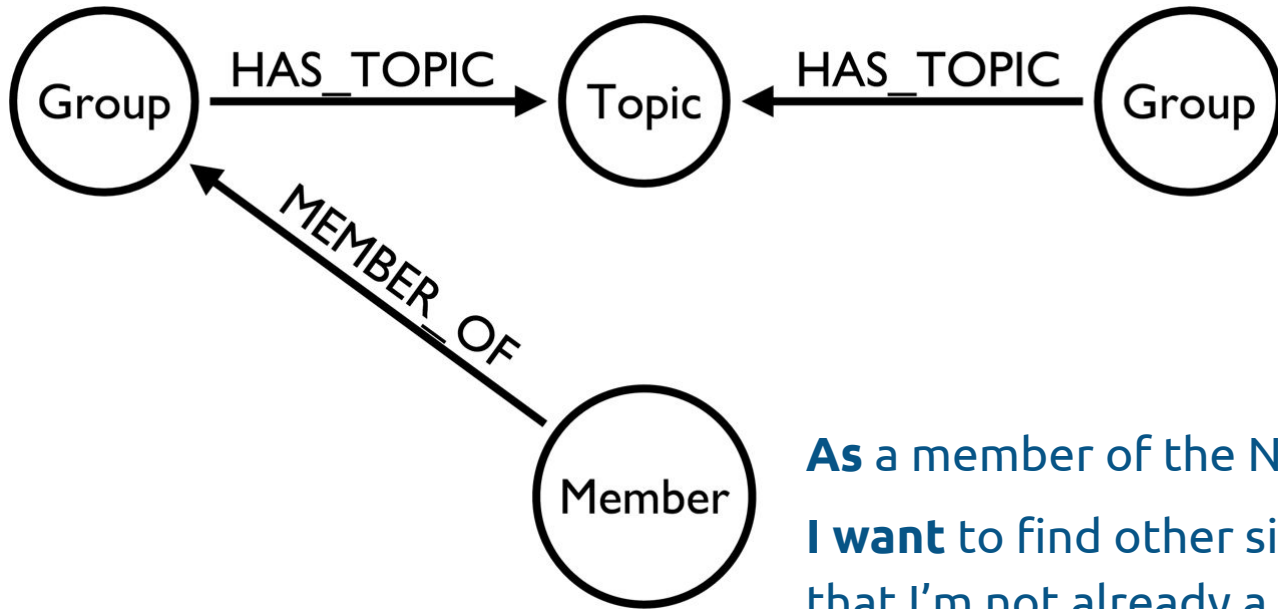
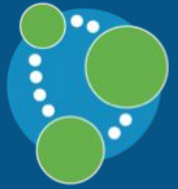
## Interests

Hide my interests from others · Edit

Software Development · Test Driven Development · Open Source · Data Visualization · Natural Language Processing · Data Mining · Data Science · Machine Learning · Data Analytics · Neo4j · Graph Databases · NoSQL



# Exclude groups I'm a member of



**As** a member of the Neo4j London group  
**I want** to find other similar meetup groups  
that I'm not already a member of  
**So that** I can join those groups

# members.csv



id	name	joined
103929052	A	1378461129000
11337881	Abhishek Shivkumar	1421419313000
39676622	Ali Syed	1395723669000
2773509	Amit	1407935487000
30225872	Attila Sztupak	1378812292000
12882650	Cathy White	1423566263000
109548702	Danny Bickson	1378196635000



# Create members



```
LOAD CSV WITH HEADERS FROM "file:///path/to/members.csv" AS row
WITH DISTINCT row.id AS id, row.name AS name
MERGE (member:Member {id: id})
ON CREATE SET member.name = name
```



# Members and groups



id	groupId
103929052	10087112
11337881	10087112
39676622	10087112
2773509	10087112
30225872	10087112
12882650	10087112
109548702	10087112



# Connect members and groups



```
LOAD CSV WITH HEADERS FROM "file:///path/to/members.csv" AS row
WITH row WHERE NOT row.joined is null
MATCH (member:Member {id: row.id})
MATCH (group:Group {id: row.groupId})
MERGE (member)-[:MEMBER_OF {joined: toint(row.joined)}]->(group)
```





# Exclude groups I'm a member of



```
MATCH (group:Group {name: "Neo4j - London User Group"})
      -[:HAS_TOPIC]->(topic)<-[:HAS_TOPIC]-(otherGroup:Group)
RETURN otherGroup.name,
       COUNT(topic) AS topicsInCommon,
       EXISTS((:Member {name: "Mark Needham"})
             -[:MEMBER_OF]->(otherGroup)) AS alreadyMember,
       COLLECT(topic.name) AS topics
ORDER BY topicsInCommon DESC
LIMIT 10
```

# Exclude groups I'm a member of



```
$ MATCH (group:Group {name: "Neo4j - London User Group"}) -[:HAS_TOPIC]->(topic)-[:HAS_TOPIC]-(otherGroup:Group) RETURN otherGroup.name, COUNT(topic)...
```



Rows



Code

otherGroup.name	topicsInCommon	alreadyMember	topics
Python for Quant Finance	8	false	[New Technology, Cloud Computing, Big Data, Data Analytics, Data Visualization, Computer programming, Open Source, Data Mining]
Analytics.Club London	7	false	[Cloud Computing, Big Data, Data Analytics, Data Visualization, Open Source, NoSQL, Data Mining]
Hadoop Users Group UK	7	true	[Cloud Computing, Big Data, Data Analytics, Java, Open Source, NoSQL, Data Mining]
London PostgreSQL Meetup Group	7	true	[New Technology, Cloud Computing, Web Development, Big Data, Database Development, Open Source, NoSQL]
Closed scaling group	7	false	[Cloud Computing, Web Development, Big Data, Computer programming, Java, Open Source, NoSQL]
Couchbase London	7	false	[Cloud Computing, Big Data, Database Development, Data Analytics, Open Source, NoSQL, Data Mining]
Business, Analytics and Data Science	7	false	[New Technology, Big Data, Data Analytics, Data Visualization, Computer programming, Open Source, Data Mining]
Agile Data - London Meetup	6	false	[New Technology, Cloud Computing, Big Data, Data Analytics, Open Source, NoSQL]
London Data Visualization	6	false	[Big Data, Data Analytics, Data Visualization, Graph Databases, Neo4j, NoSQL]
import.io users group	6	true	[New Technology, Cloud Computing, Big Data, Data Analytics, Data Visualization, Data Mining]

Returned 10 rows in 48 ms.

# Exclude groups I'm a member of



```
MATCH (group:Group {name: "Neo4j - London User Group"})
      -[:HAS_TOPIC]->(topic)<-[:HAS_TOPIC]-(otherGroup:Group)
WHERE NOT( (:Member {name: "Mark Needham"})
           -[:MEMBER_OF]->(otherGroup) )
RETURN otherGroup.name,
       COUNT(topic) AS topicsInCommon,
       COLLECT(topic.name) AS topics
ORDER BY topicsInCommon DESC
LIMIT 10
```



# Exclude groups I'm a member of



```
$ MATCH (group:Group {name: "Neo4j - London User Group"}) -[:HAS_TOPIC]->(topic)-[:HAS_TOPIC]-(otherGroup:Group) WHERE NOT((:Member {name: "Mark Need...
```



Rows



Code

otherGroup.name	topicsInCommon	topics
Python for Quant Finance	8	[New Technology, Cloud Computing, Big Data, Data Analytics, Data Visualization, Computer programming, Open Source, Data Mining]
Closed scaling group	7	[Cloud Computing, Web Development, Big Data, Computer programming, Java, Open Source, NoSQL]
Analytics.Club London	7	[Cloud Computing, Big Data, Data Analytics, Data Visualization, Open Source, NoSQL, Data Mining]
Couchbase London	7	[Cloud Computing, Big Data, Database Development, Data Analytics, Open Source, NoSQL, Data Mining]
Business, Analytics and Data Science	7	[New Technology, Big Data, Data Analytics, Data Visualization, Computer programming, Open Source, Data Mining]
Data Science & Business Analytics London Meetup	6	[Big Data, Data Analytics, Data Visualization, Open Source, NoSQL, Data Mining]
Big Data Developers in London	6	[Cloud Computing, Big Data, Data Analytics, Data Visualization, NoSQL, Data Mining]
Big Data Jobs	6	[New Technology, Cloud Computing, Big Data, Data Analytics, Open Source, Data Mining]
London Data Visualization	6	[Big Data, Data Analytics, Data Visualization, Graph Databases, Neo4j, NoSQL]
Big Data Week London Meetup	6	[Big Data, Data Analytics, Data Visualization, Open Source, NoSQL, Data Mining]

Returned 10 rows in 23 ms.



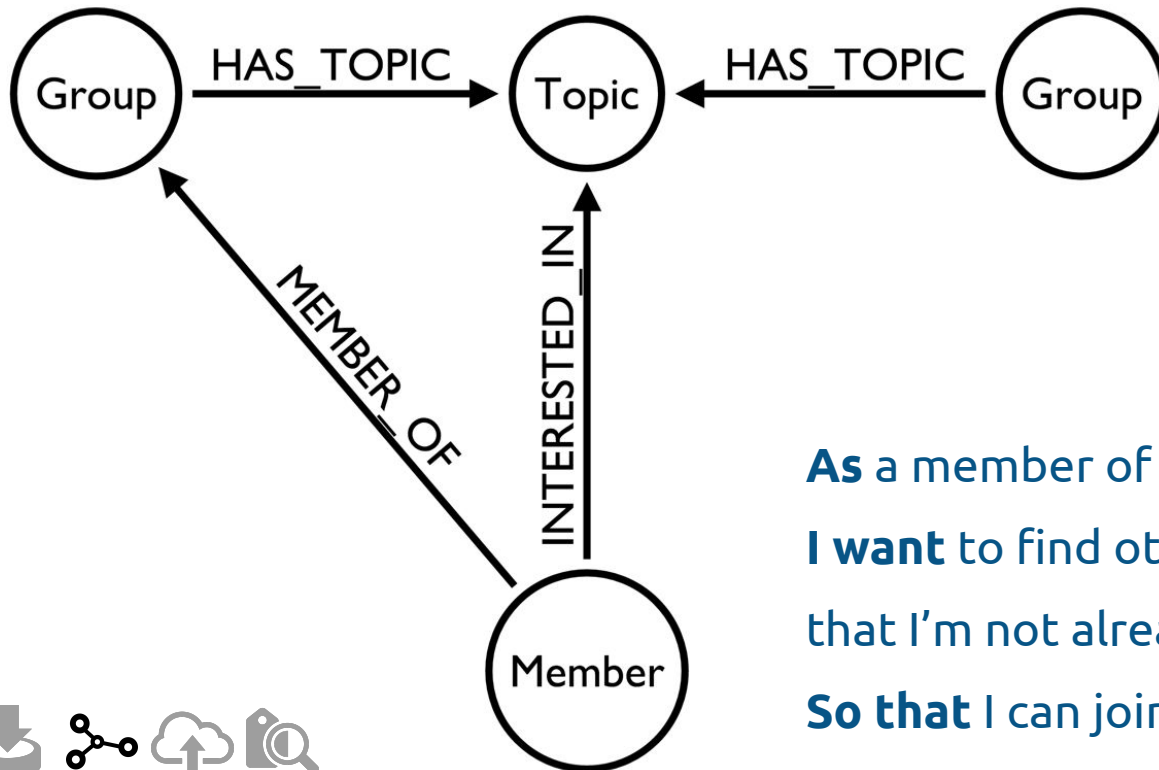
# Find my similar groups



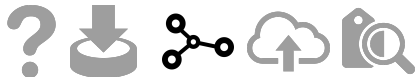
**As** a member of several meetup groups  
**I want** to find other similar meetup groups  
that I'm not already a member of  
**So that** I can join those groups



# Find my similar groups



**As** a member of several meetup groups  
**I want** to find other similar meetup groups  
that I'm not already a member of  
**So that** I can join those groups



# Members and topics



id	topics
103929052	18062;563;16575;20923;3833;108403;1307;10099
11337881	1372;1512;49585;24553;417;24778;25584;23005
39676622	
2773509	
30225872	48471;22792;58162;1762
12882650	563;3833;9696;659;1621,48471;22792
109548702	21681;30928;18062;5532,55324;15167;108403



# Connect members and topics



```
USING PERIODIC COMMIT 10000
```

```
LOAD CSV WITH HEADERS FROM "file:///path/to/members.csv" AS row
```

```
WITH split(row.topics, ";") AS topics, row.id AS memberId
```

```
UNWIND topics AS topicId
```

```
MATCH (member:Member {id: memberId})
```

```
MATCH (topic:Topic {id: topicId})
```

```
MERGE (member)-[:INTERESTED_IN]->(topic)
```





# Find my similar groups



```
MATCH (member:Member {name: "Mark Needham"})  
  -[:INTERESTED_IN]->(topic),  
  (member)-[:MEMBER_OF]->(group)-[:HAS_TOPIC]->(topic)
```

```
WITH member, topic, COUNT(*) AS score  
MATCH (topic)<-[:HAS_TOPIC]-(otherGroup)  
WHERE NOT (member)-[:MEMBER_OF]->(otherGroup)  
RETURN otherGroup.name,  
       COLLECT(topic.name),  
       SUM(score) as score  
ORDER BY score DESC
```



# Find my similar groups



```
S MATCH (member:Member {name: "Mark Needham"}) -[:INTERESTED_IN]->(topic), (member)-[:MEMBER_OF]->(group) -[:HAS_TOPIC]->(topic) WITH member, topic, C...
```

Rows	otherGroup.name	COLLECT(topic.name)	score
	Data Science & Business Analytics London Meetup	[Machine Learning, Open Source, NoSQL, Data Visualization, Data Science, Data Mining, Data Analytics]	79
	Data Science Lab	[Machine Learning, Open Source, NoSQL, Data Visualization, Data Science, Data Mining, Data Analytics]	79
	Big Data Week London Meetup	[Machine Learning, Open Source, NoSQL, Data Visualization, Data Science, Data Mining, Data Analytics]	79
	Business, Analytics and Data Science	[Machine Learning, Open Source, Data Visualization, Data Science, Data Mining, Data Analytics]	72
	Data & Analytics Innovation & Entrepreneurship	[Machine Learning, Open Source, Data Visualization, Data Science, Data Mining, Data Analytics]	72
	PyData London Meetup	[Machine Learning, Open Source, Data Visualization, Data Science, Data Mining, Data Analytics]	72
	Big Data Developers in London	[Machine Learning, NoSQL, Data Visualization, Data Science, Data Mining, Software Development, Data Analytics]	72
	Data Scientist	[Machine Learning, Open Source, Data Visualization, Data Science, Data Mining, Data Analytics]	72
	Advanced Data Visualisation London	[Machine Learning, Open Source, Data Visualization, Data Science, Data Mining, Data Analytics]	72
	London Data Science #ODSC	[Machine Learning, Open Source, Data Visualization, Data Science, Software Development, Data Analytics]	70
	Agile Data - London Meetup	[Machine Learning, Open Source, NoSQL, Data Science, Software Development, Data Analytics]	66
	London DataTech Startups	[Machine Learning, NoSQL, Data Visualization, Data Science, Data Mining, Data Analytics]	64
	DataKind UK	[Machine Learning, Open Source, Data Visualization, Data Science, Data Analytics]	62
	Python for Quant Finance	[Machine Learning, Open Source, Data Visualization, Data Mining, Data Analytics]	61
	KNIME User Group UK	[Open Source, Data Visualization, Data Science, Data Mining, Data Analytics]	61
	Analytics Network	[Machine Learning, Data Visualization, Data Science, Data Mining, Data Analytics]	57

Returned 322 rows in 40 ms.

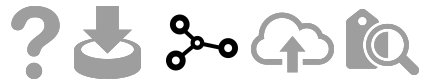


**Interests**

# What am I actually interested in?



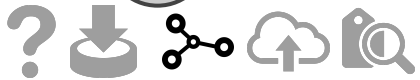
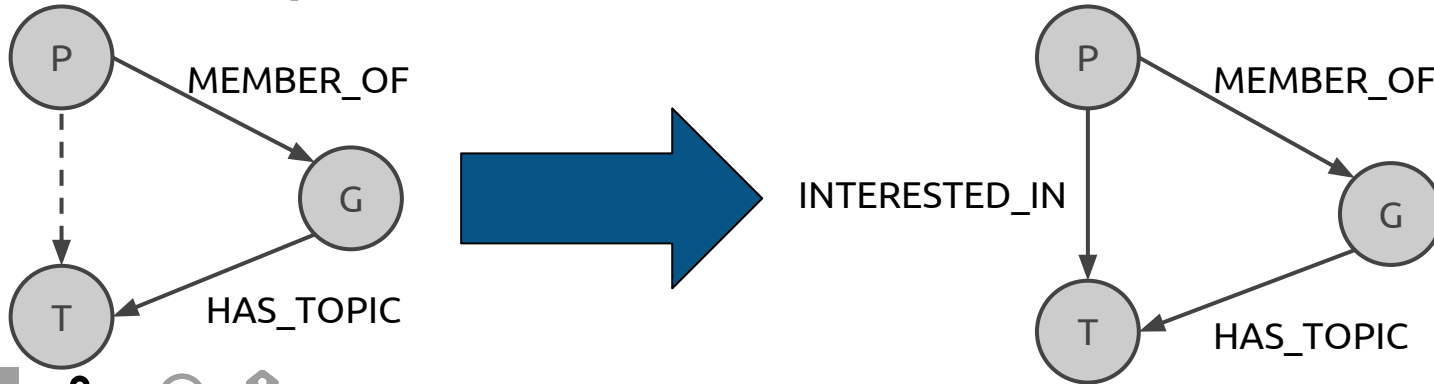
There's an **implicit** INTERESTED\_IN relationship between the topics of groups I belong to but don't express an interest in. Let's make it **explicit**



# What am I actually interested in?



There's an **implicit** INTERESTED\_IN relationship between the topics of groups I belong to but don't express an interest in. Let's make it **explicit**



# What am I actually interested in?



```
MATCH (m:Member)-[:RSVPD {response:"yes"}]->(event)
      <-[:HOSTED_EVENT]->()-[:HAS_TOPIC]->(topic)
```

```
WITH m, topic,
      COUNT(*) AS times
```

```
WHERE times > 5
```

```
RETURN m.name, topic.name, times
ORDER BY times DESC
```



# What am I actually interested in?



```
MATCH (m:Member)-[:RSVPD {response:"yes"}]->(event)
      <-[:HOSTED_EVENT]->()-[:HAS_TOPIC]->(topic)
```

```
WITH m, topic,
      COUNT(*) AS times,
      COLLECT(event.name) AS events
```

```
WHERE times > 5
```

```
AND NOT (m)-[:INTERESTED_IN]->(topic)
MERGE (m)-[:INTERESTED_IN]->(topic)
```



# What am I actually interested in?



```
$ MATCH (member:Member {name: "Mark Needham"}) -[:INTERESTED_IN]->(topic), (member)-[:MEMBER_OF]->(group)-[:HAS_TOPIC]->(topic) WITH member, topic, CO...
```

otherGroup.name	COLLECT(topic.name)	score
Business, Analytics and Data Science	[Machine Learning, New Technology, Big Data, Open Source, Data Science, Data Visualization, Computer programming, Predictive Analytics, Data Mining, Data Analytics, Big Data Analytics]	116
Big Data Developers in London	[Machine Learning, Big Data, Cloud Computing, Data Science, NoSQL, Data Visualization, Predictive Analytics, Data Mining, Data Analytics, Big Data Analytics, Software Development]	111
Python for Quant Finance	[Machine Learning, New Technology, Big Data, Cloud Computing, Open Source, Data Visualization, Computer programming, Predictive Analytics, Data Mining, Data Analytics]	110
DataKind UK	[Machine Learning, New Technology, Big Data, Cloud Computing, Open Source, Data Science, Data Visualization, Predictive Analytics, Data Analytics, Big Data Analytics]	109
Advanced Data Visualisation London	[Machine Learning, New Technology, Big Data, Open Source, Data Science, Data Visualization, Predictive Analytics, Data Mining, Data Analytics, Big Data Analytics]	109
Big Data Week London Meetup	[Machine Learning, Big Data, Open Source, Data Science, NoSQL, Data Visualization, Predictive Analytics, Data Mining, Data Analytics, Big Data Analytics]	108
London Data Science #ODSC	[Machine Learning, Big Data, Open Source, Data Science, Data Visualization, Computer programming, Predictive Analytics, Data Analytics, Big Data Analytics, Software Development]	106
Agile Data - London Meetup	[Machine Learning, New Technology, Big Data, Cloud Computing, Open Source, Data Science, NoSQL, Data Analytics, Big Data Analytics, Software Development]	106

Returned 364 rows in 106 ms.





**Finally, Events!**

# Now - let's recommend events!



MONDAY, NOVEMBER 30

2:30 PM

LondonR



Yes

## LondonR Meeting (and Workshop)

209 R Users going

9:00 AM

OpenSource & Agile Community Events

## CSM Course: Advanced Certified ScrumMaster Course by Martine Devos

5 OpenSource and Agile developers going

10:00 AM

OpenSource & Agile Community Events

## iOSCon 2015 - The Conference for iOS and Swift Developers

9 OpenSource and Agile developers going

7:00 PM

Cleanweb London

## COP21 Cleanweb and the Climate

62 Members going



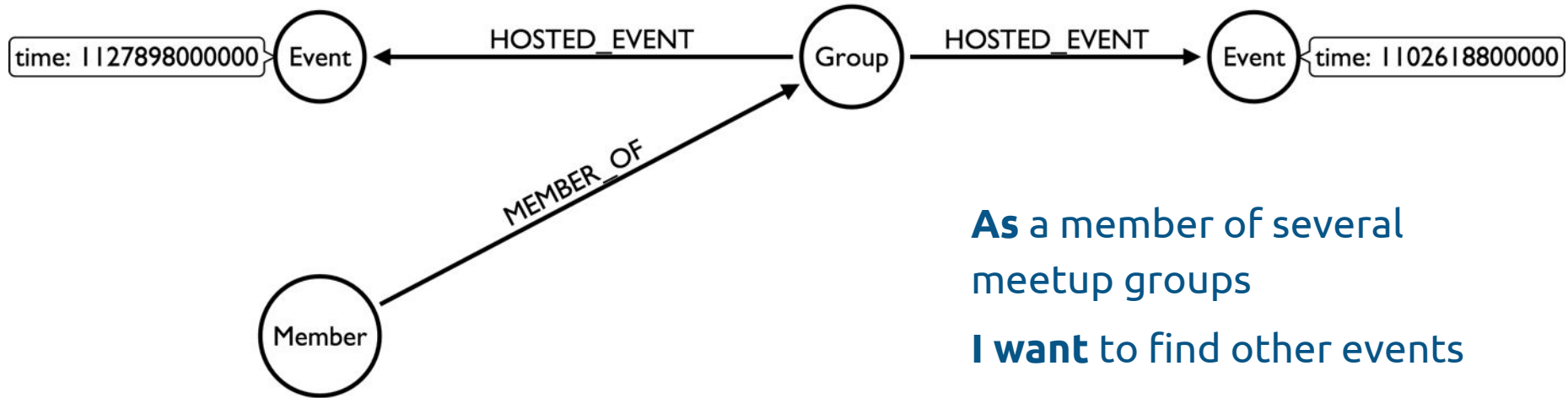
# Events in my groups



**As** a member of several meetup groups  
**I want** to find other events hosted by  
those groups  
**So that** I can attend those events



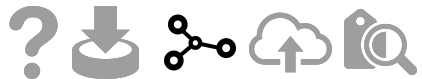
# Events in my groups



**As** a member of several  
meetup groups

**I want** to find other events  
hosted by those groups

**So that** I can attend those  
events



# Events



id	name	time	utc_offset
3261890	London Web Design October Meetup	1097776800000	3600000
3492560	London Web Design November Meetup	1100199600000	0
3683911	London Web Design December Meetup	1102618800000	0
4339054	The London Web Design March Meetup	1113413400000	3600000
4825171	The London PHP January Meetup	1136487600000	0
4795898	January Meetup	1137006000000	0
4826924	The London PHP February Meetup	1138906800000	0
4832622	The London Web Design February Meetup	1140030000000	0
8646860	JAWAWUG BOF 40 JQueryLib	1221672600000	3600000
8689280	PHP London October Meetup	1222972200000	3600000
8730923	The London Cloud Computing October Meetu	1223488800000	3600000
8879609	JWUG BOF41 Web Applications and RESTful	1224523800000	3600000
8921257	OSGi for the Web Developer followed by f	1225217700000	0



# Create events



```
CREATE INDEX ON :Event(id)
```

```
CREATE INDEX ON :Event(time)
```

```
LOAD CSV WITH HEADERS FROM "file:///events.csv" AS row  
MERGE (event:Event {id: row.id})  
ON CREATE SET event.name = row.name,  
              event.time = toint(row.time),  
              event.utcOffset = toint(row.utc_offset)
```



# Events and groups



id	group_id
3261890	163876
3492560	163876
3683911	163876
3857967	163876
4339054	163876
4572794	163876
4709866	163876
4772985	163876
4785678	163876
4825171	218194
4826924	218194
4832622	163876
4846072	218194



# Connect events and groups



```
LOAD CSV WITH HEADERS FROM "file:///events.csv" AS row
MATCH (group:Group {id: row.group_id})
MATCH (event:Event {id: row.id})
MERGE (group)-[:HOSTED_EVENT]->(event)
```





# Events in my groups



```
WITH 24.0*60*60*1000 AS oneDay
MATCH (member:Member {name: "Mark Needham"}),
      (member)-[:MEMBER_OF]->(group),
      (group)-[:HOSTED_EVENT]->(futureEvent)
WHERE futureEvent.time >= timestamp()
RETURN group.name, futureEvent.name,
       round((futureEvent.time - timestamp()) / oneDay) AS days
ORDER BY days
LIMIT 10
```



# Events in my groups



```
$ WITH 24.0*60*60*1000 AS oneDay MATCH (member:Member {name: "Mark Needham"})-[membership:MEMBER_OF]->(group), (group)-[:HOSTED_EVENT]->(futureEvent) ...
```



	group.name	futureEvent.name	days
Rows	OpenSource & Agile Community Events	iOSCon 2015 - The Conference for iOS and Swift Developers	2
</> Code	OpenSource & Agile Community Events	CSM Course: Advanced Certified ScrumMaster Course by Martine Devos	2
	Hands-on Big data science workshop	Starting Up With Graph, Real Time Analytics And Deep Learning	2
	LondonR	LondonR Meeting (and Workshop)	2
	Neo4j - London User Group	Modelling a recommendation engine: A worked example	3
	LJC - London Java Community	Getting Your System to Production and Keeping it There - Eoin Woods	3
	OpenSource & Agile Community Events	Damjan Vujnovic's Advanced JavaScript Workshop (3-days)	4
	GOTO Nights - London, UK	Eleanor McHugh - Xmas Cracker	4
	OpenSource & Agile Community Events	Martine Devos' Certified Scrum Product Owner Course	5
	OpenSource & Agile Community Events	Clojure eXchange 2015	5
	Returned 10 rows in 20 ms.		



# Events in my groups



```
$ WITH 24.0*60*60*1000 AS oneDay MATCH (member:Member {name: "Mark Needham"})-[membership:MEMBER_OF]->(group), (group)-[:HOSTED_BY]
```



Rows



Code

group.name	futureEvent.name
OpenSource & Agile Community Events	iOSCon 2015 - The Conference for iOS and Swift Developers
OpenSource & Agile Community Events	CSM Course: Advanced Certified ScrumMaster Course by Martine Devos
Hands-on Big data science workshop	Starting Up With Graph, Real Time Analytics And Deep Learning
LondonR	LondonR Meeting (and Workshop)
Neo4j - London User Group	Modelling a recommendation engine: A worked example
LJC - London Java Community	Getting Your System to Production and Keeping it There - Eoin Woods
OpenSource & Agile Community Events	Damjan Vujnovic's Advanced JavaScript Workshop (3-days)
GOTO Nights - London, UK	Eleanor McHugh - Xmas Cracker
OpenSource & Agile Community Events	Martine Devos' Certified Scrum Product Owner Course
OpenSource & Agile Community Events	Clojure eXchange 2015

Returned 10 rows in 20 ms.

MONDAY, NOVEMBER 30

2:30 PM  
**Yes** LondonR  
**LondonR Meeting (and Workshop)**  
209 R Users going

9:00 AM OpenSource & Agile Community Events  
**CSM Course: Advanced Certified ScrumMaster Course by Martine Devos**  
5 OpenSource and Agile developers going

10:00 AM OpenSource & Agile Community Events  
**iOSCon 2015 - The Conference for iOS and Swift Developers**  
9 OpenSource and Agile developers going

7:00 PM Cleanweb London  
**COP21 Cleanweb and the Climate**  
62 Members going

8:00 PM London Virtual Data Science Meetup  
**Free Online Event: Applying Machine Learning and AI for Business**  
46 Members going **4 spots left!**



# Events in my groups



```
$ WITH 24.0*60*60*1000 AS oneDay MATCH (member:Member {name: "Mark Needham"})-[membership:MEMBER_OF]->(group), (group)-[:HOS
```

group.name	futureEvent.name
OpenSource & Agile Community Events	iOSCon 2015 - The Conference for iOS and Swift Developers
OpenSource & Agile Community Events	CSM Course: Advanced Certified ScrumMaster Course by Martine Devos
Hands-on Big data science workshop	Starting Up With Graph, Real Time Analytics And Deep Learning
LondonR	LondonR Meeting (and Workshop)
Neo4j - London User Group	Modelling a recommendation engine: A worked example
LJC - London Java Community	Getting Your System to Production and Keeping it There - Eoin Woods
OpenSource & Agile Community Events	Damjan Vujnovic's Advanced JavaScript Workshop (3-days)
GOTO Nights - London, UK	Eleanor McHugh - Xmas Cracker
OpenSource & Agile Community Events	Martine Devos' Certified Scrum Product Owner Course
OpenSource & Agile Community Events	Clojure eXchange 2015

Returned 10 rows in 20 ms.

TUESDAY, DECEMBER 1	
6:15 PM Yes	LJC - London Java Community <b>Getting Your System to Production and Keeping it There - Eoin Woods</b> 78 Java Developers going
6:30 PM Yes	Neo4j - London User Group <b>Modelling a recommendation engine: A worked example</b> 119 People going
6:15 PM No	Deep Learning London Meetup <b>Deep Learning Winning Kaggle Competitions</b> 258 Neurons going <b>42 spots left!</b>
6:30 PM	London Continuous Delivery <b>December 2015: Festive Lightning Talks!</b> 100 members going



# Layered recommendations



We can improve our recommendation by weighting different attributes:

- events in my groups
- events I've previously attended
- topics I'm interested in
- events my peers attend



# Events in my groups



We can improve our recommendation by weighting different attributes:

- ▶ **events in my groups**
- ▶ events I've previously attended
- ▶ topics I'm interested in
- ▶ events my peers attend



# Events in my groups



```
WITH 24.0*60*60*1000 AS oneDay
```

```
MATCH (member:Member {name: "Mark Needham"})
```

```
MATCH (futureEvent:Event) WHERE futureEvent.time >= timestamp()
```

```
MATCH (futureEvent)<-[:HOSTED_EVENT]-(group)
```

```
RETURN group.name,
```

```
futureEvent.name,
```

```
EXISTS((group)<-[:MEMBER_OF]-(member)) AS isMember,
```

```
round((futureEvent.time - timestamp()) / oneDay) AS days
```

```
ORDER BY isMember DESC, days
```



# Events in my groups



```
$ WITH 24.0*60*60*1000 AS oneDay MATCH (member:Member {name: "Mark Needham"}) MATCH (futureEvent:Event) WHERE futureEvent.time >= timestamp() MATCH (f...
```



	group.name	futureEvent.name	isMember	days
Rows	OpenSource & Agile Community Events	CSM Course: Advanced Certified ScrumMaster Course by Martine Devos	true	1
</> Code	OpenSource & Agile Community Events	iOSCon 2015 - The Conference for iOS and Swift Developers	true	1
	Hands-on Big data science workshop	Starting Up With Graph, Real Time Analytics And Deep Learning	true	2
	LondonR	LondonR Meeting (and Workshop)	true	2
	OpenSource & Agile Community Events	Damjan Vujnovic's Advanced JavaScript Workshop (3-days)	true	3
	Neo4j - London User Group	Modelling a recommendation engine: A worked example	true	3
	LJC - London Java Community	Getting Your System to Production and Keeping it There - Eoin Woods	true	3
	OpenSource & Agile Community Events	Clojure eXchange 2015	true	4
	GOTO Nights - London, UK	Eleanor McHugh - Xmas Cracker	true	4
	OpenSource & Agile Community Events	Martine Devos' Certified Scrum Product Owner Course	true	4

Returned 10 rows in 70 ms.





# + previous events attended

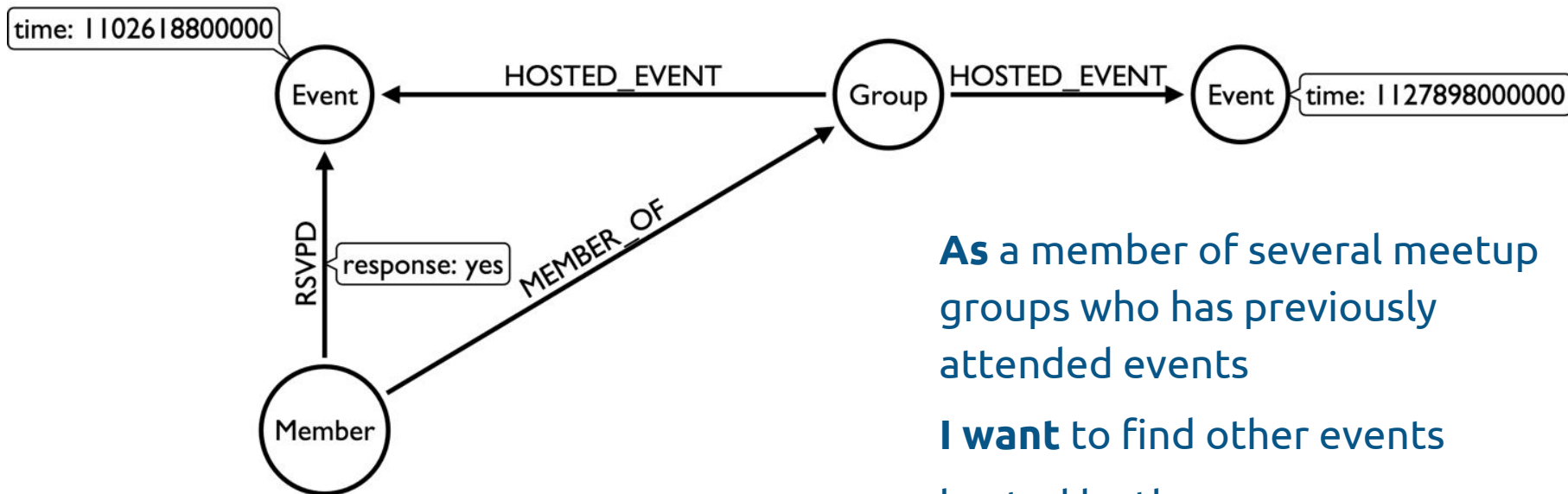


We can improve our recommendation by weighting different attributes:

- ▶ events in my groups
- ▶ **events I've previously attended**
- ▶ topics I'm interested in
- ▶ events my peers attend



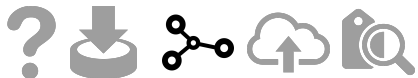
# + previous events attended



**As** a member of several meetup groups who has previously attended events

**I want** to find other events hosted by those groups

**So that** I can attend those events



# RSVPs



rsvp_id	event_id	member_id	guests	response	created	mtime
654924042	100056812	65110402	0	yes	1358436329000	1358436329000
666200862	100056812	32158012	0	yes	1359212092000	1359212092000
655045942	100056812	45574682	0	yes	1358442847000	1358442847000
654946622	100056812	64073592	0	yes	1358437486000	1358437486000
696456002	100056812	70201982	0	yes	1361279846000	1361279846000
689115982	100056812	12434405	0	yes	1360748670000	1360748670000
654924112	100056812	34168592	0	no	1358436332000	1358436332000
654925662	100056812	3401490	0	no	1358436413000	1360361799000
656439652	100056812	12252389	0	no	1358533048000	1361197297000
689112692	100056812	76908802	0	yes	1360748069000	1360748069000
690924922	100056812	10704191	0	yes	1360876122000	1360876122000
690834812	100056812	71296302	0	yes	1360871204000	1360871204000
691120252	100056812	71730512	0	yes	1360888294000	1360888294000



# Create RSVPs



```
LOAD CSV WITH HEADERS FROM "file:///rsvps.csv" AS row
MATCH (member:Member {id: row.member_id})
MATCH (event:Event {id: row.event_id})
MERGE (member)-[rsvp:RSVPD {id: row.rsvp_id}]->(event)
ON CREATE SET rsvp.created = toint(row.created),
              rsvp.lastModified = toint(row.mtime),
              rsvp.response = row.response;
```



# + previous events attended



```
WITH 24.0*60*60*1000 AS oneDay
MATCH (member:Member {name: "Mark Needham"})
MATCH (futureEvent:Event) WHERE futureEvent.time >= timestamp()
MATCH (futureEvent)<-[:HOSTED_EVENT]-(group)
```

```
WITH oneDay, group, futureEvent, member, EXISTS((group)<-[:MEMBER_OF]-(member)) AS isMember
OPTIONAL MATCH (member)-[rsvp:RSVPD {response: "yes"}]->(pastEvent)<-[:HOSTED_EVENT]-(group)
WHERE pastEvent.time < timestamp()
```

```
RETURN group.name,
       futureEvent.name,
       isMember,
       COUNT(rsvp) AS previousEvents,
       round((futureEvent.time - timestamp()) / oneDay) AS days

ORDER BY days, previousEvents DESC
```



# + previous events attended



```
S WITH 24.0*60*60*1000 AS oneDay MATCH (member:Member {name: "Mark Needham"}) MATCH (futureEvent:Event) WHERE futureEvent.time >= timestamp() MATCH (f...
```

	group.name	futureEvent.name	isMember	days	previousEvents
Rows	Neo4j - London User Group	Modelling a recommendation engine: A worked example	true	1	94
</> Code	LJC - London Java Community	Getting Your System to Production and Keeping it There - Eoin Woods	true	1	1
	Artists and Scientists Get-Together (London)	Ebola diagnostics in Sierra Leone	false	1	0
	UK Azure User Group	Scaling for the future with Cassandra and Azure	false	1	0
	NW London Hackers & Makers	Open Night @ Create Space	false	1	0
	FREE Marketing, Analytics & Digital Skills in London	Get OD on Digital, join Outreach Digital! - Monthly Meeting For New Volunteers	false	1	0
	OpenSource & Agile Community Events	Damjan Vujnovic's Advanced JavaScript Workshop (3-days)	true	1	0
	Artists and Scientists Get-Together (London)	Animal Showoff - Grant Museum After Hours	false	1	0
	PyData London Meetup	Xmas PyData!	false	1	0
	London Groovy & Grails User Group	GGUG Meeting	false	1	0
	This Happened London	This Happened London #23	false	1	0
	Couchbase London	December 2015 meet-up	false	1	0
	Pivotal London - Cloud Native Apps Meetup	Spring Cloud - What's the Scoop?	false	1	0
	OpenSource & Agile Community Events	Clojure eXchange 2015	true	2	0
	GOTO Nights - London, UK	Eleanor McHugh - Xmas Cracker	true	2	0
	London µServices (Microservices) User Group	Monitoring Microservices	false	2	0

Returned 359 rows in 187 ms.



# RSVP\_YES vs RSVPD



I was curious whether refactoring  
RSVPD {response: "yes"} to RSVP\_YES would have  
any impact as Neo4j is optimised for **querying  
by unique relationship types.**



# RSVP\_YES vs RSVPD



```
MATCH (m:Member)-[rsvp:RSVPD {response:"yes"}]->(event)
MERGE (m)-[rsvpYes:RSVP_YES {id: rsvp.id}]->(event)
ON CREATE SET rsvpYes.created = rsvp.created,
              rsvpYes.lastModified = rsvp.lastModified;
```

```
MATCH (m:Member)-[rsvp:RSVPD {response:"no"}]->(event)
MERGE (m)-[rsvpYes:RSVP_NO {id: rsvp.id}]->(event)
ON CREATE SET rsvpYes.created = rsvp.created,
              rsvpYes.lastModified = rsvp.lastModified;
```





# RSVP\_YES vs RSVPD



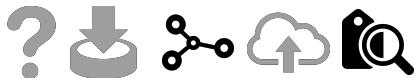
RSVPD {response: "yes"}

Cypher version: CYPHER 2.3,  
planner: COST.  
688635 total db hits in 232 ms.

VS

RSVP\_YES

Cypher version: CYPHER 2.3,  
planner: COST.  
559866 total db hits in 207 ms.



# + my topics



We can improve our recommendation by weighting different attributes:

- events in my groups
- events I've previously attended
- **topics I'm interested in**
- events my peers attend



# + my topics



```
WITH 24.0*60*60*1000 AS oneDay
MATCH (member:Member {name: "Mark Needham"})
MATCH (futureEvent:Event) WHERE futureEvent.time >= timestamp()
MATCH (futureEvent)<-[:HOSTED_EVENT]-(group)

WITH oneDay, group, futureEvent, member, EXISTS((group)<-[:MEMBER_OF]-(member)) AS isMember
OPTIONAL MATCH (member)-[rsvp:RSVPD {response: "yes"}]->(pastEvent)<-[:HOSTED_EVENT]-(group)
WHERE pastEvent.time < timestamp()

WITH oneDay, group, futureEvent, member, isMember, COUNT(rsvp) AS previousEvents
OPTIONAL MATCH (futureEvent)<-[:HOSTED_EVENT]-(group)-[:HAS_TOPIC]->(topic)<-[:INTERESTED_IN]-(member)

RETURN group.name, futureEvent.name, isMember, previousEvents,
       COUNT(topic) AS topics, round((futureEvent.time - timestamp()) / oneDay) AS days

ORDER BY days,previousEvents DESC, topics DESC
```



# + my topics



```
$ WITH 24.0*60*60*1000 AS oneDay MATCH (member:Member {name: "Mark Needham"}) MATCH (futureEvent:Event) WHERE futureEvent.time >= timestamp() MATCH (f...
```

	group.name	futureEvent.name	isMember	days	previousEvents	topics
Rows	Neo4j - London User Group	Modelling a recommendation engine: A worked example	true	1	94	7
Code	LJC - London Java Community	Getting Your System to Production and Keeping it There - Eoin Woods	true	1	1	2
	PyData London Meetup	Xmas PyData!	false	1	0	6
	Couchbase London	December 2015 meet-up	false	1	0	4
	NW London Hackers & Makers	Open Night @ Create Space	false	1	0	2
	This Happened London	This Happened London #23	false	1	0	1
	OpenSource & Agile Community Events	Damjan Vujnovic's Advanced JavaScript Workshop (3-days)	true	1	0	1
	FREE Marketing, Analytics & Digital Skills in London	Get OD on Digital, join Outreach Digital! - Monthly Meeting For New Volunteers	false	1	0	1
	Artists and Scientists Get-Together (London)	Animal Showoff - Grant Museum After Hours	false	1	0	1
	Artists and Scientists Get-Together (London)	Ebola diagnostics in Sierra Leone	false	1	0	1
	Pivotal London - Cloud Native Apps Meetup	Spring Cloud - What's the Scoop?	false	1	0	1
	London Groovy & Grails User Group	GGUG Meeting	false	1	0	1
	UK Azure User Group	Scaling for the future with Cassandra and Azure	false	1	0	0
	Web Analytics Wednesday - London	WAW Christmas Party and Analytics networking	false	2	0	3
	GOTO Nights - London, UK	Eleanor McHugh - Xmas Cracker	true	2	0	2

Returned 15 rows in 283 ms.



# + events my friends are attending

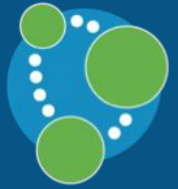


We can improve our recommendation by weighting different attributes:

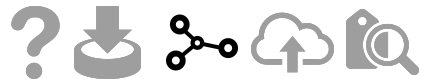
- events in my groups
- events I've previously attended
- topics I'm interested in
- **events my peers attend**



# + events my friends are attending



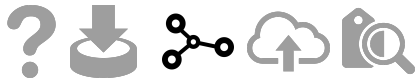
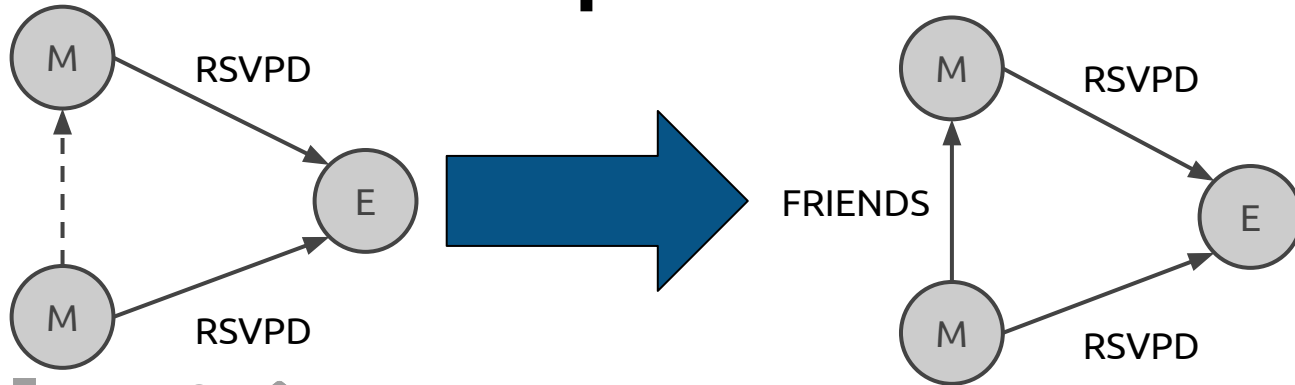
There's an **implicit** FRIENDS relationship between people who attended the same events. Let's make it **explicit**.



# + events my friends are attending



There's an **implicit** FRIENDS relationship between people who attended the same events. Let's make it **explicit**.



# + events my friends are attending



```
MATCH (m1:Member)
WHERE NOT m1:Processed

WITH m1 LIMIT {limit}
MATCH (m1)-[:RSVP_YES]->(event:Event)<-[:RSVP_YES]-(m2:Member)

WITH m1, m2, COLLECT(event) AS events, COUNT(*) AS times
WHERE times >= 5

WITH m1, m2, times, [event IN events | SIZE((event)<-[:RSVP_YES]-())] AS attendances

WITH m1, m2, REDUCE(score = 0.0, a IN attendances | score + (1.0 / a)) AS score

RETURN ID(m1) AS m1, ID(m2) AS m2, score
```





# + events my friends are attending



```
UNWIND {rows} AS row
```

```
MATCH (m1), (m2)
```

```
WHERE ID(m1) = row.m1 AND ID(m2) = row.m2
```

```
MERGE (m1)-[friendsRel:FRIENDS]-(m2)
```

```
SET friendsRel.score = row.score
```

```
SET m1:Processed
```

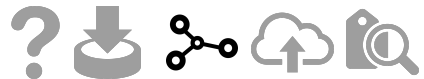
```
rows
[
  ...
  {
    "m1": 12345,
    "m2": 678912,
    "score": 0.23471
  },
  ...
]
```



# Bidirectional relationships



- ▶ You may have noticed that we didn't specify a direction when creating the relationship  
`MERGE (m1) - [:FRIENDS] - (m2)`
- ▶ FRIENDS is a bidirectional relationship. We only need to create it once between two people.
- ▶ We ignore the direction when querying



# + events my friends are attending



```
WITH 24.0*60*60*1000 AS oneDay
MATCH (member:Member {name: "Mark Needham"})
MATCH (futureEvent:Event) WHERE futureEvent.time >= timestamp()
MATCH (futureEvent)<-[:HOSTED_EVENT]-(group)

WITH oneDay, group, futureEvent, member, EXISTS((group)<-[:MEMBER_OF]-(member)) AS isMember
OPTIONAL MATCH (member)-[rsvp:RSVPD {response: "yes"}]->(pastEvent)<-[:HOSTED_EVENT]-(group)
WHERE pastEvent.time < timestamp()

WITH oneDay, group, futureEvent, member, isMember, COUNT(rsvp) AS previousEvents
OPTIONAL MATCH (futureEvent)<-[:HOSTED_EVENT]-(topic)-[:HAS_TOPIC]->(topic)<-[:INTERESTED_IN]-(member)

WITH oneDay, group, futureEvent, member, isMember, previousEvents, COUNT(topic) AS topics
OPTIONAL MATCH (member)-[:FRIENDS]-(Member)-[rsvpYes:RSVP_YES]->(futureEvent)

RETURN group.name, futureEvent.name, isMember, round((futureEvent.time - timestamp()) / oneDay) AS days,
       previousEvents, topics, COUNT(rsvpYes) AS friendsGoing
ORDER BY days, friendsGoing DESC, previousEvents DESC
LIMIT 15
```



# + events my friends are attending



```
$ WITH 24.0*60*60*1000 AS oneDay MATCH (member:Member {name: "Mark Needham"}) MATCH (futureEvent:Event) WHERE futureEvent.time >= timestamp() MATCH (f...
```

	group.name	futureEvent.name	isMember	days	previousEvents	topics	friendsGoing
Rows	Neo4j - London User Group	Modelling a recommendation engine: A worked example	true	1	94	7	5
Code	LJC - London Java Community	Getting Your System to Production and Keeping it There - Eoin Woods	true	1	1	2	2
	PyData London Meetup	Xmas PyData!	false	1	0	6	2
	Pivotal London - Cloud Native Apps Meetup	Spring Cloud - What's the Scoop?	false	1	0	1	1
	Artists and Scientists Get-Together (London)	Ebola diagnostics in Sierra Leone	false	1	0	1	0
	FREE Marketing, Analytics & Digital Skills in London	Get OD on Digital, join Outreach Digital! - Monthly Meeting For New Volunteers	false	1	0	1	0
	This Happened London	This Happened London #23	false	1	0	1	0
	UK Azure User Group	Scaling for the future with Cassandra and Azure	false	1	0	0	0
	Artists and Scientists Get-Together (London)	Animal Showoff - Grant Museum After Hours	false	1	0	1	0
	Couchbase London	December 2015 meet-up	false	1	0	4	0
	NW London Hackers & Makers	Open Night @ Create Space	false	1	0	2	0
	London Groovy & Grails User Group	GGUG Meeting	false	1	0	1	0
	OpenSource & Agile Community Events	Damjan Vujnovic's Advanced JavaScript Workshop (3-days)	true	1	0	1	0
	OpenSource & Agile Community Events	Clojure eXchange 2015	true	2	0	1	1
	Firefox OS London	Firefox OS Add-ons: Hands-on Session	false	2	0	2	0

Returned 15 rows in 281 ms.



# Scoring



We're using a simple count based scoring ordering.

In a production system we might apply **something more sophisticated** e.g. log or Pareto function

# Real time recommendations



## Chunked HTTP RSVP Stream

```
GET /2/rsvps
```

```
json
```

```
Host: stream.meetup.com
```

```
public
```

```
stream version 2
```

Live HTTP stream of RSVPs within public Meetup groups. This method uses [chunked transfer encoding](#) to maintain a persistent connection with the client. This connection will only be terminated for server maintenance or a connection error.

### Request Parameters

This method does not require authentication, or any parameters. Applications should only need a single connection to the stream, and at most 10 connections are allowed per client IP address.

<code>api_version</code>	2
<code>since_count</code>	Request that some number of recent messages be sent immediately, if available. May not be specified in the same request as <code>since_mtime</code> .
<code>since_mtime</code>	Return recent RSVP with an mtime greater than the supplied time, in milliseconds since the epoch



# Real time recommendations



```
{
  "venue": {
    "venue_id": 14544952
  },
  "response": "no",
  "guests": 0,
  "member": { "member_id": 54585732 },
  "rsvp_id": 1579878700,
  "mtime": 1448705224460,
  "event": {
    "event_id": "226676071",
  },
  "group": {
    "group_id": 8501832,
  }
}
```



# Real time recommendations



```
import requests
import json

def stream_meetup():
    r = requests.get('http://stream.meetup.com/2/rsvps', stream=True)
    for raw_rsvp in r.iter_lines():
        if raw_rsvp:
            yield raw_rsvp
```





# Real time recommendations



```
from py2neo import authenticate, Graph

authenticate("localhost:7474", "neo4j", "test")
graph = Graph()

group_ids = []
group_query = "MATCH (g:Group) RETURN g.id AS groupId"
for row in graph.cypher.execute(group_query):
    group_ids.append(int(row["groupId"]))
```



# Real time recommendations



```
for rsvp in stream_meetup():
    if rsvp["group"]["group_id"] in group_ids:
        params = { "rsvp_id": str(rsvp["rsvp_id"]),
                   "event_id": str(rsvp["event"]["event_id"]),
                   "member_id": str(rsvp["member"]["member_id"]),
                   "response": rsvp["response"],
                   "mtime": rsvp["mtime"] }
        graph.cypher.execute("""
            MATCH (event:Event {id: {event_id}})
            MATCH (member:Member {id: {member_id}})
            MERGE (member)-[rsvpRel:RSVPD {id: {rsvp_id}}]->(event)
            ON CREATE SET rsvpRel.created = toint({mtime})
            ON MATCH SET rsvpRel.lastModified = toint({mtime})
            SET rsvpRel.response = {response}""", params)
```



# What could we do next?



- ▶ Comments sentiment analysis
  - do people actually like the events they go to?
- ▶ Topic ontology
  - how are topics related? e.g. Neo4j, Cassandra, MongoDB are part of NoSQL
- ▶ Event similarity based on descriptions
  - use automated topic derivation to derive categories

# What could we do next?



- ▶ Social network
  - what events do our twitter/Facebook friends attend?
- ▶ Location
  - do we favour events in a certain part of town?
- ▶ Day of the week
  - do we only go to events on certain days of the week?
  - do we go to different events on weekdays vs weekend?

# Why Neo4j for recommendations?



## ▶ Real time

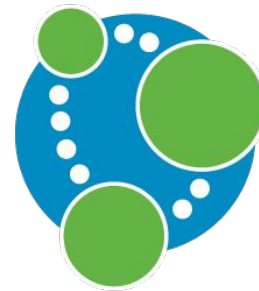
- take into account what you've just done

## ▶ Flexibility

- bring information from different sources and evolve data model as needed for use-cases
- easily combine collaborative + content filtering in a single query

## ▶ Intuitive query language

- focus on describing the domain problem. Even non technical users can read our queries.



**That's all for today!**  
**Questions? :-)**

Michael Hunger @mesirii  
created by

Mark Needham @markhneedham

<https://github.com/neo4j-meetups/modeling-worked-example>

# Graph Connect Europe 2016



- ▶ 26th April 2016
- ▶ **HERE** QEll Centre, Westminster, London
- ▶ <http://www.graphconnect.com>
- ▶ Use **QCON50** to get 50% off

# Data Dump

