Graph Community Detection

neo4j

Community Detection

Goal: Divide the network into **communities**.

A graph has **community structure** if nodes are joined together in tightly knit groups, between which there are only looser connections.
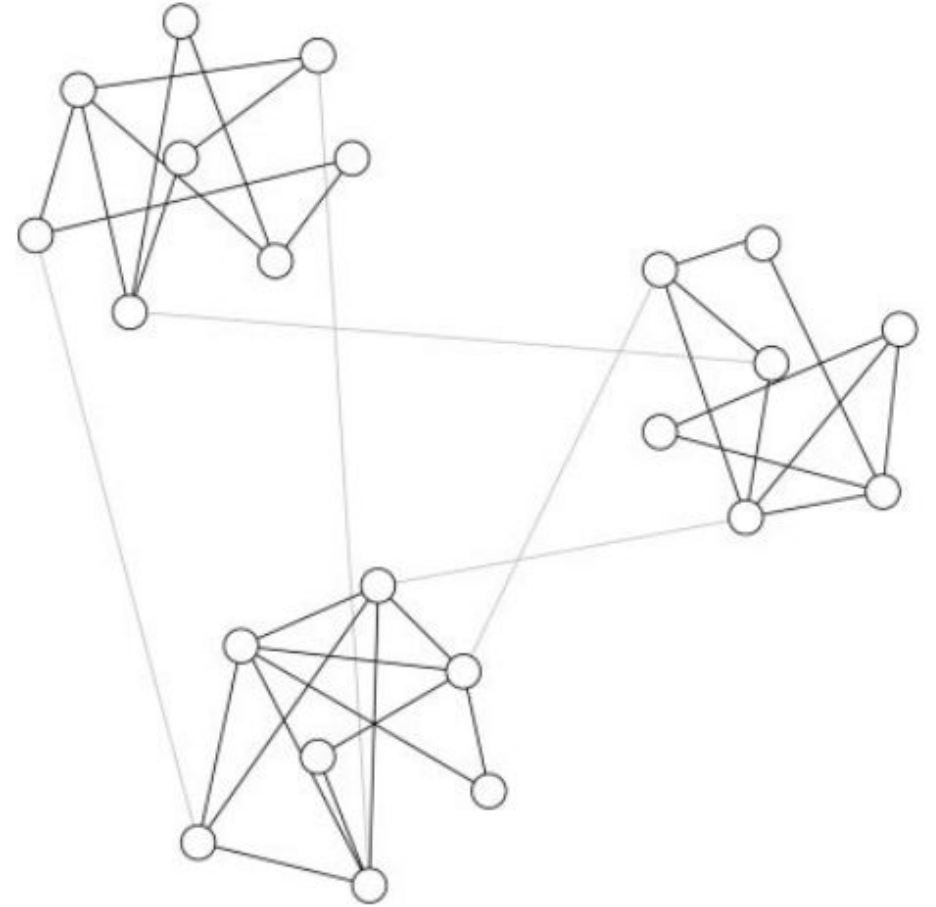

Fig. 1. A schematic representation of a network with community structure. In this network there are three communities of densely connected vertices (circles with solid lines), with a much lower density of connections (gray lines) between them.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC122977/

Community Detection

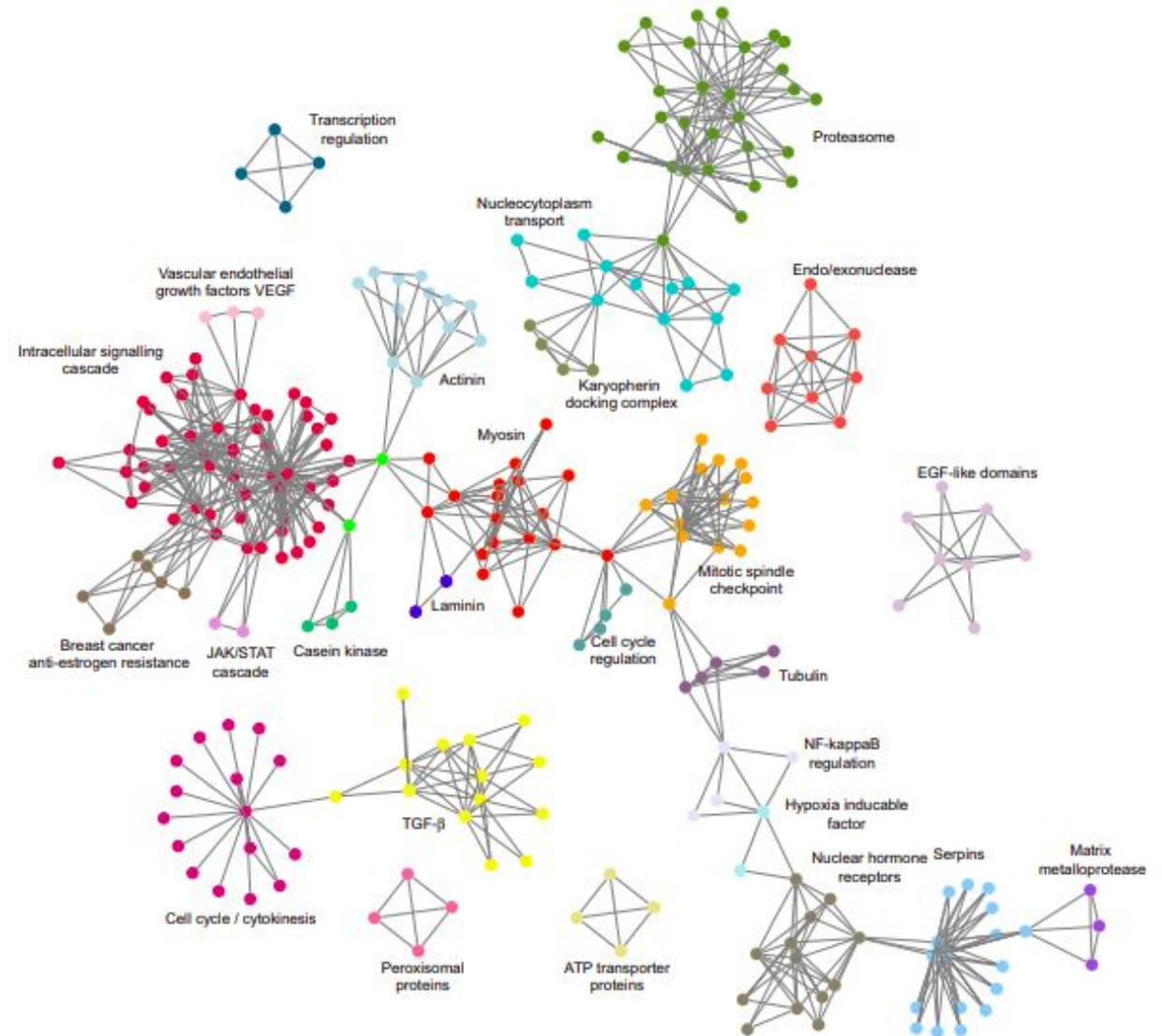**Example:** Detecting "communities" of proteins via protein-protein interactions.

Figure 3
Identifying protein communities by cluster analysis. The communities identified by *k*-clique analysis performed on the predicted genome-wide rat protein network. The communities are distinguished by different colours and labelled by the overall function or the dominating protein class. Note that proteins, particularly at community edges, can belong to more than two communities, although this is not shown. A complete list of protein names is included as supplementary material [see Additional file 1]. The graph was created by Graphviz [61].

Community Detection

Other Examples:

- Zombie Account Detection on social media (https://arxiv.org/abs/2101.00922)

- Finding related Git repositories (https://arxiv.org/abs/2002.02707)

- Understanding epidemic spreading and estimating the probability of a pandemic (https://arxiv.org/abs/1909.09695) (Submitted to arxiv.org in September of 2019!)

The Louvain Algorithm

A community detection algorithm that looks to optimize **modularity** - a measure of the interconnectedness of a subgraph, compared to the expected interconnectedness if edges were distributed at random.

First optimizes locally at each node and then combines communities into new nodes and repeats.

Source: https://arxiv.org/abs/0803.0476

Example using Neo4j

In order to apply the Louvain algorithm in Neo4j, we will need to first install the Graph Data Science Library Plugin.
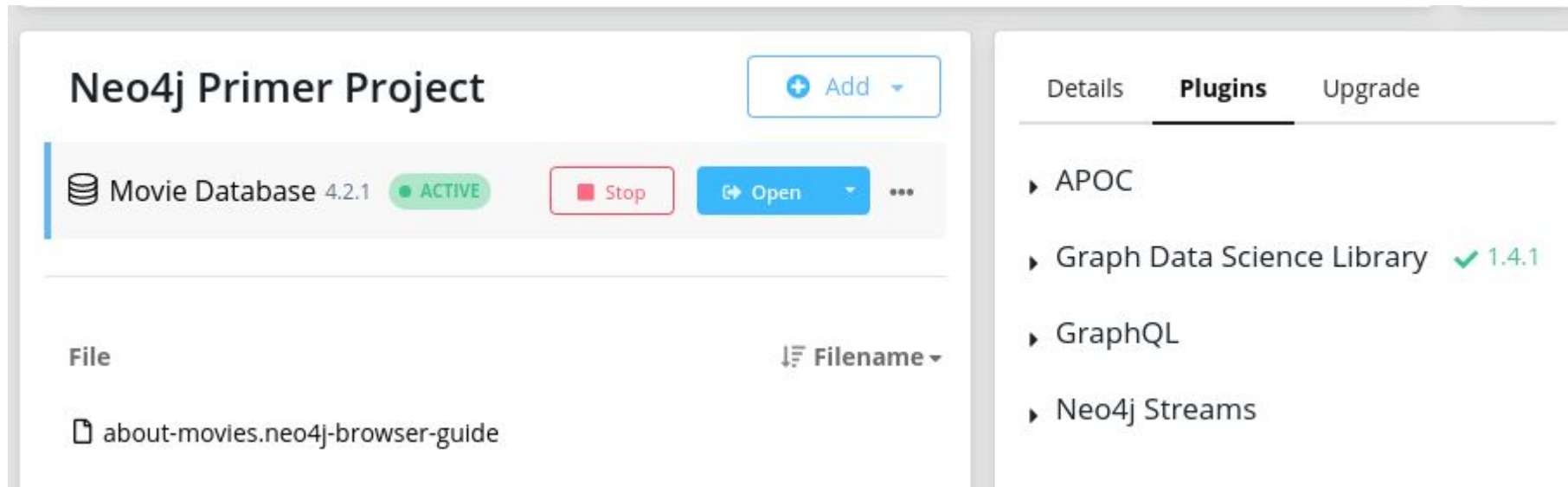
Example using Neo4j

For this example, we'll be using (a cleaned version of) the Enron email dataset.

This is a corpus of around a half-million emails that were made public by the Federal Energy Regulatory Commission during its investigation.

Example using Neo4j

	sender	recipient	num_messages
0	Brent Hendry AT ENRON_DEVELOPMENT@CCMAIL	sara shackleton	20
1	Carol St Clair	brant reves	34
2	Carol St Clair	brent hendry	31
3	Carol St Clair	bstclair dellnet	36
4	Carol St Clair	david dupre	25

The dataset that we'll be using contains a sender, a recipient, and the number of messages sent from by sender to that recipient.

It has been filtered to just pairs with at least 20 messages sent.

Example using Neo4j

	sender	recipient	num_messages
0	Brent Hendry AT ENRON_DEVELOPMENT@CCMAIL	sara shackleton	20
1	Carol St Clair	brant reves	34
2	Carol St Clair	brent hendry	31
3	Carol St Clair	bstclair dellnet	36
4	Carol St Clair	david dupre	25

s: Person {name}

WROTE {num_messages}

r: Person {name}
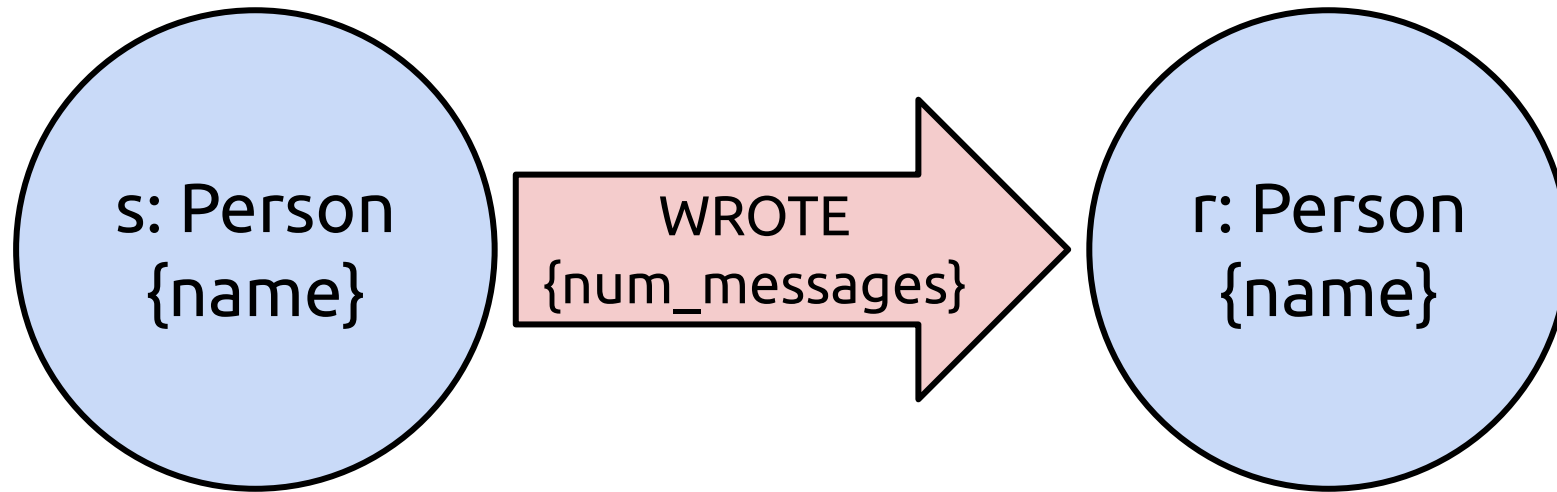
Neo4j - Loading Data From csv

First, we need to create a new database.

```
neo4j$ create database enron
```

Then switch to that database.
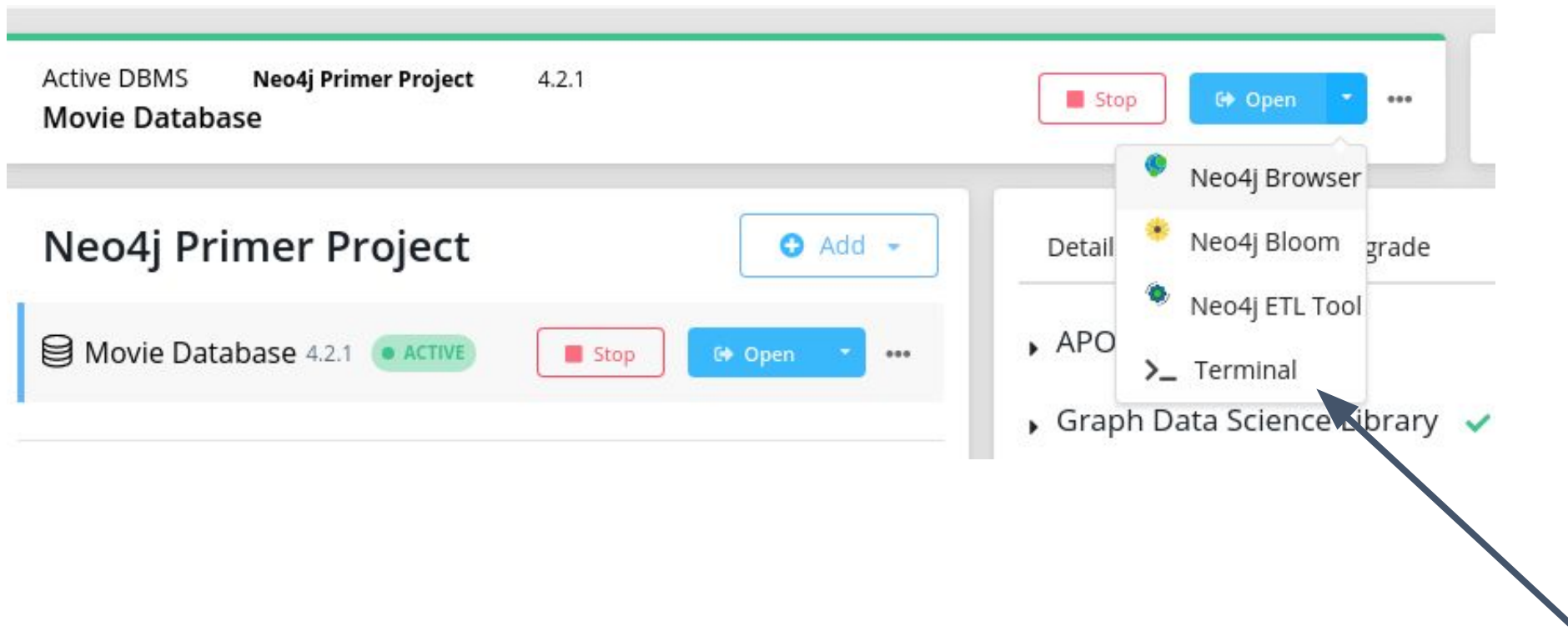
```
neo4j$ :use enron
```

Neo4j - Loading Data From csv

Before importing, we should create a constraint on the nodes:

    CREATE CONSTRAINT ON (p:Person) ASSERT p.name IS UNIQUE;

Neo4j - Loading Data From csv

Neo4j will be looking for the data in its import folder. Probably the easiest way to locate it is by opening the neo4j terminal:

Neo4j - Loading Data From csv

Neo4j will be looking for the data in its import folder. Probably the easiest way to locate it is by opening the neo4j terminal:

Once you have the path to the import directory, you can copy your csv there:

Neo4j - Loading Data From csv

Then we can import:

```
:auto USING PERIODIC COMMIT
LOAD CSV WITH HEADERS
FROM 'file:///enron.csv' AS line
WITH line

MERGE (from:Person {name: line.sender})
MERGE (to:Person {name: line.recipient})

CREATE (from)-[:WROTE {num_messages: toInteger(line.num_messages)}]->(to);
```

Neo4j - Preparing to Apply the Algorithm

Now, we need to add our nodes to the graph catalog:

```
CALL gds.graph.create(
    'enron',
    'Person',
    {
        WROTE: {
            orientation: 'UNDIRECTED'
        }
    },
    {
        relationshipProperties: 'num_messages'
    }
)
```

Neo4j - Applying the Algorithm

Finally, we can call the algorithm on our graph.

```
CALL gds.louvain.stream('myGraph', { relationshipWeightProperty: 'num_messages' })
YIELD nodeId, communityId
RETURN gds.util.asNode(nodeId).name AS name, communityId
ORDER BY name ASC
```

You can export the results as csv for easier analysis.

Results
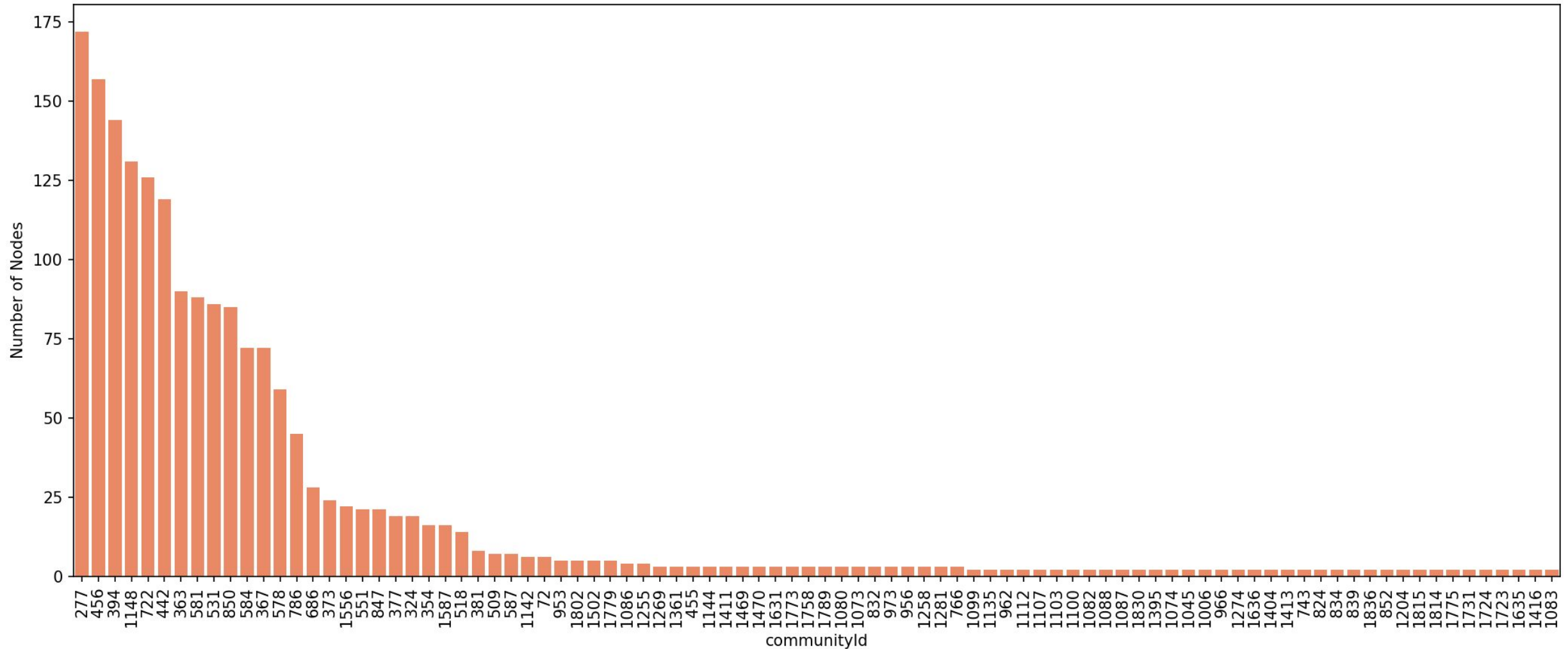
The algorithm uncovers 90 communities. (Your results may vary due to the random nature of the algorithm.)

Results

The algorithm uncovers 90 communities. (Your results may vary due to the random nature of the algorithm.)

A lot of these are sender/recipient pairs that only ever were connected to each other.

Results

There is not a whole lot of information about individuals at Enron, but we can borrow from https://arxiv.org/abs/1410.2759

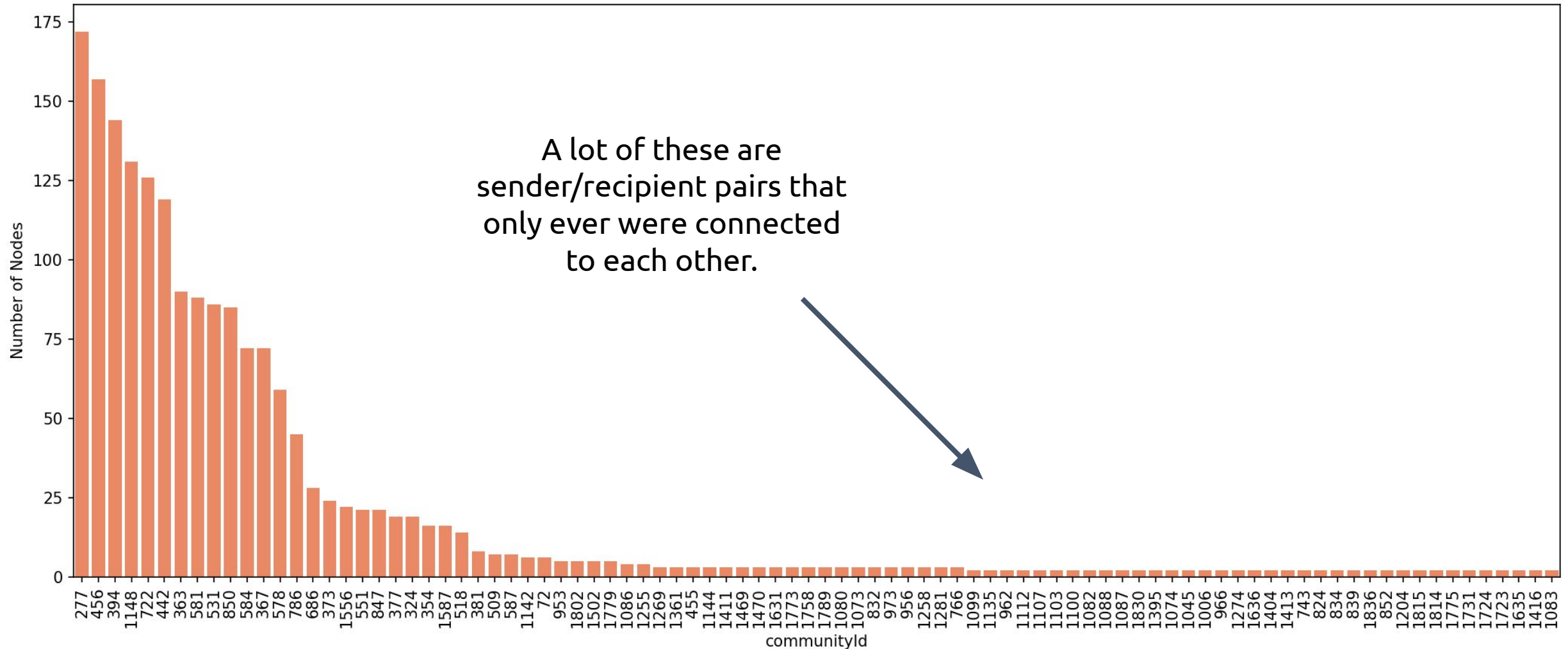| Name	Department/Title
Susan Bailey	ENA Legal/Legal Specialist
Marie Heard	ENA Legal/Legal Specialist
Tana Jones	ENA Legal/Legal Specialist
Stephanie Panus	ENA Legal/Legal Specialist
Elizabeth Sager	ENA Legal/VP & General Assistant Counsel
Sara Shackleton	ENA Legal/General Counsel Assistant
Robert Badeer	ENA West Power/Mgr Trading
Jeff Dasovich	Reg. and Gov. Affairs/Director
Mary Hain	Reg. and Gov. Affairs/Director
Steven J. Kean	Enron/VP & Chief of Staff
Richard Shapiro	Reg. and Gov. Affairs/VP
James D. Steffes	Reg. and Gov. Affairs/VP
Lindy Donoho	ETS/Employee
Michelle Lokay	ETS/Director
Mark McConnell	ETS/Director
Kimberly Watson	ETS/Director
Drew Fossum	ETS/VP & Gen. Cnsl.
Steven Harris	ETS/VP
Kevin Hyatt	ETS/Director
Susan Scott	ETS/Counsel

Results

There is not a whole lot of information about individuals at Enron, but we can borrow from https://arxiv.org/abs/1410.2759

Name	Department/Title
Susan Bailey	ENA Legal/Legal Specialist
Marie Heard	ENA Legal/Legal Specialist
Tana Jones	ENA Legal/Legal Specialist
Stephanie Panus	ENA Legal/Legal Specialist
Elizabeth Sager	ENA Legal/VP & General Assistant Counsel
Sara Shackleton	ENA Legal/General Counsel Assistant
Robert Badeer	ENA West Power/Mgr Trading
Jeff Dasovich	Reg. and Gov. Affairs/Director
Mary Hain	Reg. and Gov. Affairs/Director
Steven J. Kean	Enron/VP & Chief of Staff
Richard Shapiro	Reg. and Gov. Affairs/VP
James D. Steffes	Reg. and Gov. Affairs/VP
Lindy Donoho	ETS/Employee
Michelle Lokay	ETS/Director
Mark McConnell	ETS/Director
Kimberly Watson	ETS/Director
Drew Fossum	ETS/VP & Gen. Cnsl.
Steven Harris	ETS/VP
Kevin Hyatt	ETS/Director
Susan Scott	ETS/Counsel

	name	communityId
501	elizabeth sager	456
1127	marie heard	456
1538	sara shackleton	456
1646	stephanie panus	456
1679	susan bailey	456
1706	tana jones	456

Results

There is not a whole lot of information about individuals at Enron, but we can borrow from https://arxiv.org/abs/1410.2759

Name	Department/Title
Susan Bailey	ENA Legal/Legal Specialist
Marie Heard	ENA Legal/Legal Specialist
Tana Jones	ENA Legal/Legal Specialist
Stephanie Panus	ENA Legal/Legal Specialist
Elizabeth Sager	ENA Legal/VP & General Assistant Counsel
Sara Shackleton	ENA Legal/General Counsel Assistant
Robert Badeer	ENA West Power/Mgr. Trading
Jeff Dasovich	Reg. and Gov. Affairs/Director
Mary Hain	Reg. and Gov. Affairs/Director
Steven J. Kean	Enron/VP & Chief of Staff
Richard Shapiro	Reg. and Gov. Affairs/VP
James D. Steffes	Reg. and Gov. Affairs/VP
Lindy Donoho	ETS/Employee
Michelle Lokay	ETS/Director
Mark McConnell	ETS/Director
Kimberly Watson	ETS/Director
Drew Fossum	ETS/VP & Gen. Cnsl.
Steven Harris	ETS/VP
Kevin Hyatt	ETS/Director
Susan Scott	ETS/Counsel

	name	communityId
733	james steffes	277
768	jeff dasovich	277
1160	mary hain	277
1436	richard shapiro	277
1661	steven kean	277

Results

There is not a whole lot of information about individuals at Enron, but we can borrow from https://arxiv.org/abs/1410.2759

Name	Department/Title
Susan Bailey	ENA Legal/Legal Specialist
Marie Heard	ENA Legal/Legal Specialist
Tana Jones	ENA Legal/Legal Specialist
Stephanie Panus	ENA Legal/Legal Specialist
Elizabeth Sager	ENA Legal/VP & General Assistant Counsel
Sara Shackleton	ENA Legal/General Counsel Assistant
Robert Badeer	ENA West Power/Mgr Trading
Jeff Dasovich	Reg. and Gov. Affairs/Director
Mary Hain	Reg. and Gov. Affairs/Director
Steven J. Kean	Enron/VP & Chief of Staff
Richard Shapiro	Reg. and Gov. Affairs/VP
James D. Steffes	Reg. and Gov. Affairs/VP
Lindy Donoho	ETS/Employee
Michelle Lokay	ETS/Director
Mark McConnell	ETS/Director
Kimberly Watson	ETS/Director
Drew Fossum	ETS/VP & Gen. Cnsl.
Steven Harris	ETS/VP
Kevin Hyatt	ETS/Director
Susan Scott	ETS/Counsel

	name	communityId
467	drew fossum	394
971	kevin hyatt	394
991	kimberly watson	394
1067	lindy donoho	394
1140	mark mcconnell	394
1229	michelle lokay	394
1660	steven harris	394
1688	susan scott	394