# Understanding Trolls with Efficient Analytics of Large Graphs in Neo4j

David Allen, Amy E.Hodler, Michael Hunger, Martin Knobloch, William Lyon, Mark Needham, Hannes Voigt

> BTW Rostock Feb 2019





# Michael Hunger

Director Neo4j Labs at Neo4j @mesirii | michael@neo4j.com



# **1. Graph Databases vs. Graph Processing**

- 2. Neo4j Graph Platform
- 3. Neo4j Graph Algorithms
- 4. Application in SNA on Twitter Troll Dataset

# Why graphs?



# The world is a graph – everything is connected



- people, places, events
- companies, markets
- countries, history, politics
- sciences, art, teaching
- technology, networks, machines, applications, users
- software, code, dependencies, architecture, deployments
- criminals, fraudsters and their behavior



# What are people using Neo4j for?



# Neo4j - Transforming 100s of Large Enterprises For Over 14 Years

MARRIOTT

# Real-time promotion recommendations

FORTUNE 50

RETAIL

- Record "Cyber Monday" sales
- About 35M daily transactions
- Each transaction is 3-22 hops
- Queries executed in 4ms or less
- Replaced IBM Websphere commerce

#### Marriott's Real-time Pricing Engine

- 300M pricing operations per day
- 10x transaction throughput on half the hardware compared to Oracle
- Replaced Oracle database





- Large postal service with over 500k employees
- Neo4j routes 7M+ packages daily at peak, with peaks of 5,000+ routing operations per second.







**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 labeled property graph model



# **Property Graph Model Components**



#### Nodes

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







# **Property Graph Model Components**



#### Nodes

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

#### Relationships

• Relate nodes by type and direction



# **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.



### Summary of the graph building blocks



- **Nodes** Entities and complex value types
- **Relationships** Connect entities and structure domain
- **Properties** Entity attributes, relationship qualities, metadata
- Labels Group nodes by role

# **Neo4j is a Graph Platform**





### Neo4j is a database



## Neo4j is a graph platform





# **Graph Querying**







A pattern matching query language made for graphs

- Declarative
- Expressive
- Pattern Matching

Formal specification, SIGMOD paper:

https://homepages.inf.ed.ac.uk/libkin/papers/sigmod18.pdf

### **Cypher: Express Graph Patterns**





### **Cypher: CREATE Graph Patterns**





### **Cypher: MATCH Graph Patterns**





### **Cypher: Query Planner**





# **Cypher: Query Plan**

- different planners
- e.g. IDP planner
- different runtimes
- e.g. bytecode compiled



# openCypher / GQL



- open source graph query language specification and reference implementation
- Multi-Vendor effort to standardize a Graph Query Language, see: gglstandards.org

GQL is a proposed new international standard language for property graph querying. The idea of a <u>standalone graph query</u> <u>language to complement SQL</u> was raised by ISO SC32/ WG3 members in early 2017, and is echoed in the <u>GQL manifesto</u> of May 2018.

GQL supporters aim to develop a next-generation declarative graph query language that builds on the foundations of SQL and integrates proven ideas from the existing <u>openCypher, PGQL, and G-CORE</u> languages.

GQL will incorporate this prior work, as part of an expanded set of features including regular path queries, graph compositional queries (enabling views) and schema support.

# A graph query example



### A social recommendation







## A social recommendation





```
MATCH (person:Person)-[:IS_FRIEND_OF]->(friend),
        (friend)-[:LIKES]->(restaurant),
        (restaurant)-[:LOCATED_IN]->(loc:Location),
        (restaurant)-[:SERVES]->(type:Cuisine)
WHERE person.name = 'Philip'
AND loc.location='New York'
AND type.cuisine='Sushi'
RETURN restaurant.name
```

### A social recommendation



Sushi restaurants in New York, New York that my friends like

f

Sushi restaurants in New York, New York that my friends like



-

# **Graph Algorithms**



#### Jona hai Portes Julia Hartley-Brewer The Spectator Vote Leave elmer Charlotte Visio **Bernard** Jenkin Tanya Beckett Daniel Hannan The Andrew Marr Show Chris Mason Guido Fawkes **BBC** Radio 4 Today Reuters UK The Sun Tim Montgomerie a ConservativesIN **BBC News (UK)** Adrian Monck Autora No. Rober Kate Hoey **Toby Young** John Rentoul Jim Pickard s Out HM Treasury **ITV News** -James Murray Keiran Pede LEAVE.EU Dr.Patinia HuffPost UK Dr Liam Fox MP Nicholas Soames Euro Guido Margham QC Jill Rutter Damian Collins ter Esq. lain Martin THEOR Chelaticate Isabel Hardman CapX Ladbrokes Politics lan Paisley Matt Chorie-Faisal Islam LBC Lines Allo (a) Will Straw Paul Stronger In Paul Quigley Stron Source: John Swain - Twitter Analytics Right Relevance Talk steve echards **Yvette Cooper** Manualhow



#### **Example Workflow Pipeline**



#### Example Workflow Pipeline





# Usage

- 1. Call as Cypher procedure
- 2. Pass in specification (Label, Prop, Query) and configuration
- 3. ~.stream variant returns (a lot) of results

CALL algo.<name>.stream('Label','TYPE',{conf})
YIELD nodeId, score

4. non-stream variant writes results to graph returns statistics

```
CALL algo.<name>('Label','TYPE',{conf})
```



# **Cypher Projection**

Pass in Cypher statement for node- and relationship-lists.

```
CALL algo.<name>(
    'MATCH ... RETURN id(n)',
    'MATCH (n)-->(m)
    RETURN id(n) as source,
        id(m) as target', {graph:'cypher'})
```



# **Design Considerations**

- Ease of Use Call as Procedures
- Parallelize everything: load, compute, write
- Efficiency: Use direct access, efficient datastructures, provide high-level API
- Scale to billions of nodes and relationships
   Use up to hundreds of CPUs and Terabytes of RAM

# Architecture

- 1. Load Data in parallel from Neo4j
- 2. Store in efficient data structures
- 3. Run Graph Algorithm in parallel using Graph API
- 4. Write data back in parallel



# Scale: 144 CPU

1 [	37 [	73 [
2 [	38 [	74 []]
3 [	39 [	75 [
4 [	40 [	76 [
5 [	41 [	77
6 [	42 [	78 [
7 [	43 [	79 [
8 [	44 [	80 [
9 [	45 [	81 []
10 [	46 [	82 [
11 [	47 [	83
12 [	48 [	84 [
13 [	49 [	85 [11111111111
14 [	50 [	86 [1111]
15 [	51 [	87 []]]]]]]]]]]
16 [	52 [	88 [11111]
17 [	53 [	89 11111111111111
18 [	54 []	90 [
10 [	55 [11111111111111111111111111111111111	91 [[[[[[[[[[[[[
20 [ 0 0%]	56 [	92 [111111111111111
21 [	57 [	93
22 [	58 [	94
23 [	50 [	95
24 [	60 [	96
25 [	61 [	97 [
	62 [	
20 [ 0.06]	62 [ ] ] ] ] ] ] ] ] ] ] ] ] ] ] ] ] ] ]	90 [
	05 [ ] ] ] [ ] ] ] ] ] ] ] ] ] ] ] ] ] ]	100[
20 [ 0.06]	CE [	101
29 [ ] ] ] ] ] ] ] ] ] ] ] ] ] ] ] ] ] ]	03 [	102[
30 [	00 [	102[
31 [           10.1%]	67 [	103
32 [	68 [	104
55 [	09 [	105[
34 [                       27.7%]	70 [	106
35 [	/1 [	10/1111111111111111111111
30 LIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	// L	108
Mem[		Tasks: 58, 949 thr; 93 running
Swp[	0K/0K]	Load average: 12.87 27.10 32.33
		Uptime: 162 days(!), 01:28:15

$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	21.9%]	109[////////////////////////////////////
25.2%       111[       64.9%]         45.5%       112[       88.3%]         35.5%       113[       76.1%]         29.5%       114[       79.4%]         28.8%       115[       52.9%]         18.8%       116[       67.1%]         1.9%       127[       63.7%]         1.9%       127[       63.7%]         1.9%       127[       63.7%]         1.9%       126[       76.1%]         7.7%       126[       76.1%]         7.7%       126[       76.1%]         7.7%       126[       76.1%]         7.7%       126[       76.1%]         7.7%       126[       76.1%]         7.7%       126[       76.1%]         1.0%       126[       76.1%]         7.7%       126[       76.6%]         11.0%       126[       71.6%]         11.0%       126[       71.6%]         11.0%       126[       71.6%]         12.3%       127[       68.2%]         31.0%       136[       83.8%]         0.0%       132[       60.6%]         0.0%]       136[       60.6%] <th>3.9%]</th> <th>110[                                   </th>	3.9%]	110[
45.5%       112[       76.1%]         35.5%       113[       76.1%]         29.5%       114[       52.9%]         18.8%       115[       67.7%]         19.4%       117[       67.7%]         54.8%       120[       87.7%]         10.4%       120[       86.5%]         11.9%       120[       86.5%]         12.12%       120[       86.5%]         12.4%       120[       86.5%]         12.4%       120[       86.5%]         12.4%       120[       86.5%]         12.4%       120[       76.1%]         7.7%       122[       76.1%]         7.7%       122[       76.1%]         7.7%       122[       76.1%]         7.7%       122[       76.1%]         7.7%       122[       76.1%]         7.7%       122[       76.6%]         12.4%       123[       76.6%]         13.0%       124[       76.6%]         0.6%       126[       71.6%]         11.6%       126[       71.6%]         12.3%       127[       78.4%]         0.6%       132[       79.4%]	25.2%]	111[
35.5%       113[       76.1%]         29.5%       114[       79.4%]         28.8%       15[       79.4%]         18.8%       16[       67.1%]         1.9%       177[       63.7%]         21.3%       18[       67.7%]         21.3%       19[       57.4%]         0.0%       120[       57.4%]         1.9%       177[       85.5%]         21.2%       121[       76.1%]         7.7%       122[       59.6%]         19.4%       123[       76.1%]         11.0%       124[       76.1%]         12.13%       126[       76.1%]         13.0%       126[       76.1%]         11.0%       126[       77.4%]         12.13%       127[       80.0%]         13.0%       128[       80.0%]         0.0%       131[       83.8%]         11.0%       133[       83.8%]         13.0%       134[       97.4%]         13.0%       134[       91.0%]         0.0%       133[       83.8%]         13.13[       60.0%]       60.6%]         0.0%]       134[       77.4%]	45.5%]	112[
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	35.5%]	113[///////////////////////////////////
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	29.5%]	114[///////////////////////////////////
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	28.8%]	115[
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	18.8%]	116[
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	1.9%]	117[
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	21.3%]	118[
0.0%       120[       76.1%]         7.7%       122[       76.1%]         19.4%       123[       76.8%]         19.4%       122[       76.8%]         11.0%       124[       76.8%]         12.7%       125[       80.0%]         19.4%       126[       71.6%]         21.3%       127[       88.0%]         1.0%       126[       71.6%]         10.0%       130[       83.0%]         0.0%       131[       83.0%]         0.0%       132[       91.0%]         0.0%       133[       83.0%]         0.0%       133[       60.6%]         0.0%       133[       60.6%]         0.0%       133[       60.6%]         0.0%       136[       60.6%]         0.0%       138[       69.2%]         0.0%       138[       69.2%]         0.0%       139[       77.6%]         38.1%       144[       73.5%]         0.0%       144[       73.5%]	54.8%]	119[
21.2%       121[       7.6.1%]         7.7%       122[       59.6%]         19.4%       123[       66.2%]         11.0%       124[       66.2%]         12.7%       125[       68.0%]         0.6%       126[       71.6%]         11.0%       124[       68.2%]         31.0%       126[       71.6%]         11.0%       126[       71.6%]         31.0%       126[       71.6%]         31.0%       126[       71.6%]         31.0%       136[       83.8%]         0.0%       131[       84.5%]         0.0%       132[       84.5%]         0.0%       132[       84.5%]         0.0%       133[       60.6%]         0.0%       135[       69.2%]         0.0%       136[       69.2%]         0.0%       136[       69.2%]         0.0%       136[       69.2%]         0.0%       136[       69.2%]         0.0%       136[       69.2%]         0.0%       136[       69.2%]         0.0%       136[       69.2%]         0.0%       136[       69.2%] <th>0.0%]</th> <th>120[                                    </th>	0.0%]	120[
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	21.2%]	121[
19.4%       122[       76.8%]         11.0%]       124[       76.8%]         29.7%]       125[       80.0%]         0.0%]       126[       71.6%]         13.0%]       126[       71.6%]         13.0%]       126[       71.6%]         10.0%]       126[       79.4%]         11.0%]       130[       83.8%]         0.0%]       131[       83.8%]         0.0%]       132[       91.0%]         0.0%]       133[       83.9%]         0.0%]       133[       60.6%]         0.0%]       136[       60.6%]         0.0%]       136[       60.6%]         0.0%]       136[       69.2%]         0.0%]       139[       77.5%]         0.0%]       139[       77.5%]         0.0%]       142[       73.5%]         0.0%]       144[       74.5%]	7.7%]	122[
11.0%       124[       66.2%]         29.7%       125[       71.6%]         12.0%       126[       71.6%]         13.0%       126[       71.6%]         11.0%       126[       71.6%]         11.0%       126[       71.6%]         11.0%       126[       71.6%]         11.0%       126[       71.6%]         11.0%       136[       83.8%]         0.0%       131[       81.5%]         0.0%       132[       84.5%]         0.0%       133[       83.9%]         0.0%       135[       91.0%]         0.0%       135[       91.0%]         0.0%       135[       60.6%]         0.0%       136[       69.2%]         0.0%       136[       69.2%]         0.0%       136[       72.9%]         0.0%       136[       72.9%]         0.0%       136[       72.9%]         0.0%       136[       72.9%]         0.0%       144[       73.5%]         0.0%       144[       73.5%]	19.4%]	123[
29.7%       125[         0.0%       126[         21.3%       127[         31.0%       128[         31.0%       128[         31.0%       128[         31.0%       128[         31.0%       128[         31.1%       68.2%]         0.0%       131[         0.0%       132[         0.0%       132[         0.0%       134[         0.0%       135[         0.0%       135[         0.0%       136[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       140[         0.0%       140[         0.0%       140[         0.0%       144[ <th>11.0%]</th> <th>124[                                    </th>	11.0%]	124[
0.0%1       126[       71.6%]         21.3%1       127[       68.2%]         31.0%1       28[       88.0%]         44.8%1       129[       79.4%]         11.0%1       30[       83.8%]         0.0%1       133[       83.3%]         0.0%1       133[       84.5%]         0.0%1       133[       83.9%]         0.0%1       136[       66.6%]         0.0%1       136[       66.6%]         0.0%1       136[       66.6%]         0.0%1       136[       66.6%]         0.0%1       136[       66.6%]         0.0%1       137[       77.5%]         0.0%1       136[       77.5%]         0.0%1       142[       77.5%]         0.0%1       144[       84.5%]	29.7%]	125[
21.3%       127[       68.2%]         31.0%       128[       88.0%]         44.8%       129[       88.0%]         11.0%       130[       88.3%]         0.0%       131[       83.5%]         0.0%       132[       83.5%]         0.0%       133[       91.0%]         0.0%       135[       91.0%]         13.1       60.6%]       91.0%]         0.0%       135[       60.6%]         0.0%       136[       69.2%]         0.0%       136[       69.2%]         0.0%       136[       69.2%]         0.0%       136[       72.5%]         0.0%       136[       72.5%]         0.0%       136[       72.5%]         0.0%       136[       69.2%]         0.0%       141[       73.5%]         0.0%       144[       73.5%]         0.0%]       144[       84.5%]	0.0%]	126[
31.0%       128[         44.8%       129[         11.0%       130[         11.0%       130[         0.0%       131[         0.0%       132[         0.0%       132[         0.0%       132[         0.0%       132[         0.0%       136[         0.0%       136[         0.0%       136[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       137[         0.0%       140[         77.5%       140[         73.5%       140[         73.5%       144[         0.0%       144[	21.3%]	127[
44.8%       129[       79.4%         11.0%       130[       83.8%]         0.0%i       131[       83.8%]         0.0%i       131[       87.7%]         0.0%i       133[       91.0%]         0.0%i       133[       91.0%]         0.0%i       136[       60.6%]         0.0%i       136[       60.6%]         0.0%i       137[       87.7%]         0.0%i       139[       77.7%]         0.0%i       142[       73.5%]         0.0%i       144[       84.5%]	31.0%]	128[
11.0%       130[       83.0%]         0.0%       131[       87.7%]         0.0%       132[       83.5%]         0.0%       132[       83.5%]         0.0%       132[       91.0%]         135[       91.0%]       91.0%]         0.0%       135[       91.0%]         0.0%       135[       91.0%]         0.0%       135[       91.0%]         0.0%       136[       60.6%]         0.0%       138[       69.2%]         0.0%       138[       69.2%]         0.0%       136[       72.5%]         0.0%       141[       73.5%]         0.0%       142[       77.5%]         38.1%       43[       90.4%]	44.8%]	129[
0.0%       131[       87.7%]         0.0%       132[       84.5%]         0.0%       133[       83.9%]         30.3%       134[       91.0%]         0.0%       135[       91.0%]         0.0%       135[       91.0%]         0.0%       136[       91.0%]         0.0%       137[       91.0%]         0.0%       137[       91.0%]         0.0%       137[       91.0%]         0.0%       137[       92.0%]         0.0%       140[       72.9%]         4.5%       144[       73.5%]         0.0%       142[       90.4%]         0.0%       144[       84.5%]	11.0%]	130[
0.0%1       132[       84.5%1         0.0%1       133[       83.9%1         0.0%1       134[       91.0%1         0.0%1       135[       91.0%1         0.0%1       136[       60.6%1         0.0%1       137[       60.6%1         0.0%1       137[       87.1%1         0.0%1       138[       69.2%1         0.0%1       139[       77.5%1         0.0%1       142[       77.5%1         0.0%1       144[       84.5%]	0.0%]	131[
0.0%       133[       134[       91.0%]         30.3%       134[       91.0%]       91.0%]         0.0%       135[       91.0%]       91.0%]         41.3%       136[       60.6%]       60.6%]         0.0%       137[       60.0%]       60.2%]         0.0%       139[       60.0%]       69.2%]         0.0%       139[       72.9%]         4.5%       141[       73.5%]         0.0%       142[       77.6%]         38.1%       143[       90.4%]         9.0%       144[       84.5%]	0.0%]	132[
30.3%       134[       91.0%]         0.0%]       135[       91.0%]         41.3%       136[       91.0%]         0.0%]       135[       91.0%]         0.0%]       137[       91.0%]         0.0%]       137[       91.0%]         0.0%]       137[       91.0%]         0.0%]       136[       91.0%]         0.0%]       140[       72.9%]         4.5%]       141[       73.5%]         0.0%]       142[       91.0%]         0.0%]       144[       91.0%]	0.0%]	133[
0.0%1       135[       (1) <t< th=""><th>30.3%]</th><th>134[                                    </th></t<>	30.3%]	134[
41.3%       136[       60.6%]         0.0%       137[       87.1%]         28.6%       138[       69.2%]         0.0%       139[       87.7%]         0.0%       140[       72.9%]         4.5%]       141[       73.5%]         0.0%       142[       77.6%]         38.1%       143[       90.4%]         0.0%       144[       84.5%]	0.0%]	135[
0.0%       137[       87.1%]         28.6%       138[       69.2%]         0.0%       139[       69.2%]         0.0%       140[       72.9%]         4.5%       141[       73.5%]         0.0%       142[       77.6%]         38.1%       143[       90.4%]         0.0%       144[       84.5%]	41.3%]	136[
28.6%       138[       69.2%]         0.0%       139[       87.7%]         0.0%1       140[       72.9%]         4.5%1       141[       73.5%]         0.0%1       142[       77.6%]         38.1%       143[       90.4%]         0.0%1       144[       90.4%]         8.1%1       144[       84.5%]	0.0%]	137[
0.0%       139[       77%]         0.0%       140[       72.9%]         4.5%]       141[       73.5%]         0.0%       142[       77.5%]         38.1%       143[       90.4%]         0.0%       144[       84.5%]	28.6%]	138[
0.0%         140[         72.9%]           4.5%         141[         73.5%]           0.0%         142[         77.6%]           38.1%         143[         90.4%]           0.0%         144[         84.5%]	0.0%]	139[
4.5%         141[         73.5%]           0.0%         142[         77.6%]           38.1%         143[         90.4%]           0.0%         144[         84.5%]	0.0%]	140[
0.0%         142[         77.6%]           38.1%         143[         90.4%]           0.0%         144[         91.11	4.5%]	141[
38.1%         143[         98.4%]           0.0%         144[         84.5%]	0.0%]	142[
0.0%] 144[	38.1%]	143[
	0.0%]	144[

# Neo4j Graph Platform with Neo4j Algorithms vs. Apache Spark's GraphX



#### Neo4j provides same order of magnitude performace

#### Twitter 2010 Dataset

- 1.47 Billion Relationships
- 41.65 Million Nodes

#### Spark GraphX results publicly available

- Amazon EC2 cluster running 64-bit Linux
- 128 CPUs with 68 GB of memory, 2 hard disks

#### Neo4j Configuration

- Physical machine running 64-bit Linux
- 128 CPUs with 55 GB RAM, SSDs

# **Compute At Scale – Payment Graph**

3,000,000,000 nodes and 18,000,000,000 relationships (600G) PageRank (20 iterations) on 1 machine, 20 threads, 700G RAM

call algo.pageRank('Account','SENT',{graph:'big', iterations:20,write:false});
+-----+
| nodes | iterations | loadMillis | computeMillis |
+-----+
| 3000000096 | 20 | 0 | 9845756 |
+-----+
1 row
9845794 ms -> 2h 44m

# **Evaluation**

Graph		#Nodes [M]	#Relationships [M]	Avg. out degree	Disk size [GB]
Pokec	(PK)	1.63	30.62	18.75	0.99
cit-patents	(CP)	3.77	16.52	4.38	0.58
Graphs500-23	(G5)	4.61	129.33	28.05	4.17
soc-LifeJournal1	(LJ)	4.85	68.99	14.23	2.27
DBPedia	(DP)	11.47	116.60	10.16	3.87
Twitter-2010	(TW)	41.65	1468.37	35.25	47.60
Friendster	(FR)	65.61	1806.07	27.53	58.94

Tab. 2: Graph datasets used in measurements.

# **Evaluation**



# **Twitter Troll Analysis**





SECTIONS 🗸

NIGHTLY NEWS MSNBC MEET THE PRESS DATELINE C TODAY

Q

TECH > SOCIAL MEDIA

GADGETS INTERNET SECURITY INNOVATION MOBILE

TECH DEC 20 2017, 11:11 AM ET

# Russian trolls went on attack during key election moments

by BEN POPKEN

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



TECH > SOCIAL MEDIA

GADGETS INTERNET SECURITY INNOVATION MOBILE

TECH DEC 20 2017, 11:11 AM ET

# Russian trolls went on attack during key election moments

by BEN POPKEN



Ben Popken 🤡 @bpopken



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.

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



http://www.lyonwj.com/2017/11/12/scraping-russian-twitter-trolls-python-neo4j/

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



# Your typical American Citizen?



@LeroyLovesUSA

#### **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 News Publication?

## Your typical Local Political Party?



Tennessee @TEN\_GOP

Unofficial Twitter of Tennessee Republicans. Covering breaking news, national politics, foreign policy and more. #MAGA #2A

@TEN\_GOP

## Your typical Russian Troll



#### @LeroyLovesUSA



### Your typical Russian Troll

@ClevelandOnline

## Your typical Russian Troll



@TEN\_GOP

### **IRA - Internet Research Agency**

n	р	r	

4

9

 $\sim$ 

**V DONATE Q SIGN IN** SHOP

III NEWS X ARTS & LIFE J MUSIC  $\Omega$  SHOWS & PODCASTS **Q** SEARCH

NATIONAL SECURITY

The Russia Investigations: Mueller Indicts The 'Internet Research Agency'

February 17, 2018 · 7:00 AM ET

#### PHILIP EWING

This week in the Russia investigations: A major new indictment from the special counsel's office that charges thirteen individuals and three companies and shakes up the political rhetoric as new facts are revealed in the sprawling imbroglio.

Justice Department special counsel Robert Mueller prefers to let his work do the talking for him. On Friday, he delivered a stemwinder.

Thirteen Russians and three Russian entities were indicted by a federal grand jury in connection with the attack on the 2016 election. The indictment lays out a number of detailed allegations against the Internet Research Agency located in St. Petersburg and against individuals who owned, controlled, funded or worked for the organization.



NATIONAL SECURITY Grand Jury Indicts Russians Linked To Interference In 2016 Election

#### **National Security**

U.S. Cyber Command operation disrupted Internet access of Russian troll factory on day of 2018 midterms



The building that housed the Internet Research Agency in St. Petersburg, shown in 2018. (Dmitri Lovetsky/AP)

1 MATCH

2 (u:User {screen\_name: "LeroyLovesUSA"})-[:POSTED]->(t:Tweet)-[:HAS\_TAG]->(ht:Hashtag {key: "thanksobama"})

3 RETURN \*





2 RETURN t.dayOfYear AS day, COUNT(\*) AS num ORDER BY day

### Hashtags

- Use of hashtags to gain visibility and insert into conversation
- @WorldOfHashtags

   #RejectedDebateTopics

https://www.nbcnews.com/tech/social-media/russian-trolls-went-attack-duringkey-election-moments-n827176



#### totalRetweets, totalReplies and originalContent





- 1 MATCH (tr:Troll)-[:POSTED]->(tw:Tweet) WITH tr, tw
- 2 OPTIONAL MATCH (tw)-[:RETWEETED]-(rt:Tweet)
- 3 OPTIONAL MATCH (tw)-[:IN\_REPLY\_TO]-(irp:Tweet)
- 4 RETURN distinct tr.screen\_name as screen\_name, count(tw) as totalTweets,
- 5 count(rt) as totalRetweets, count(irp) as totalReplies,
- 6 (count(tw) (count(rt) + count(irp))) as originalContent
- 7 ORDER BY totalTweets DESC;

tweets vs. hour of day



# **Inferred Relationships**

1 MATCH (r1:Troll)-[:POSTED]->(t1:Tweet)<-[:RETWEETED]-(t2:Tweet)<-[:POSTED]-(r2:Troll)</pre>



# **Inferred Relationships**



MATCH (r1:Troll)-[:POSTED]->(:Tweet)
 <-[:RETWEETED]-(:Tweet)<-[:POSTED]-(r2:Troll)
WITH r1,r2, count(\*) as freq where freq > 5
RETURN r1,r2, apoc.create.vRelationship(r1, 'AMPLIFIED', {freq:freq}, r2) as rel

# **Inferred Relationships**



CREATE (r2)-[:AMPLIFIED {weight:freq}]->(r1)

# Weighted In-Degree Centrality

match (t:Troll)<-[r:AMPLIFIED]-() with t, sum(r.weight) as total return t.screen\_name, total order by total desc limit 5

]	"t.screen_name"	"total"
	"TEN_GOP"	239
	"NotRitaHart"	107
]	"GiselleEvns"	104
e	"tpartynews"	98
	"DaileyJadon"	84

# PageRank on Inferred AMPLIFIED Graph

# PageRank on Inferred AMPLIFIED Graph

match (t:Troll) where exists (t.pagerank) return t.screen\_name, t.pagerank order by t.pagerank desc limit 5

"t.screen_name"	"t.pagerank"
"TEN_GOP"	10.3859635
「  "TheFoundingSon"	8.28164400000002
"GiselleEvns"	6.4624315
"tpartynews"	6.3289815
"ChrixMorgan"	4.231436500000001

# **Graph Visualization**

Based on metrics computed by graph algorithms



### **Graph Visualization**

Centrality & community detection AMPLIFIED relationships

Node size  $\rightarrow$  PageRank Color  $\rightarrow$  community detection Rel Thickness  $\rightarrow$  weight



# **Graph Visualization**

#### neovis.js

Graph visualizations powered by vis.js with data from Neo4j.



#### var config = { container\_id: "viz", server\_url: "bolt://localhost:7687", server\_user: "neo4j", server\_password: "sorts-swims-burglaries", labels: { //"Character": "name", "Character": { "caption": "name", "size": "pagerank", "community": "community" //"sizeCypher": "MATCH (n) WHERE $id(n) = \{id\}$ MATCH (n)-[r]-() RETURN sum(r.weight) }, relationships: { "INTERACTS": { "thickness": "weight", "caption": false } }, initial\_cypher: "MATCH (n)-[r:INTERACTS]->(m) RETURN n,r,m" }; viz = new NeoVis.default(config); viz.render();

#### https://github.com/neo4j-contrib/neovis.js

# 2 days later - IRA taken to court and indicted

### Feb 14:

#### **JE NEWS**

Twitter Deleted 200,000 Russian Troll Tweets. Read Them Here.

#### GET THE DATA:

- Regular reader? Download streamlined spreadsheet (29 mb) with just usernames, tweet and timestamps. We recommend you right click on links and select "save link as" or similar, otherwise it may take a long time to load in your browser.
- ► View full data for ten influential accounts in Google Sheets
- Researcher? Download tweets.csv (50 mb) and users.csv with full underlying data
- Explore a graph database in Neo4j

### Feb 16:

#### IN THE UNITED STATES DISTRICT COURT FOR THE DISTRICT OF COLUMBIA

UNITED STATES OF AMERICA	*	CRIMINAL NO.
v.	*	
INTERNET DESEARCH AGENOVIIC	*	(18 U.S.C. §§ 2, 371, 1349, 1028A)
INTERNET RESEARCH AGENCY LLC	*	
A/K/A MEDIASINTEZ LLC A/K/A		
GLAVSET LLC A/K/A MIXINFU	*	
NOVINEO LLC	*	
CONCORD MANACEMENT AND	*	
CONCORD MANAGEMENT AND	*	
CONSULTING LLC,	*	
VENCENIN VINTOROVICH	*	
PRICOZUNI	*	
MINUAL IVANOVICH DVSTDOV	*	
MIKHAIL IVANOVICH BISTROV,	*	
A/K/A MIKHAIL ADD AMOV	*	
ALEKSANDRA VURVEVNA	*	
KENIOVA	*	
ANNA VI ADISI AVOVNA	*	
BOGACHEVA	*	
SERGEV PAVI OVICH POLOZOV	*	
MARIA ANATOI VEVNA BOVDA	*	
A/K/A MARIA ANATOI VEVNA	*	
BELVAEVA	*	
ROBERT SERGEVEVICH BOVDA	*	
DZHEVKHUN NASIMLOGUV	*	
ASLANOV A/K/A JAYHOON	*	
ASLANOV A/K/A JAY ASLANOV	*	
VADIM VI ADIMIROVICH	*	
PODKOPAEV	*	
GLEB IGOREVICH VASILCHENKO	*	
IRINA VIKTOROVNA KAVERZINA	*	
and	*	
VLADIMIR VENKOV	******	

https://www.nbcnews.com/tech/social-media/now-available-more-200-000-deleted-russian-troll-tweets-n844731

# **Surprising Takeaways**

- Amplifying w/ retweets
- Used social media automation tools
  - Not necessarily live responses
- Meddling in elections is just another 9-5 job
- Data availability
- See <u>lyonwj.com</u> for code, etc.
- https://www.nbcnews.com/tech/social-media/russian-trolls-went-attack-during-key-election-mome nts-n827176

## neo4jsandbox.com





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

# Thank You

# **Questions?**