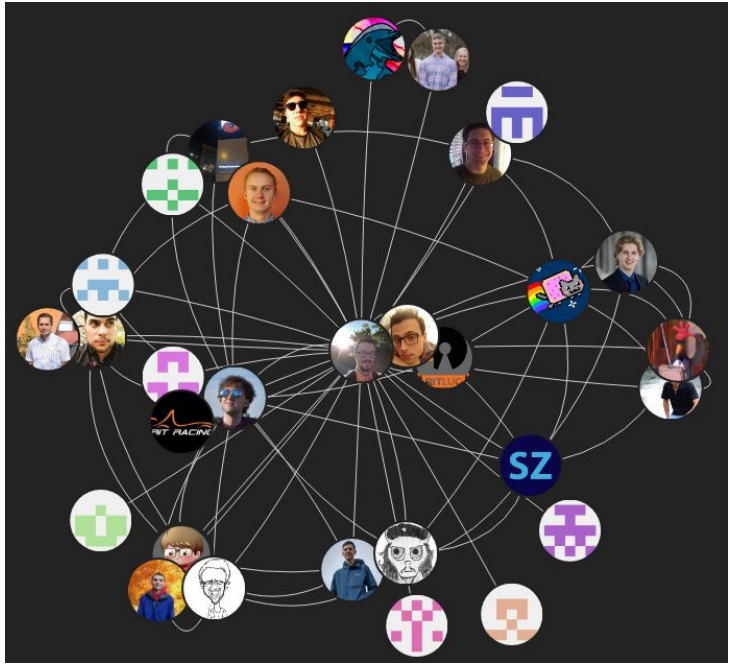
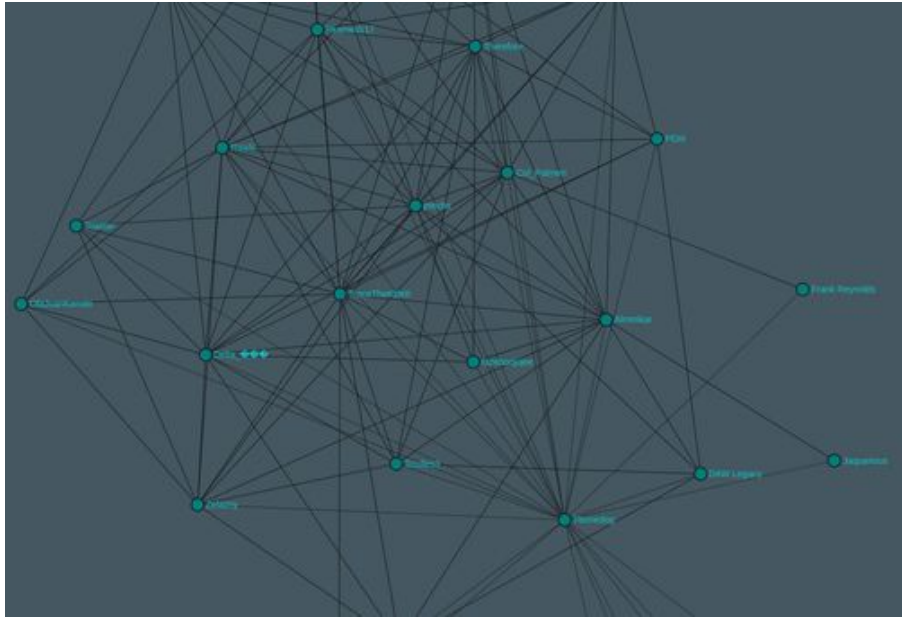
Graph Databases

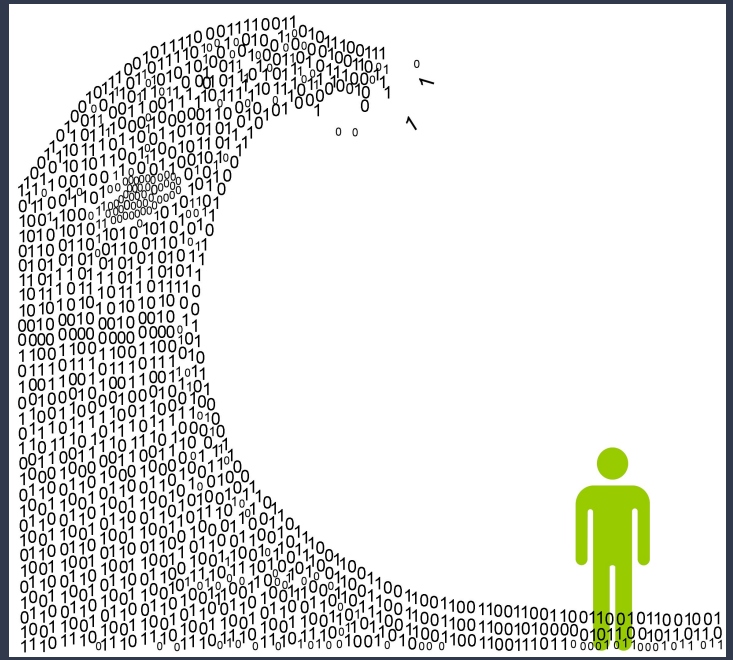
Sign In:
bit.ly/ritlug-mar-01

About Me





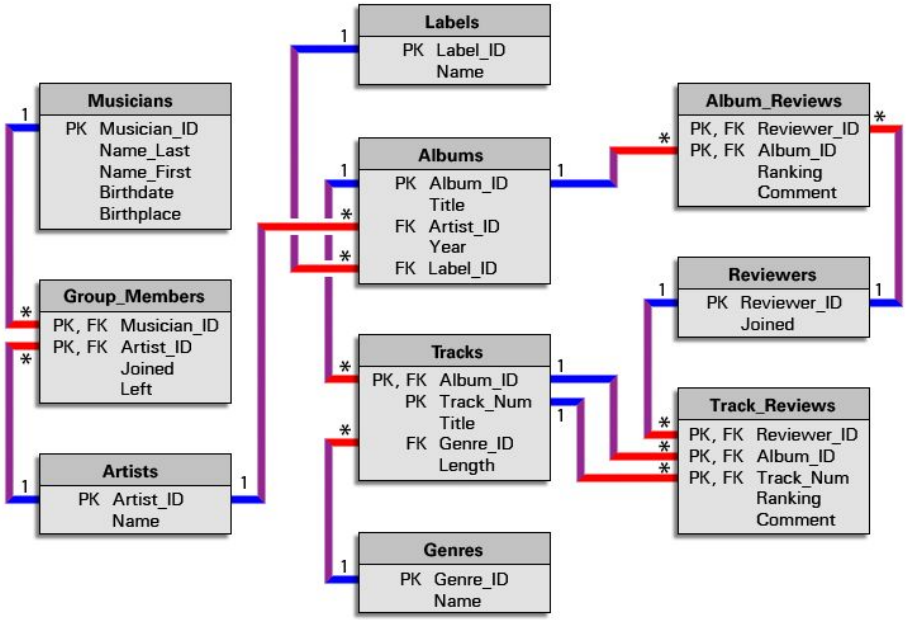
What even is a databases?



- MySQL
- PostgreSQL
- AWS Redshift
- DB2
- Derby
- Exasol
- Sybase ASE
- ClickHouse
- SQL Server
- Azure SQL Database
- Oracle
- H2
- HSQL
- SQLite
- MariaDB
- Cassandra

Relational Databases

- Types of things are tables where the columns are properties.
- Objects are stored as rows in a table.
- You use foreign key constraints to represent relations.

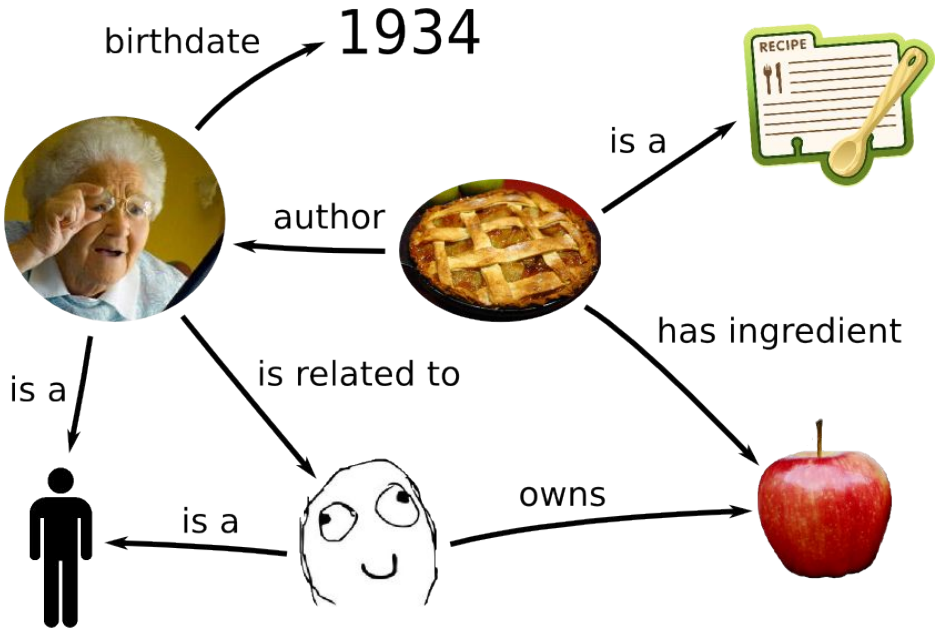


What about that graph database thing?

- Graph databases use graph theory and math stuff.

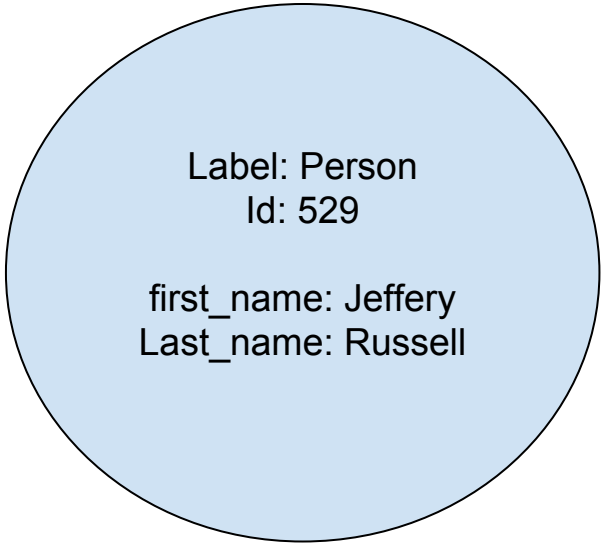
Graph Databases are Actually Pretty Simple

- Things in a database are “Nodes” or “Vertices” and relations are “Edges”.



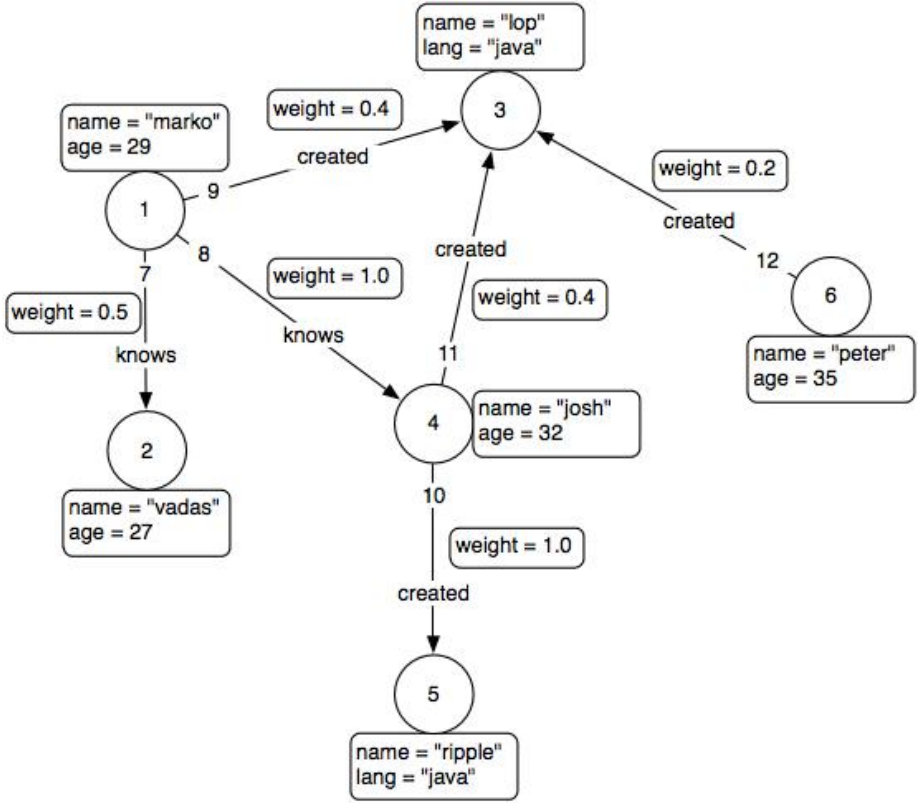
Nodes

- The type of data in a node is specified by the label.
- Nodes can have as many properties as you wish.
- Nodes typically have auto assigned unique ids.



Edges

- Edges represent relationships in a database.
- Unlike SQL which solely relies on foreign key restraints, you can add attributes to your relationships.
- There are both directional and bidirectional relationships.



NoSQl vs SQL

Strengths

Graph Databases

- Does not have a rigid schema which allows you to add more relations as needed.
- Good for answering questions that you didn’t expect when you first created the database.
- Very efficient at calculating indirect relationships.
- Structure is conceptually simpler.

SQL

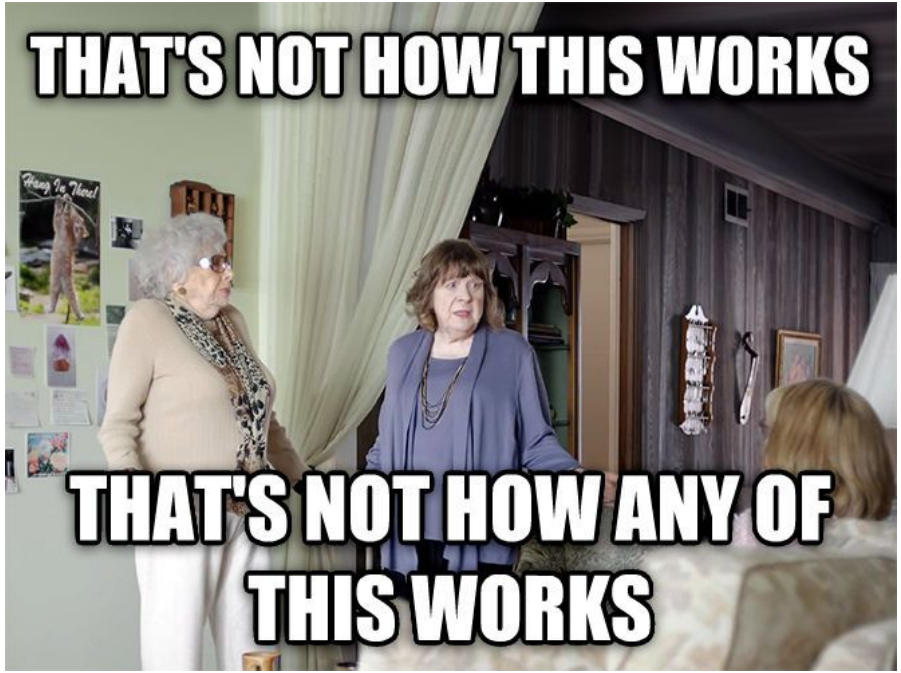
- Very efficient for data with stringent structures.
- Computationally efficient
- Well known
- Easy to host SQL servers

Example of a Graph DB Strength

How would you use SQL for the following problems?

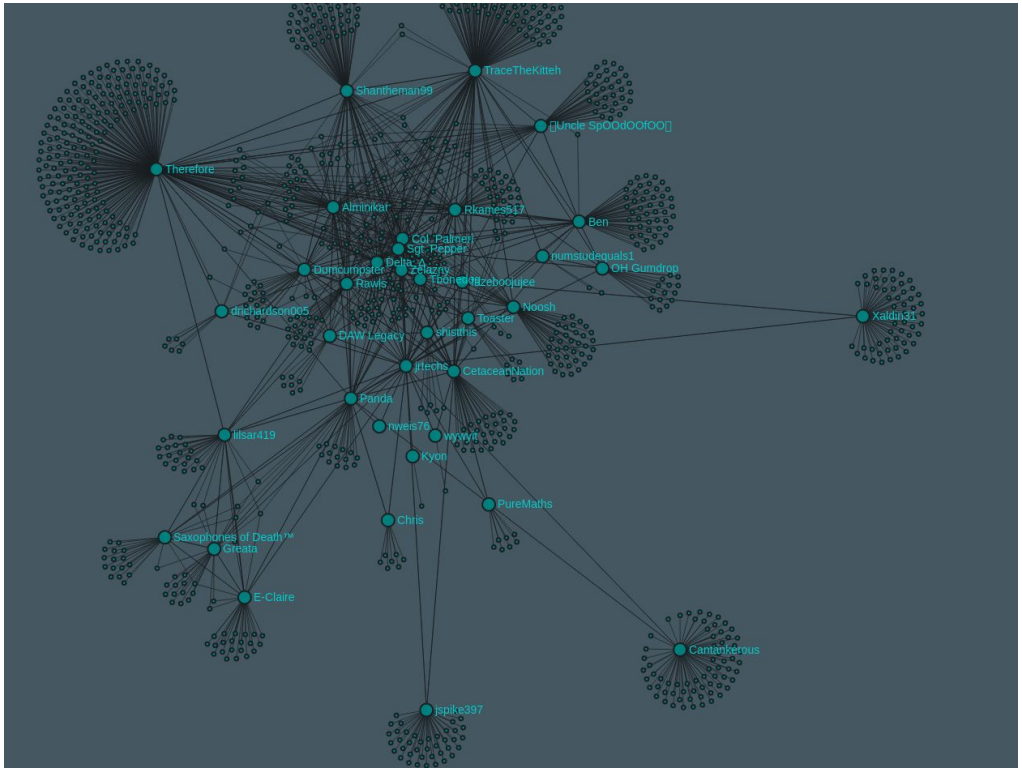
“How many people from last week’s hockey game lives in NRH, has a 4.0 GPA, is from Texas and, used a ticket from the Den?”

“Find all of the friends of friends of a person on Steam”



Takeaways

- Relational databases work well with sets.
- Graph databases work well with paths.
- Graph databases help you extract information when the meaning is in the relationships



OrientDB

TITAN

blazegraph™

APACHE HBASE

OLTP

Gremlin

$G = (V, E)$

Graph Databases

OLAP

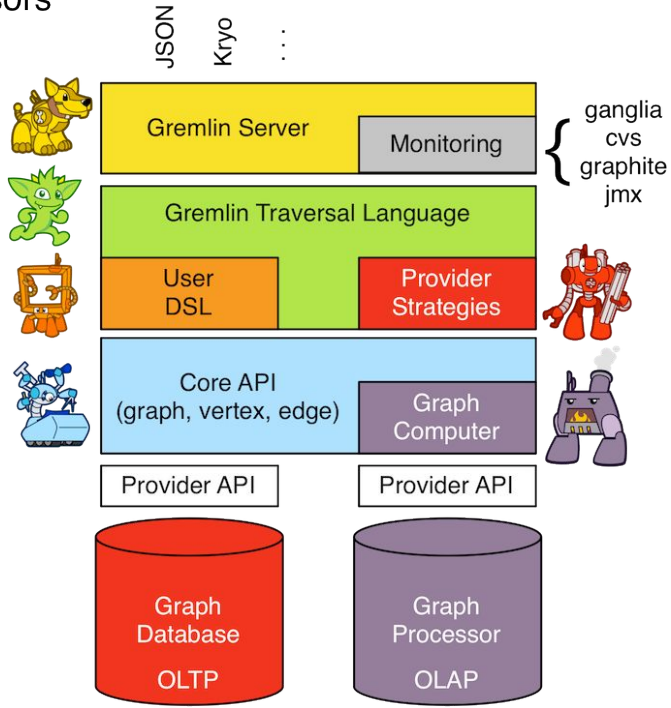
Graph Processors

APACHE GIRAPH

hadoop

Spark

- Gremlin is a graph traversal language. IE: SQL for graphs.
- Gremlin is a part of Apaches Tinkerpop which is an open sourcegraph computing framework.
- Apache has a **ton** of open source projects.



SQL	Gremlin
SELECT * **FROM** users	g.V().hasLabel("user")
SELECT name **FROM** users	g.V().hasLabel("user").values("name")
SELECT name, email **FROM** users	g.V().hasLabel("user").values("name", "email")

SQL

```
SELECT name FROM users
ORDER BY score ASC
```

Gremlin

```
g.V().hasLabel("user").order()
.by("score", incr).valueMap("name")
```

SQL

```
SELECT * FROM USERS WHERE id='42'
```

Gremlin

```
g.V().hasLabel("user").has('id', 42)
```

SQL

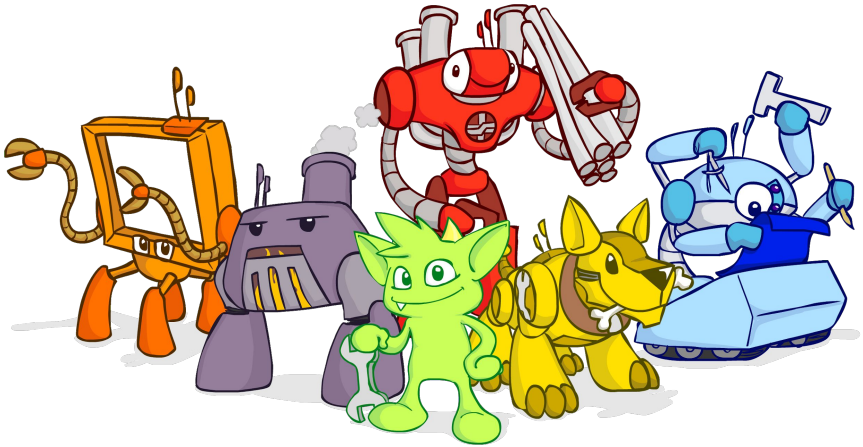
```
INSERT INTO users(name, email)
VALUES ('Jeff', 'hello@world')
```

Gremlin

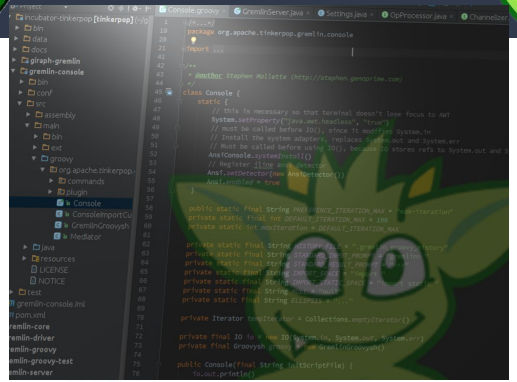
```
g.addV('user').property('name', 'Jeff')
.property('email', 'hello@world')
```

Resources

- http://tinkerpop.apache.org/
- https://jrtechs.net/java/gremlin-in-10-minutes
- http://sql2gremlin.com/
- https://neo4j.com/

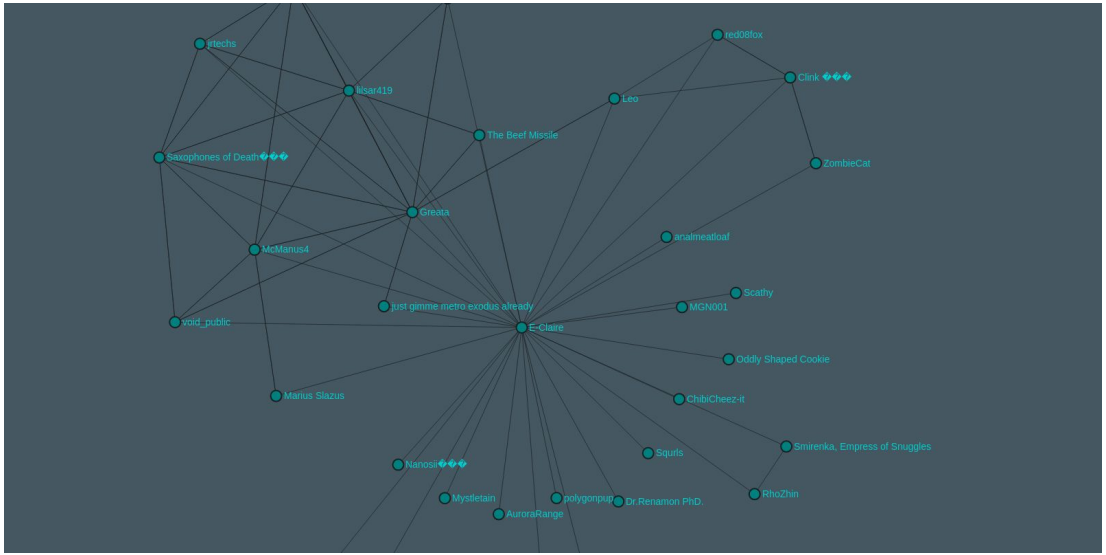



smallette@ubuntu: ~
smallette@ubuntu:~$ bin/gremlin.sh
\,,,/
(o o)
-----oOOo-(3)-oOOo-----
plugin activated: tinkerpop.server
plugin activated: tinkerpop.gephi
plugin activated: tinkerpop.utilities
plugin activated: tinkerpop.tinkergraph
gremlin>




^7

Did I mention Projects?



https://jrtechs.net/steam/

https://github-graphs.com/

Questions?

RITLUG