

# From The Panama Papers to Russian Twitter Trolls

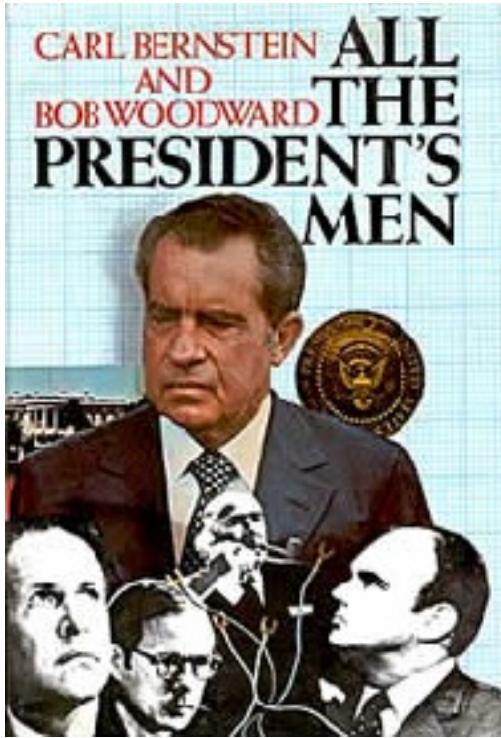


How Graph Databases are Revealing Hidden Relationships and Exposing Corruption

Mark Quinsland  
[neo4j.com](http://neo4j.com)

Big Data Day - LA  
U.S.C. Aug 2018

Alternate Title:



# ‘Follow the Money’

You just never know where it will lead.

## About Me



Mark Quinsland  
Field Engineer,  
Neo4j

[mark.quinsland@neo4j.com](mailto:mark.quinsland@neo4j.com)  
LinkedIn: mquinsland

# Neo4j - The World's Leading Graph Database

## Industry's Largest Dedicated Investment in Graphs



- Creator of the Neo4j Graph Platform
- ~200 employees
- HQ in Silicon Valley, other offices include London, Munich, Paris and Malmö (Sweden)
- \$80M in funding from Fidelity, Sunstone, Conor, Creandum, and Greenbridge Capital
- Over 10M+ downloads
- 270+ enterprise subscription customers with over half with >\$1B in revenue

### Adoption

- 7/10 Top Retail Firms
- 12/25 Top Financial Firms
- 8/10 Top Software Vendors

### Ecosystem

- 720+ Startups in program
- 270+ Enterprise customers
- 100+ Partners
- 53K+ Meet up members
- 450+ Events per year

### Customers



COMCAST



Lufthansa



### Partners



structr

# The Impact of Connected Data

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

---

**Gartner**

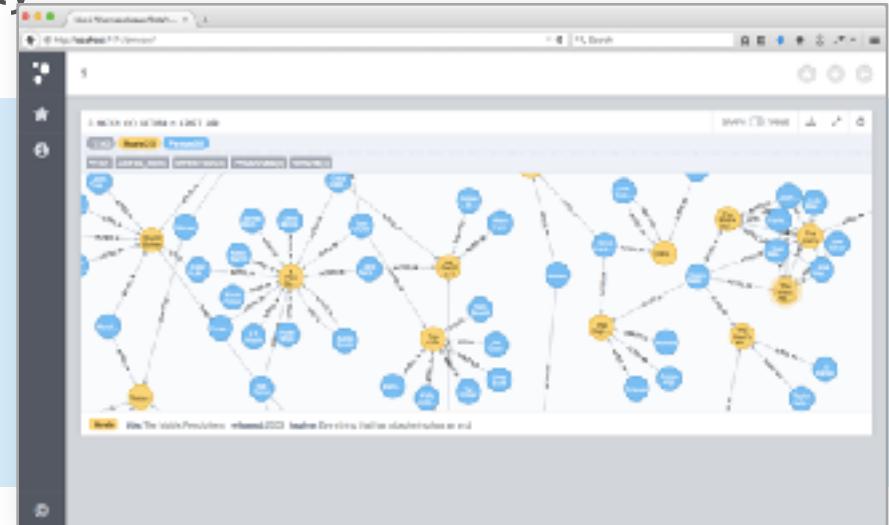
# Neo4j - The #1 Platform for Connected Data

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

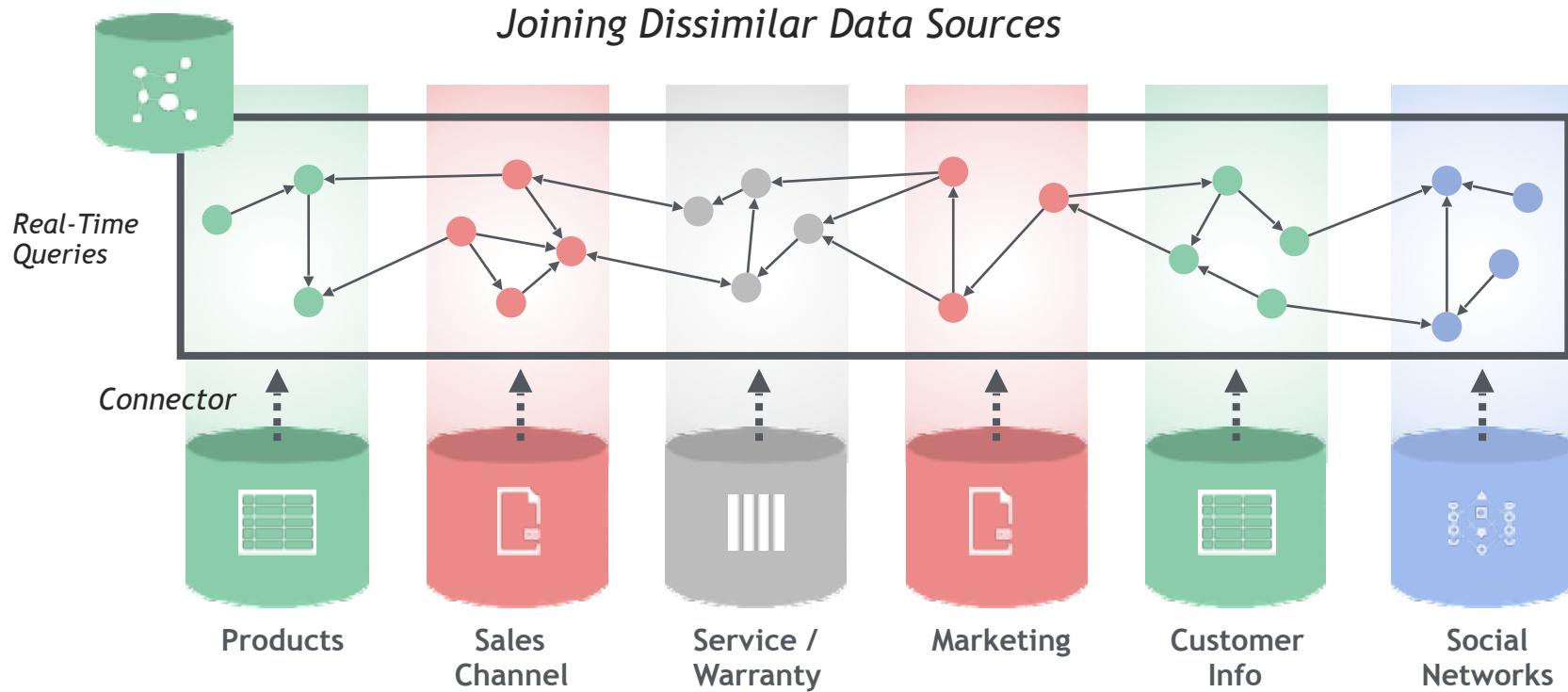
- **Store and query** data relationships
- **Traverse** any levels of depth in real-time
- **Add and connect** new data on the fly

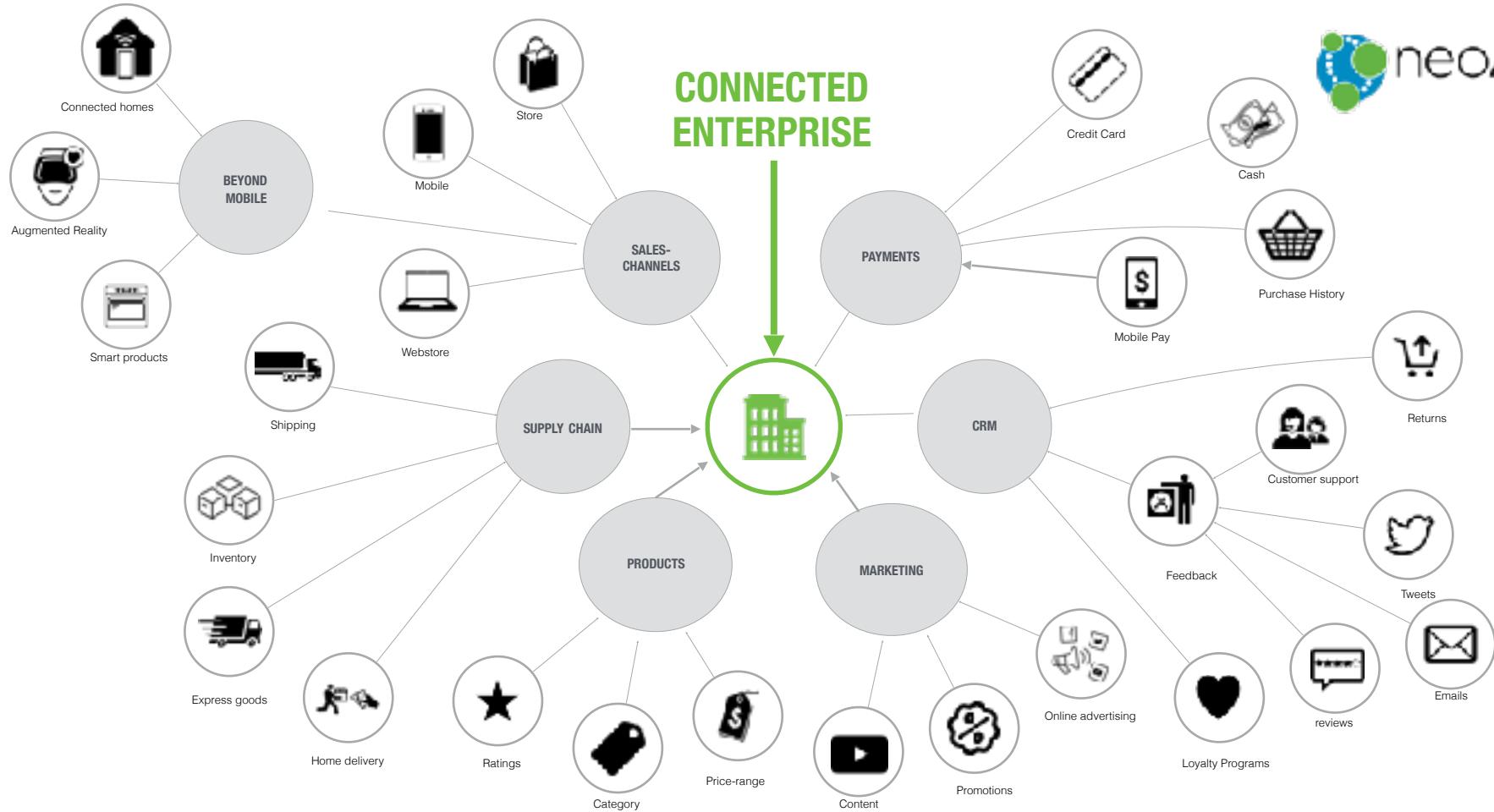
**Designed, built and tested *natively* for graphs from the start to ensure:**

- Performance
- ACID Transactions
- Agility
- Developer Productivity
- Hardware Efficiency



# Neo4j Simplifies Cross-Domain Analytics



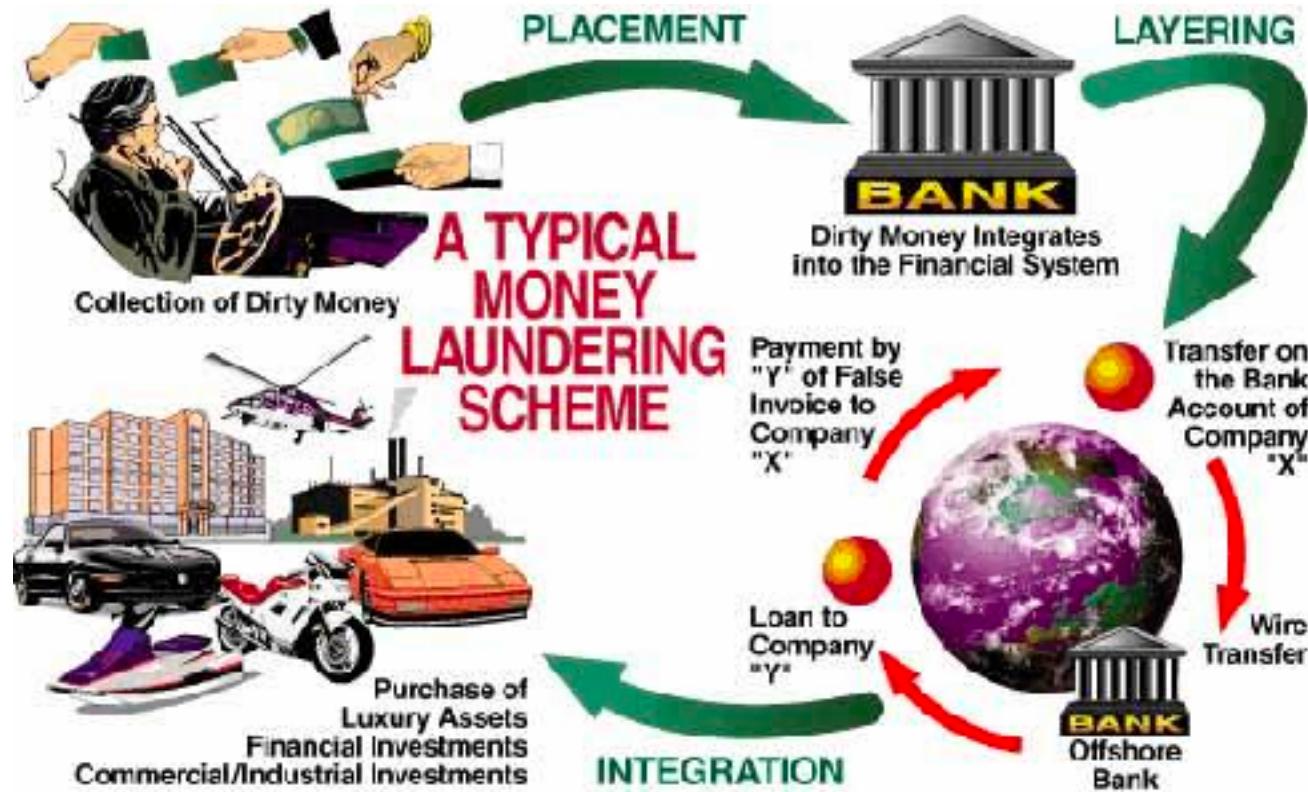


# Money Laundering 101:

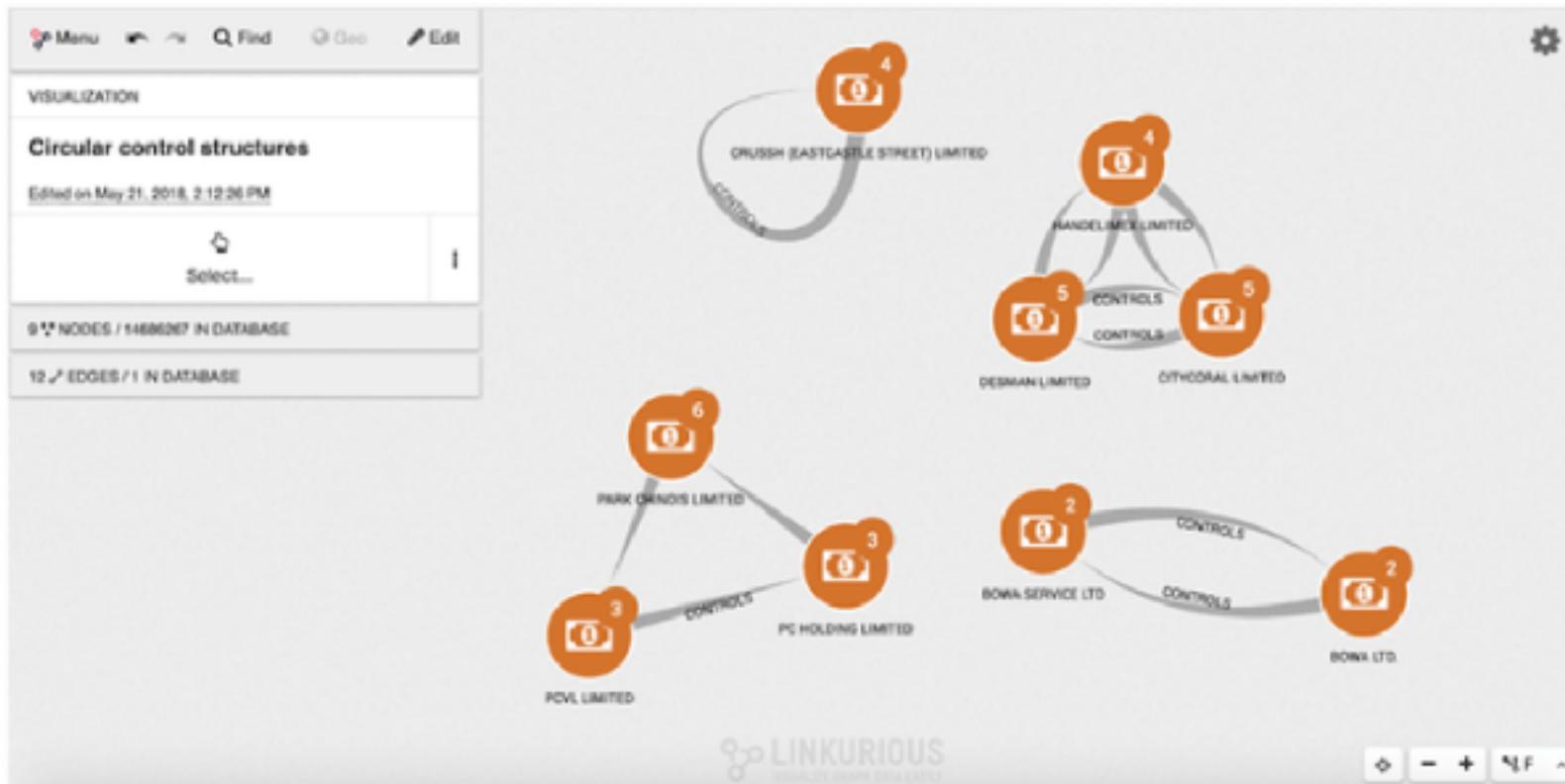


<http://www.financial-forensics.net/~financ2q/images/HiRes.jpg>

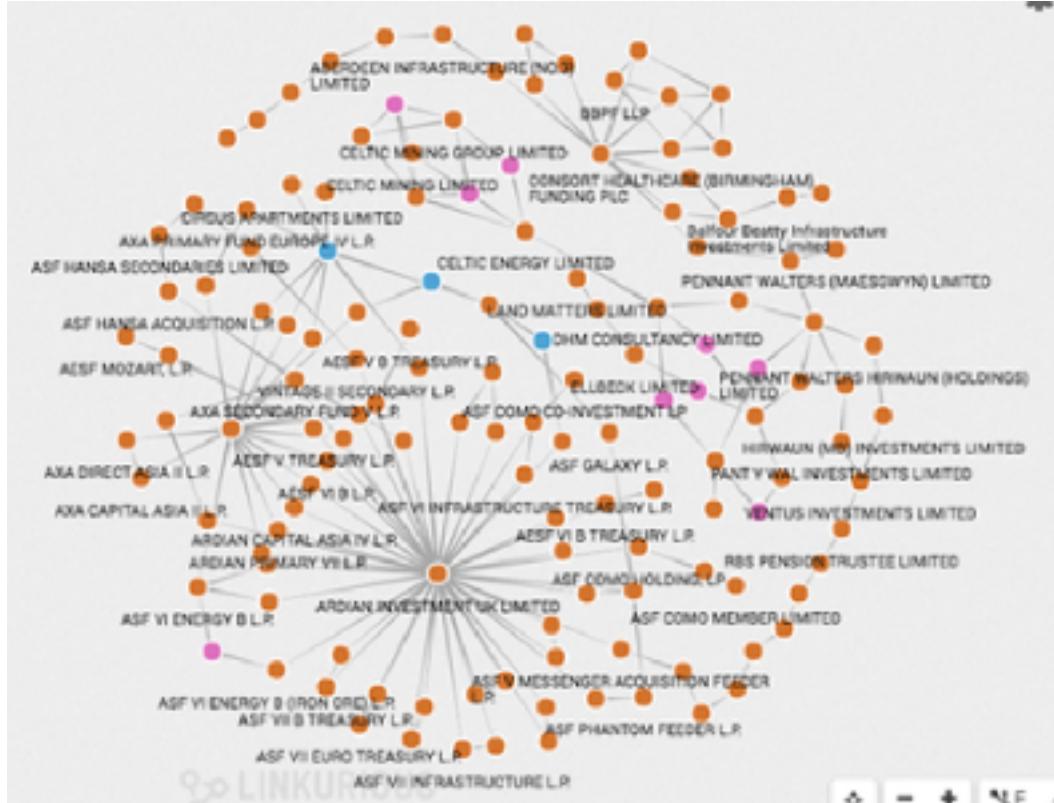
# Money Laundering 101: Step 1- Placement



# Shell Company Layering Can be Simple...



# ... Or Complex



<https://www.globalwitness.org/en-gb/campaigns/corruption-and-money-laundering/anonymous-company-owners/companies-we-keep/#chapter-1/section-1>



# The Panama Papers



- 40+ years of data (2.6 TB) leaked from Panamanian law firm
- 11.5 Million Emails, PDFs, passports, etc
- Covering 200k Shell Companies from 200+ countries
- Heads of State, Banks, Politicians, Criminals, Royalty, etc
- Investigated by 200+ investigative journalists from 70+ countries
- Technical assistance provided by Neo4j

# Winner of 2017 Pulitzer Prize



## Panama Papers investigation wins Pulitzer prize

Guardian and other media organisations collaborated on series of stories exposing offshore secrets of rich and famous



<https://www.theguardian.com/world/2017/apr/11/panama-papers-investigation-wins-pulitzer-prize>  
<https://neo4j.com/blog/icij-neo4j-unravel-panama-papers/>

# Heads of State Were Exposed



Country leaders Politicians/public officials

Filters All regions Country

The image displays a grid of 14 portraits of world leaders and public figures, each with their name and title below it. The portraits are arranged in three rows: the top row has six portraits, the middle row has six portraits, and the bottom row has two portraits. The names and titles are as follows:

- President of Argentina
- Former prime minister of Georgia
- Iceland's prime minister
- Ex-prime minister of Iraq
- Former prime minister of Jordan
- Former prime minister of Qatar
- Former Emir of Qatar
- King of Saudi Arabia
- Former president of Sudan
- UAE President, Abu Dhabi emir
- Convicted former Ukraine prime minister
- President of Ukraine
- Former prime minister of Mongolia
- Australia's prime minister

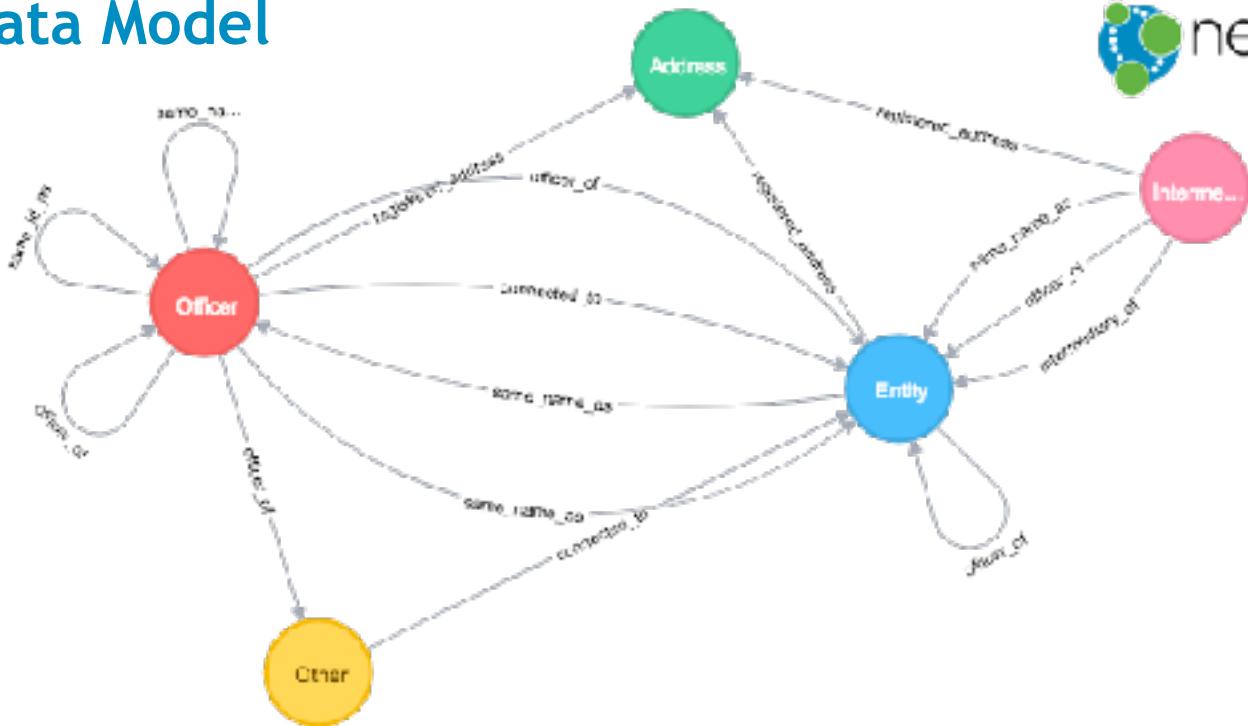
# And Their Relatives & Friends

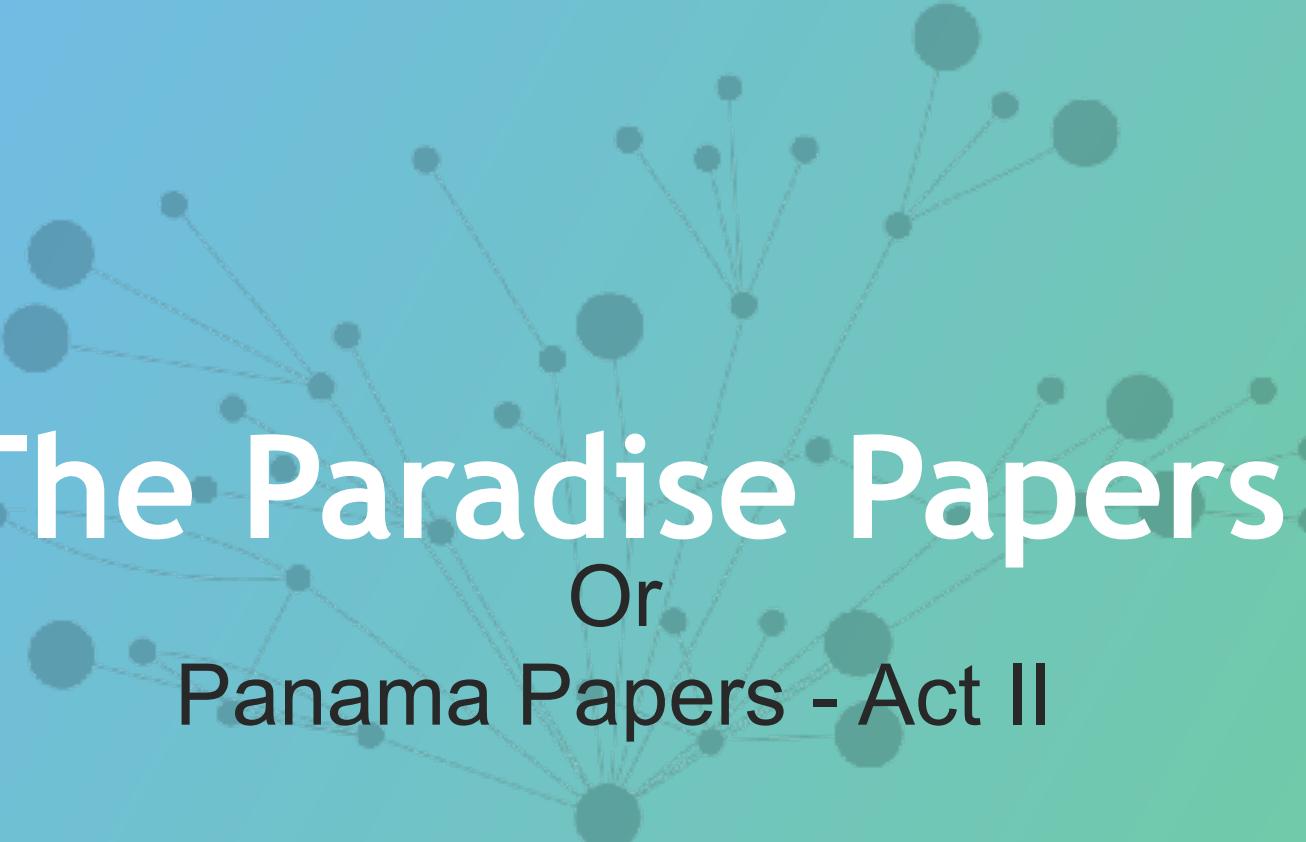


Relatives/associates of country leaders



# Simple Graph Data Model





# The Paradise Papers

Or

## Panama Papers - Act II

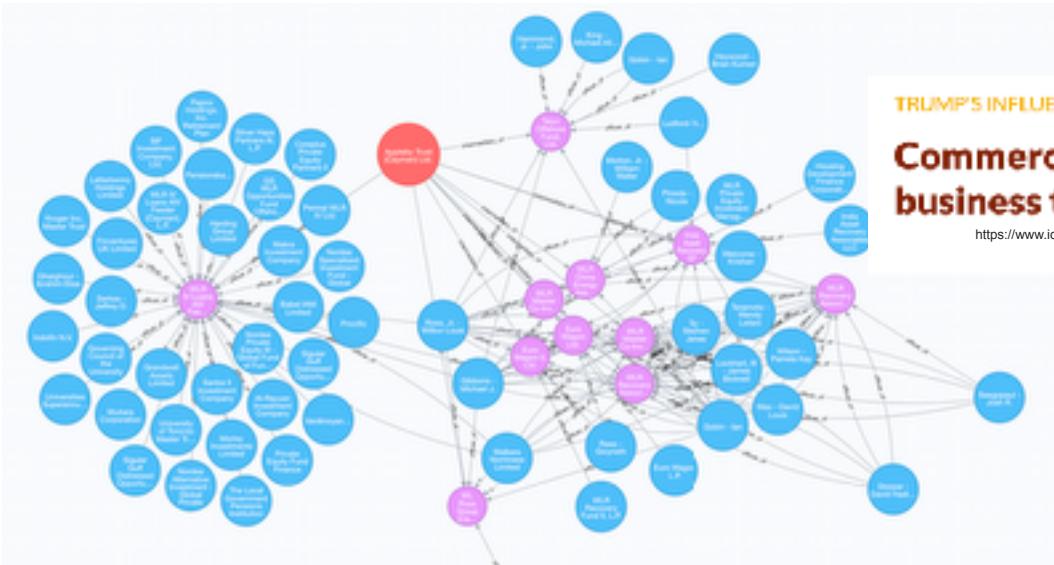
# Paradise Papers (2017)



A screenshot of the ICIJ Offshore Leaks Database website. At the top, it says "OFFSHORE LEAKS DATABASE". Below that is a section titled "ICIJ'S INVESTIGATIONS" with a sub-section for "PARADISE PAPERS" featuring a collage of documents and a person sitting at a desk. Another section below it is "THE PANAMA PAPERS" with a similar collage. At the bottom, there's a section titled "SECRETSALE" with a collage of documents and a person's face. The footer says "DATABASE POWERED BY" with logos for "LINKURIOUS" and "neo4j".

- 1.4TB leaked from Bahamian law firm
- 13.4 million records
- 290k Shell companies
- 380 journalists in 67 countries
- Technical assistance provided by Neo4j

# Highly-Connected US Cabinet Officials



## TRUMP'S INFLUENCERS

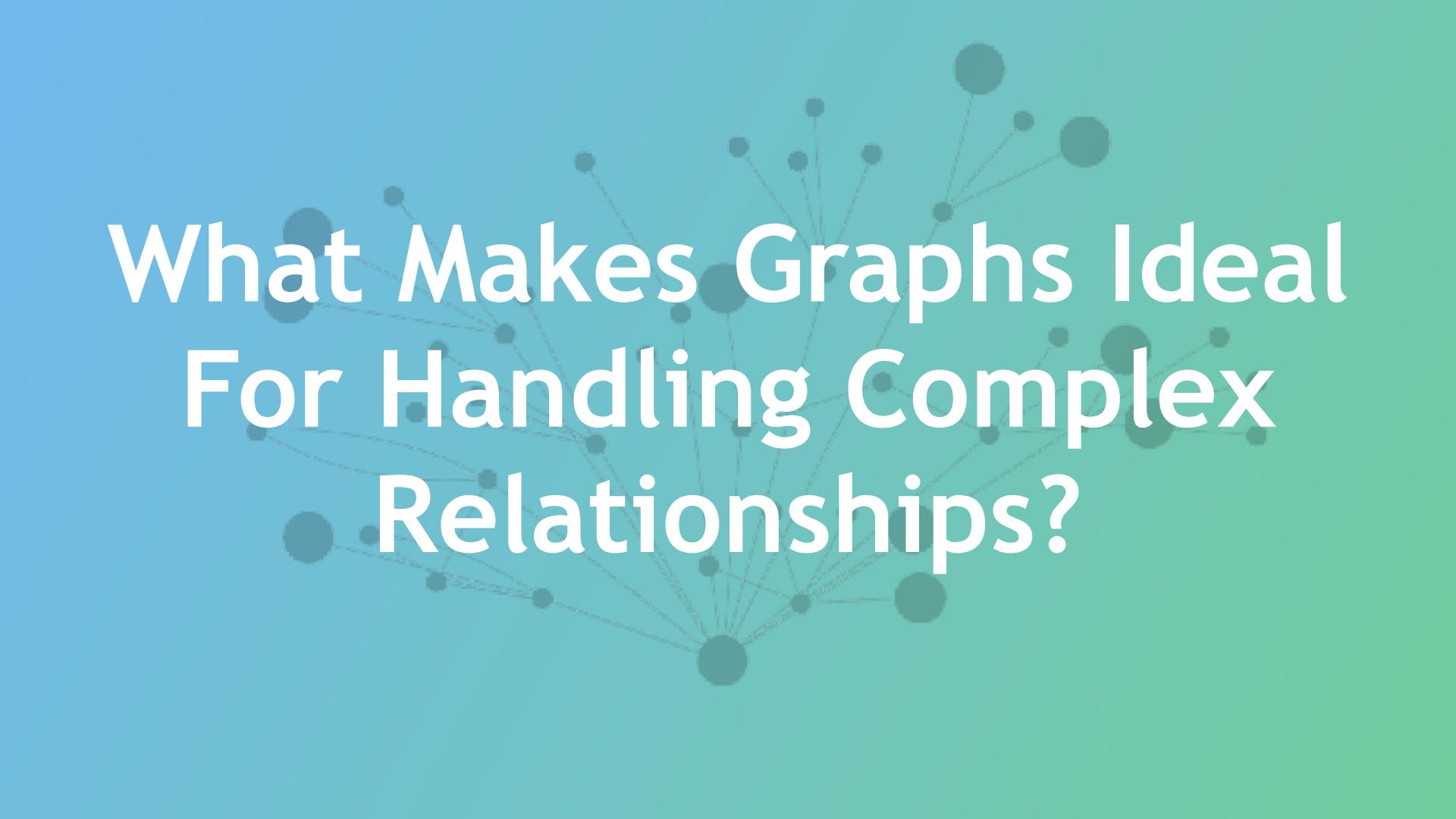
**Commerce Secretary Wilbur Ross benefits from business ties to Putin's inner circle**

<https://www.icij.org/investigations/paradise-papers/donald-trumps-commerce-secretary-wilbur-ross-and-his-russian-business-ties/>

// Wilbur Ross's connections in the Paradise Papers

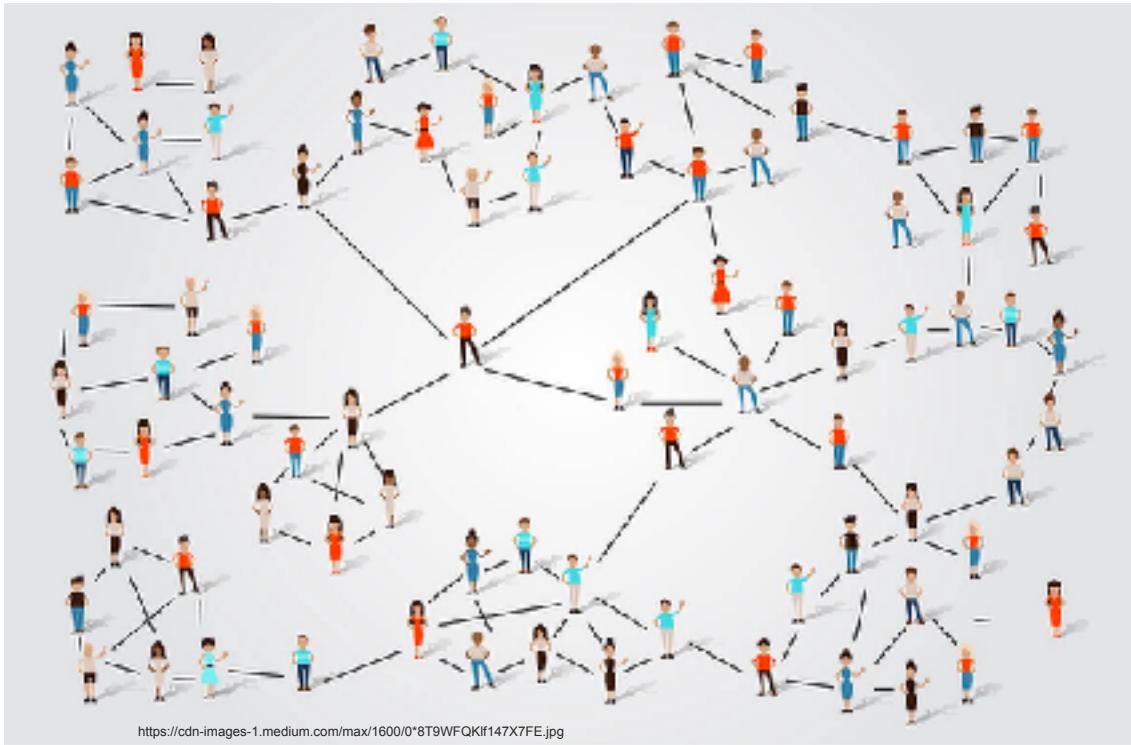
```
MATCH (o:Officer)-->(e:Entity)-[:intermediary_of]-(i:Intermediary)
WHERE o.name CONTAINS "Ross"
MATCH (e)--(o2:Officer)
RETURN *
```

<https://neo4j.com/blog/analyzing-paradise-papers-neo4j/>

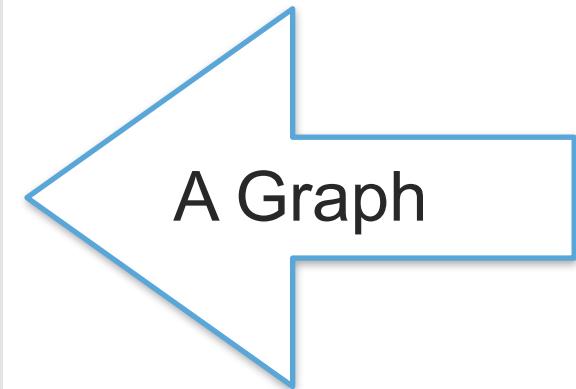


# What Makes Graphs Ideal For Handling Complex Relationships?

# Six Degrees of Separation Game

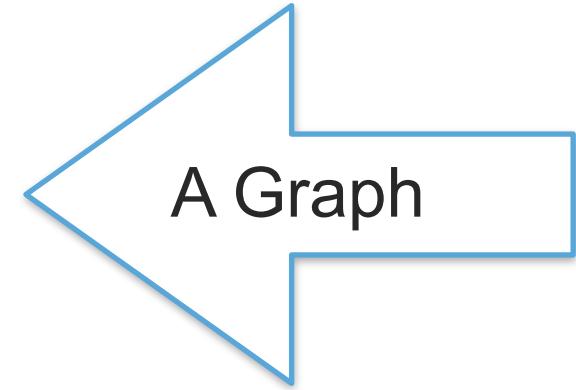
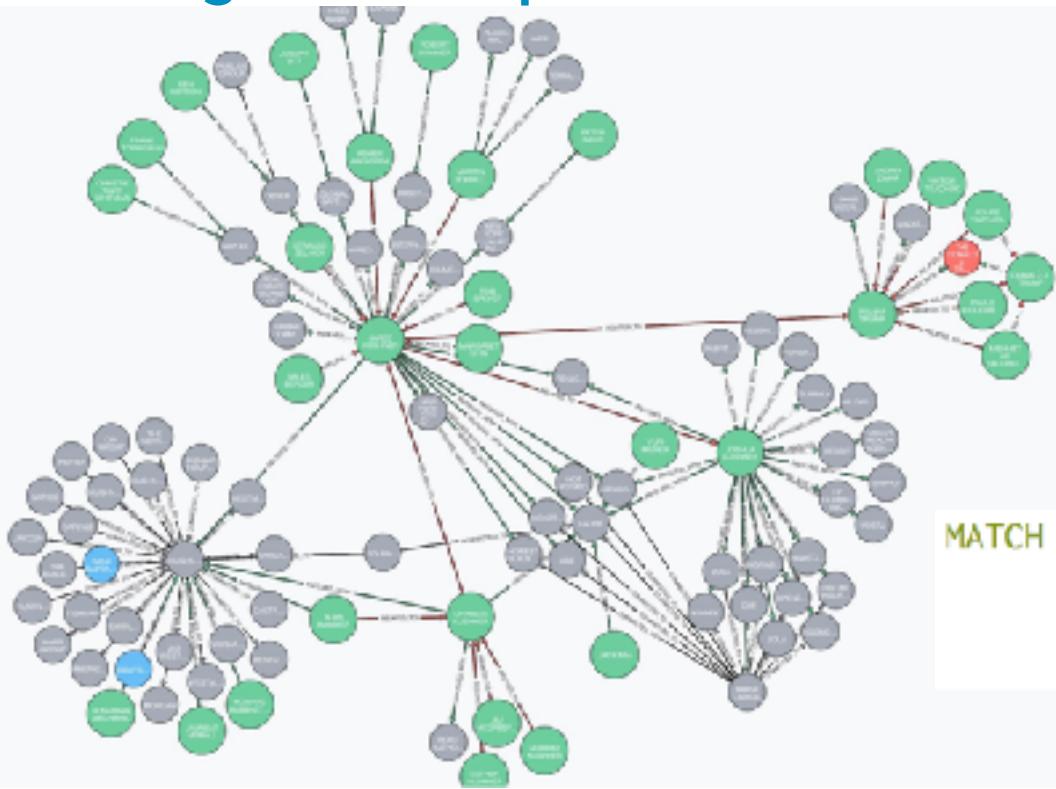


[https://cdn-images-1.medium.com/max/1600/0\\*8T9WFQKlf147X7FE.jpg](https://cdn-images-1.medium.com/max/1600/0*8T9WFQKlf147X7FE.jpg)



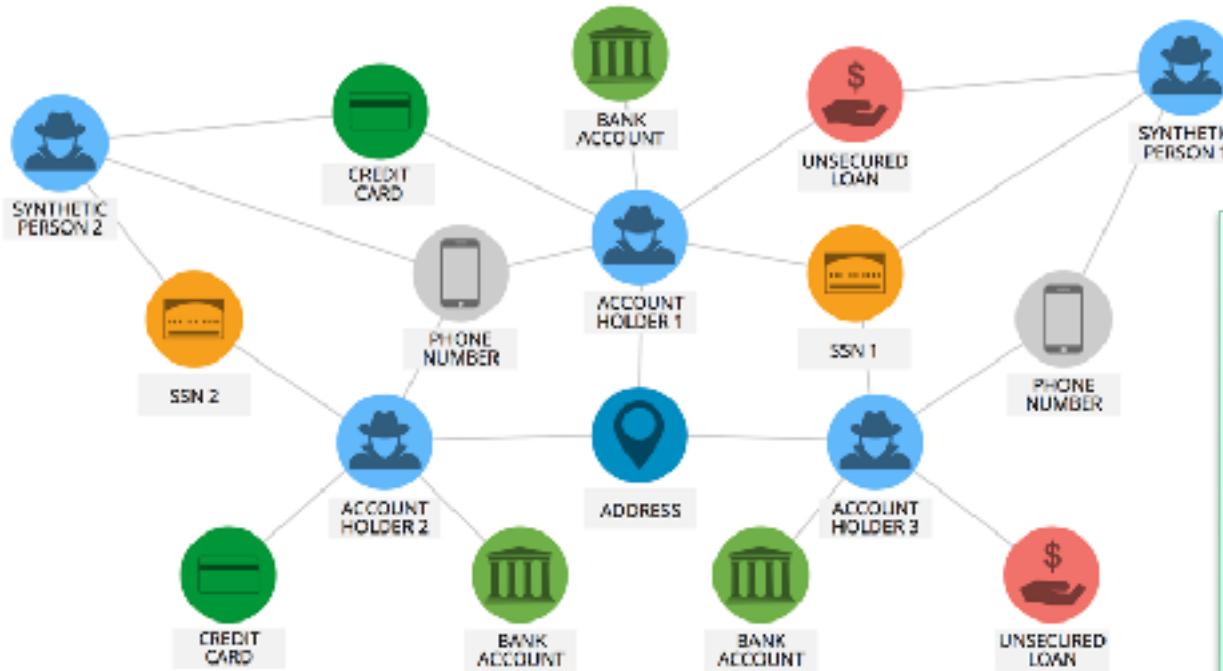
A Graph

# Two Degrees of Separation



```
MATCH network = (p:Person)-[*..2]-(others)
WHERE p.name='Jared Kushner'
RETURN network
```

# Graphs can easily join data from multiple domains

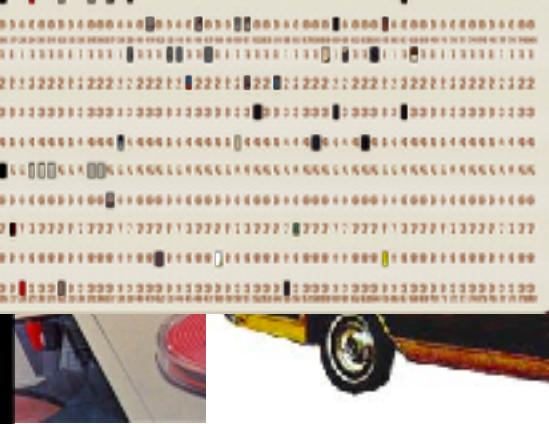
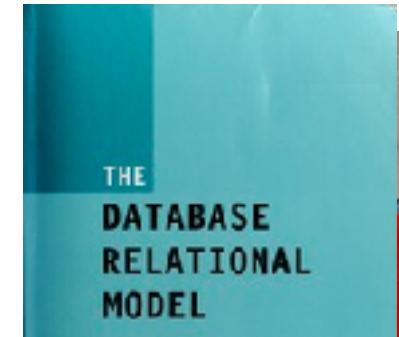
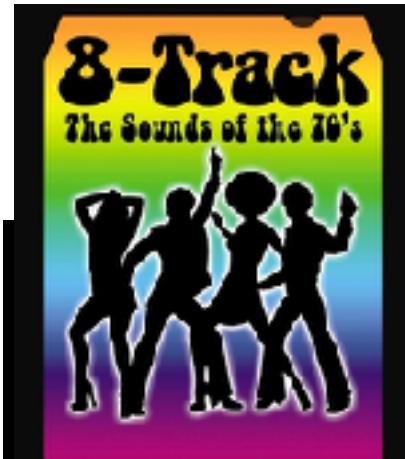


While many databases struggle with semi-structured and highly-connected data, Neo4j takes it in stride.

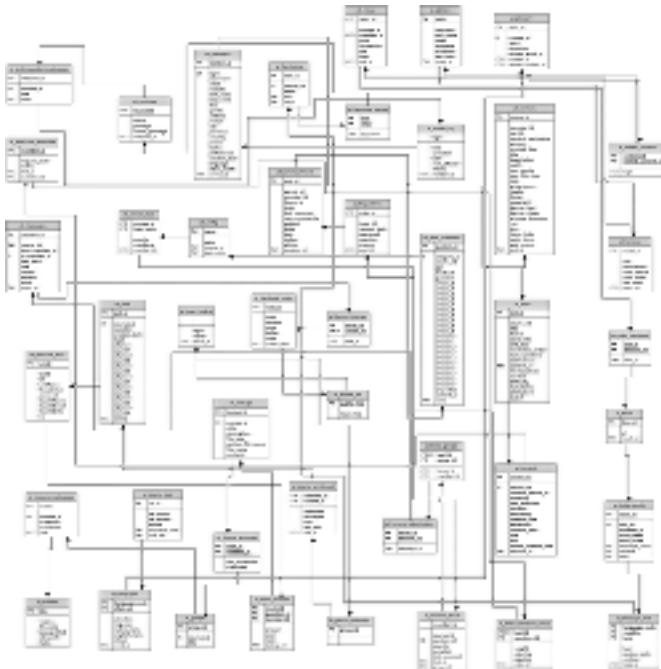
# Why do Relational Databases Fail?



# RDBMS and Other Technical Relics from the 70's



# Relational Databases



- Optimized for Storage Efficiency
- Relationship properties are hidden
- Even simple schema changes can be PITA
- OK for small number of joins, but then ....



# Complex SQL Queries vs Cypher

Find all direct reports and how many people they manage, up to three levels down

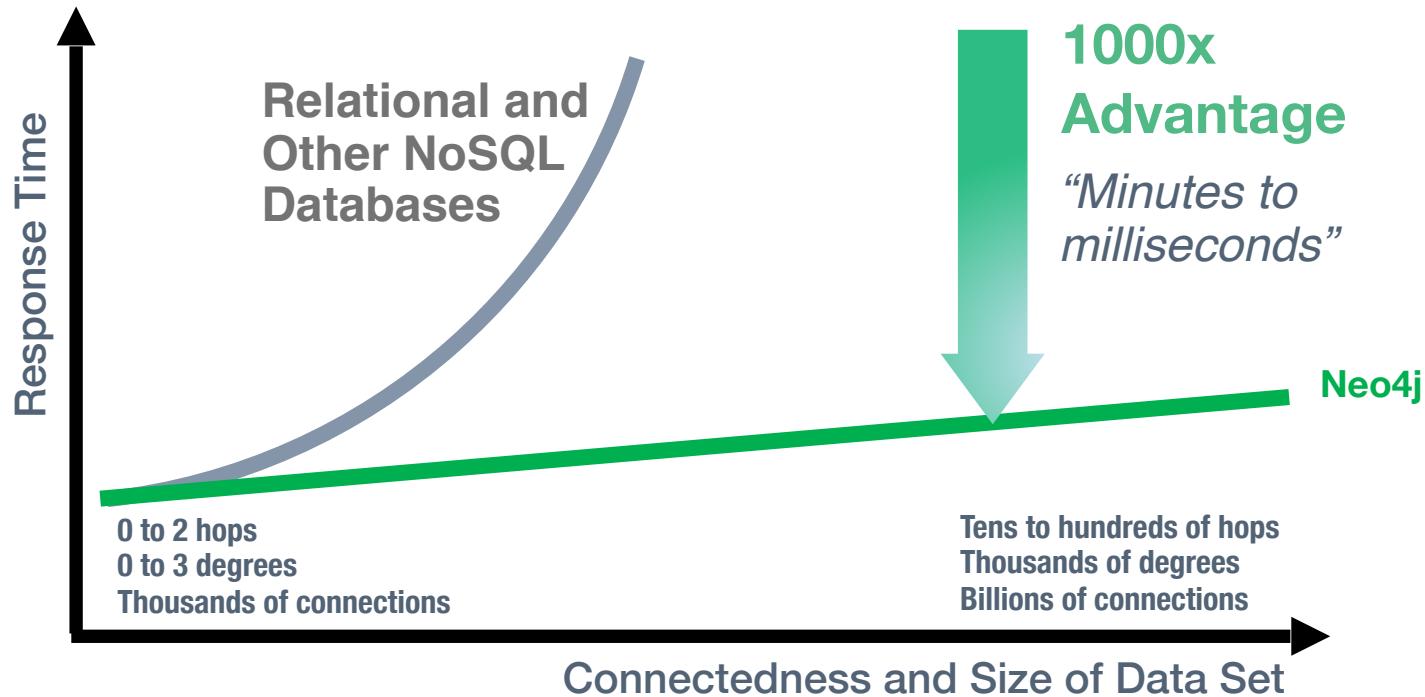
```
(SELECT T.directReports AS directReports, sum(T.count) AS count
FROM (
    SELECT manager.pid AS directReports, 0 AS count
    FROM person
    WHERE manager.manager = "John Doe"
    UNION
    SELECT manager.pid AS directReports, count(manager.directly_manages) AS count
    FROM person, reportee
    WHERE manager.pid = (SELECT id FROM person WHERE name = "John Doe")
    GROUP BY directReports
    UNION
    SELECT manager.pid AS directReports, count(L1Reportees.directly_manages) AS count
    FROM person, reportee
    ON manager.directly_manages = reportee.pid
    WHERE manager.pid = (SELECT id FROM person WHERE name = "John Doe")
    GROUP BY directReports
    UNION
    SELECT manager.pid AS directReports, count(L2Reportees.directly_manages) AS count
    FROM person, reportee
    ON manager.directly_manages = L1Reportees.pid
    WHERE manager.pid = (SELECT id FROM person WHERE name = "John Doe")
    GROUP BY directReports
    UNION
    SELECT manager.pid AS directReports, count(L3Reportees.directly_manages) AS count
    FROM person, reportee
    ON manager.directly_manages = L2Reportees.pid
    WHERE manager.pid = (SELECT id FROM person WHERE name = "John Doe")
    GROUP BY directReports
    ) AS T
    GROUP BY directReports)
    UNION
    (SELECT directReports AS directReports, sum(T.count) AS count
    FROM (
        SELECT manager.directly_manages AS directReports, 0 AS count
        FROM person
        WHERE manager.manager = "John Doe"
        UNION
        SELECT L1Reportees.directly_manages AS directReports, count(L1Reportees.directly_manages) AS count
        FROM person, reportee
        WHERE manager.pid = (SELECT id FROM person WHERE name = "John Doe")
        GROUP BY directReports
        UNION
        SELECT L2Reportees.directly_manages AS directReports, count(L2Reportees.directly_manages) AS count
        FROM person, reportee
        ON manager.directly_manages = L1Reportees.pid
        WHERE manager.pid = (SELECT id FROM person WHERE name = "John Doe")
        GROUP BY directReports
        UNION
        SELECT L3Reportees.directly_manages AS directReports, count(L3Reportees.directly_manages) AS count
        FROM person, reportee
        ON manager.directly_manages = L2Reportees.pid
        WHERE manager.pid = (SELECT id FROM person WHERE name = "John Doe")
        GROUP BY directReports
        ) AS T
        GROUP BY directReports)
        UNION
        (SELECT directReports AS directReports, sum(T.count) AS count
        FROM (
            SELECT manager.directly_manages AS directReports, 0 AS count
            FROM person
            WHERE manager.manager = "John Doe"
            UNION
            SELECT L1Reportees.directly_manages AS directReports, count(L1Reportees.directly_manages) AS count
            FROM person, reportee
            ON manager.directly_manages = L2Reportees.pid
            WHERE manager.pid = (SELECT id FROM person WHERE name = "John Doe")
            GROUP BY directReports
            UNION
            SELECT L2Reportees.directly_manages AS directReports, count(L2Reportees.directly_manages) AS count
            FROM person, reportee
            ON manager.directly_manages = L3Reportees.pid
            WHERE manager.pid = (SELECT id FROM person WHERE name = "John Doe")
            GROUP BY directReports
            ) AS T
            GROUP BY directReports)
            UNION
            (SELECT directReports AS directReports, sum(T.count) AS count
            FROM (
                SELECT manager.directly_manages AS directReports, 0 AS count
                FROM person
                WHERE manager.manager = "John Doe"
                UNION
                SELECT L1Reportees.directly_manages AS directReports, count(L1Reportees.directly_manages) AS count
                FROM person, reportee
                ON manager.directly_manages = L3Reportees.pid
                WHERE manager.pid = (SELECT id FROM person WHERE name = "John Doe")
                GROUP BY directReports
                UNION
                SELECT L3Reportees.directly_manages AS directReports, count(L3Reportees.directly_manages) AS count
                FROM person, reportee
                ON manager.directly_manages = (SELECT id FROM person WHERE name = "John Doe")
                WHERE manager.pid = (SELECT id FROM person WHERE name = "John Doe")
                GROUP BY directReports
                ) AS T
                GROUP BY directReports)
```

## Cypher Query

```
MATCH (boss)-[:MANAGES*0..3]->(sub),
      (sub)-[:MANAGES*1..3]->(report)
WHERE boss.name = "John Doe"
RETURN sub.name AS Subordinate,
       count(report) AS Total
```

- Less time writing queries
- Less time debugging queries
- Code that's easier to read

# “Minutes to Milliseconds” Real-Time Query Performance



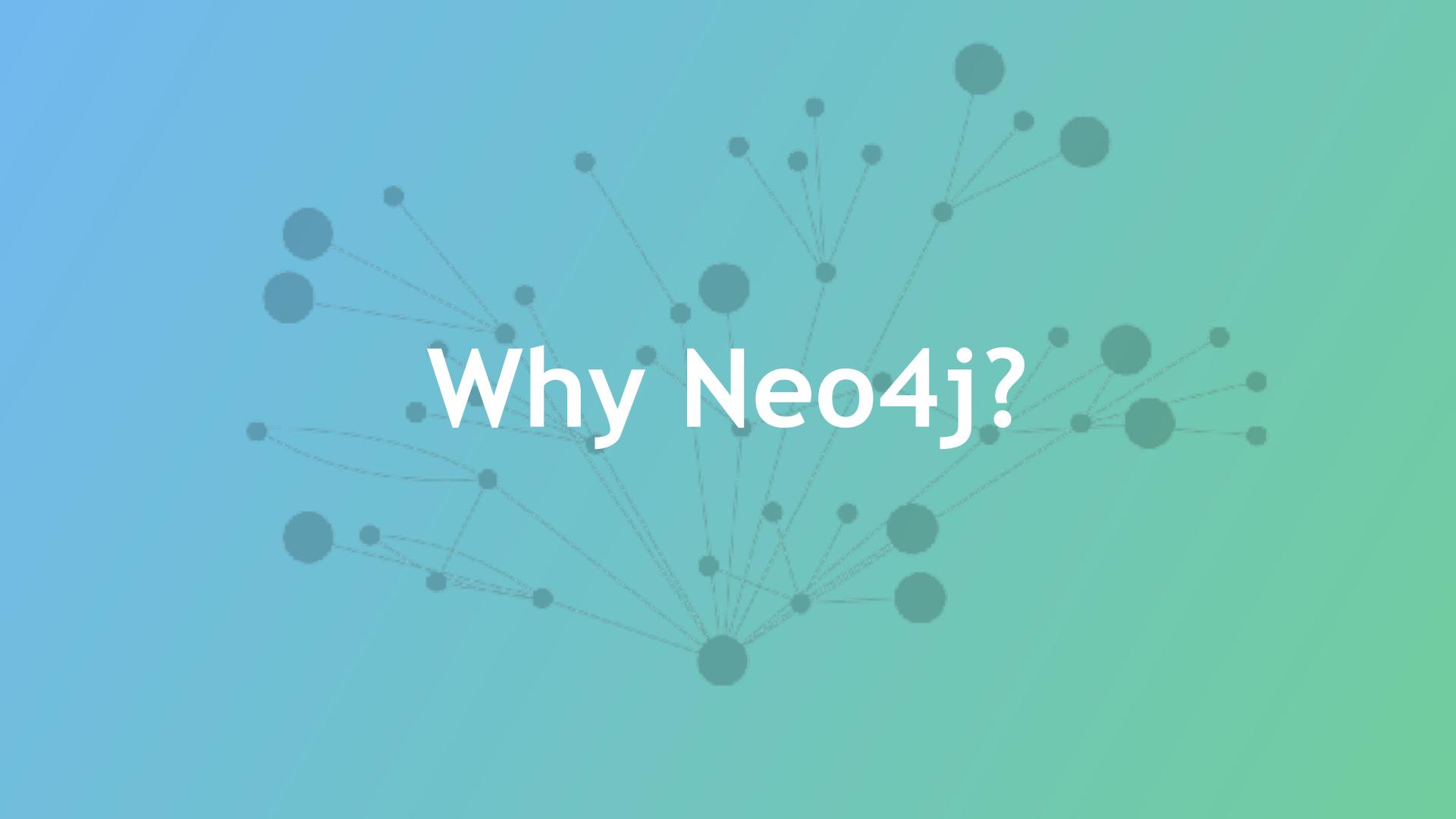
# Neo4j Replaces MySQL at eBay



We found Neo4j to be literally thousands of times faster than our prior MySQL solution, with queries that require 10-100 times less code. Today Neo4j provides eBay with functionality that was previously impossible”

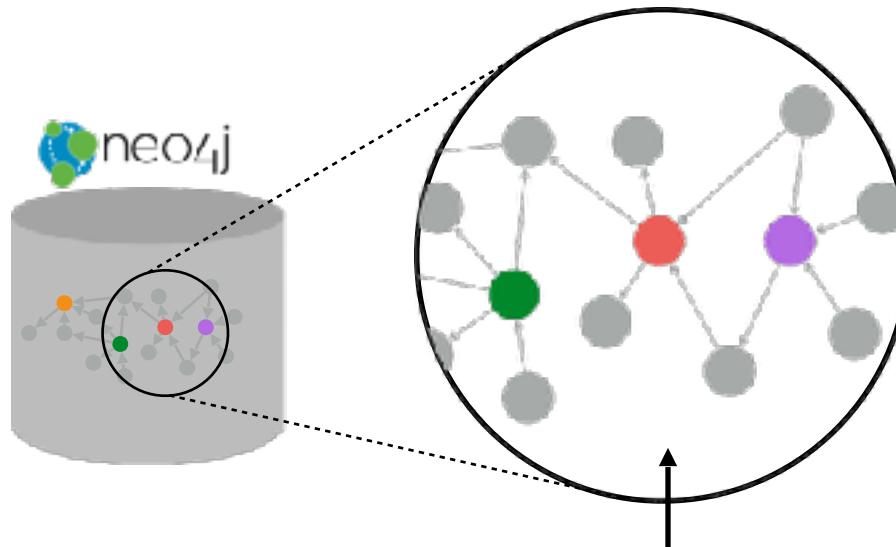
— Volker Pacher,  
Senior Developer, eBay

<https://neo4j.com/case-studies/ebay-shopbot>



# Why Neo4j?

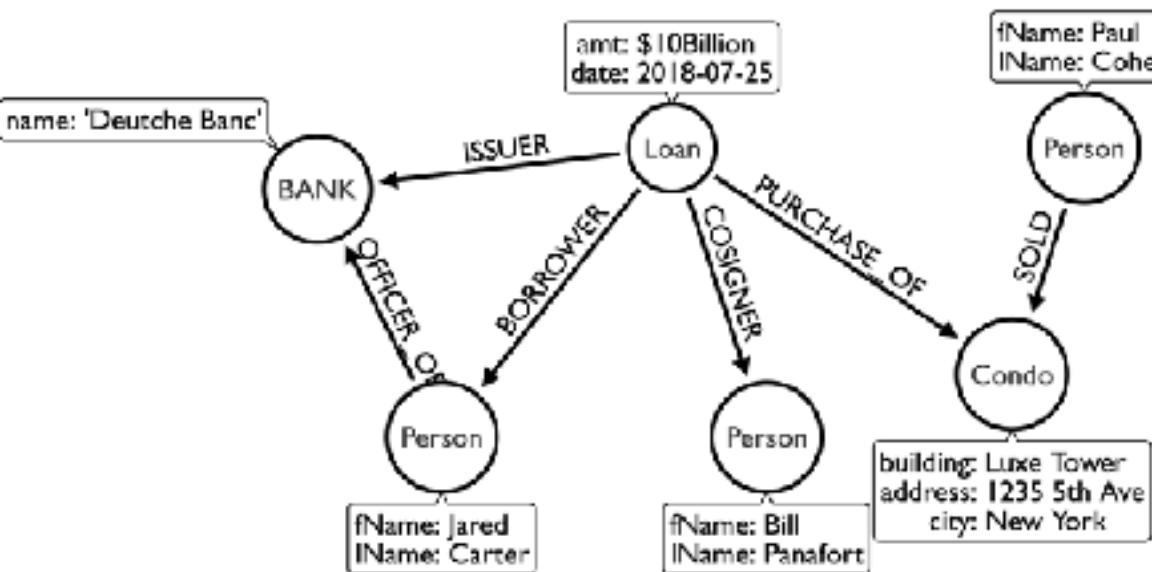
# Index-Free Adjacency:



**At Read Time:**  
Lightning-fast retrieval of data and relationships via pointer chasing

**At Write Time:**  
data is *connected*  
as it is stored

# Property Graph Model Components



## Nodes

- Represent the objects in the graph
- Can be *labeled*

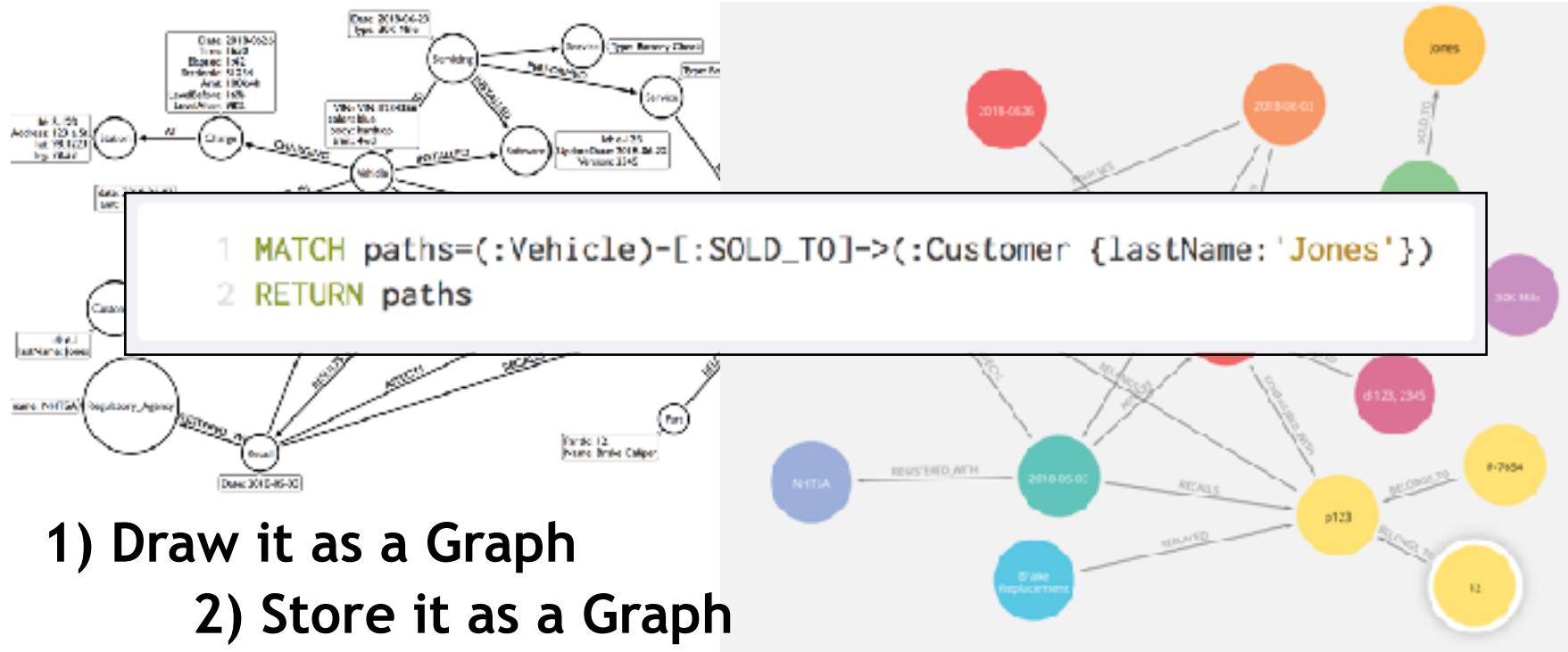
## Relationships

- Relate nodes by *type* and *direction*

## Properties

- Name-value pairs that can go on nodes and relationships.

# The Whiteboard Model is the Physical Model

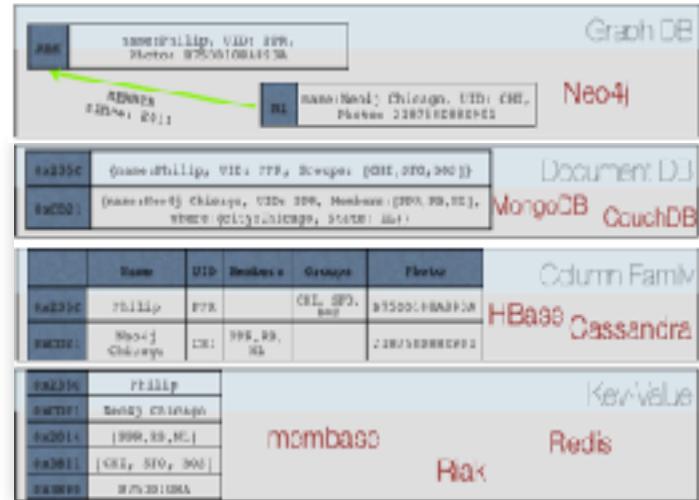


- 1) Draw it as a Graph
- 2) Store it as a Graph
- 3) Query it as a Graph

# NoSQL Databases Don't Handle Relationships Well



- **Collection-oriented** not **Connection-oriented**
- **No data structures** to model or store relationships
- **No query constructs** to support data relationships
- **Relating data requires complex “JOIN logic”** in the application (client-side ‘joins’)
- **No ACID support** for transactions



... making NoSQL databases inappropriate when data relationships are valuable

# Neo4j Replaces Cassandra at Adobe



- Replaced 48 Cassandra servers (50TB data) with 3 Neo4j Servers (40GB data)
- Ingestion reduced from 12 minutes to 106 ms.
- Signup reduced from 1.4 seconds to 400 ms.
- Several improved and new features
- Significantly reduced infrastructure demands

<https://neo4j.com/blog/this-week-in-neo4j-moving-adobe-behance-from-cassandra-to-neo4j-new-go-driver-emil-on-the-new-stack-makers-podcast/>

# Russian Twitter Trolls



# Neo4j Helps Uncover Russian Twitter Trolls



SECTIONS ▾

NIGHTLY NEWS MSNBC MEET THE PRESS DATELINE ☁ TODAY



TECH > SOCIAL MEDIA

GADGETS INTERNET SECURITY INNOVATION MOBILE

TECH DEC 20 2017, 11:11AM ET

## Russian trolls went on attack during key election moments

by BEN POPKEN



Ben Popken   
@bpopken

Following



@mdavidallen

Huge props and thank you to @neo4j and their @mdavidallen and @lyonwj for helping compile and analyze the deleted twitter data, surfacing trends and uncovering new angles.



@lyonwj

<https://www.nbcnews.com/pages/author/ben-popken>

<https://www.nbcnews.com/tech/social-media/russian-trolls-went-attack-during-key-election-moments-n827176>

# Indicted !

IN THE UNITED STATES DISTRICT COURT  
FOR THE DISTRICT OF COLUMBIA



UNITED STATES OF AMERICA	*	CRIMINAL NO.
v.	*	
INTERNET RESEARCH AGENCY LLC	*	(18 U.S.C. §§ 2, 371, 1349, 1028A)
A/K/A MEDIASINTEZ LLC A/K/A	*	
GLAVSET LLC A/K/A MIXINFO	*	
LLC A/K/A AZIMUT LLC A/K/A	*	
NOVINFO LLC,	*	
CONCORD MANAGEMENT AND	*	
CONSULTING LLC,	*	
CONCORD CATERING,	*	
YEVGENIY VIKTOROVICH	*	
PRIGOZHIN,	*	
MIKHAIL IVANOVICH BYSTROV,	*	
MIKHAIL LEONIDOVICH BURCHIK	*	
A/K/A MIKHAIL ABRAMOV,	*	
ALEKSANDRA YURYEVNA	*	
KRYLOVA,	*	
ANNA VLADISLAVOVNA	*	
BOGACHEVA,	*	
SERGEY PAVLOVICH POLOZOV,	*	
MARIA ANATOLYEVNA BOVDA	*	
A/K/A MARIA ANATOLYEVNA	*	
BELYAEVA,	*	
ROBERT SERGEYEVICH BOVDA,	*	
DZHEYKHUN NASIMI OGLY	*	
ASLANOV A/K/A JAYHOON	*	
ASLANOV A/K/A JAY ASLANOV,	*	
VADIM VLADIMIROVICH	*	
PODKOPAEV,	*	
GLEB IGOREVICH VASILCHENKO,	*	

# Russia Twitter Trolls



- 2752 Twitter accounts tied to Russia's Internet Research Agency
- Accounts suspended by Twitter
  - Data deleted
- What were they tweeting about?

user id	handle
905074658055453760	10_gop
2928776986	1488reasons
2936964756	1D_Nicole_
78783801	10RussiaPNOM
1860330774	1Erik_Lee
638742761615091041	1orenafawa1
2537164155	2kunegibon
823484273787294720	2ndHalfOnion
80712875	2silverbaker
80723575	30ToMarsFreedom
3644473036	450Usage
2261081126	4ckineeworks
746870305280847168	4new1837
4036537452	4MySquad
713885792550845793	69Revolvers
2534875924	6Drux
3936736149	71blajam11
2533198949	756shawn
2616792207	__Judith__D_
2096610923	__Sophia__
2598468901	__akihew
2644063209	__Amy_Mccann_
2753302910	__AnnaSwanson
2540858370	__Becker_Genaid
2426927403	__beglov
2659290936	__Ben_Santos

# Wayback API

user_id	handle
905074659055453780	10_gop
2528776986	148reasons
253664756	1D_Nicole_
78793801	10RussellNDM
180330774	1Erik_Lee
835742751515991041	1Korenatova1
2537164155	2Burgleben
823484279787294720	2ndHalfOnion
80712875	2othrbaker
80723575	30ToMoreFreedom
3644473036	499Usage
2961091126	4dsinevans
746870305280847168	4ever1937
4036537452	4MySquad
713685792550845780	68showwogers
2554675924	8Dux
3936736149	71bla_jam11
2533198949	756shawn
2616792207	__Judithe_D_
2556510923	__Sophia_
2596468901	_ashaw
2644663209	_Amy_Mccann_
2703352910	_AnneSwanson
2540858370	_Becker_Candid
2426927403	_beglov
2559290936	_Ben_Santos
2408864074	_Billy_Moyer_
2659366248	_Edward_Connor_
2844885589	_Giordi_Pope_
2502814810	_GeorgeSchultz_
2449237434	_GregorVan
2822970817	_Howard_Good_



[http://archive.org/wayback/  
available?url=http://  
twitter.com/TEN\\_GOP](http://archive.org/wayback/available?url=http://twitter.com/TEN_GOP)



```
{  
    "url": "https://twitter.com/TEN_GOP",  
    "archived_snapshots": {  
        "- closest": {  
            "status": "200",  
            "available": true,  
            "url": "https://web.archive.org/web/20170818065026/https://twitter.com/TEN_GOP",  
            "timestamp": "20170818065026"  
        }  
    }  
}
```



# Scraping Internet Archive...



```
(*suggestion_details*:[],*tweet_id*:  
"898272614224769824",*scribe_document*:"tweet")>  
:before  
+div class="tweet js-stream-tweet js-actionable-tweet js-profile-sugr-actionable disclosure-type-context"  
original-tweet js-original-tweet  
  
bus-card bus-context  
* data-tweet-id="898272614224769824" data-item-id="898272614224769824" data-permalink-pref="TEN_GOP/status/898272614224769824" data-conversation-id="898272614224769824" data-tweet-name="898272614224769824-e405e87-bc9-41f7-bfd-88edfb81c" data-retweet-id="898272614224769824-e405e87-bc9-41f7-bfd-88edfb81c" data-retweeter="TEN_GOP" data-screen-name="TEN_GOP" data-name="Tennessee" data-user-id="4224729994" data-you-folllow="false" data-follows-you="false" data-you-block="false" data-reply-to-user-json:[{"id":4224729994,"screen_name":"TEN_GOP","name":"Tennessee","emojified_name":1,"text": "Tennessee","emojified_text_as_html":1,"Tennessee"}])> data-disclosure-type bus-card> :before  
+div class="context"></div>  
+div class="context">  
+div class="stream-item-header"></div>  
+div class="js-tweet-text-container">  
+p class="TweetTextSize TweetTextSize--normal js-tweet-text-tweet-text" lang="en" data-aria-label="Text">  
"About a week ago a horde of Moroccans landed on the  
beach in Spain. This week a terror attack by a Moroccan  
in "#Barcelona".<a href="https://web/281760104556021" href="https://twitter.com/hashtag/Barcelona&src=tash" data-query-source="hashtag_click" class="twitter-hashtag-link pretty-link is-new" dir="ltr"><u>#Barcelona</u>">
```

# Scraping Internet Archive...

```
1  [
2    {
3      "tweet_id": 561931644785811457,
4      "screen_name": "AlwaysHungryBae",
5      "permalink": "/AlwaysHungryBae/status/561931644785811457",
6      "tweet_text": "Happy Super Bowl Sunday \n#superbowlfood pic.twitter.com/s6rvMtdLor",
7      "user_id": "2882138845",
8      "timestamp": 1422888918888,
9      "hashtags": [
10        {
11          "tag": "superbowlfood",
12          "archived_url": "/web/20150603094250/https://twitter.com/hashtag/superbowlfood?src=hash"
13        }
14      ],
15      "links": [
16        {
17          "url": "pic.twitter.com/s6rvMtdLor",
18          "archived_url": "http://web.archive.org/web/20150603094250/http://t.co/s6rvMtdLor"
19        },
20        {
21          "url": "https://pbs.twimg.com/media/B8xh2FFC0AE-wxU.jpg:large",
22          "archived_url": "/web.archive.org/web/20150603094258/https://twitter.com/AlwaysHungryBae/status/561931644785811457/photo/1"
23        }
24      ]
25    }
26  ]
```

# Loading Tweets Into Neo4j

```
{  
    "tweet_id": "4505104479551247",  
    "screen_name": "AlyoshaKapchuk",  
    "user_id": "1000000000000000",  
    "tweet_text": "Happy Soviet Day! Sender: https://twitter.com/dm7dilition",  
    "user_id": "2021560000000000",  
    "timestamp": "2013-08-01T10:00:00Z",  
    "hashtags": [  
        {"tag": "#opensofice"},  
        {"archived_url": "https://webcache.googleusercontent.com/search?q=cache:twinkie%2Bopensofice"}  
    ],  
    "links": [  
        {"url": "http://pic.twitter.com/0zCm", "archived_url": "https://webcache.googleusercontent.com/search?q=cache:twinkie%2Bhttp://pic.twitter.com/0zCm"},  
        {"url": "https://t.co/2qM40000004250", "archived_url": "https://webcache.googleusercontent.com/search?q=cache:twinkie%2Bhttps://t.co/2qM40000004250"}  
    ]  
}
```



```
1 WITH $tweetArr AS tweets  
2 UNWIND tweets AS tweet  
3  
4 MERGE (u:User {user_id: tweet.user_id})  
5 ON CREATE SET u.screen_name = tweet.screen_name  
6  
7 MERGE (t:Tweet {tweet_id: tweet.tweet_id})  
8 ON CREATE SET t.text = tweet.tweet_text,  
     t.permalink = tweet.permalink  
9  
10 MERGE (u)-[:POSTED]->(t)  
11  
12  
13 FOREACH (ht IN tweet.hashtags |  
14     MERGE (h:Hashtag {tag: ht.tag })  
15     ON CREATE SET h.archived_url = ht.archived_url  
16     MERGE (t)-[:HAS_TAG]->(h)  
17 )  
18  
19 FOREACH (link IN tweet.links |  
20     MERGE (l:Link {url: link.url})  
21     ON CREATE SET l.archived_url = link.archived_url  
22     MERGE (t)-[:HAS_LINK]->(l)  
23 )
```

# Success!

345k Tweets, 41k Users (454 Russian Trolls)

```
1 WITH $tweetArr AS tweets
2 UNWIND tweets AS tweet
3
4 MERGE (u:User {user_id: tweet.user_id})
5 ON CREATE SET u.screen_name = tweet.screen_name
6
7 MERGE (t:Tweet {tweet_id: tweet.tweet_id})
8 ON CREATE SET t.text = tweet(tweet.text)
9 t_permalink = tweet permalink
10
11 MERGE (u)-[:POSTED]-(t)
12
13 FOREACH (ht IN tweet.hashtags |
14 MERGE (h:Hashtag {tag: ht.tag})
15 ON CREATE SET h.archived_url = ht.archived_url
16 MERGE (t)-[:HAS_TAG]->(h)
17 )
18
19 FOREACH (link IN tweet.links |
20 MERGE (l:Link {url: link.url})
21 ON CREATE SET l.archived_url = link.archived_url
22 MERGE (t)-[:HAS_LINK]->(l)
23 )
```



# Extending Neo4j with Natural Language Processing



Lucy Barton  
@LucyBartonUSA

We hung out a huge banner right on the Arlington Memorial Bridge. Goodbye to murderer @BarackObama #ThanksObama

323 362 8:59 AM · 10 Mar 2013

Basic Tweet



Use NLP to extract  
Hashtags  
from Tweet



Tweet Plus HashTags

# How to combine open source NLP tools w/ Neo4j?



## GraphAware Natural Language Processing

build passing This [Neo4j](#) plugin offers Graph Based Natural Language Processing capabilities.



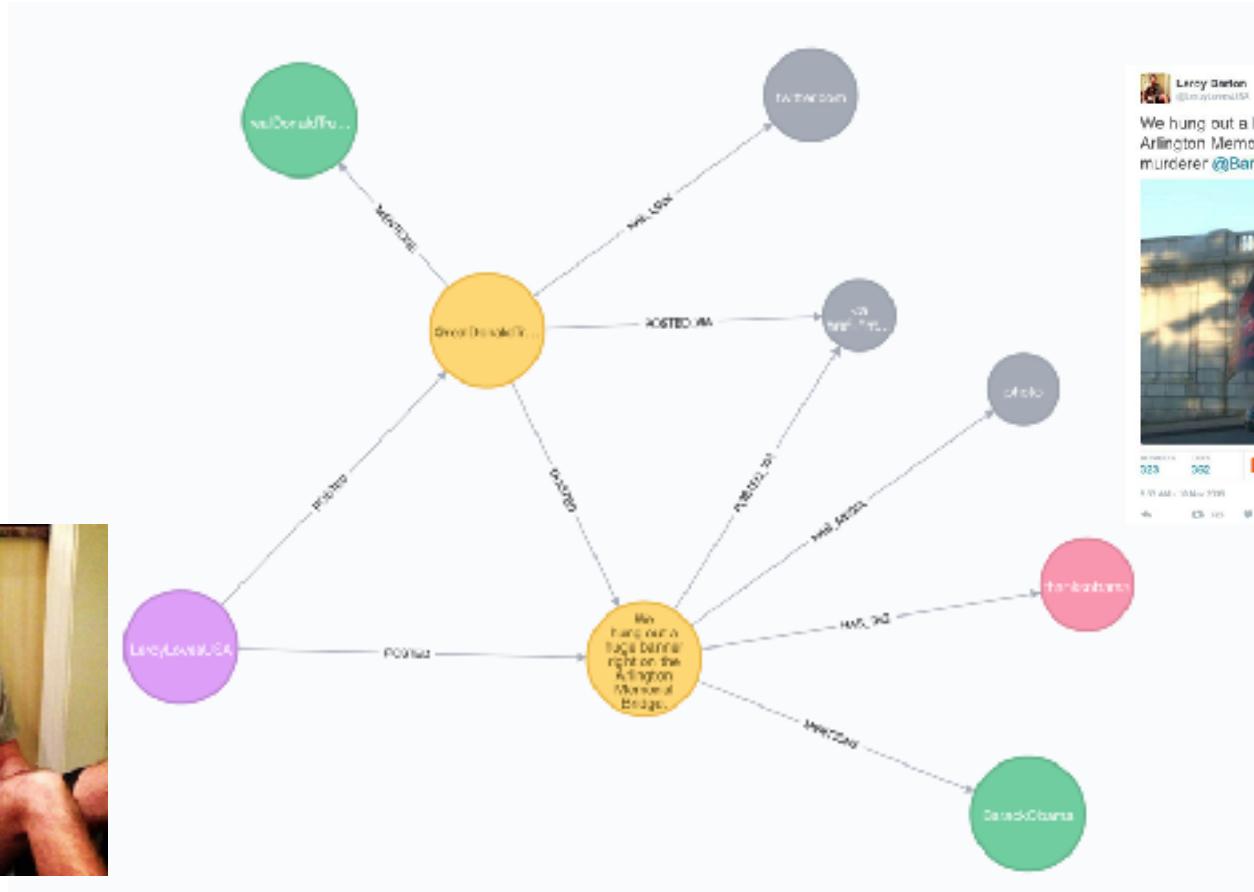
```
1 MATCH (tw:Tweet {lang: "en"})
2 CALL ga.nlp.annotate({text: tw.text, id: id(tw)})
3 YIELD result
4 MERGE (tw)-[:HAS_ANNOTATED_TEXT]->(result)
5 RETURN count(result)
```

<https://github.com/graphaware/neo4j-nlp>

```
1 MATCH
```

```
2 (u:User {screen_name: "LeroyLovesUSA"})-[:POSTED]->(t:Tweet)-[:HAS_TAG]->(ht:Hashtag {key: "thanksobama"})
```

```
3 RETURN *
```



Your typical American Citizen?



@TroyLOVEUSA

RUSSIAN

Your typical Local News Publication?

Cleveland Online

@OnlineCleveland

Breaking news, weather, traffic  
and more for Cleveland. DM us  
anytime. RTs not endorsements

City of Cleveland, USA

@ClevelandOnline

Your typical local Political Party?

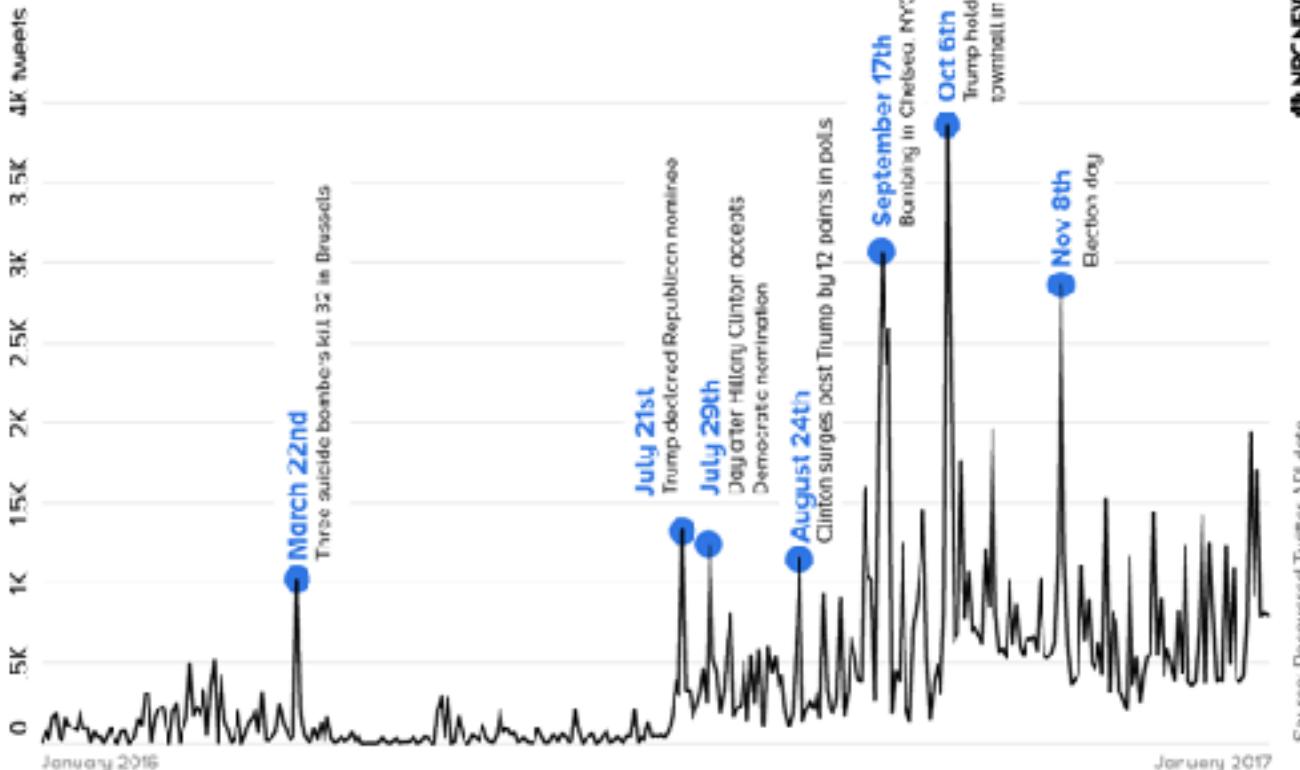


University of Tennessee  
Republican Young Leadership  
Team - Knoxville, Tennessee  
2012-2013

@TEN\_GOP

TROLLS!

## Russian Troll Volume Spiked During 2016 Campaign Events



NBC NEWS

# Graph Algorithms

Harness Graph Theory to Understand Your Data



# Centrality

Measure of importance

## PageRank

- Importance and number of connected nodes
- How many friends - and how important are their friends?

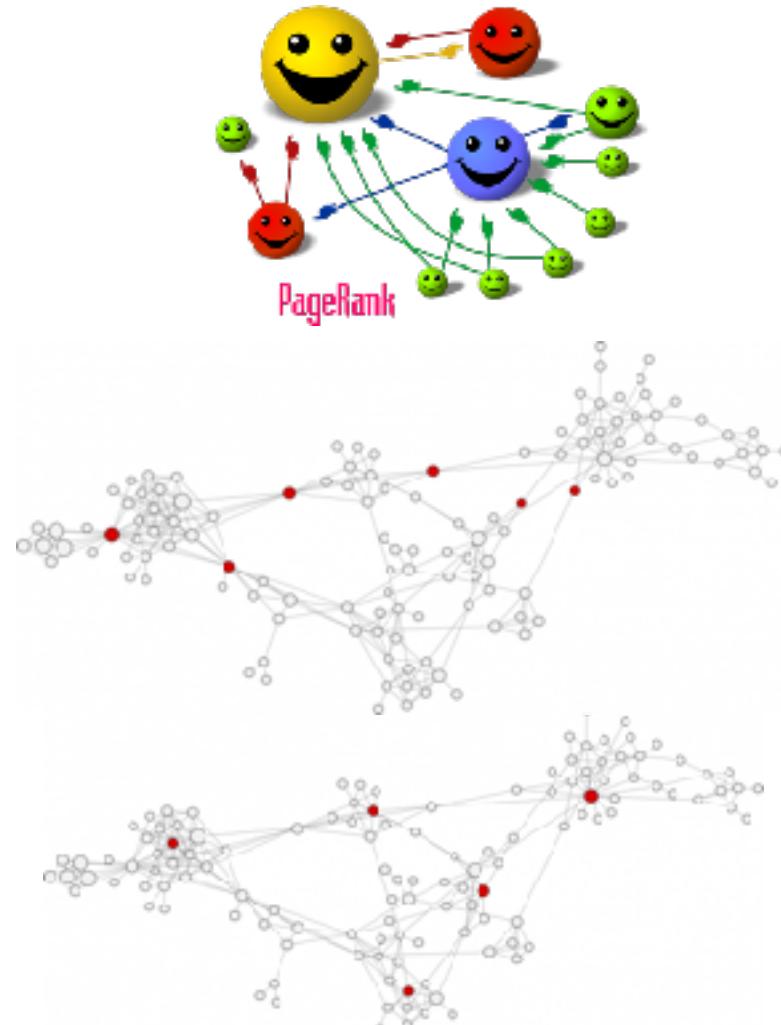
## Betweenness Centrality

- Number of shortest paths connecting all pairs in the network
- Identifies bottlenecks and points of vulnerability

## Closeness Centrality

- Inverse of distance to all other nodes in the network
- Identifies optimal nodes for quickly reaching all other nodes

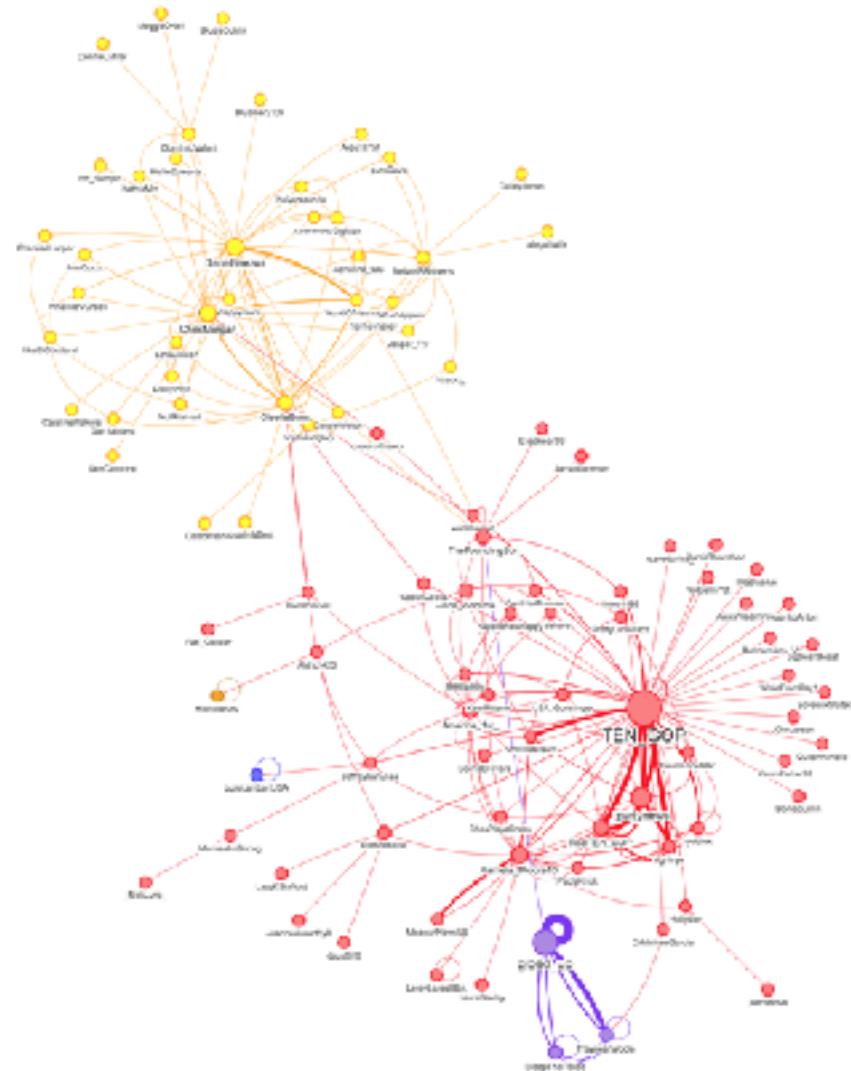
<https://neo4j.com/blog/graph-algorithms-neo4j-15-different-graph-algorithms-and-what-they-do/>



# Graph Visualization

Centrality & community detection  
AMPLIFIED relationships

Node size → PageRank  
Color → community detection  
Rel Thickness → weight



# Knowledge Graphs



# Neo4j Knowledge Graph Aids Mission to Mars



Using Neo4j saved “well over two years of work and \$1,000,000 of taxpayer funds”

— David Meza,  
Chief Knowledge Architect, NASA



<https://neo4j.com/blog/nasa-lesson-learned-database-using-neo4j-linkurious/>

# Airbnb's Dataportal Tackles 'Tribal Knowledge'



“Our hope is that any employee, regardless of role, can easily find or discover data and feel confident about its trustworthiness and relevance”

<https://neo4j.com/blog/democratizing-data-discovery-airbnb/>

# neo4jsandbox.com

neo4j sandbox

Products | Solutions | Methods | Customer | About | Contact | Search

Launch a New Sandbox.

Each sandbox allows you to interact with our sample services, or customize code.

**New Services**

- Twitter Trolls  Launch Service
- Facebook Trolls  Launch Service

**Services**

- Twitter Trolls  Launch Service
- Instagram Trolls  Launch Service

**Resource Centers**

- Twitter Trolls  Contains consolidated real-time communication - long duration streams
- Facebook Trolls  Contains unified communication streams for instant user management

**Tools**

- Twitter Trolls  Allows individual Twitter users to analyze their own streams
- Facebook Trolls  Allows analysis of a single Facebook account streams

**Connected Graphs**

- Twitter Trolls  Loads data from TwitterStreams
- Facebook Trolls  Loads data from FacebookStreams

**Resources**

- Twitter Trolls  Begins monitoring it and associates Trump Administrator using the default form (incorrect)
- Facebook Trolls  GraphCentral Europe 2D+3D simulation game

Russian Twitter Trolls    Get Started    **Details**    Data Model    Code    Advanced ▾



**Neo4j Browser:** <https://10-0-1-194-33031.neo4jsandbox.com/>  
**Direct Neo4j HTTP:** <http://54.237.227.207:33031/browser/>

**Username:** neo4j  
**Password:** recognition-bins-procurement 

**IP Address:** 54.237.227.207  
**HTTP Port:** 33031  
**Bolt Port:** 33030

**Expires:** 8 days, 19 hours, 42 minutes

<https://hackernoon.com/six-ways-to-explore-the-russian-twitter-trolls-database-in-neo4j-6e52394c38f1>

# Neo4j's Experts:



**David Allen**  
Technology Partner Architect, Neo4j



**William Lyon**

Software Developer, Neo4j  
*(Developer Relations, integrations, GraphQL)*

[will@neo4j.com](mailto:will@neo4j.com)

@lyonwj

lyonwj.com

# Neo4j Data Journalism Program:



Blog Support Company News Contact Us

Download



PRODUCTS

SOLUTIONS

PARTNERS

CUSTOMERS

LEARN

DEVELOPERS

Search

## Neo4j Data Journalism Accelerator Program

Graph Databases are an amazing tool to discover connections in real-world data. Neo4j has a long history of being used by journalists, including the International Consortium of Investigative Journalists (ICIJ). The ICIJ used the Panama Papers and Swiss Leaks datasets to investigate and draw connections between people, shell companies and firms which helped create them ([Database Download](#), [Why Graph Databases](#)).

### What is the Accelerator Program?

The Neo4j Developer Relations team wants to work with interested data journalists to help them understand how their data would be modeled as a graph database and what types of queries might be helpful to explore the graph. Drawing new connections between existing information and new sources allows new stories to be uncovered and told.

### Short Application Process

We have a quick application process for us to understand your goals and your data. Our Engineers will work with selected applicants over the coming weeks to help you start being successful using Neo4j – either the community or enterprise editions. The only thing we ask is that you mention your usage of Neo4j in any resulting stories.

GET STARTED NOW

<https://neo4j.com/graph-database-data-journalism-accelerator-program/>

# Questions?



Mark Quinsland  
[neo4j.com](http://neo4j.com)

Big Data Day - LA  
U.S.C. Aug 2018