Hej!

@ryguyrg

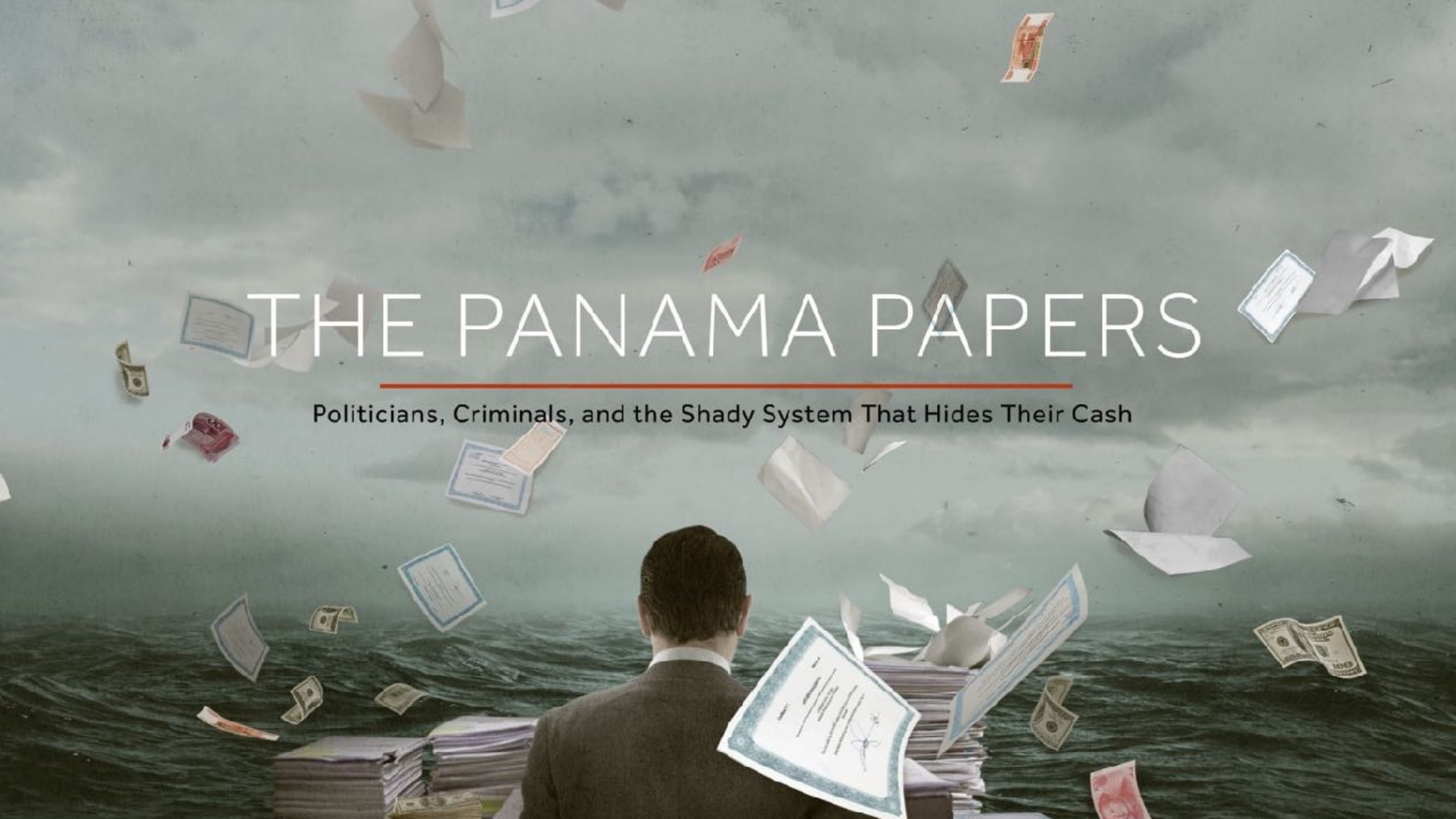
neo4j

ABOUT ME

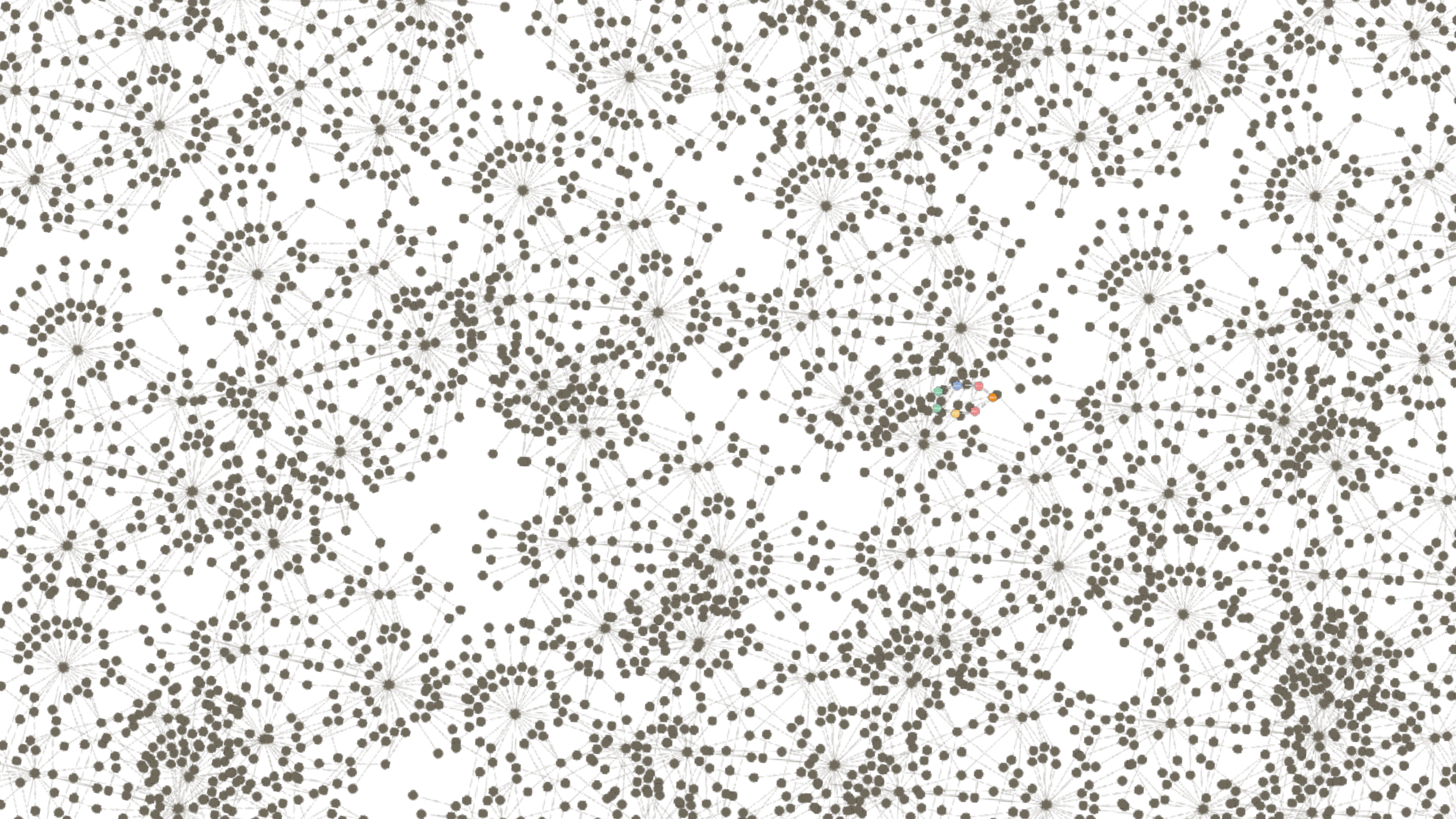
- Developed web apps for 5 years including e-commerce, business workflow, more.
- Worked at Google for 8 years on Google Apps, Cloud Platform
- Technologies: Python, Java, BigQuery, Oracle, MySQL, OAuth

ryan@neo4j.com
@ryguyrg

Carpe ~~Diem~~
Data


THE PANAMA PAPERS
Politicians, Criminals, and the Shady System That Hides Their Cash


TAX RESIDENCE CERTIFICATE
APPLEBY
Certificate of Incorporation



Why are YOU here today, hopefully

Power of Graph Algorithms to Understand Your Data

~~Power of Graph Algorithms to Understand Your Data~~

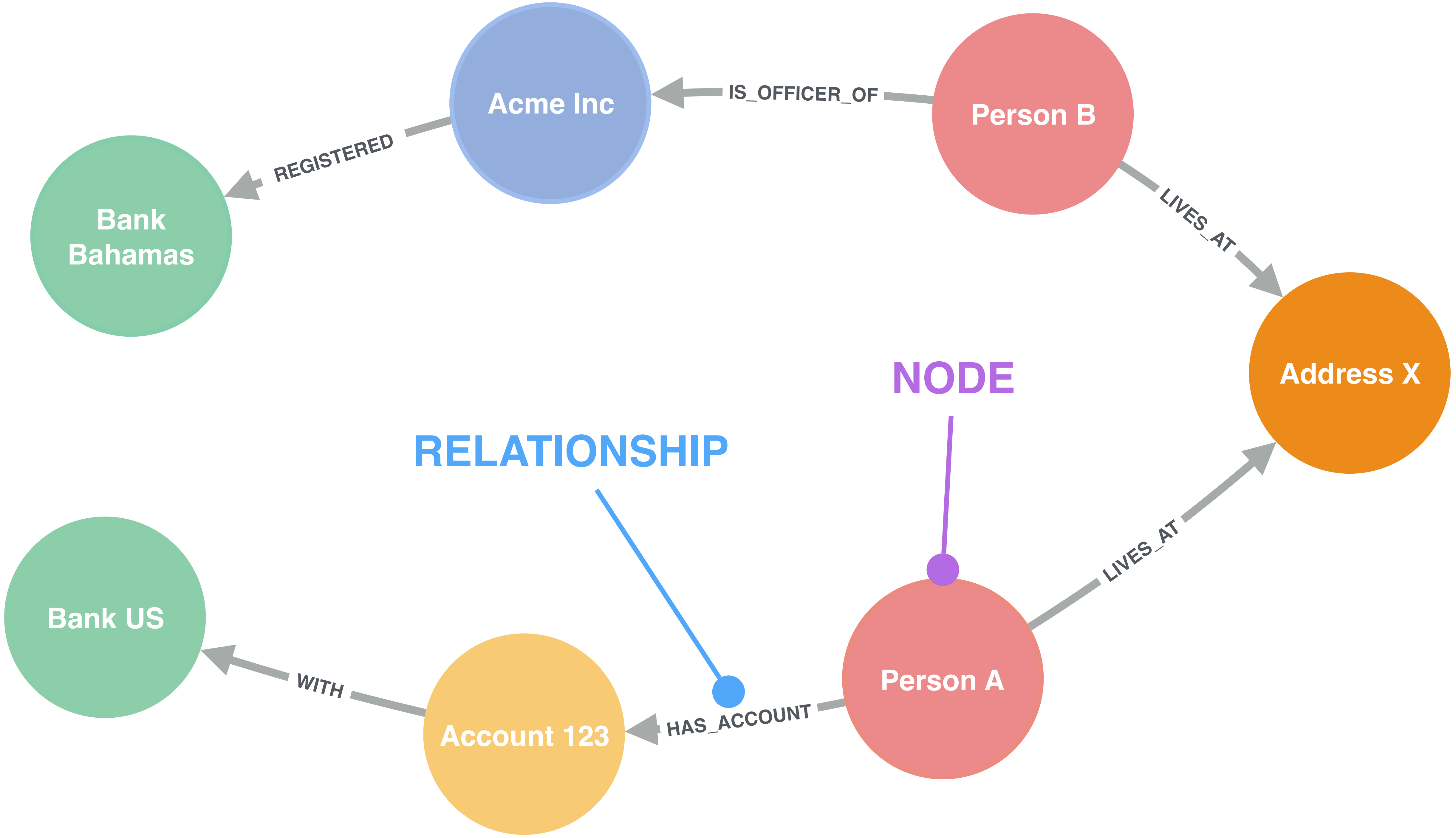
Graph Algorithms on ACID

Graph Algorithms on ACID

Graph Algorithms

+

ACID-compliant
native graph database


Acme Inc
IS_OFFICER_OF
Person B
REGISTERED
Bank Bahamas
LIVES_AT
Address X
NODE
RELATIONSHIP
LIVES_AT
Bank US
WITH
Account 123
HAS_ACCOUNT
Person A
neo4j

graph algorithms

All Videos Books Images Shopping More Settings Tools

About 1,020,000 results (0.32 seconds)

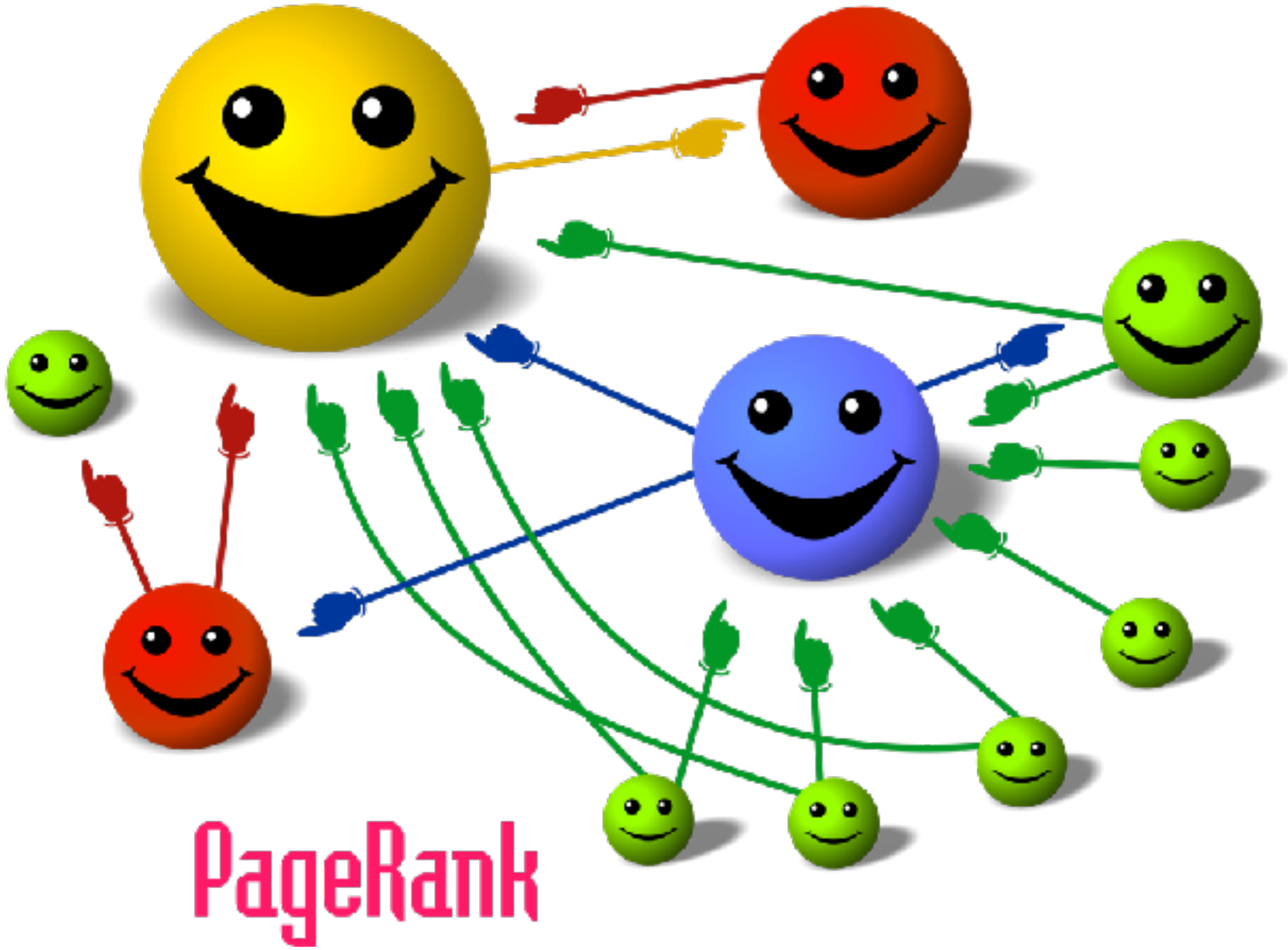
About this result • Feedback

[PDF] Basic Graph Algorithms - Stanford University
https://web.stanford.edu/class/cs97si/06-basic-graph-algorithms.pdf
Jun 29, 2015 - Graphs. Adjacency Matrix and Adjacency List. Special Graphs. Depth-First and .... The most **basic graph algorithm** that visits nodes of a graph.

Category:Graph algorithms - Wikipedia
https://en.wikipedia.org/wiki/Category:Graph_algorithms
T. Tarjan's off-line lowest common ancestors algorithm. Tarjan's strongly connected components algorithm. Theta* **Topological** sorting. **Transitive closure**. **Transitive** reduction. **Travelling salesman problem. Tree traversal.**

[PDF] Graph Algorithms - users.cs.umn.edu
www-users.cs.umn.edu/~karypis/parbook/Lectures/AG/chap10_slides.pdf
In a **weighted** graph, the weight of a **subgraph** is the **sum** of the weights of the edges in the **subgraph**. A minimum spanning tree (MST) for a **weighted** undirected graph is a spanning tree with minimum weight. An undirected graph and its minimum spanning tree. Prim's algorithm for finding an MST is a greedy algorithm.

Graphs - Algorithms, 4th Edition
algs4.cs.princeton.edu/40graphs/
4.1 Undirected **Graphs** introduces the **graph** data type, including depth-first ... and two classic **algorithms** for solving it: Dijkstra's **algorithm** and Bellman-Ford.


PageRank

Restructuring Transactional Data for Link Analysis in the FinCEN AI System*

Henry G. Goldberg** and Raphael W.H. Wong

U.S. Department of the Treasury, Financial Crimes Enforcement Network (FinCEN),
2070 Chain Bridge Road, Vienna VA 22182
**Current address: National Association of Securities Dealers (NASD) Regulation, Inc.,
9513 Key West Avenue, Rockville MD 20850
goldberh@nasd.com, wongr@fincen.treas.gov

Abstract

Due to the nature and costs of data collection, many real-world databases consist of large numbers of independent transactions. Finding evidence of structured groups of entities reflected in this data is a task aptly suited to Link Analysis. However, the databases usually must be restructured to allow effective search and analysis of the linkage structures hidden in the original transactions. The FinCEN AI System (FAIS) [Senator 1995] is an example of such an application. We briefly discuss the process of database restructuring and show how it is used to support the discovery and analysis of evidence of money laundering in a database of cash transactions.

Introduction

Transactional Databases

In our modern world, much of human activity is initially recorded in terms of individual transactions. Both the

that might be used to select out the relevant ones for further analysis. In situations where we are looking for fraud in transactional data, the fraudulent activities are likely to be camouflaged to look like normal activities. The actors involved may spend much of their time in normal, uninteresting, activities and only occasionally in fraudulent ones. [Goldberg 1997] It is clear to many analysts who work with this data that the inter-relationships among transactions hold the key to select the proper subset and to understand the implicit activities hidden in the data -- activities which are incompletely reflected in the recorded transactions. The analysis of these relationships, called Link Analysis, is a vital technique used by law enforcement and intelligence analysts the world over. [Andrews 1990]

Finding Hidden Structure

Statistical and other data mining methods (which often are formulated to model and characterize populations of similar instances) can analyze sets of transactions to show

Anti Money Laundering

Personalized Product Recommendations with Neo4j

Recommendations

Personalized product recommendations can increase conversions, improve sales rates and provide a better experice for users. In this Neo4j Browser guide, we'll take a look at how you can generate graph-based real-time personalized product recommendations using a dataset of movies and movie ratings, but these techniques can be applied to many different types of products or content.

Graph-Based Recommendations

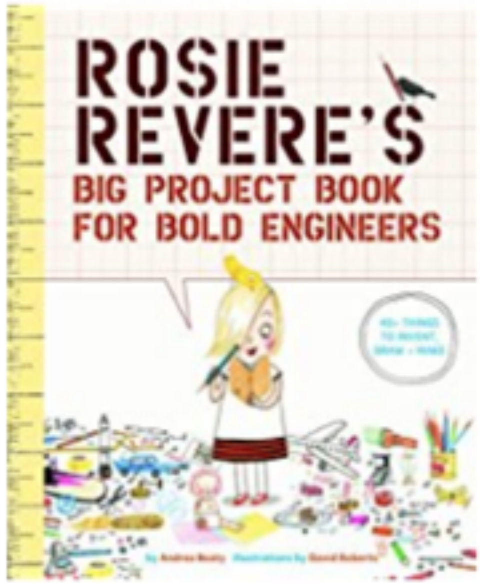
Generating personalized recommendations is one of the most common use cases for a graph database. Some of the main benefits of using graphs to generate recommendations include:

1. **Performance**. Index-free adjacency allows for **calculating recommendations in real time**, ensuring the recommendation is always relevant and reflecting up-to-date information.
2. **Data model**. The labeled propety graph model allows for easily combining datasets from multiple sources, allowing enterprises to **unlock value from previously separated data silos.**

Product Recommendations

Recommended for you

See more recommendations ›



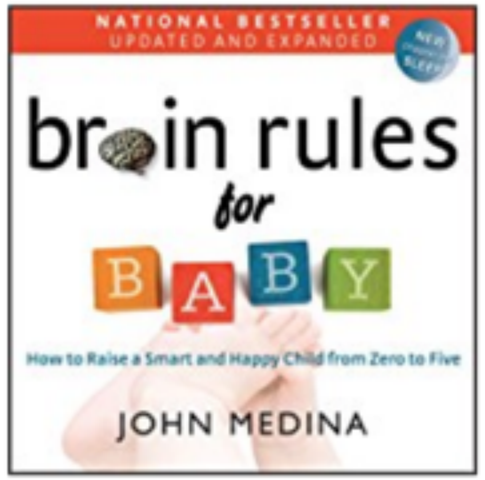
Rosie Revere's Big Project Book for Bold Engineers

Andrea Beaty, David Roberts

Paperback

24

$12.04 ✓prime



Brain Rules for Baby : How to Raise a Smart and...

John Medina, Pear Press

Audiobook

259

$14.95

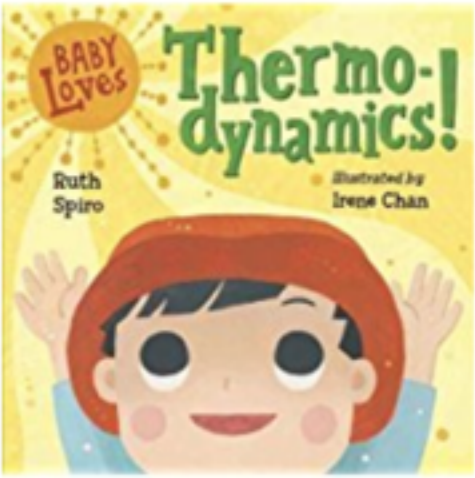
Iggy Peck's Big Project Book for Amazing...

Andrea Beaty, David Roberts

Paperback

4

$10.39 ✓prime



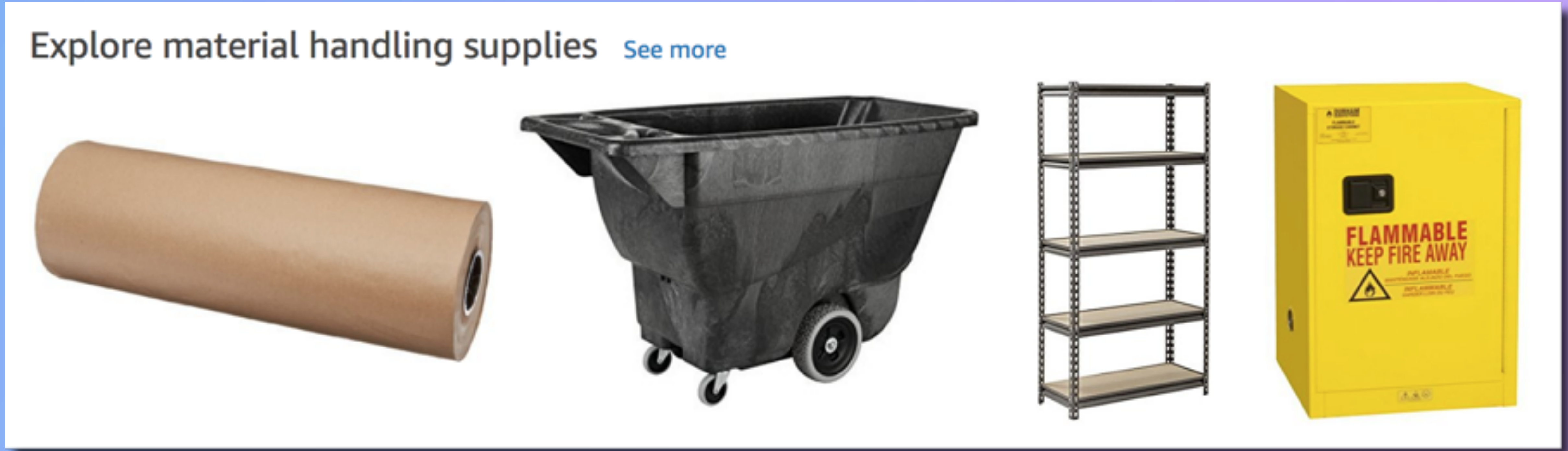
Baby Loves Thermodynamics! (Baby Loves Science)

Ruth Spiro, Irene Chan

Board book

21

$8.09 ✓prime


Explore material handling supplies
See more
FLAMMABLE
KEEP FIRE AWAY

Sports

Who Is the Best Player Ever? A Complex Network Analysis of the History of Professional Tennis

Filippo Radicchi*

Matjaz Perc, Editor

Author information ▶ Article notes ▶ Copyright and License information ▶

Abstract

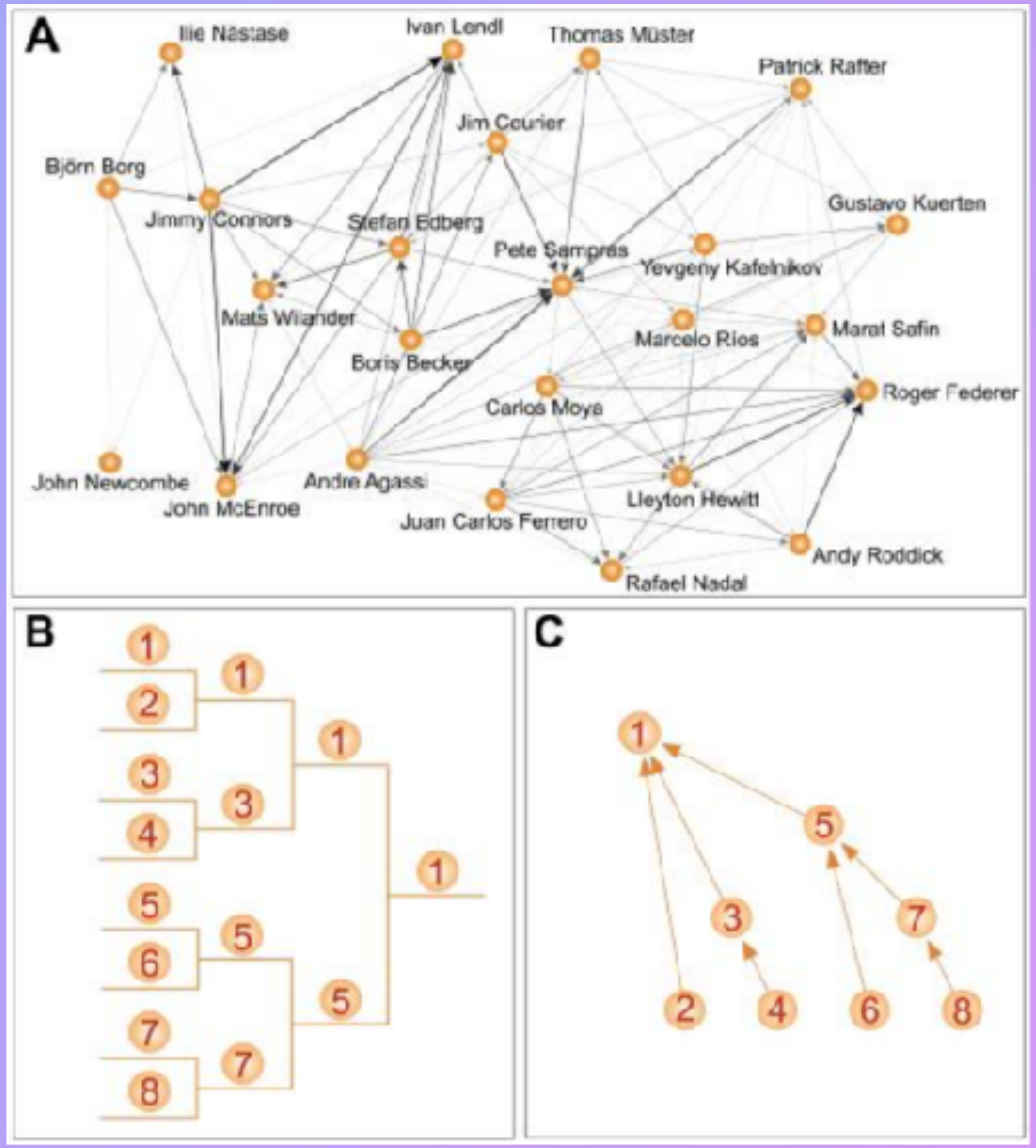
Go to:

We considered all matches played by professional tennis players between 1968 and2010, and, on the basis of this data set, constructed a directed and weighted network of contacts. The resulting graph showed complex features, typical of many real networked systems studied in literature. We developed a diffusion algorithm and applied it to the tennis contact network in order to rank professional players. *Jimmy Connors* was identified as the best player in the history of tennis according to our ranking procedure. We performed a complete analysis by determining the best players on specific playing surfaces as well as the best ones in each of the years covered by the data set. The results of our technique were compared to those of two other well established methods. In general, we observed that our ranking method performed better: it had a higher predictive power and did not require the arbitrary introduction of external criteria for the correct assessment of the quality of players. The present work provides novel evidence of the utility of tools and methods of network theory in real applications.

Introduction

Go to:

Social systems generally display complex features [1]. Complexity is present at the individual level: the behavior of humans often obeys complex dynamical patterns as for example demonstrated by the rules governing electronic correspondence [2]–[5]. At the same time, complexity is present also at the global level. This can be seen for example when social systems are mathematically represented in terms of graphs



Literature

ars TECHNICA

BIZ & IT TECH SCIENCE POLICY CARS GAMING & CULTURE FORUMS

BIZ & IT —

Novel text analysis uses PageRank to identify influential Victorian authors

Searching for the Adam and Eve of 20th century authors.

LIAT CLARK, WIRED.CO.UK - 8/18/2012, 8:47 AM

Urban Planning

Ranking Spaces for Predicting Human Movement in an Urban Environment

Bin Jiang

Department of Land Surveying and Geo-informatics
The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
Email: bin.jiang@polyu.edu.hk

(*February 2008 version 3*)

Abstract

A city can be topologically represented as a connectivity graph, consisting of nodes representing individual spaces and links if the corresponding spaces are intersected. It turns out in the space syntax literature that some defined topological metrics can capture human movement rates in individual spaces. In other words, the topological metrics are significantly correlated to human movement rates, and individual spaces can be ranked by the metrics for predicting human movement. However, this correlation has never been well justified. In this paper, we study the same issue by applying the weighted PageRank algorithm to the connectivity graph or space-space topology for ranking the individual spaces, and find surprisingly that (1) the PageRank scores are better correlated to human movement rates than the space syntax metrics, and (2) the underlying space-space topology demonstrates small world and scale free properties. The findings provide a novel justification as to why space syntax, or topological analysis in general, can be used to predict human movement. We further conjecture that this kind of analysis is no more than predicting a drunkard's walking on a small world and scale free network.

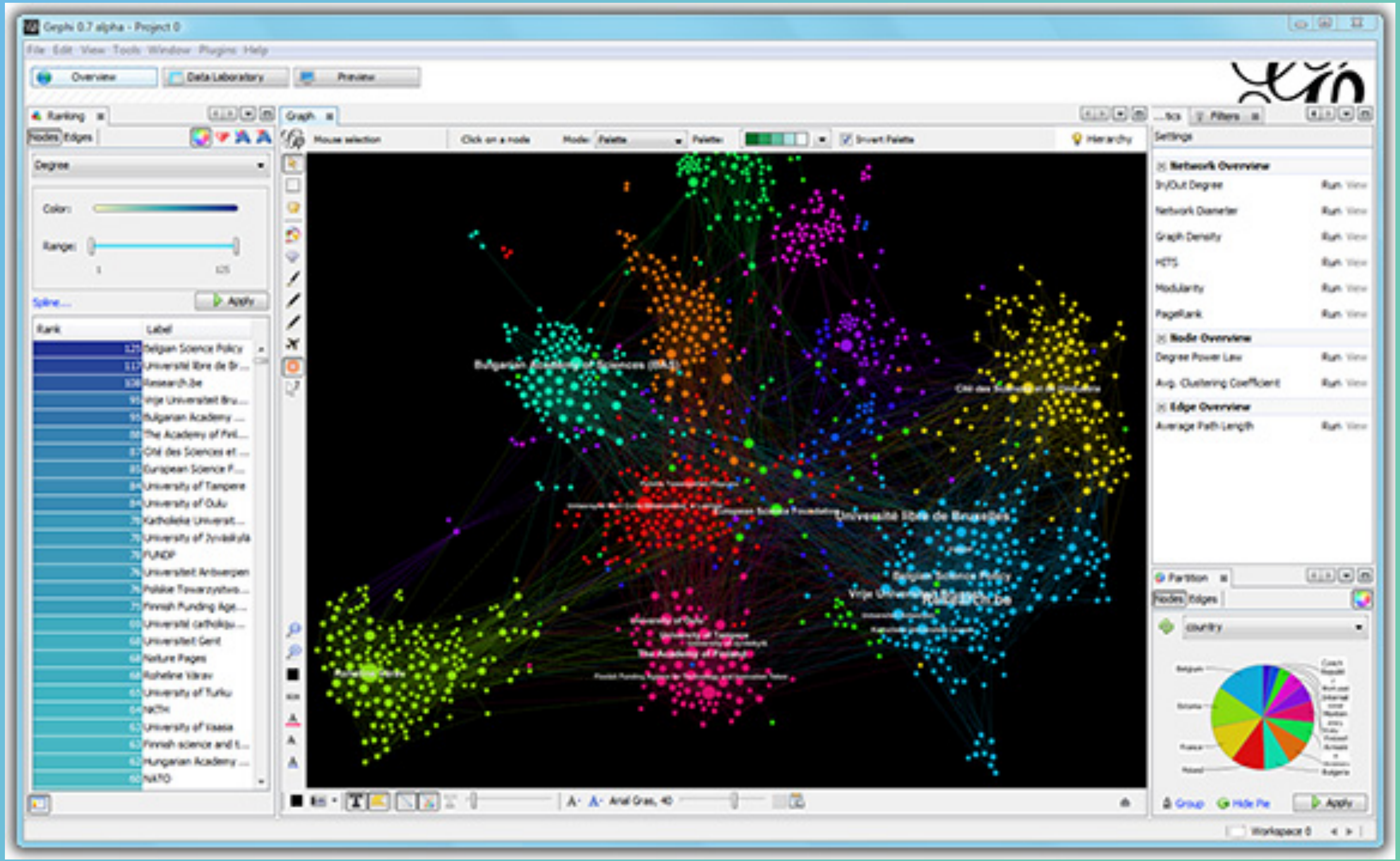
Keywords: Space syntax, topological analysis of networks, small world, scale free, human movement, and PageRank

Toxic Waste Management

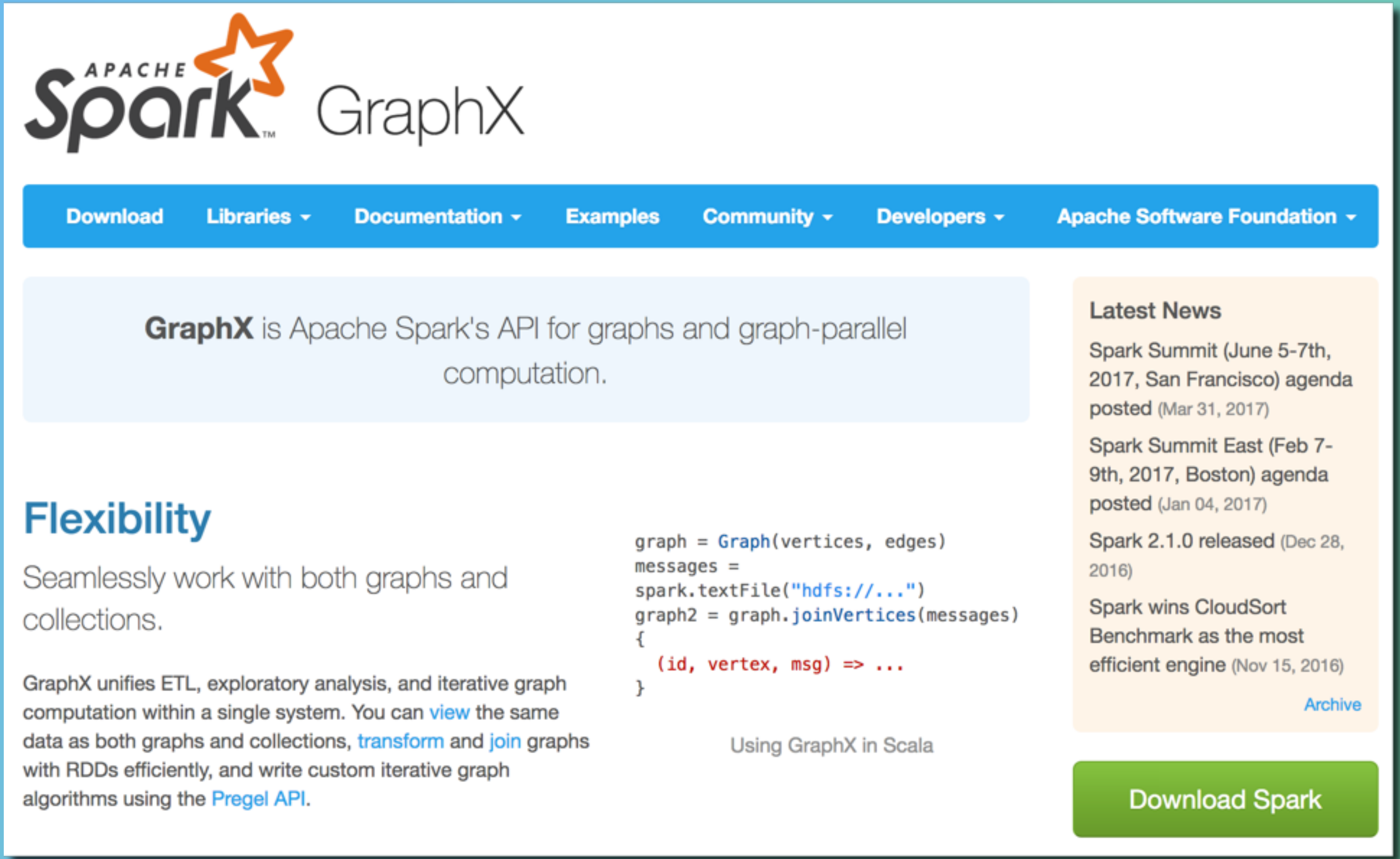
CALEB GARLING BUSINESS 02.16.12 04:20 PM

RESEARCHERS FIGHT TOXIC WASTE WITH GOOGLE PAGERANK

Historical Tooling


Gephi 0.7 alpha - Project 0
File Edit View Tools Window Plugins Help
Overview
Data Laboratory
Preview
Ranking
Nodes Edges
Degree
Color:
Range:
Apply
Rank
Label
Belgian Science Policy
Research.be
University of Tampere
University of Oulu
FUNDP
NATO
Graph
Mouse selection
Click on a node
Mode:
Palette:
Invert Palette
Hierarchy
Bulgarian Academy of Sciences (BAS)
Université libre de Bruxelles
Filters
Settings
Network Overview
In/Out Degree
Network Diameter
Graph Density
HITS
Modularity
PageRank
Node Overview
Degree Power Law
Avg. Clustering Coefficient
Edge Overview
Average Path Length
Run
View
Partition
country
Group
Hide Pie
Arial Gras, 40
Workspace 0


igraph
Products
News
On github
igraph – The network analysis package
igraph is a collection of network analysis tools with the emphasis on **efficiency**, **portability** and ease of use. igraph is **open source** and free. igraph can be programmed in **R**, **Python** and **C/C++**.
igraph R package
python-igraph
igraph C library


APACHE Spark™ GraphX
Download
Libraries
Documentation
Examples
Community
Developers
Apache Software Foundation
GraphX is Apache Spark's API for graphs and graph-parallel computation.
Flexibility
Seamlessly work with both graphs and collections.
GraphX unifies ETL, exploratory analysis, and iterative graph computation within a single system. You can view the same data as both graphs and collections, transform and join graphs with RDDs efficiently, and write custom iterative graph algorithms using the Pregel API.
graph = Graph(vertices, edges)
messages =
spark.textFile("hdfs://...")
graph2 = graph.joinVertices(messages)
{
(id, vertex, msg) => ...
}
Using GraphX in Scala
Latest News
Spark Summit (June 5-7th, 2017, San Francisco) agenda posted (Mar 31, 2017)
Spark Summit East (Feb 7-9th, 2017, Boston) agenda posted (Jan 04, 2017)
Spark 2.1.0 released (Dec 28, 2016)
Spark wins CloudSort Benchmark as the most efficient engine (Nov 15, 2016)
Archive
Download Spark

Graph Analytics
vs Graph Databases
neo4j
3
neo4j
Global
analysis
APACHE GIRAPH
Spark
PageRank, Centrality, etc.
4
neo4j
ACID
transactions
5
neo4j
Real-time
queries
6
neo4j
APACHE GIRAPH
Spark
neo4j

hadoop

hadoop

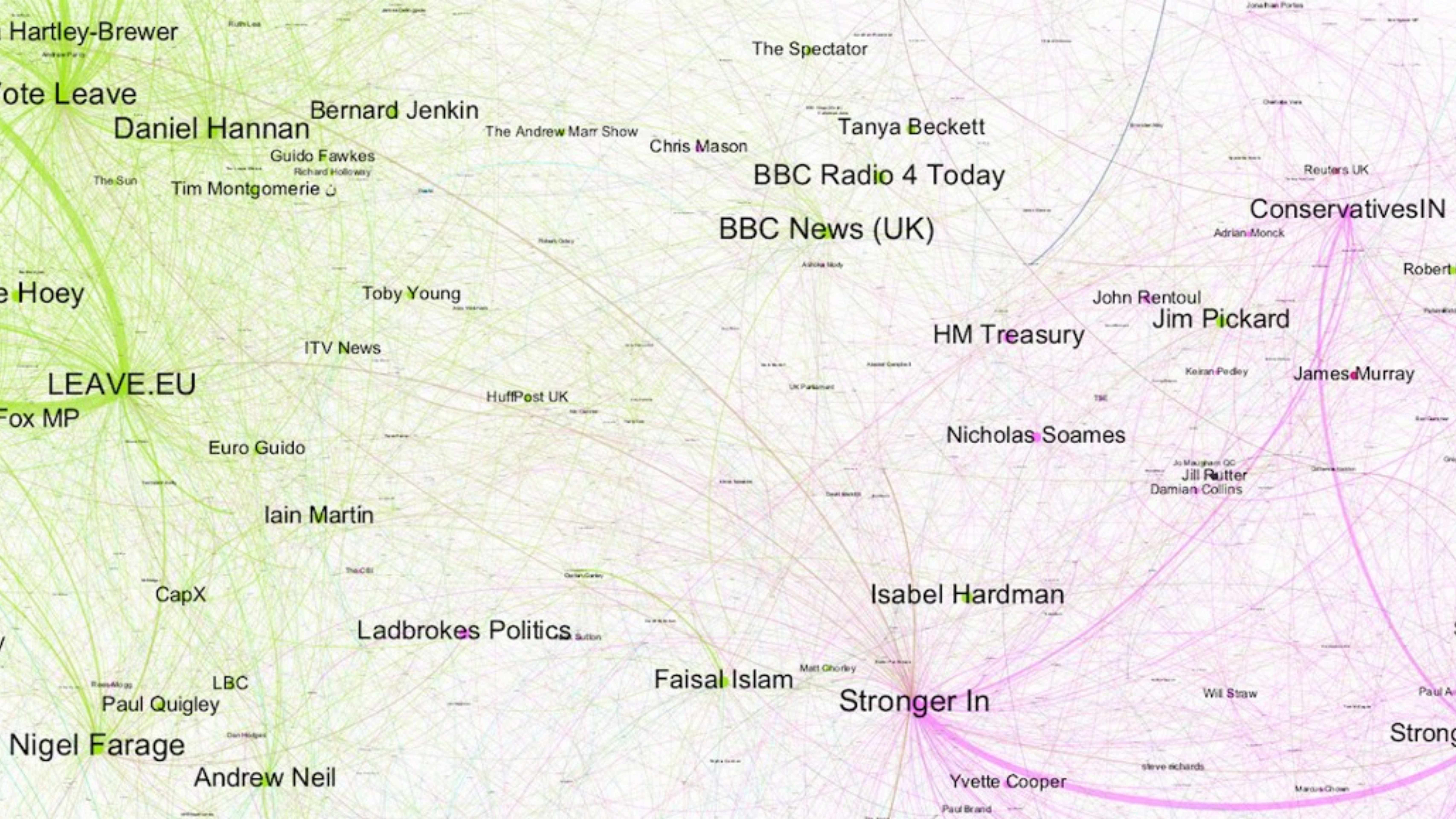
OR

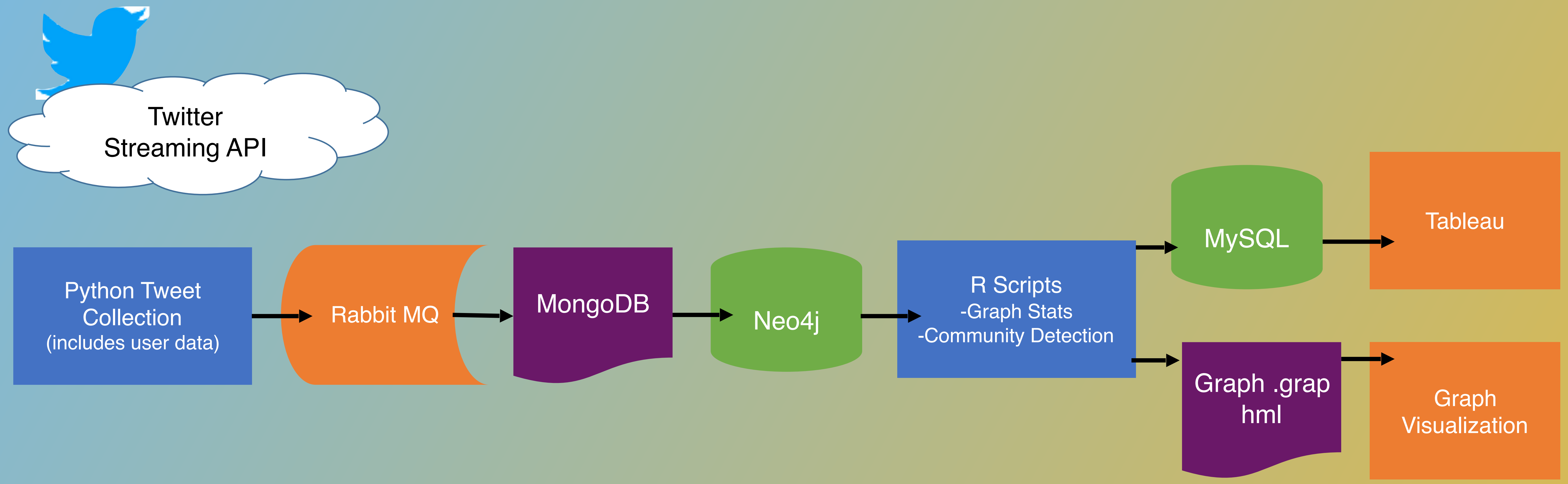
XLS

OR

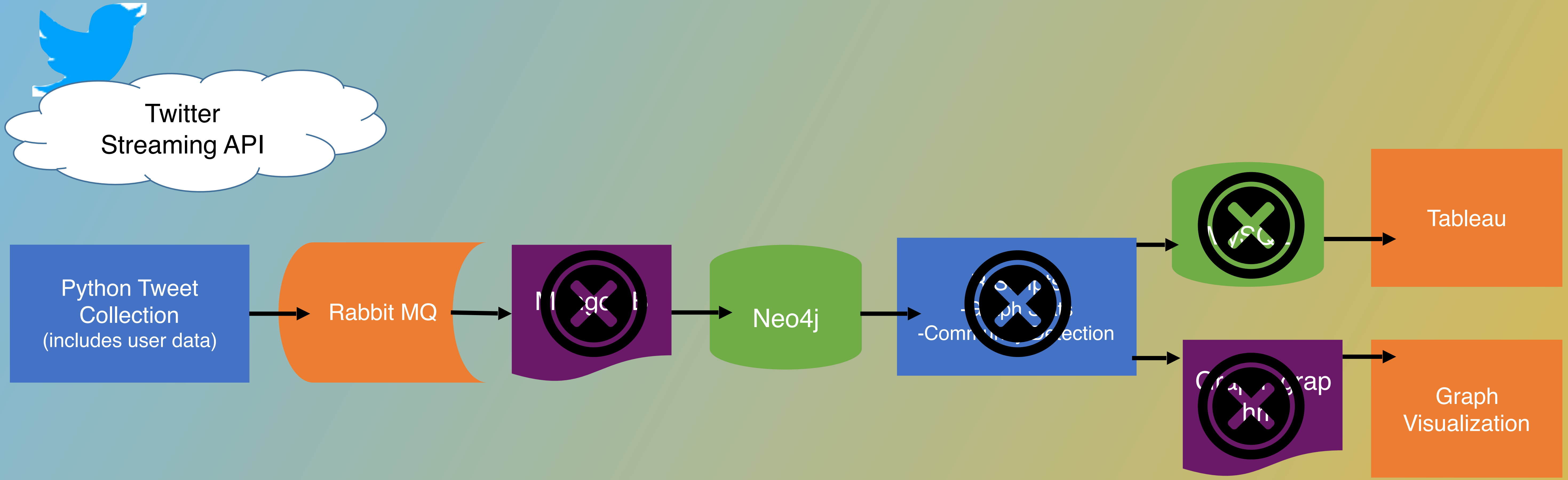
NumPy

The New World


Hartley-Brewer
ote Leave
Daniel Hannan
Bernard Jenkin
The Andrew Marr Show
Chris Mason
The Spectator
Tanya Beckett
BBC Radio 4 Today
BBC News (UK)
ConservativesIN
Reuters UK
Adrian Monck
Guido Fawkes
Richard Holloway
The Sun
Tim Montgomerie
e Hoey
Toby Young
John Rentoul
Jim Pickard
HM Treasury
ITV News
LEAVE.EU
Fox MP
HuffPost UK
Keiran Pedley
James Murray
Nicholas Soames
Euro Guido
Jill Rutter
Damian Collins
Iain Martin
CapX
Isabel Hardman
Ladbrokes Politics
Faisal Islam
Matt Chorley
LBC
Paul Quigley
Stronger In
Will Straw
Nigel Farage
Andrew Neil
Yvette Cooper
steve richards
Paul Brand
Strong


Twitter
Streaming API
Python Tweet
Collection
(includes user data)
Rabbit MQ
MongoDB
Neo4j
R Scripts
-Graph Stats
-Community Detection
MySQL
Tableau
Graph .grap
hml
Graph
Visualization

- Hit a wall with igraph/R
- Need to scale graph algorithms

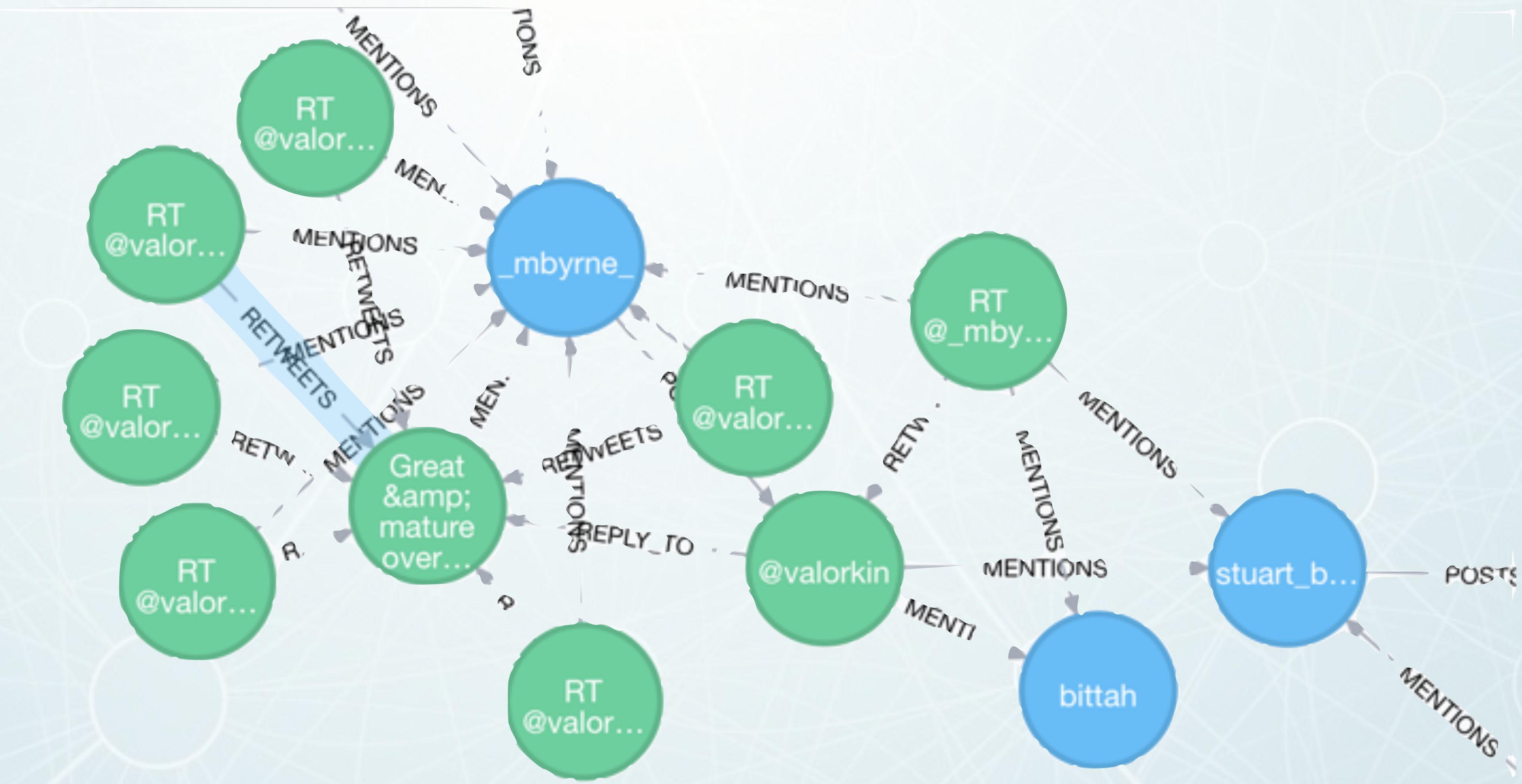

Twitter
Streaming API
Python Tweet
Collection
(includes user data)
Rabbit MQ
Neo4j
Tableau
Graph
Visualization

OPTIMIZED FOR

OLTP

GREAT FOR

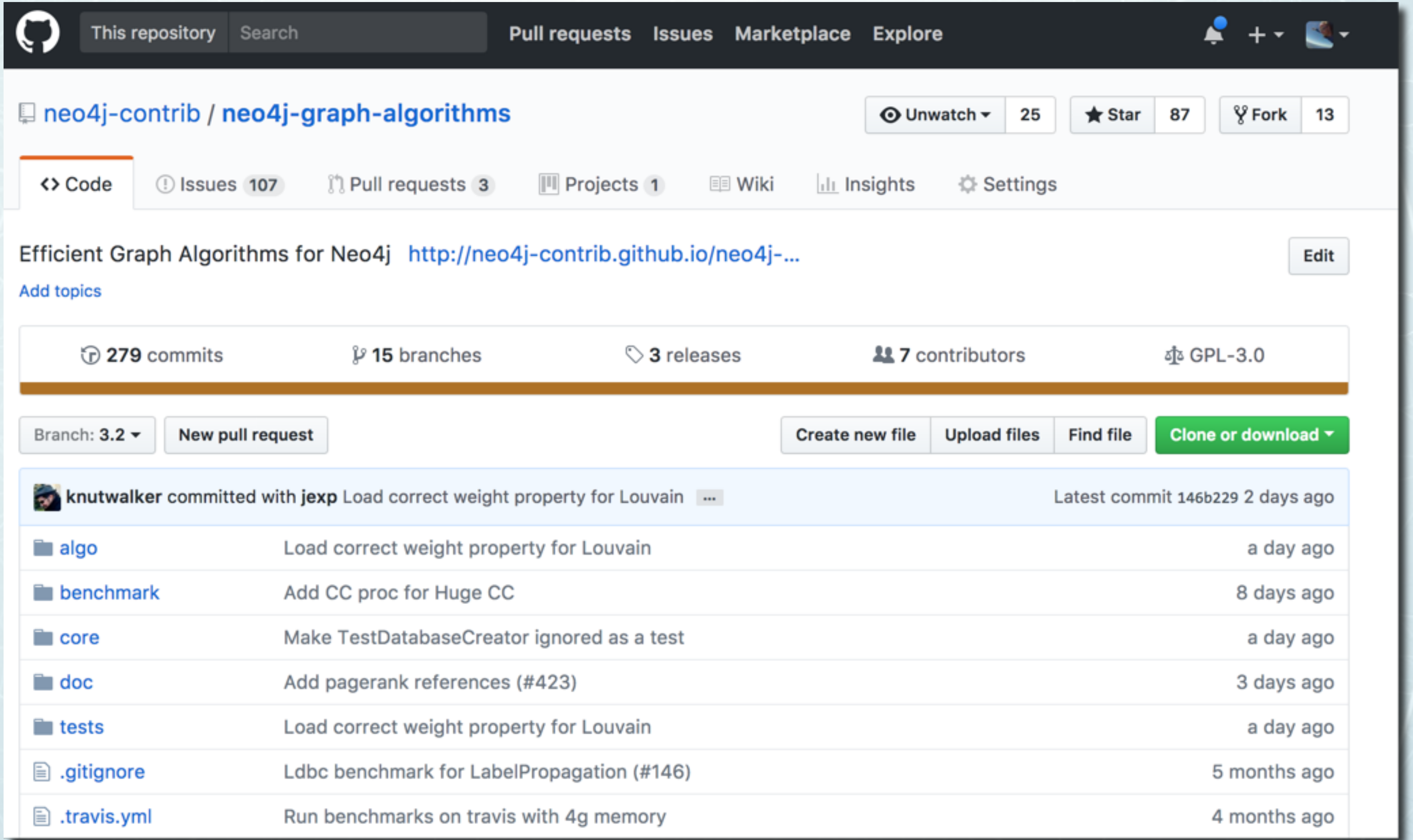
Subgraph Queries



WORKING ON

Global Queries

IN


This repository
Search
Pull requests
Issues
Marketplace
Explore
neo4j-contrib / neo4j-graph-algorithms
Unwatch
25
Star
87
Fork
13
Code
Issues 107
Pull requests 3
Projects 1
Wiki
Insights
Settings
Efficient Graph Algorithms for Neo4j http://neo4j-contrib.github.io/neo4j-...
Edit
Add topics
279 commits
15 branches
3 releases
7 contributors
GPL-3.0
Branch: 3.2
New pull request
Create new file
Upload files
Find file
Clone or download
knutwalker committed with jexp Load correct weight property for Louvain ...
Latest commit 146b229 2 days ago
algo
Load correct weight property for Louvain
a day ago
benchmark
Add CC proc for Huge CC
8 days ago
core
Make TestDatabaseCreator ignored as a test
a day ago
doc
Add pagerank references (#423)
3 days ago
tests
Load correct weight property for Louvain
a day ago
.gitignore
Ldbc benchmark for LabelPropagation (#146)
5 months ago
.travis.yml
Run benchmarks on travis with 4g memory
4 months ago

Neo4j Graph Algorithms


Neo4j
Native Graph
Database
(a)→(b)
Cypher Query
Language
neo4j
ANALYTICS
Analytics
Integrations
</>
Wide Range of
APOC Procedures
Optimized
Graph Algorithms

Finds the optimal path
or evaluates route
availability and quality

Determines the
importance of distinct
nodes in the network

Evaluates how a
group is clustered
or partitioned

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})
```

What about Virtual Graphs?

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'})
```

Supported Centrality Algos

- PageRank (baseline)
- Betweeness
- Closeness
- Degree

Supported Centrality Algos

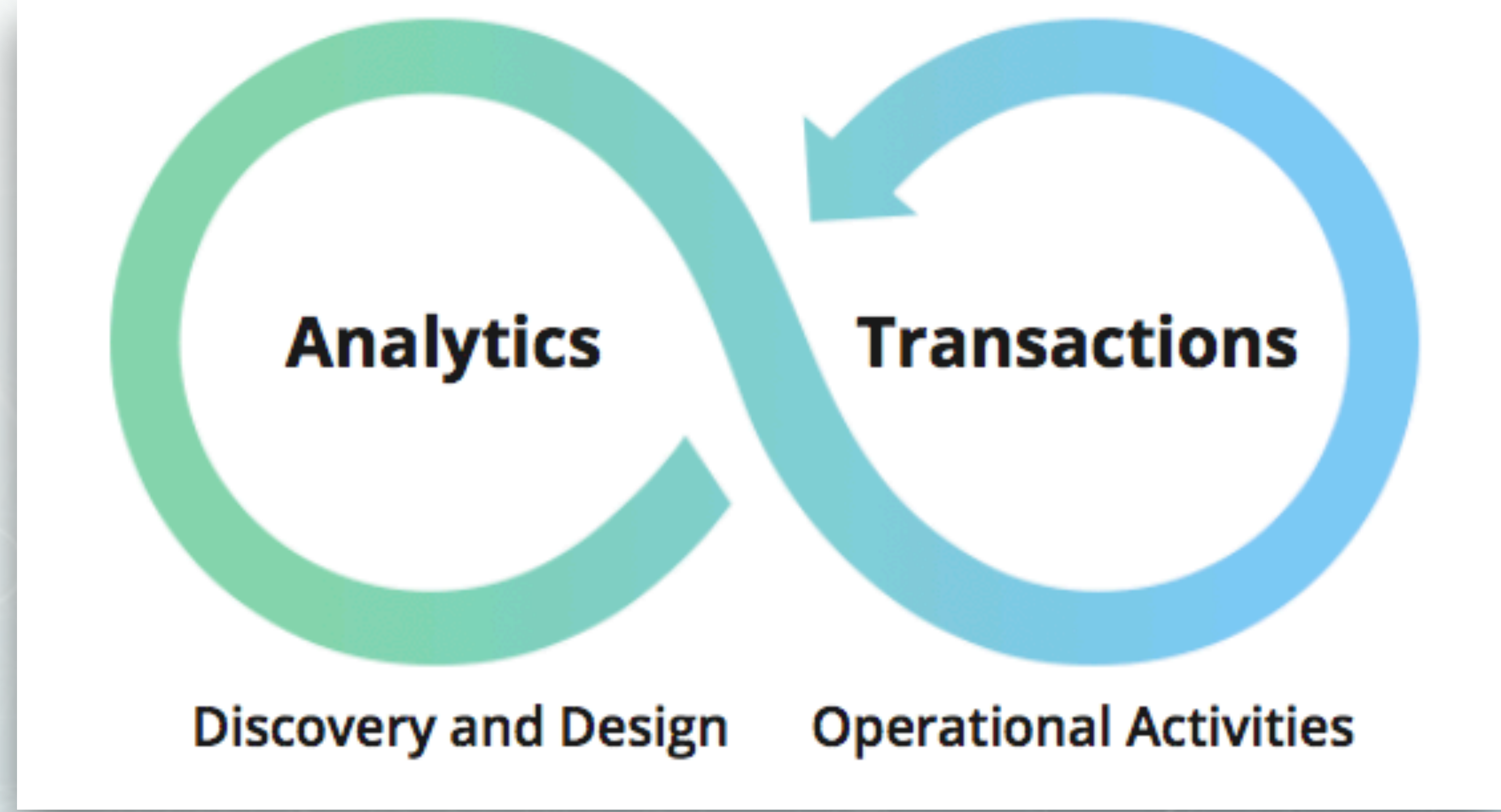
```
CALL algo.pageRank.stream
('Page', 'LINKS', {iterations:20, dampingFactor:0.85})
YIELD node, score
RETURN node, score
ORDER BY score DESC LIMIT 20



CALL algo.pageRank('Page', 'LINKS',
{iterations:20, dampingFactor:0.85,
write: true, writeProperty:"pagerank"})

YIELD nodes, loadMillis, computeMillis, writeMillis
```

Supported Pathfinding Algos

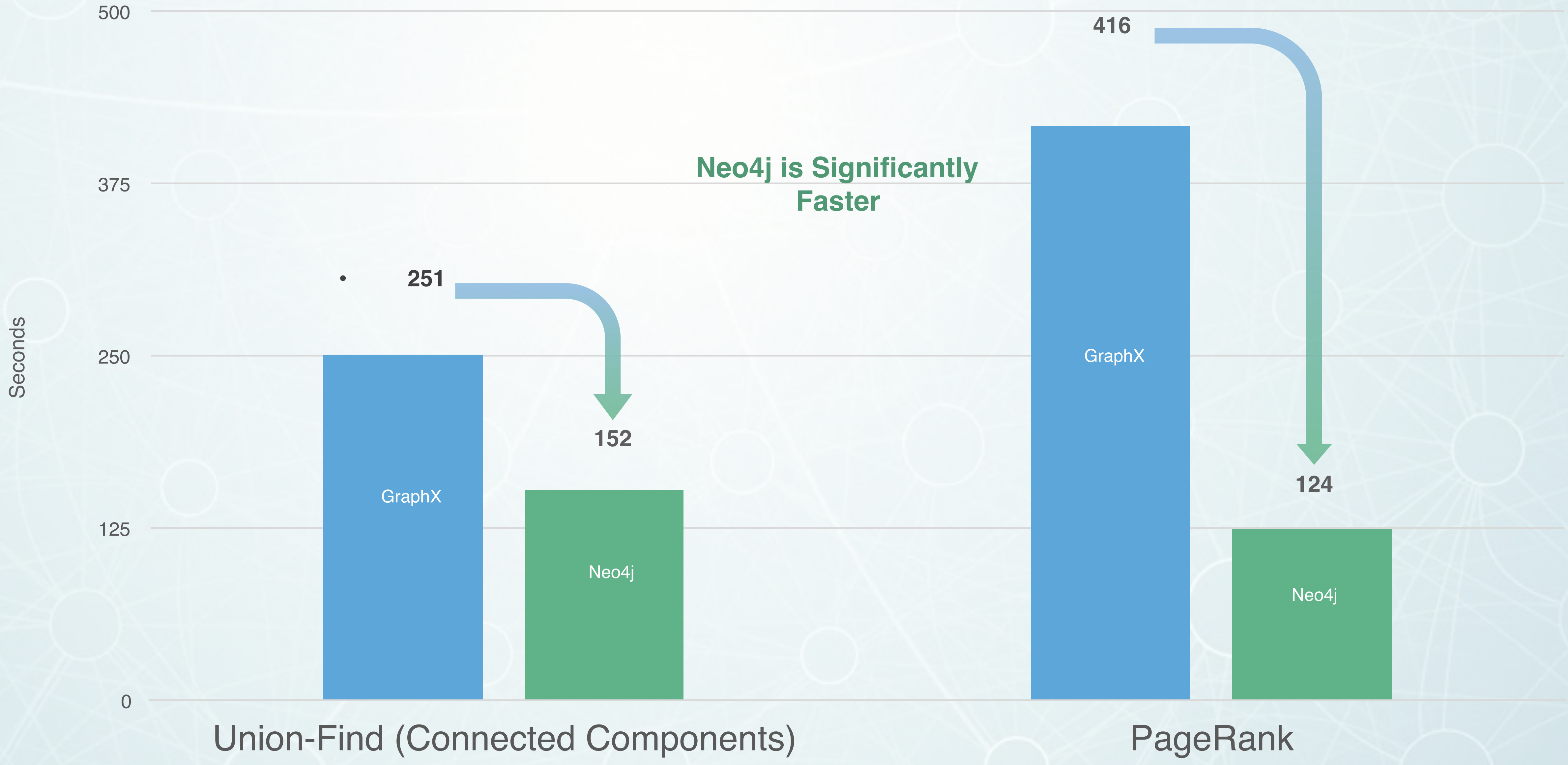
- Single Source Short Path
- All-Nodes SSP
- Parallel BFS / DFS

Goal: Iterate Quickly

- Combine data from sources into one graph
- Project to relevant subgraphs
- Enrich data with algorithms
- Traverse, collect, filter aggregate with queries
- Visualize, Explore, Decide, Export
- From all APIs and Tools



A note on Performance



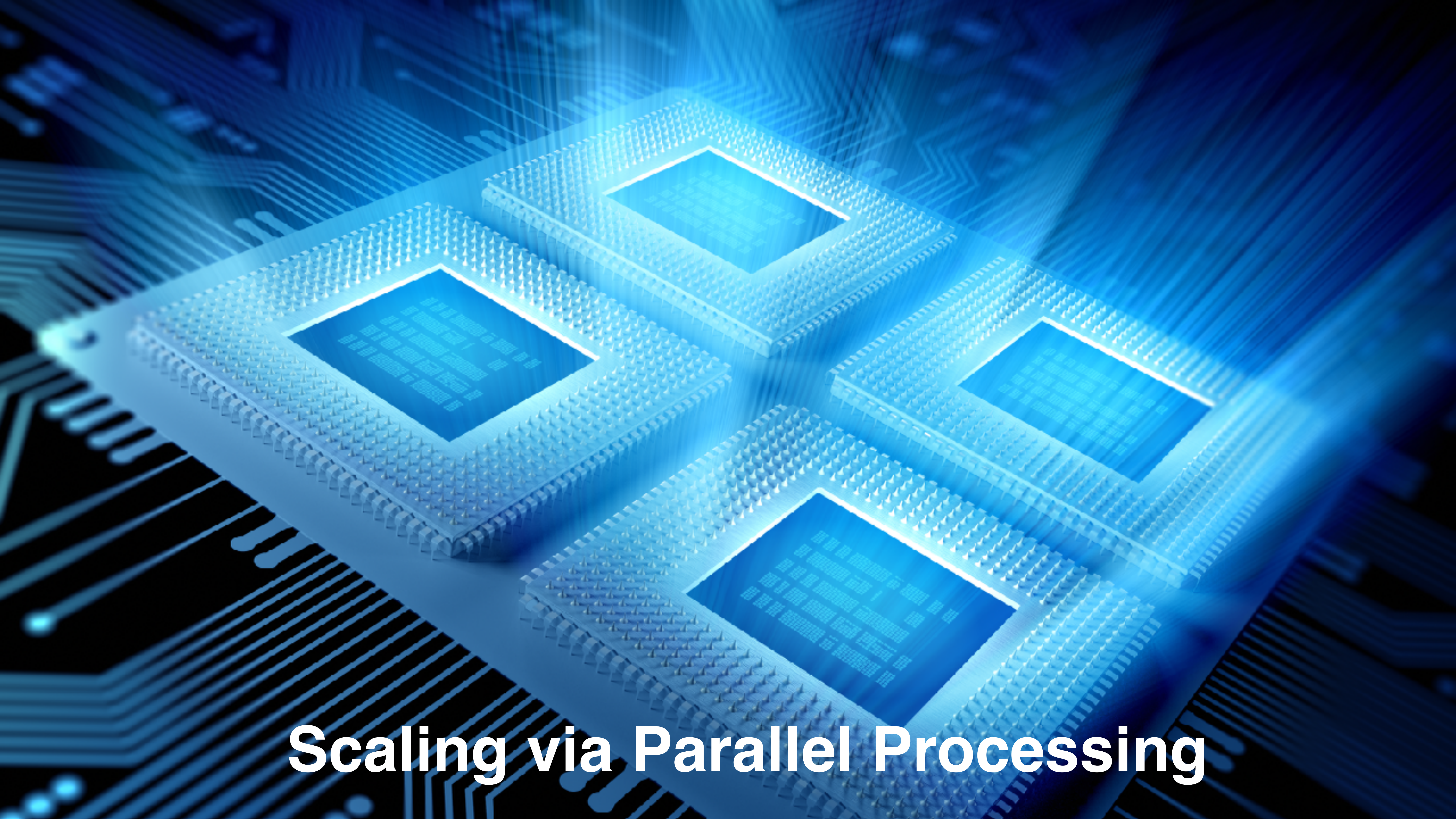
Twitter 2010 Dataset
- 1.47 Billion Relationships
- 41.65 Million Nodes

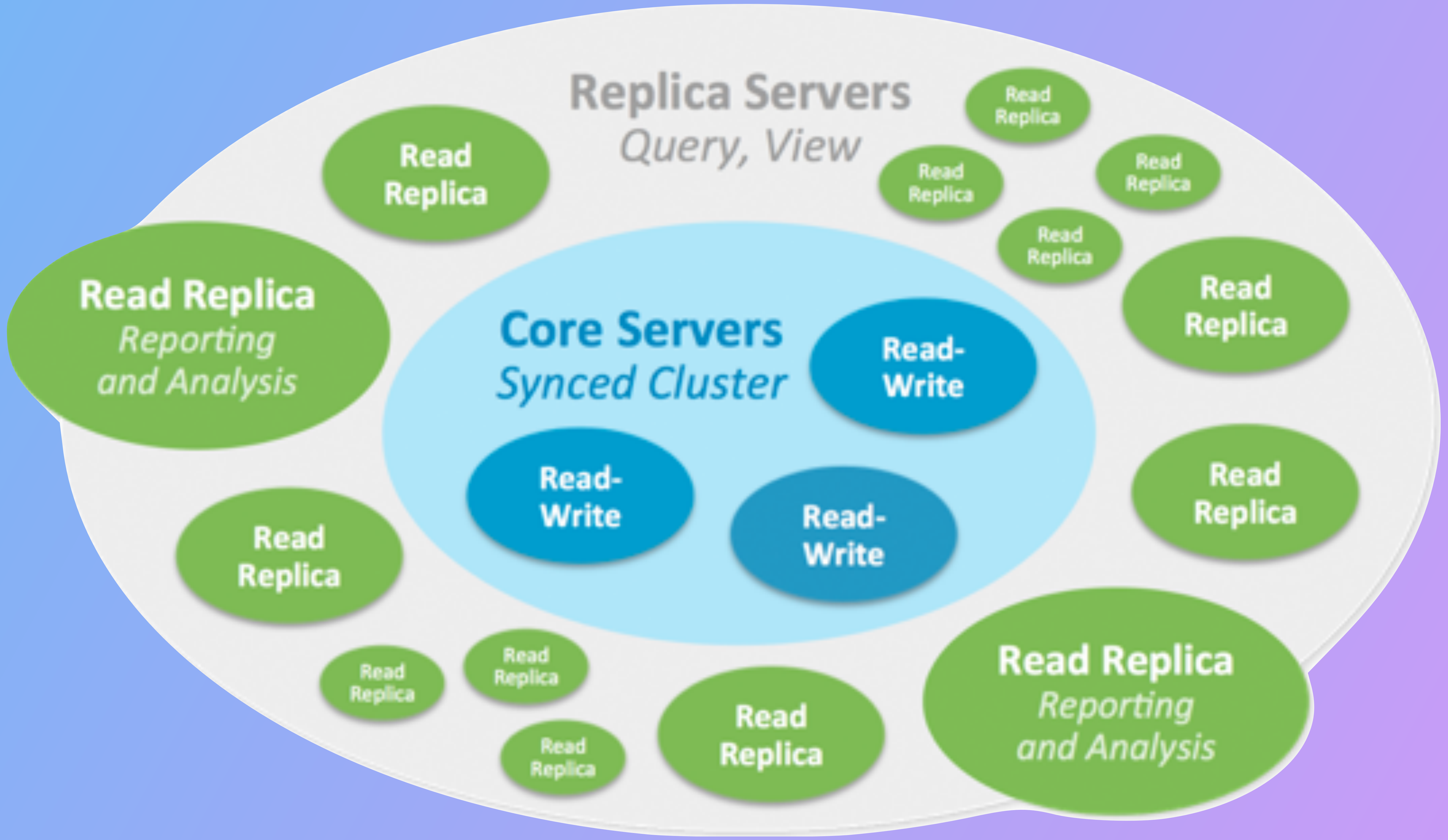
Spark GraphX results [publicly available]()
- Amazon EC2 cluster running 64-bit Linux
- 128 CPUs with 68 GB of memory, 2 hard disks

Neo4j Configuration
- Physical machine running 64-bit Linux
- 128 CPUs with 55 GB RAM, SSDs

What’s the Future Look Like?

Improved Performance & Testing


Scaling via Parallel Processing



Scaling Across the Cluster

THANK YOU!

ryan@neo4j.com
@ryguyrg

neo4j