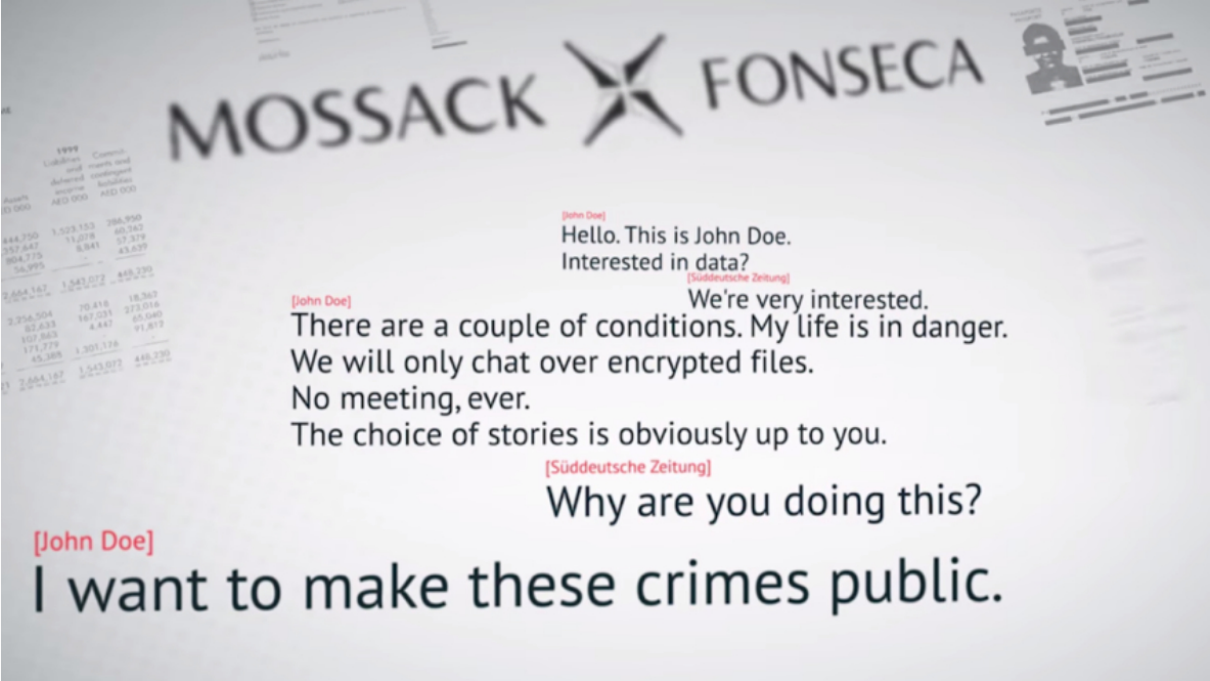
PANAMA PAPERS


MOSSACK FONSECA
[John Doe]
Hello. This is John Doe.
Interested in data?
[Süddeutsche Zeitung]
We're very interested.
[John Doe]
There are a couple of conditions. My life is in danger.
We will only chat over encrypted files.
No meeting, ever.
The choice of stories is obviously up to you.
[Süddeutsche Zeitung]
Why are you doing this?
[John Doe]
I want to make these crimes public.

EL COMER
HAARETZ
Established 1919
ISRAEL'S LEADING DAILY NEWSPAPER
e tax haven document leak exposes global corrupt
Le Monde
Panama papers
LES COMPTES
FALTER
TORONTO STAR
The
anama Papers
2.6 terabytes
12
29
740
l'Espresso
Panama Papers
Ecco i
100 n
OBAMA OVER ZIJN BUITENLANDSBELEID
'Ik had te veel vertrouwen in de Europeanen'
Knack
TOP SECRET
PANAMA PAPERS
De verborgen constructies van de rijkste Belgen
24 часа
25
Първите български имена от
eguardi
Putin


FIFA


SAY NO TO RACISM
CAPE TOWN
10
10

The scale of the leak

Volume of data compared to previous leaks

1,7 GB
Cablegate/Wikileaks (2010)

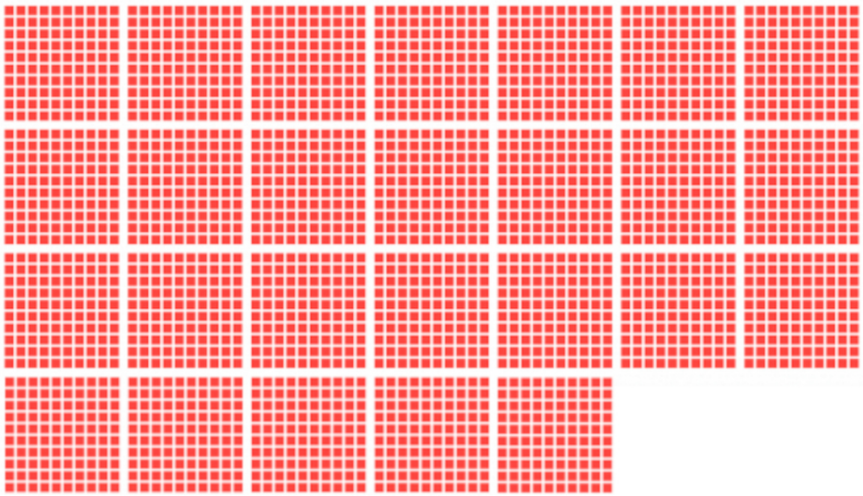
260 GB
Offshore Leaks/ICIJ (2013)

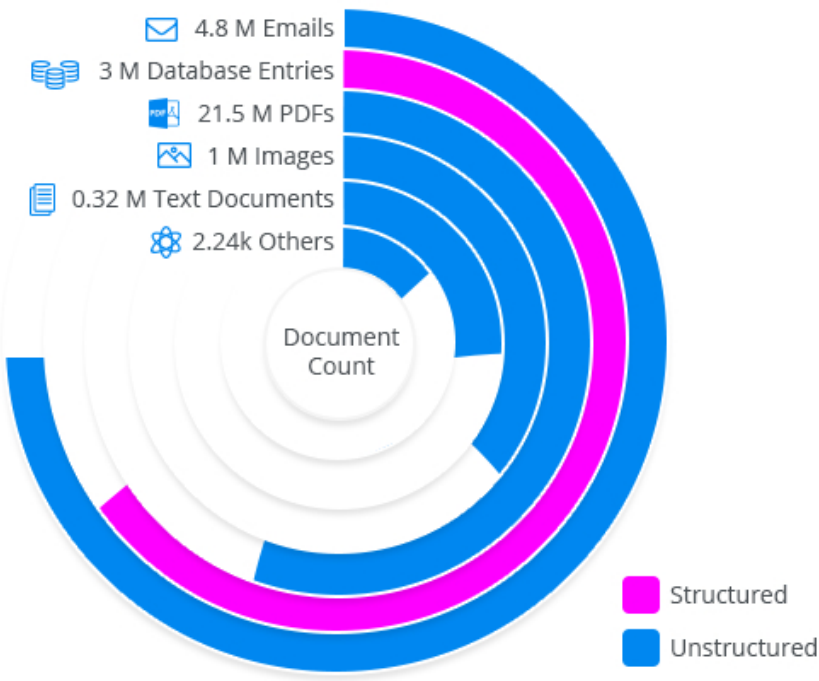
4 GB
Luxemburg Leaks/ICIJ (2014)

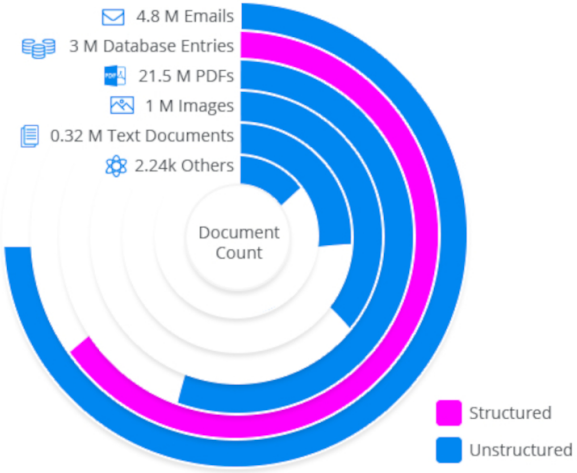
3,3 GB
Swiss Leaks/ICIJ (2015)

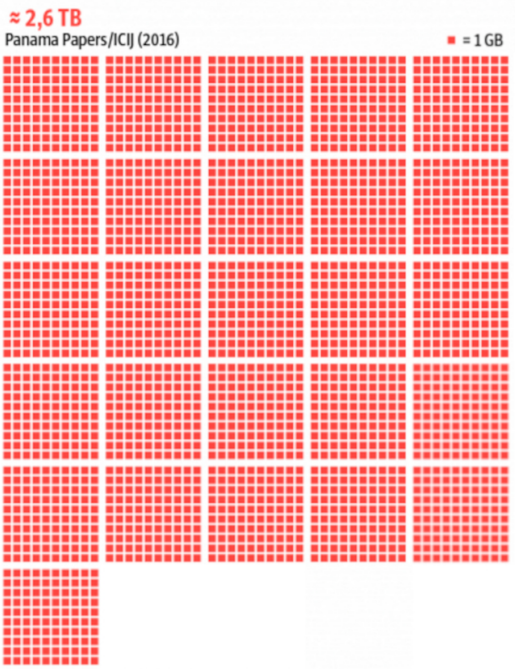
≈ 2,6 TB
Panama Papers/ICIJ (2016)

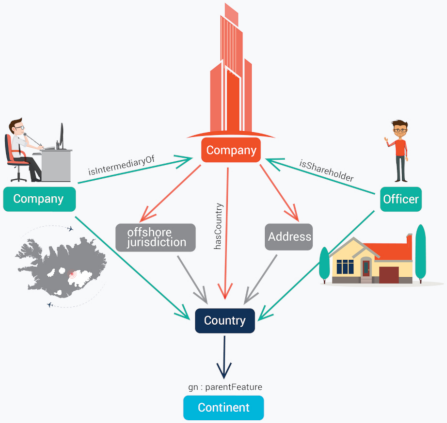
■ = 1 GB




4.8 M Emails
3 M Database Entries
21.5 M PDFs
1 M Images
0.32 M Text Documents
2.24k Others
Document Count
Structured
Unstructured


4.8 M Emails
3 M Database Entries
21.5 M PDFs
1 M Images
0.32 M Text Documents
2.24k Others
Document Count
Structured
Unstructured

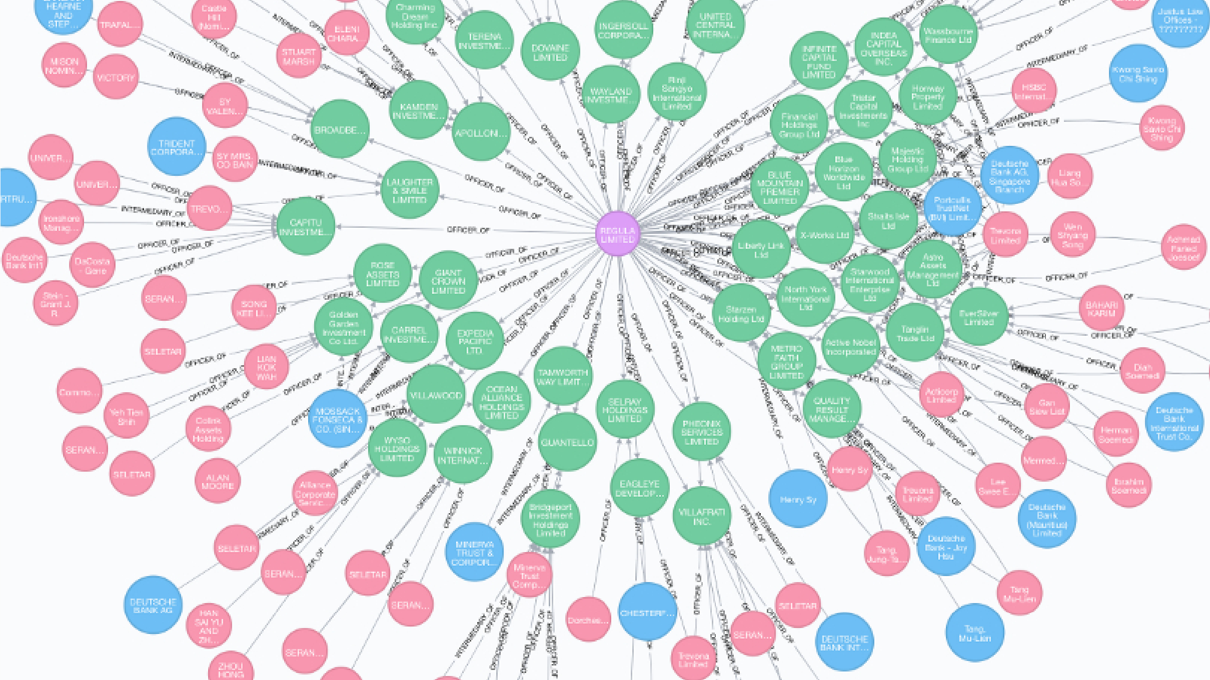

≈ 2,6 TB
Panama Papers/ICIJ (2016)
= 1 GB


Company
isIntermediaryOf
Company
offshore jurisdiction
hasCountry
Address
isShareholder
Officer
Country
gn : parentFeature
Continent


SIMILAR_N...
Officer
SECRETARY_OF
BENEFICIARY_OF
INTERMEDIARY_OF
SIMILAR_NAME_AND_ADDRESS
DIRECTOR_OF
PROTECTOR_OF
AUTHORISED_PERSON_SIGNATORY_OF
TRUST_SETTLOR_OF
SHAREHOLDER_OF
AUDITOR_OF
BENEFICIAL_OWNER_OF
SH...
SAME_NAM...
REGISTERED_ADDRESS
REGISTERED_A...
Entity
Intermediary
SIMILAR_N...
Address


```
MATCH (:Address)--(:Officer)--(e:Entity)
MATCH (e)--(:Intermediary)--(:Address)
MATCH (e)--(:Address)
WHERE o.name CONTAINS 'Emma Watson'
```

```
(:Address)--(:Officer)--(:Entity)--(:Intermediary)--(:Address)
```


REGULA LIMITED
OFFICER_OF
INTERMEDIARY_OF
DOVAINE LIMITED
INGERSOLL CORPORA...
UNITED CENTRAL INTERNA...
TERENA INVESTME...
Charming Dream Holding Inc.
WAYLAND INVESTME...
KAMDEN INVESTME...
APOLLONI...
BROADBE...
LAUGHTER & SMILE LIMITED
CAPITU INVESTME...
TRIDENT CORPORA...
ROSE ASSETS LIMITED
GIANT CROWN LIMITED
Golden Garden Investment Co Ltd
CARREL INVESTME...
EXPEDIA PACIFIC LTD
TAMWORTH WAY LIMIT...
VILLAWOOD
OCEAN ALLIANCE HOLDINGS LIMITED
SELRAY HOLDINGS LIMITED
PHEONIX SERVICES LIMITED
GUANTELLO
WYSO HOLDINGS LIMITED
WINNICK INTERNAT...
EAGLEYE DEVELOP...
VILLAFRATI INC.
Bridgeport Investment Holdings Limited
MOSSACK FONSECA & CO. (SIN...
MINERVA TRUST & CORPOR...
Minerva Trust Comp...
DEUTSCHE BANK AG
CHESTERF...
DEUTSCHE BANK INT...
INFINITE CAPITAL FUND LIMITED
INDEA CAPITAL OVERSEAS INC.
Financial Holdings Group Ltd
Tristar Capital Investments Inc
Honway Property Limited
Majestic Holding Group Ltd
Blue Horizon Worldwide Ltd
BLUE MOUNTAIN PREMIER LIMITED
X-Works Ltd
Straits Isle Ltd
Liberty Link Ltd
Starwood International Enterprise Ltd
Astro Assets Management Ltd
North York International Ltd
Starzen Holding Ltd
EverSilver Limited
Tanglin Trade Ltd
Active Nobel Incorporated
METRO FAITH GROUP LIMITED
QUALITY RESULT MANAGE...
Deutsche Bank AG, Singapore Branch
Portcullis TrustNet (BVI) Limit...
Deutsche Bank International Trust Co.
Deutsche Bank (Mauritius) Limited
Deutsche Bank - Joy Hou
Henry Sy
Tang Mu-Lien
Trevona Limited
SELETAR
SERAN...
ALAN MOORE
STUART MARSH
VICTORY
TRAFAL...
UNIVER...
TREVO...
BAHARI KARIM
Diah Soemedi
Herman Soemedi
Ibrahim Soemedi
Actiocorp Limited

Clustering Graphs

Applying a Label Propagation Algorithm to Detect Communities in Graph Databases

Name Surname

The work done

1. **Literature review** of graph theory, graph databases and clustering algorithms.
2. `dblp.org` **Dataset download, conversion & import** in ArangoDB Graph DBMS.
3. **Data transformations** to obtain vertices, edges and the complete graph.
4. **Community Detection Algorithm application** on the graph for clustering.
5. **Web Application development** to display the results of the clustering.

Graph databases

Simple — Data Complexity — Complex

Key-value

Key	Value

Examples:
Redis
Riak
DynamoDB
Aerospike

Wide-column

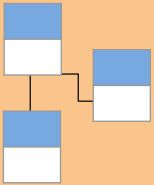
1			
	1		1
	1	1	
1	1	1	

Examples:
Cassandra
Apache HBase
Google BigTable

Document

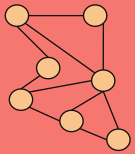
Examples:
MongoDB
DocumentDB
CouchBase

Relational

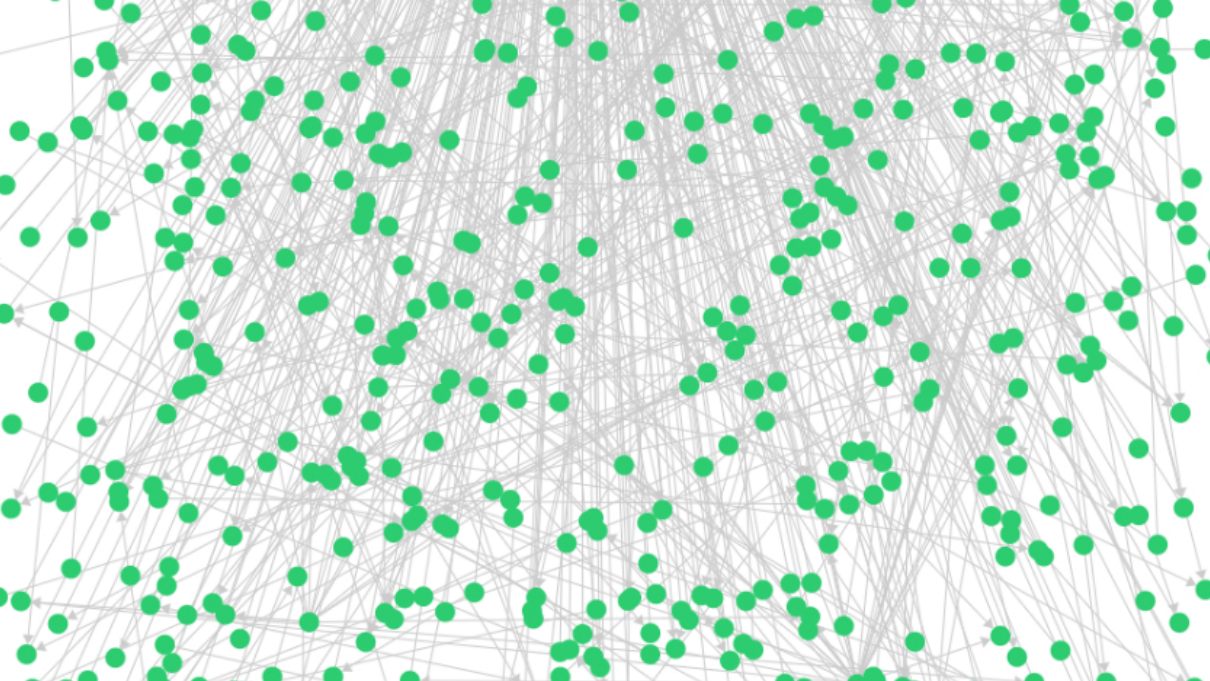


Examples:
Oracle
MySQL
Postgres
SQL Server

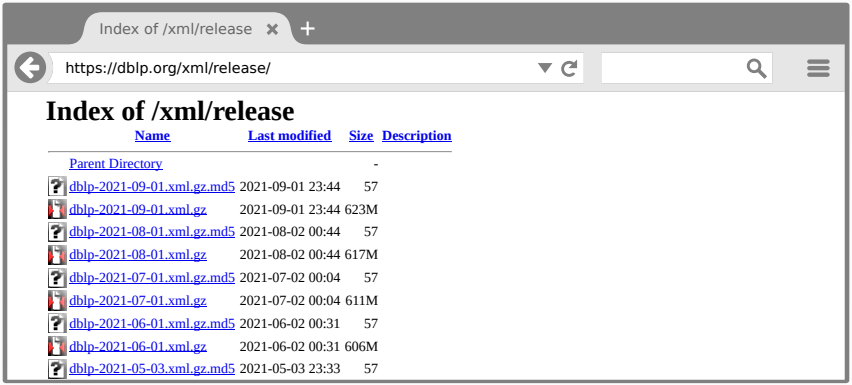
Graph



Examples:
ArangoDB
Dgraph
Neo4j
TigerGraph
CosmosDB



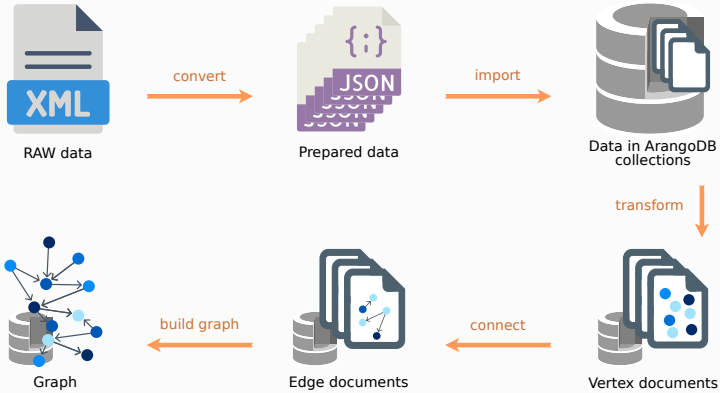
The dataset: dblp.org



The dataset:

- Compressed archive of 623 MB.
- Once extracted: Single XML file of 3.2 GB.
- 8.5 million XML entries on publications, authors, journals, institutions, citations etc.

Data conversion, import and transformations

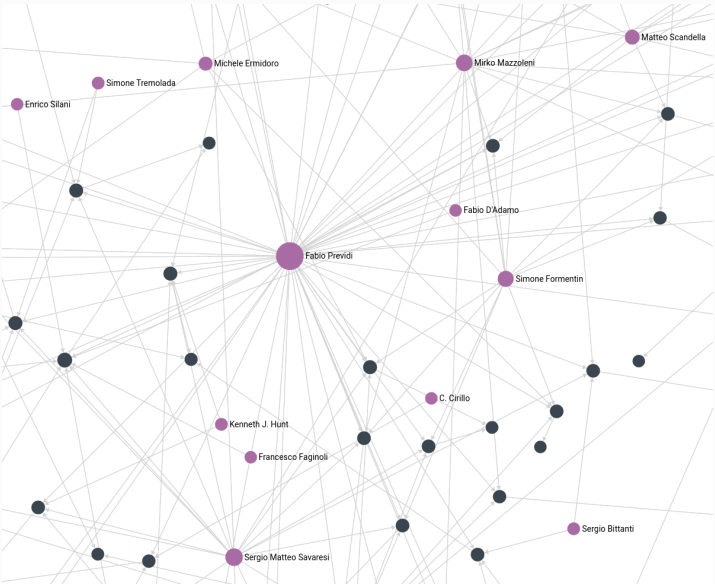


The steps:

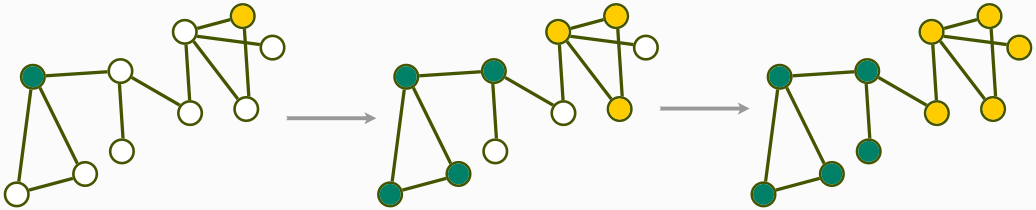
1. Conversion from XML to line JSON;
2. Import in ArangoDB collection;
3. Transform imported data to vertex documents;
4. Link vertices by edges;
5. Build the graph.

The graph

A subgraph of the obtained final graph:



Pregel Label Propagation Community Detection Algorithm



The algorithm:

1. Initially every vertex is labeled with a unique label.
2. The labels are propagated from vertex to vertex iteratively.
3. At the end of an iteration, each vertex updates its label to the one with maximum weight of neighbor vertices.
4. Convergence is reached when each vertex is labeled as the majority of its neighbors.

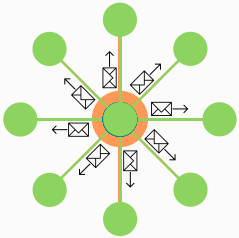
A generic Pregel algorithm

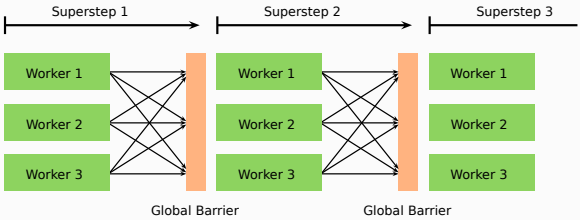
Pregel's supersteps:

Diffusion: information is propagated from vertex to neighbors.

Fusion: information is aggregated from neighbors to a set of entities

```
while not converged do
    /* Superstep */
    communicate:
        send(msg: own_value, to: neighbors)
    compute:
        own_value ← max(value_from_all_messages ∪ own_value)
end while
```

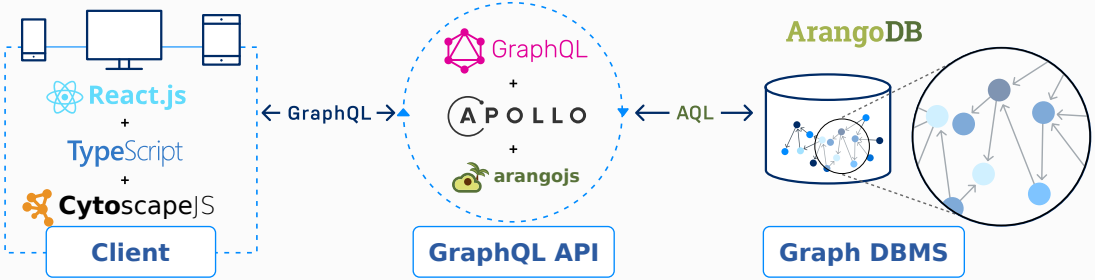




Clustering results

Vertex type	Number of vertices	Number of detected communities
author	2786113	177592
editor	43644	9837
institution	56918	25415
journal	1905	1896
publication	5662747	141939
publisher	2292	1437
school	2098	1677
series	1742	934
all types	8557459	187451

Web Application's Architecture

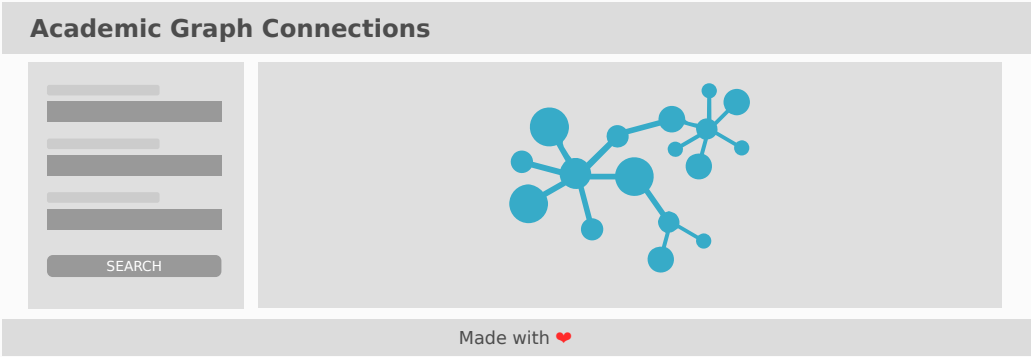


GraphQL API

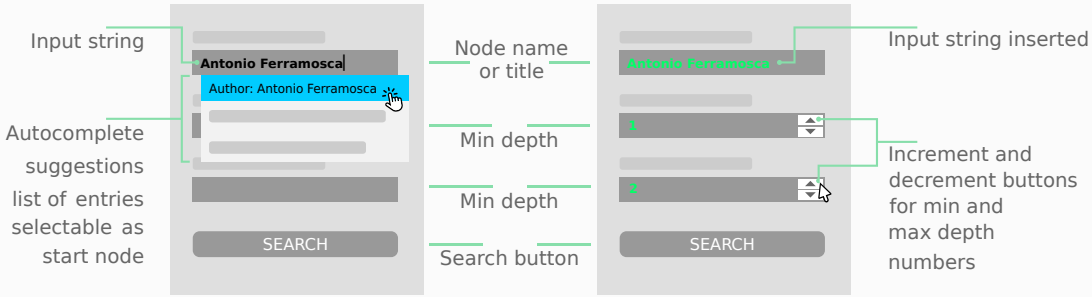
The API, initially REST then converted to a GraphQL API, is made of two resolvers:

1. A resolver function to handle the search form autocomplete suggestions.
2. A resolver function to provide collaboration graph's data.

Frontend UI Layout



Frontend UI Search Form



Results display (Searching for ”Angelo Gargantini”)

Academic Graph Connections

Node name or title

Angelo Gargantini

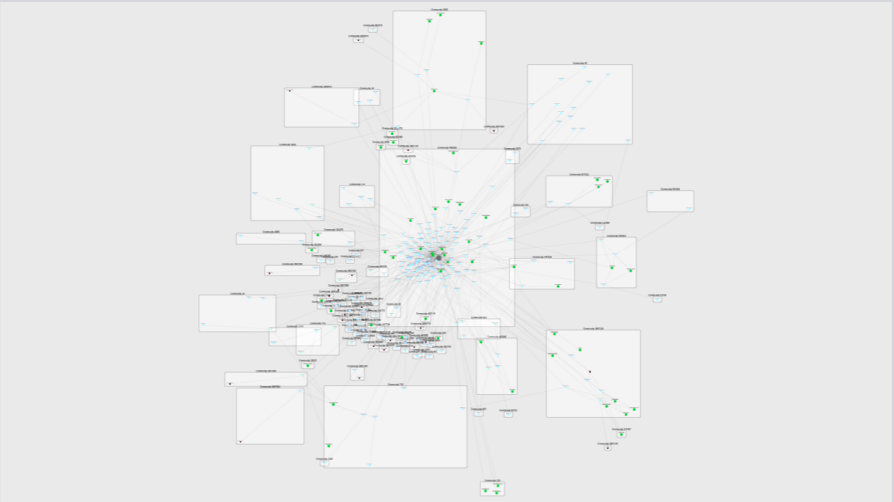
Minimum depth

1

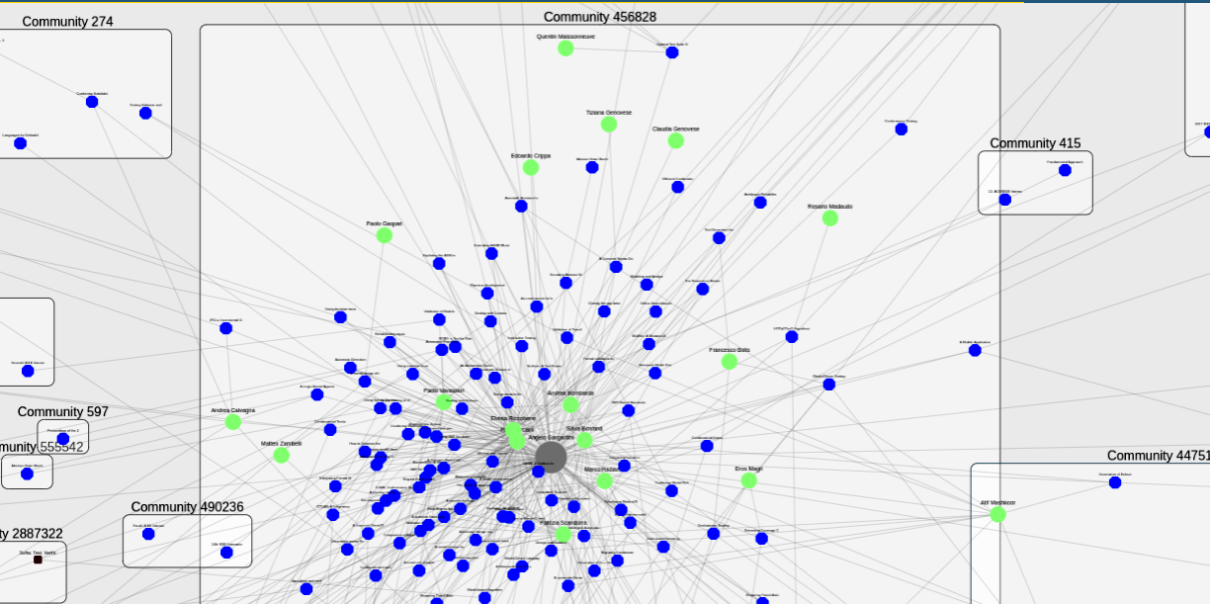
Maximum depth

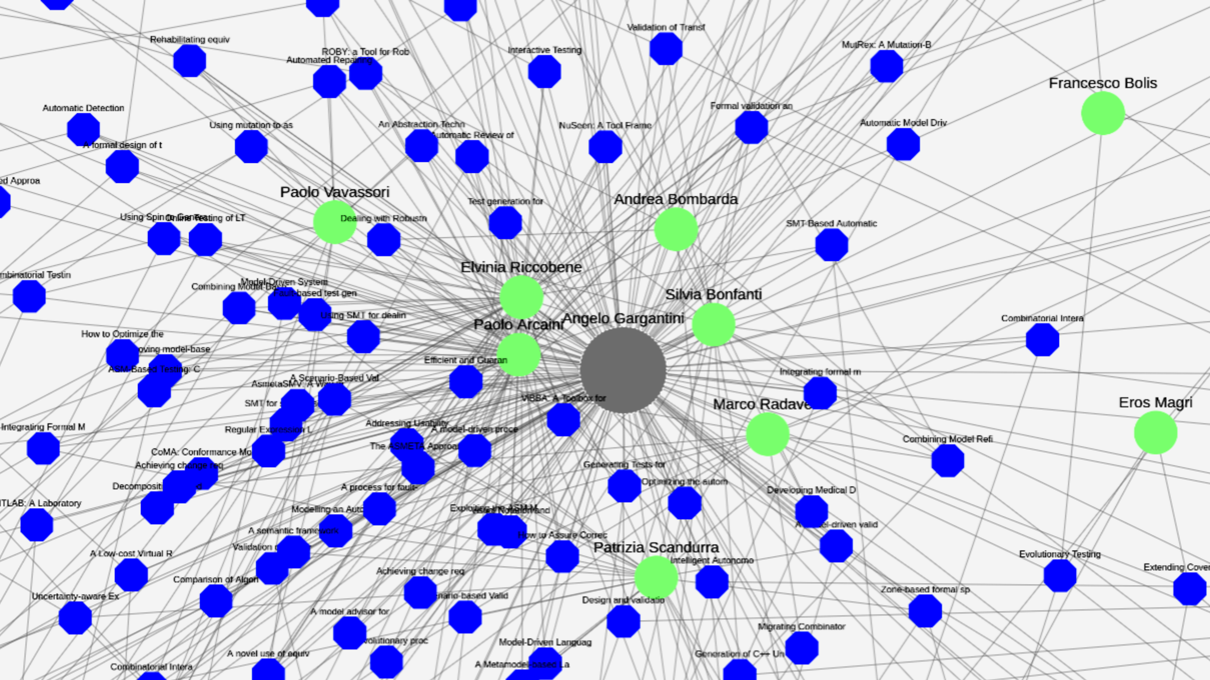
2

SEARCH

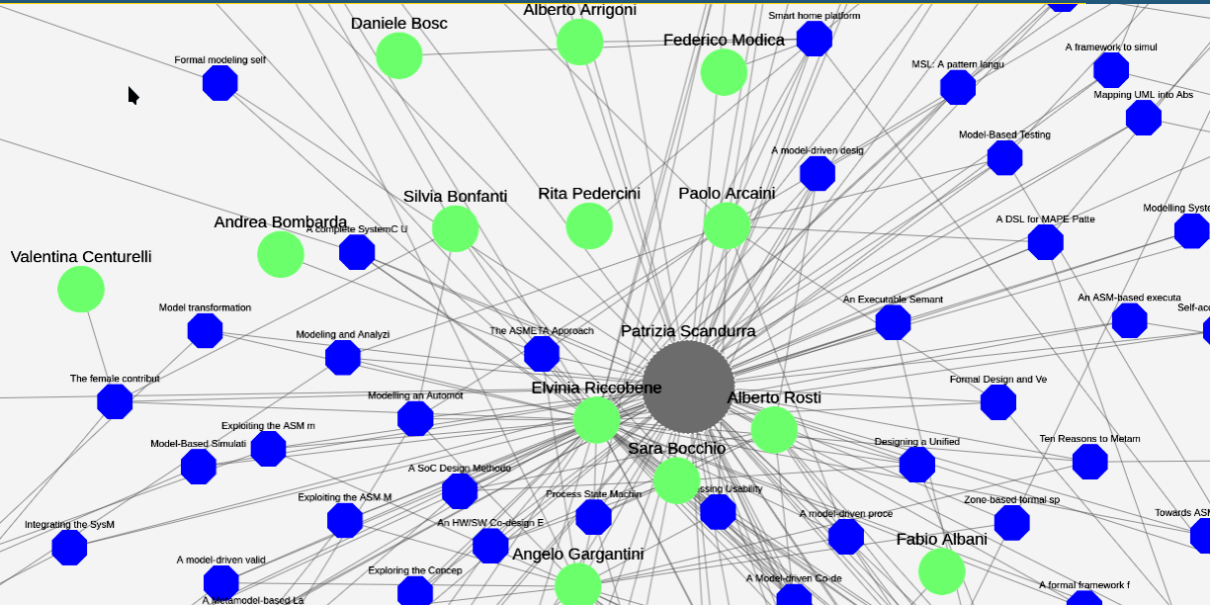


Results display (Angelo Gargantini)

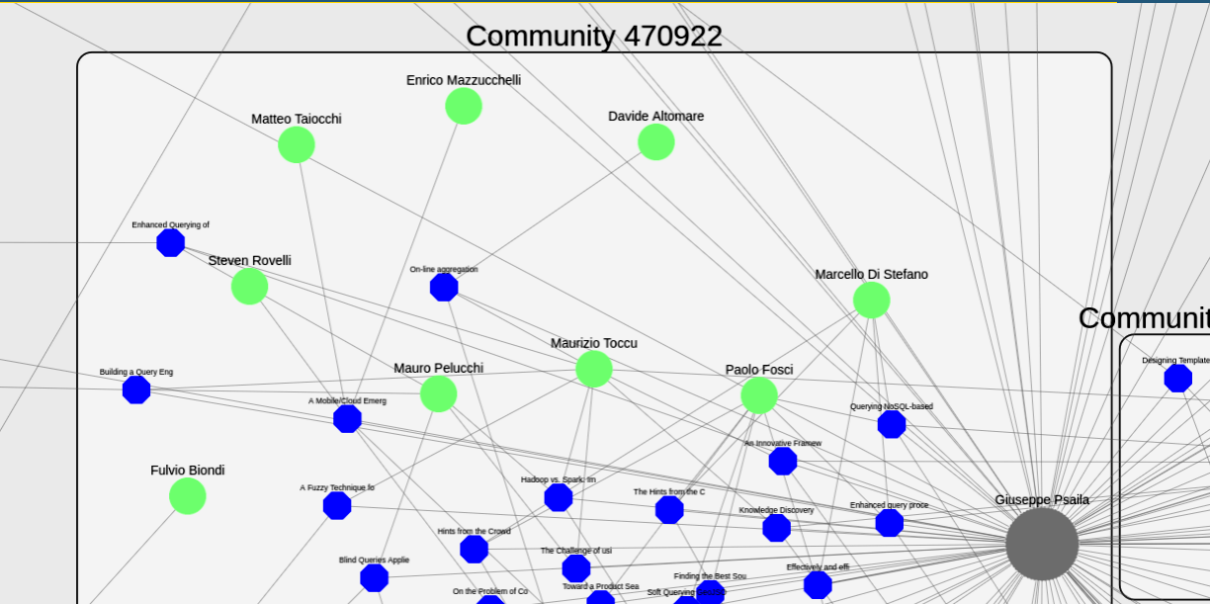



Rehabilitating equiv
ROBY: a Tool for Rob
Automated Repairing
Interactive Testing
Validation of Transf
MutRex: A Mutation-B
Francesco Bolis
Automatic Detection
Using mutation to as
An Abstraction Techn
Automatic Review of
NuSeen: A Tool Frame
Formal validation an
Automatic Model Driv
A formal design of t
ed Approa
Paolo Vavassori
Test generation for
Andrea Bombarda
Using Spin
Online Testing of LT
Dealing with Robustn
SMT-Based Automatic
mbinatorial Testin
Elvinia Riccobene
Model-Driven System
Combining Model-Ba
Fault-based test gen
Silvia Bonfanti
Using SMT for dealin
Paolo Arcaini
Angelo Gargantini
Combinatorial Intera
How to Optimize the
oving model-base
ASM-Based Testing: C
Efficient and Guaran
A Scenario-Based Val
Integrating formal m
AsmetaSMV: A W
Eros Magri
SMT for
VIBBA: A Toolbox for
Marco Radave
Integrating Formal M
Addressing Usability
Regular Expression L
A model-driven proce
Combining Model Refi
CoMA: Conformance Mo
The ASMETA Approac
Achieving change req
Generating Tests for
Decomposit
Optimizing the autom
A process for fault
Developing Medical D
TLAB: A Laboratory
Modelling an Aut
Exploiting the ASMM
Model-driven valid
A semantic framework
How to Assure Correc
Validation
A Low-cost Virtual R
Patrizia Scandurra
Evolutionary Testing
Intelligent Autonomo
Extending Cover
Comparison of Algori
Achieving change req
Uncertainty-aware Ex
nario-based Valid
Zone-based formal sp
Design and validatio
A model advisor for
Migrating Combinator
volutionary proc
Model-Driven Languag
A novel use of equiv
Generation of C++ Un
Combinatorial Intera
A Metamodel-based La

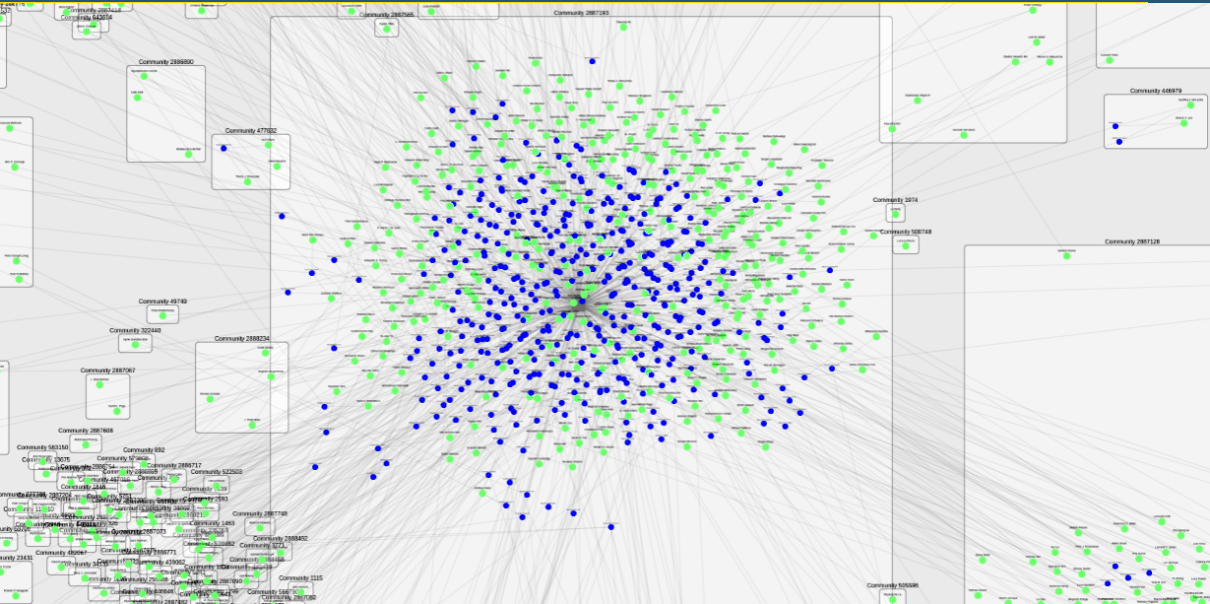
Results display (Patrizia Scandurra)



Results display (Giuseppe Psaila)



Results display (Statistical Methods & Applications Journal)



Results display (ETH Zurich)

Academic Graph Connections

Node name or title

ETH Zurich

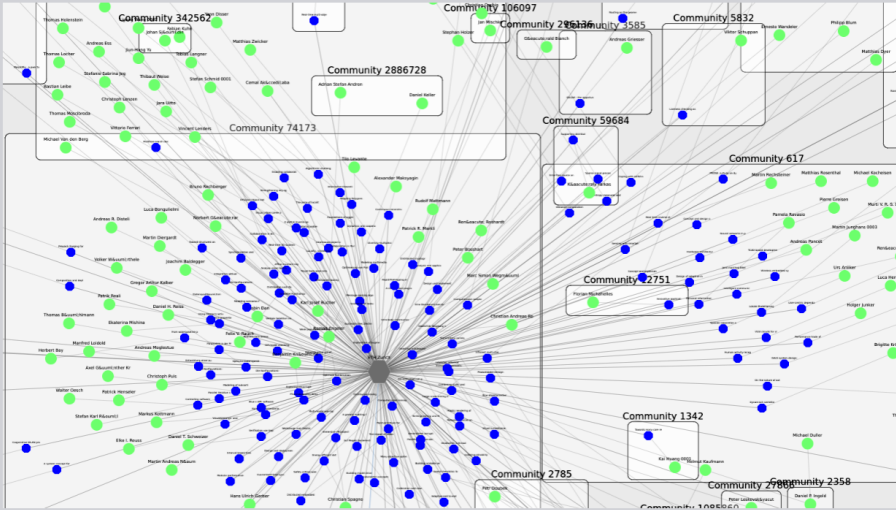
Minimum depth

1

Maximum depth

2

SEARCH



Thank you!

Questions?

Name
Surname