
neo4j

Sobre Mi !

01 **Max De Marzi**
Neo4j Ingeniero de Ventas

02 **maxdemarzi.com**

03 **@maxdemarzi**

04 **github.com/maxdemarzi**

Cerca de 200 repositorios públicos


75 %
Experiencia


50%
Técnica

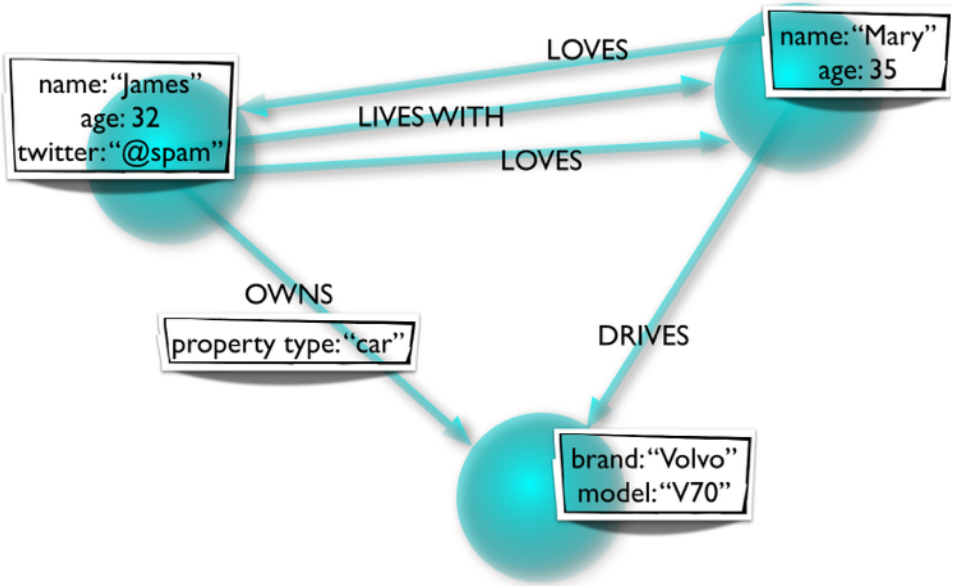

95%
No Importa

Todo lo que importa

Te vas a casa, pensando acerca de grafos

Te

Acerca de



Grafo de Propiedades

Modelo Grafo de Propiedades

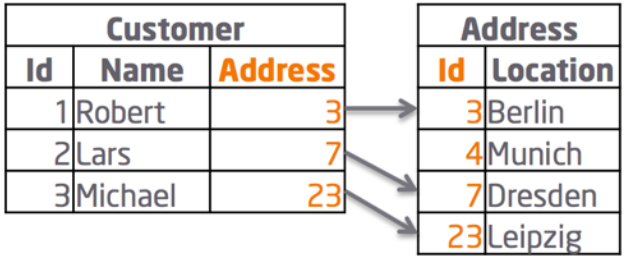
Es super simple.

Todo lo que tienes es:

- **Nodos**
- **Relaciones**
- **Propiedades**

Lo que (probablemente) ya sabes:

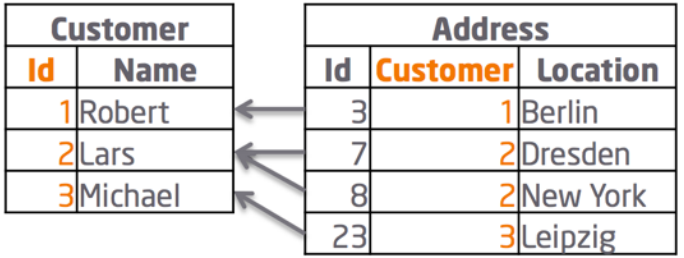
SQL Join Hell (1)



Customer		
Id	Name	Address
1	Robert	3
2	Lars	7
3	Michael	23

Address	
Id	Location
3	Berlin
4	Munich
7	Dresden
23	Leipzig

1:1 Relationship



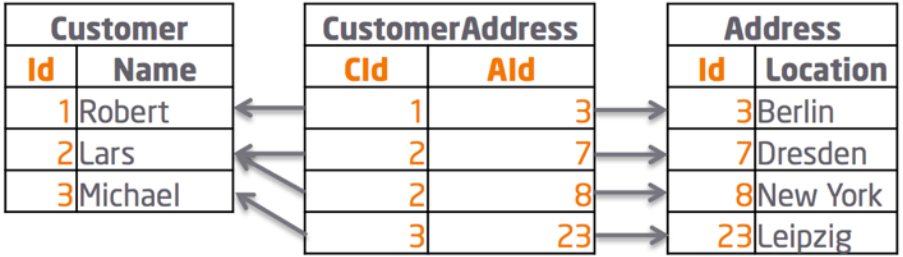
Customer	
Id	Name
1	Robert
2	Lars
3	Michael

Address		
Id	Customer	Location
3	1	Berlin
7	2	Dresden
8	2	New York
23	3	Leipzig

1:n Relationship

Clientes tienen Direcciones

Direcciones tienen Clientes



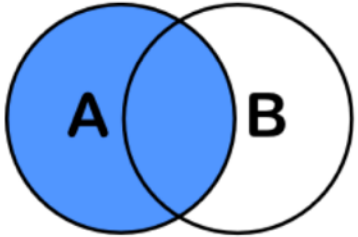
Customer	
Id	Name
1	Robert
2	Lars
3	Michael

CustomerAddress	
CId	AId
1	3
2	7
2	8
3	23

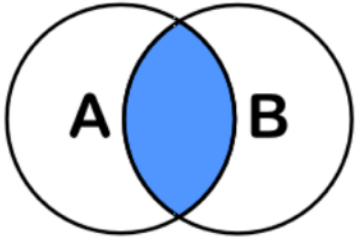
Address	
Id	Location
3	Berlin
7	Dresden
8	New York
23	Leipzig

m:n Relationship

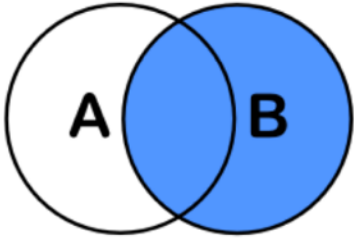
CHEATSHEET SQL JOINS



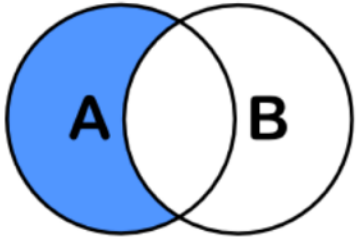

```
SELECT <auswahl>
FROM tabelleA A
LEFT JOIN tabelleB B
ON A.key = B.key
```



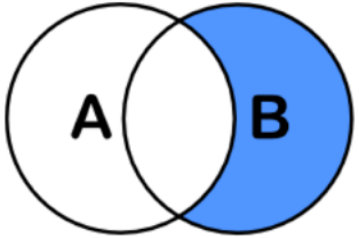

```
SELECT <auswahl>
FROM tabelleA A
INNER JOIN tabelleB B
ON A.key = B.key
```



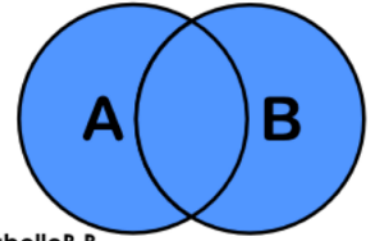

```
SELECT <auswahl>
FROM tabelleA A
RIGHT JOIN tabelleB B
ON A.key = B.key
```



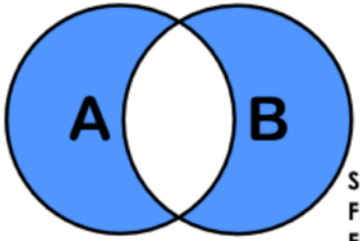

```
SELECT <auswahl>
FROM tabelleA A
LEFT JOIN tabelleB B
ON A.key = B.key
WHERE B.key IS NULL
```



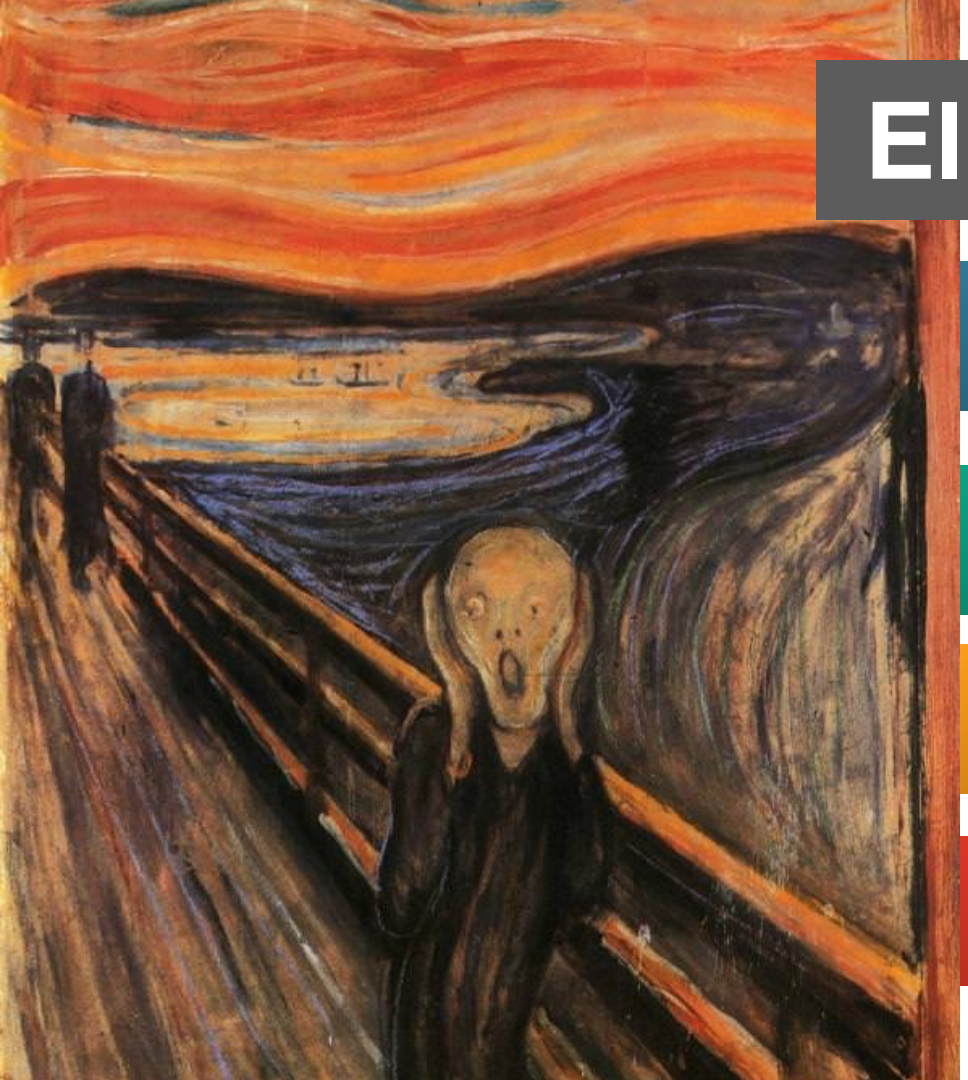

```
SELECT <auswahl>
FROM tabelleA A
RIGHT JOIN tabelleB B
ON A.key = B.key
WHERE A.key IS NULL
```




```
SELECT <auswahl>
FROM tabelleA A
FULL OUTER JOIN tabelleB B
ON A.key = B.key
```




```
SELECT <auswahl>
FROM tabelleA A
FULL OUTER JOIN tabelleB B
ON A.key = B.key
WHERE A.key IS NULL
OR B.key IS NULL
```



El Problema

1. Las uniones se ejecutan **cada vez** que se consulta la relación
2. Ejecutar una unión significa **buscar** una clave
3. **Índice B-Tree: O(log(n))**
   Tus datos crecen 10 veces, tu velocidad se reduce a la mitad
4. Más datos = más búsquedas
   Más lento

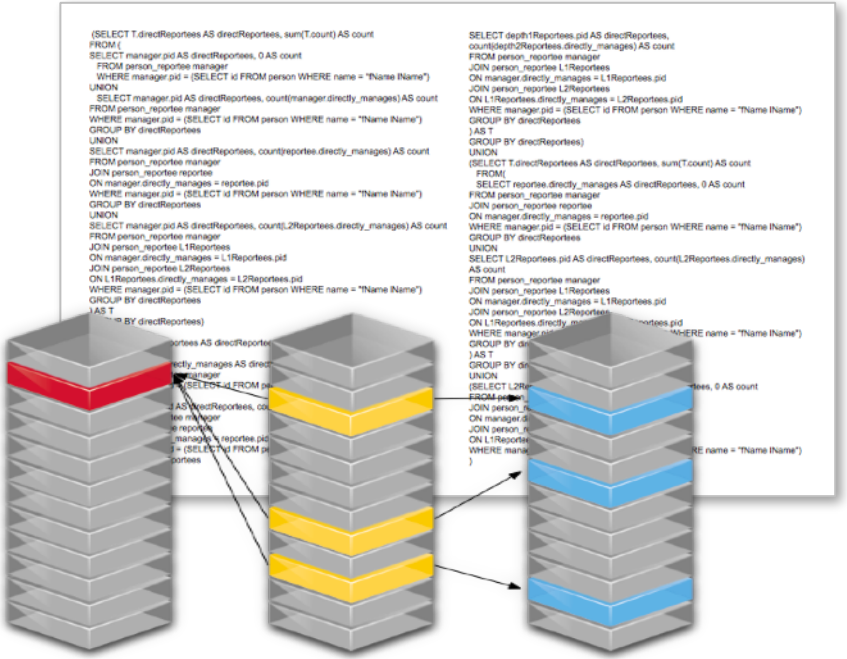
Las bases de datos **Relacionales** no pueden manejar las **Relaciones**

1 **Modelo Equivocado**
No pueden modelar o almacenar relaciones sin complejidad

2 **Más Lentos**
La velocidad cae cuando los datos o las uniones crecen

3 **Idioma Equivocado**
SQL se construyó con la teoría de conjuntos, no de grafos

4 **No es Flexible**
Nuevos tipos de datos y relaciones requieren un rediseño del esquema



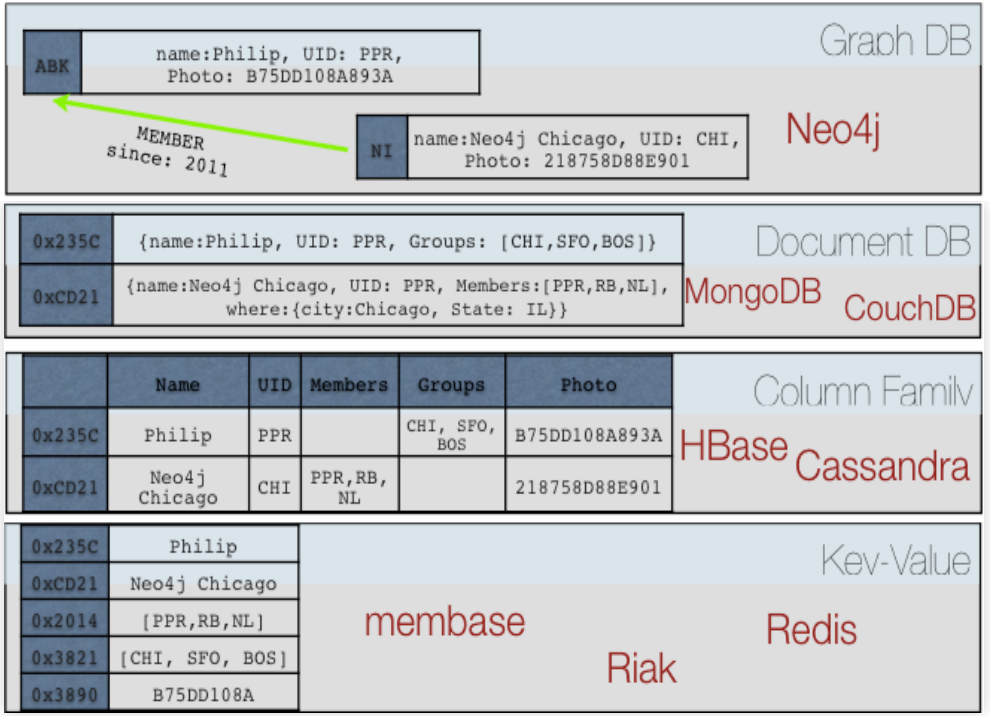
Las bases de datos **NoSQL** no pueden manejar las **Relaciones**

1 **Modelo Equivocado**
No pueden modelar o almacenar relaciones sin complejidad

2 **Más Lentos**
La velocidad cae al intentar de unir datos en la aplicación

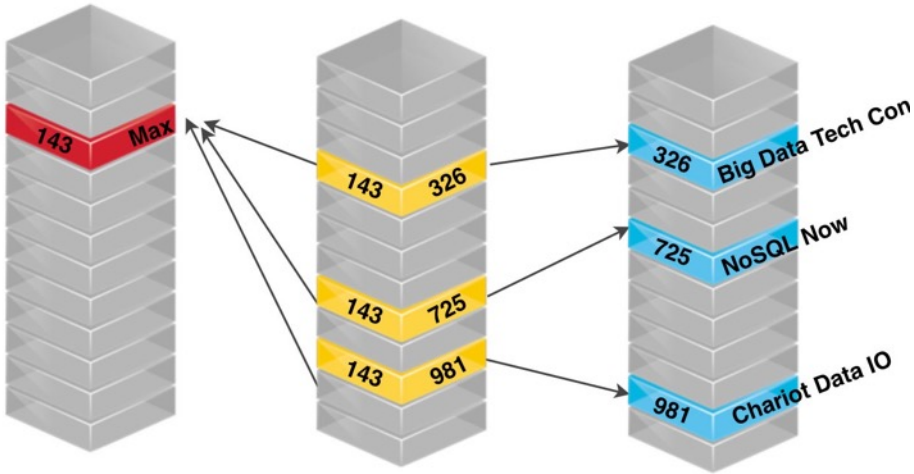
3 **Idiomas Incorrectos**
Muchos idiomas raros "casi sql" terribles para las uniones

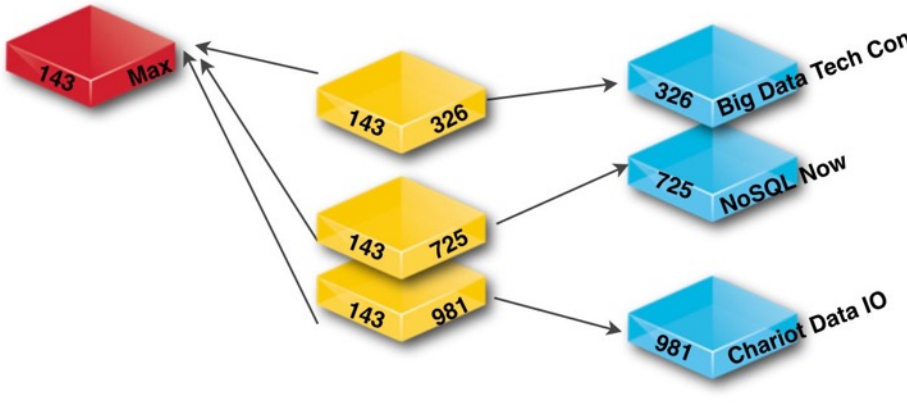
4 **No ACID (no tienen transacciones)**
Eventualmente Consistente significa Eventualmente Corrupto



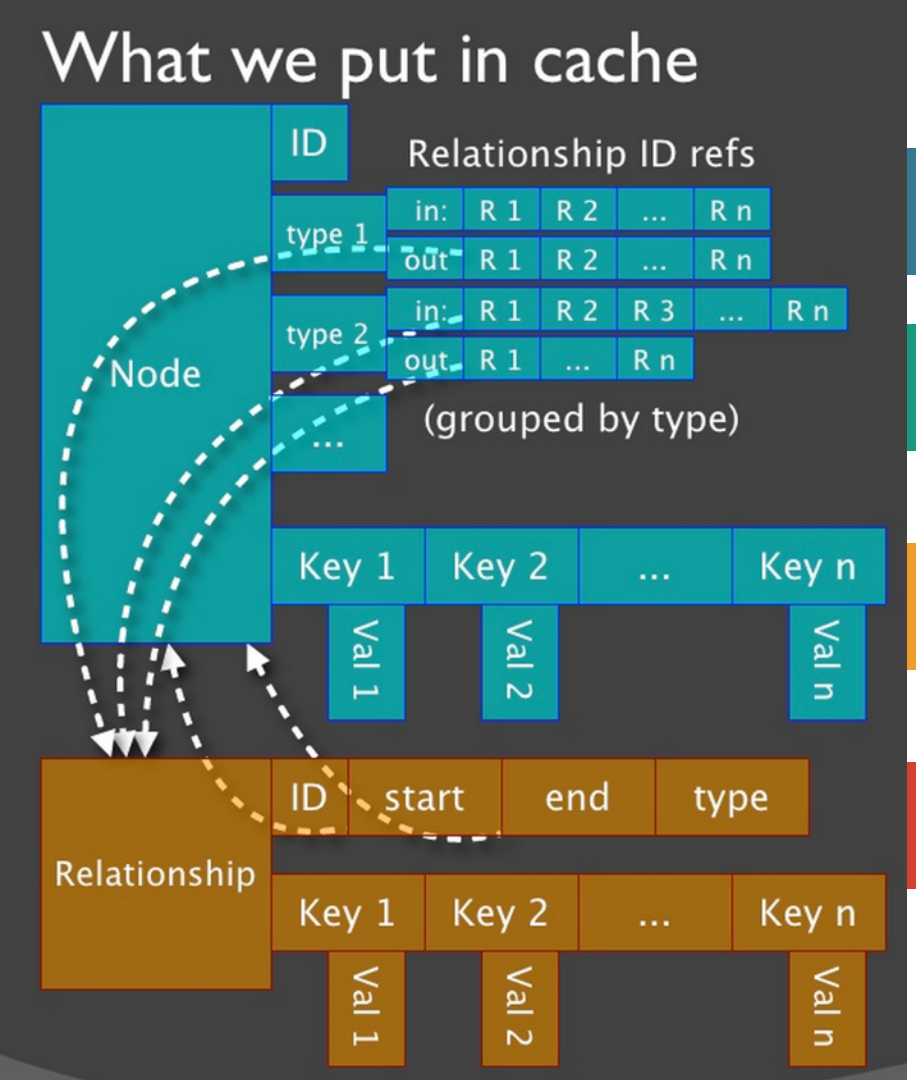
Neo4j: Misma información, diseño diferente

No más tablas, no más claves externas, no más uniones





¿Cuál es la receta secreta?



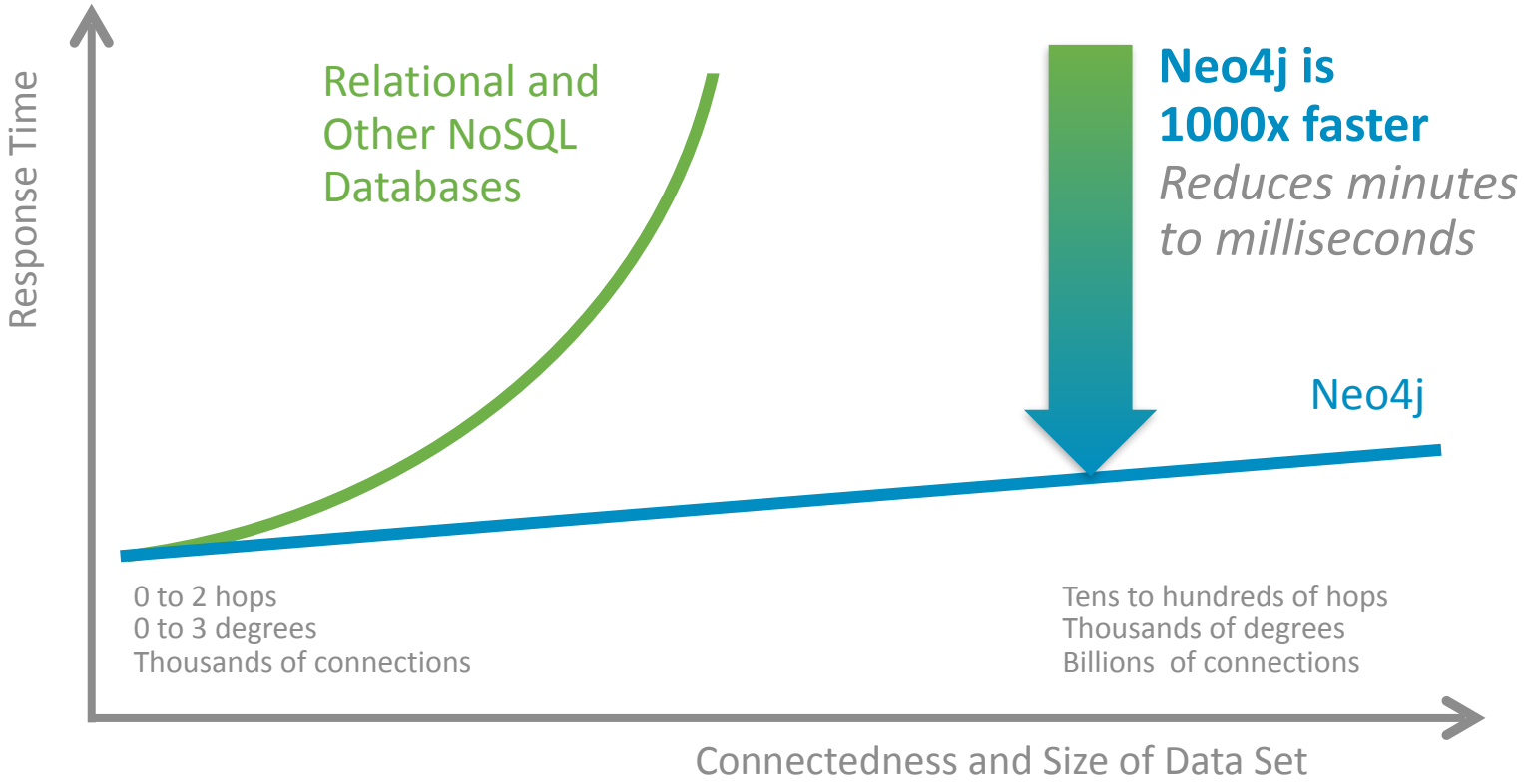
La Receta Secreta de Neo4j

1. **Acceso directo en lugar de búsquedas**
2. **Registros de tamaño fijo en dos formaciones**
3. **“Uniones” desde la creación**
4. **A través de esta estructura de datos**

Saltamos de la colección de nodos a la colección de relaciones y otra vez a nodos

Respuestas en Tiempo Real

Se quedan igual no importa que los datos crecen



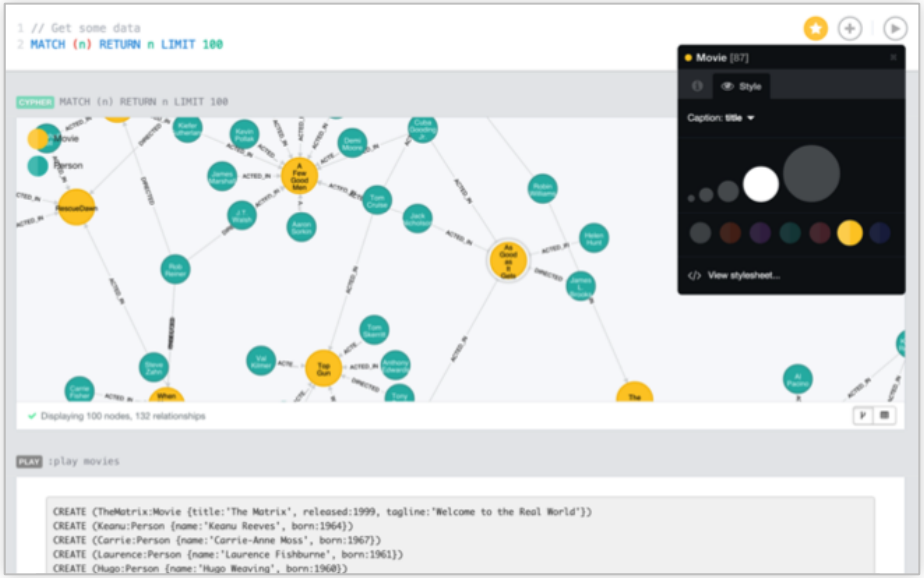
Reimagina tus datos como un Grafo

1 **Modelo Correcto**
Grafos simplifican cómo piensas

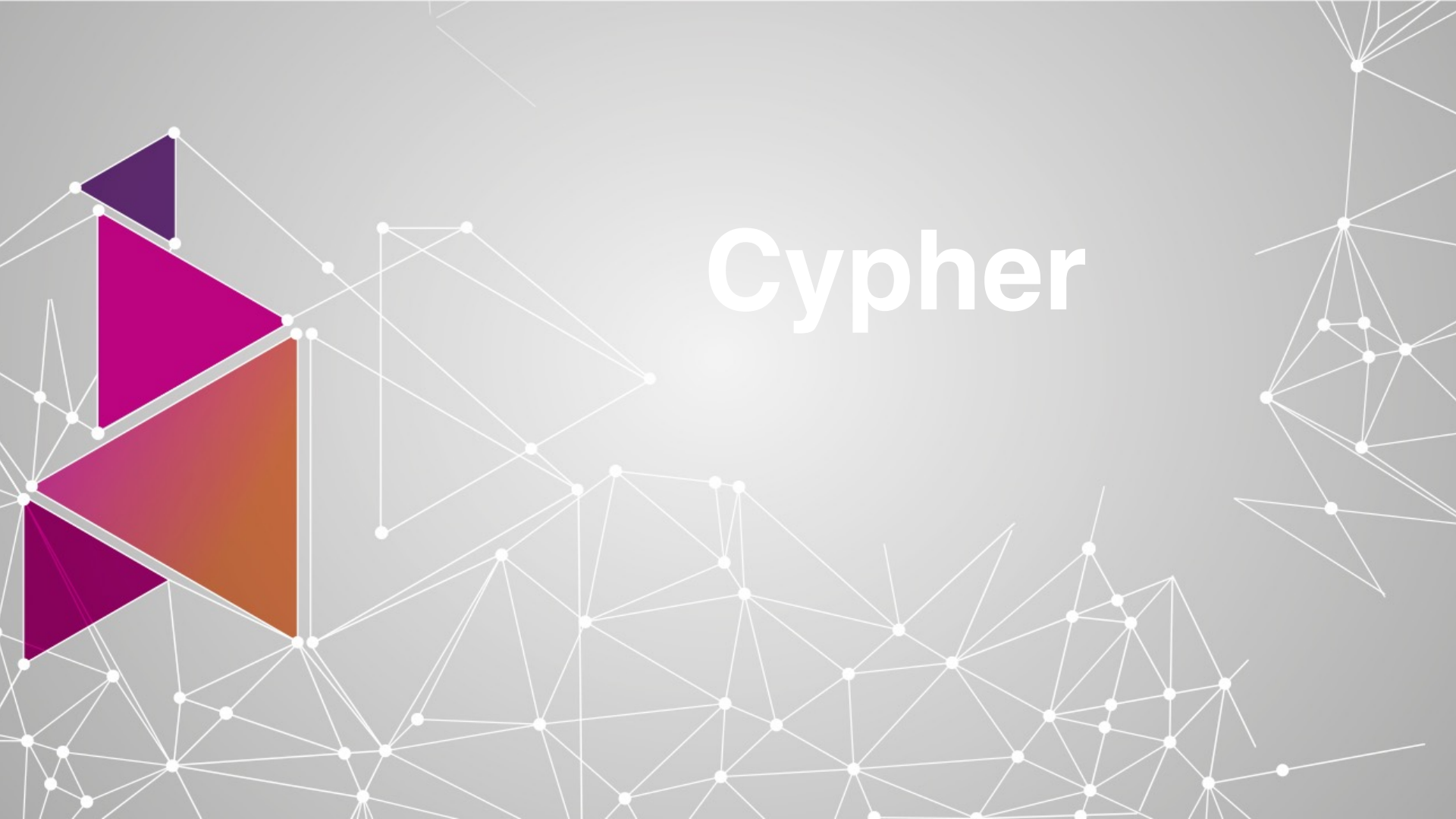
2 **Más rápido**
Encuentra relaciones en tiempo real

3 **Lenguaje Correcto**
Cypher fue creado para Grafos

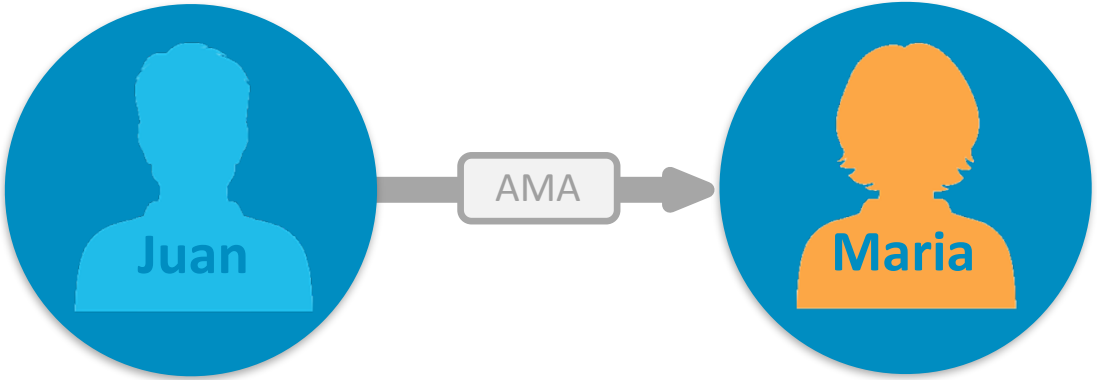
4 **Flexible y Consistente**
Desarrolla tu esquema sin problemas mientras mantiene transacciones



Ágil, alto rendimiento y escalable sin sacrificio


Cypher

Cypher: Lenguaje de Consulta Potente y Expresivo



Nodo

Nodo

```
MATCH (:Persona { nombre:“Juan”} ) -[:AMA]-> (:Persona { nombre:“Maria”} )
```

Etiqueta Propiedad

Etiqueta Propiedad

Expresa Consultas Complejas fácilmente con Cypher

SQL Query

```
(SELECT T.directReportees AS directReportees, sum(T.count) AS count
FROM (
SELECT manager.pid AS directReportees, 0 AS count
  FROM person_reportee manager
  WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
UNION
  SELECT manager.pid AS directReportees, count(manager.directly_manages) AS count
FROM person_reportee manager
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
GROUP BY directReportees
UNION
SELECT manager.pid AS directReportees, count(reportee.directly_manages) AS count
FROM person_reportee manager
JOIN person_reportee reportee
ON manager.directly_manages = reportee.pid
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
GROUP BY directReportees
UNION
SELECT manager.pid AS directReportees, count(L2Reportees.directly_manages) AS count
FROM person_reportee manager
JOIN person_reportee L1Reportees
ON manager.directly_manages = L1Reportees.pid
JOIN person_reportee L2Reportees
ON L1Reportees.directly_manages = L2Reportees.pid
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
GROUP BY directReportees
) AS T
GROUP BY directReportees)
UNION
(SELECT T.directReportees AS directReportees, sum(T.count) AS count
FROM (
SELECT manager.directly_manages AS directReportees, 0 AS count
FROM person_reportee manager
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
UNION
SELECT reportee.pid AS directReportees, count(reportee.directly_manages) AS count
FROM person_reportee manager
JOIN person_reportee reportee
ON manager.directly_manages = reportee.pid
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
GROUP BY directReportees
UNION
SELECT depth1Reportees.pid AS directReportees,
count(depth2Reportees.directly_manages) AS count
FROM person_reportee manager
JOIN person_reportee L1Reportees
ON manager.directly_manages = L1Reportees.pid
JOIN person_reportee L2Reportees
ON L1Reportees.directly_manages = L2Reportees.pid
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
GROUP BY directReportees
) AS T
GROUP BY directReportees)
UNION
(SELECT T.directReportees AS directReportees, sum(T.count) AS count
  FROM(
  SELECT reportee.directly_manages AS directReportees, 0 AS count
FROM person_reportee manager
JOIN person_reportee reportee
ON manager.directly_manages = reportee.pid
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
GROUP BY directReportees
UNION
SELECT L2Reportees.pid AS directReportees, count(L2Reportees.directly_manages)
AS count
FROM person_reportee manager
JOIN person_reportee L1Reportees
ON manager.directly_manages = L1Reportees.pid
JOIN person_reportee L2Reportees
ON L1Reportees.directly_manages = L2Reportees.pid
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
GROUP BY directReportees
) AS T
GROUP BY directReportees)
UNION
(SELECT L2Reportees.directly_manages AS directReportees, 0 AS count
FROM person_reportee manager
JOIN person_reportee L1Reportees
ON manager.directly_manages = L1Reportees.pid
JOIN person_reportee L2Reportees
ON L1Reportees.directly_manages = L2Reportees.pid
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
)
```

Cypher Query

```
MATCH (jefe)-[:MANEJA*0..3]->(empleado),
      (empleado)-[:MANEJA*1..3]->(otro)
WHERE boss.name = “Fulano de Tal”
RETURN empleado.name AS Empleado,
  count(otro) AS Total
```

Encuentra todos los empleados directos y aquellos que le reportan, hasta 3 niveles más abajo

Neo4j para Servicios Financieros

Plan

Tipos de Fraude

- Fraude de tarjeta de credito
- Identidades Sintéticas y Anillos de Fraude
- Fraude de Seguro

Tipos de Análisis

- Análisis Tradicional
- Análisis basado en grafos

Detección y Prevención de Fraude

…pero antes de entrar en eso …

- ¿Qué no es fraude?

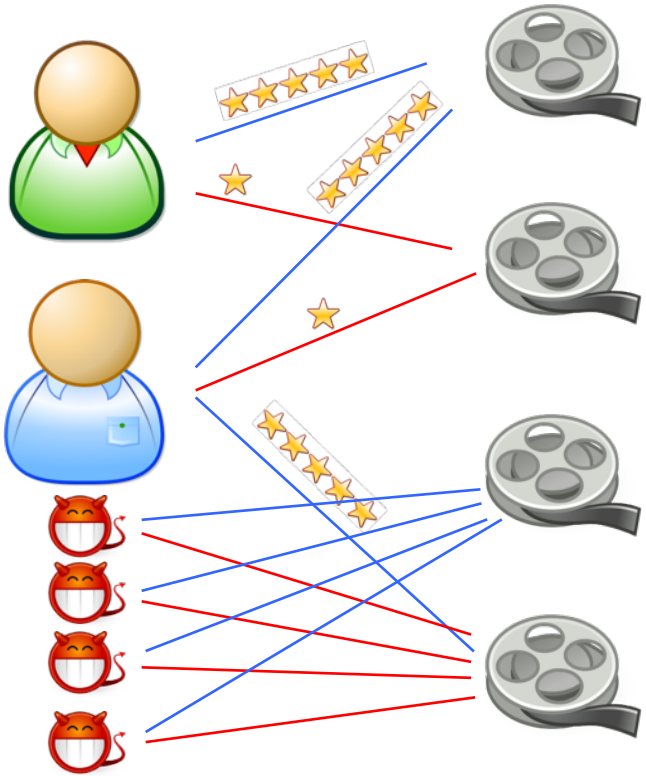
Aprende de los Expertos

- Alex Beutel, CMU
- Leman Akoglu, Stony Brook
- Christos Faloutsos, CMU



- Graph-Based User Behavior Modeling: From Prediction to Fraud Detection
- http://www.cs.cmu.edu/~abeutel/kdd2015_tutorial/

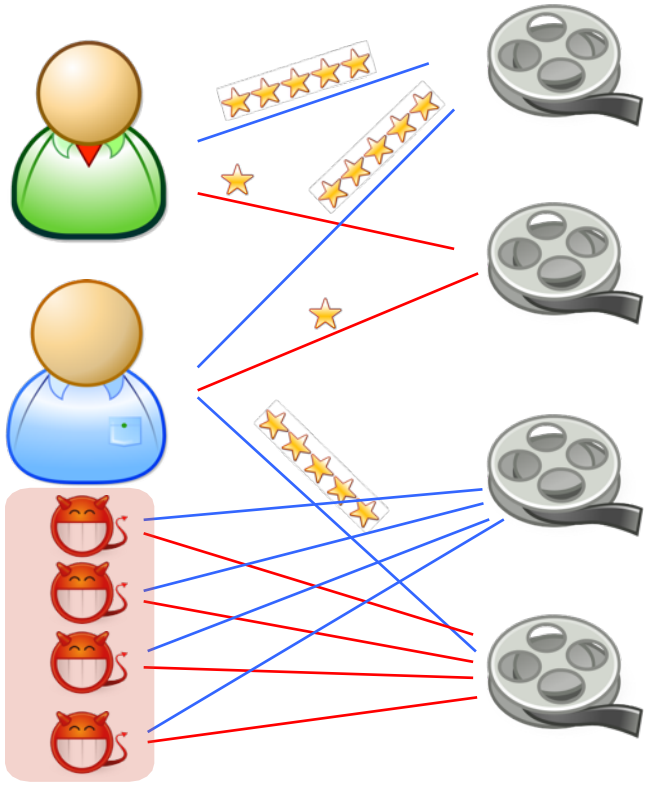
Problemas de Comportamiento del Usuario



- ¿Cómo podemos entender el comportamiento normal del usuario?

NETFLIX

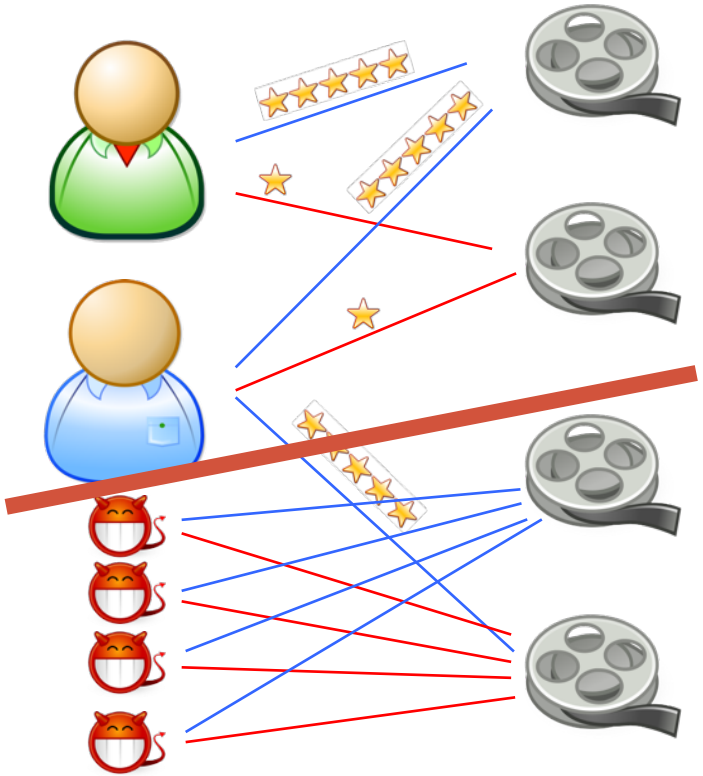
Problemas de Comportamiento del Usuario



- ¿Cómo podemos entender el comportamiento normal del usuario?
- ¿Cómo podemos encontrar un comportamiento sospechoso?

NETFLIX

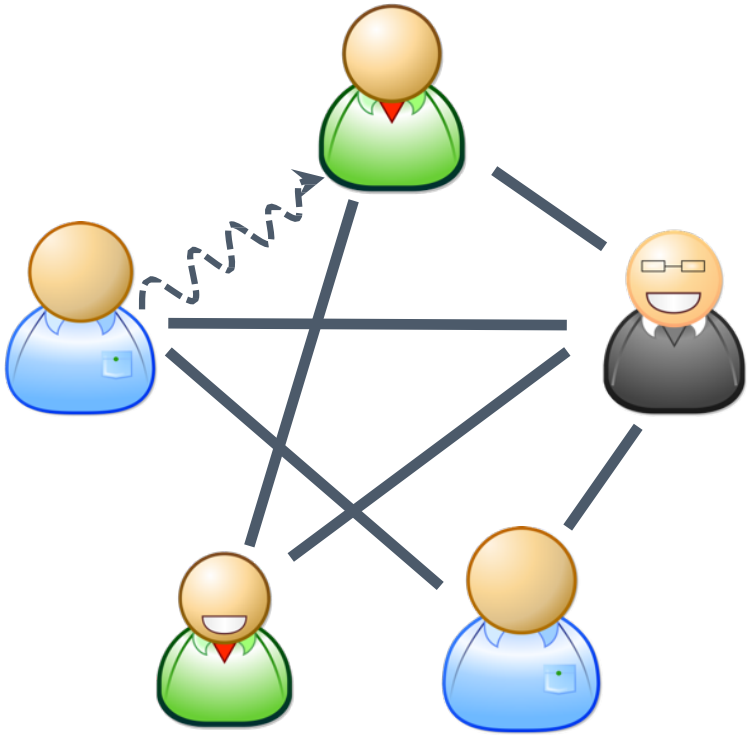
Problemas de Comportamiento del Usuario



- ¿Cómo podemos entender el comportamiento normal del usuario?
- ¿Cómo podemos encontrar un comportamiento sospechoso?
- ¿Cómo podemos distinguir los dos?

NETFLIX

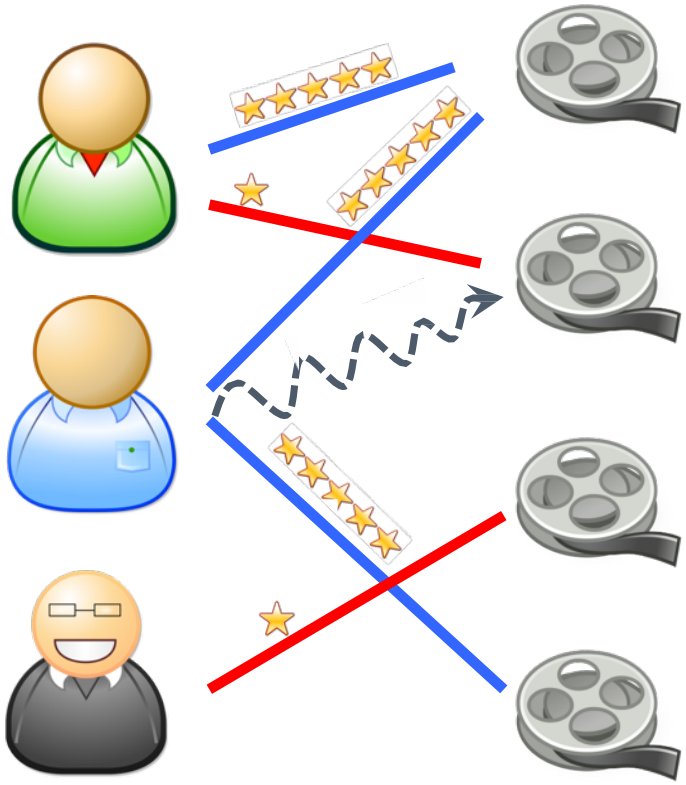
Modelando Comportamiento "Normal"



- Predecir Relaciones (usuarios similares)

Si todos tus amigos tienen 35 años de edad tu probablemente tienes 35 años de edad también

Modelando Comportamiento "Normal"



- Predecir relaciones (películas que debería ver)

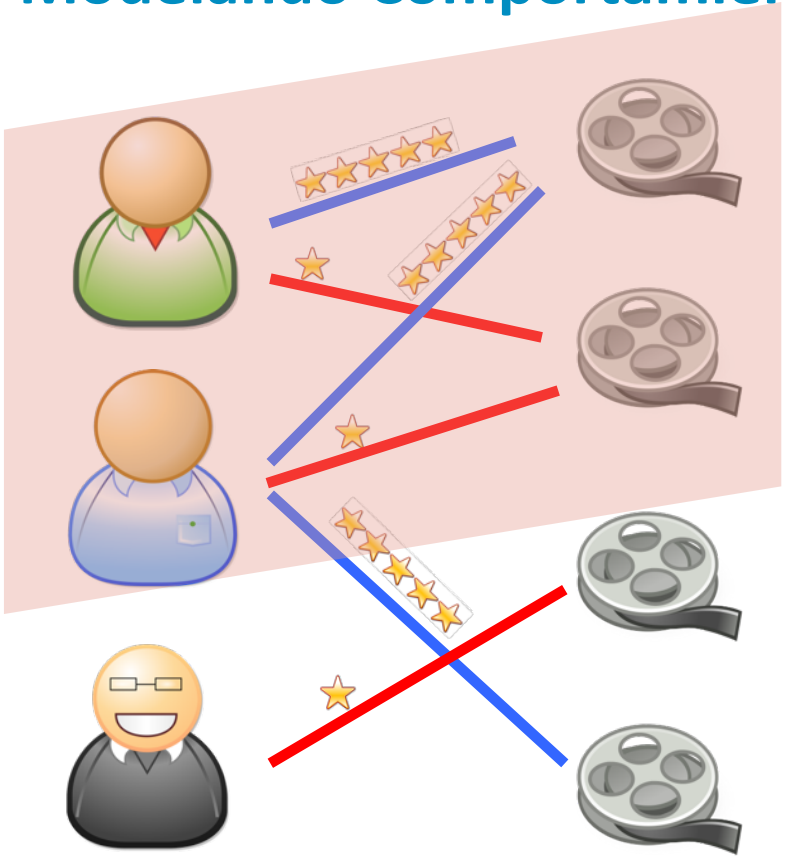
Qué acciones debes tomar?

¿Cuántas estrellas debo dar a “Amores Perros”?

```
MATCH (yo:Usuario {id:”max”})-[r1:VIO]->(m:Pelicula)
<-[r2:VIO]-(:Usuario)-[r3:VIO]->
(m2:Pelicula {titulo:”Amores Perros”})
WHERE ABS(r1.stars-r2.stars) <=1
RETURN AVG(r3.stars)
```

```
+--------------------+
| AVG(rating.stars)  |
+--------------------+
| 3.600526612751551  |
+--------------------+
```

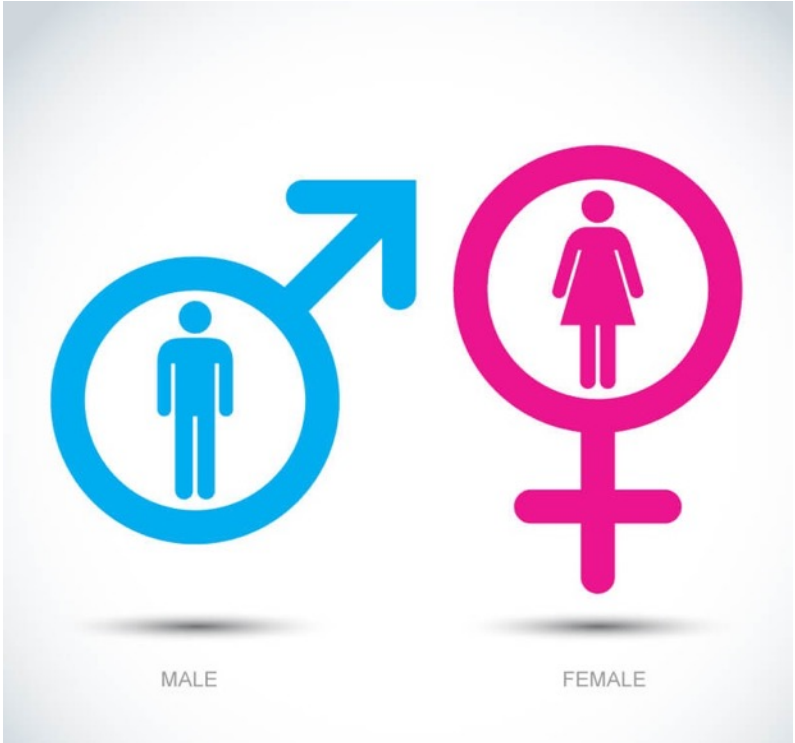
Modelando Comportamiento "Normal"



- Predecir Relaciones
- Predecir Atributos de Nodo
- Predecir Atributos de Relaciones
- Agrupamiento y detección de comunidad

Demografía: Edad

Demografía: Sexo



¿A las niñas les gustan las películas que le gustan a otras niñas?

gender	age	user 1	total	matching	percent of max
F	1	368	5292	0.069538927	1
F	18	1403	23247	0.060351873	0.867886179
F	25	2074	39240	0.05285423	0.760066813
F	35	1296	23493	0.055165368	0.793301983
F	45	559	9643	0.057969512	0.83362678
F	50	508	9568	0.053093645	0.763509706
F	56	362	5432	0.066642121	0.958342671
M	1	697	12058	0.057803948	0.831245898
M	18	3747	74656	0.050190206	0.72175698
M	25	6520	140239	0.04649206	0.668576036
M	35	3687	74902	0.04922432	0.70786712
M	45	1385	28118	0.049256704	0.708332818
M	50	1277	26396	0.048378542	0.695704471
M	56	838	15472	0.054162358	0.778878254

¡Sí! A las niñas les gustan las películas que a otras niñas les gusta

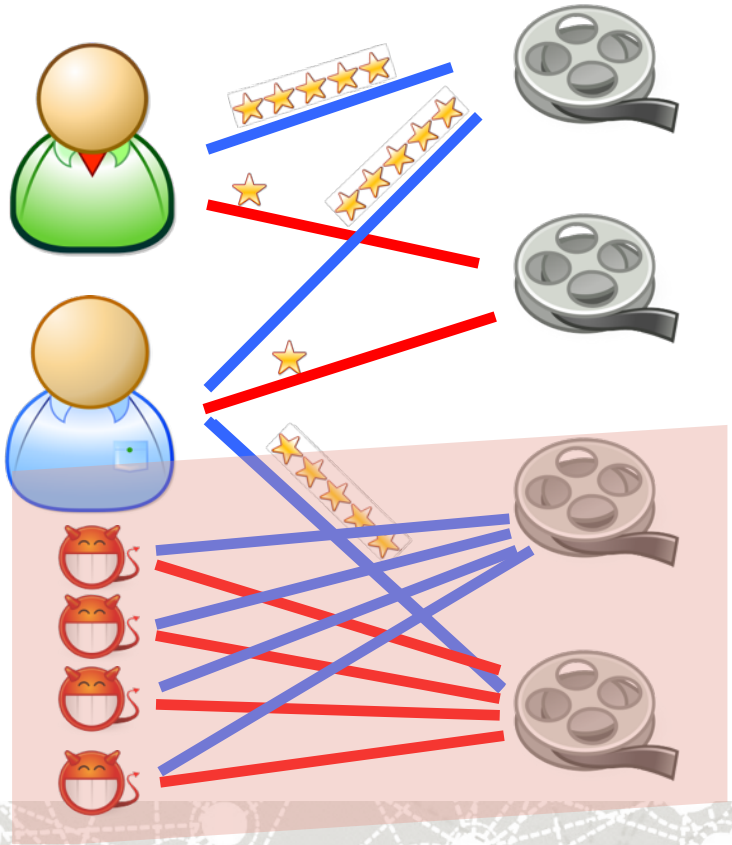
Predecir la Clasificación de peliculas puramente basado en demografía

MATCH (u:Usuario)-[r:VIO]->(m:Pelicula {titulo:”Diarios Motocicleta”})
WHERE u.edad = “20-29” AND u.sexo = "M"
RETURN AVG(r.stars)

```
+--------------------+
| AVG(rating.stars)  |
+--------------------+
| 4.142857142857143  |
+--------------------+
```

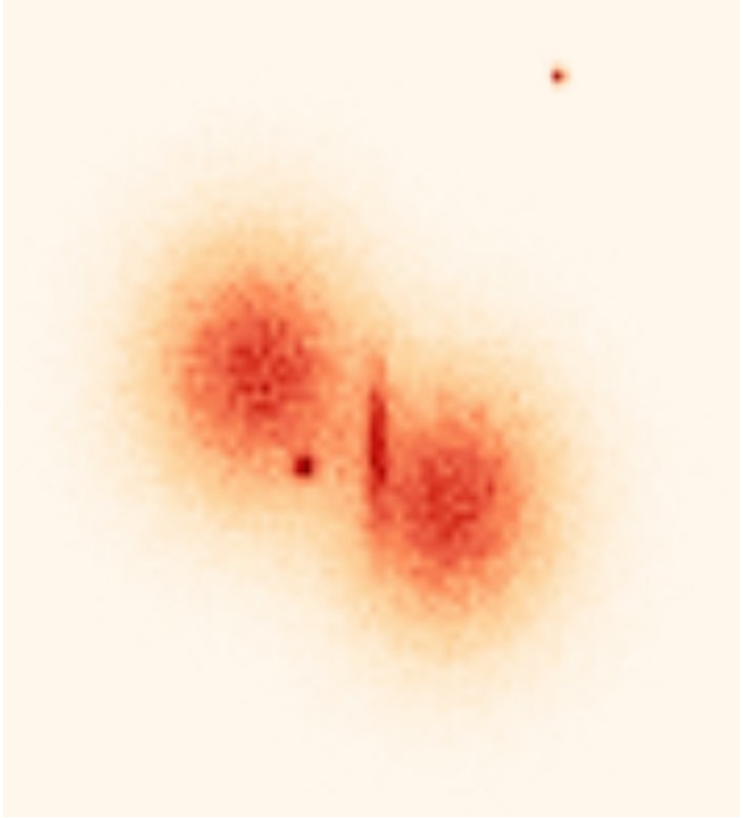
El promedio =>

Modelando Comportamiento "Anormal"



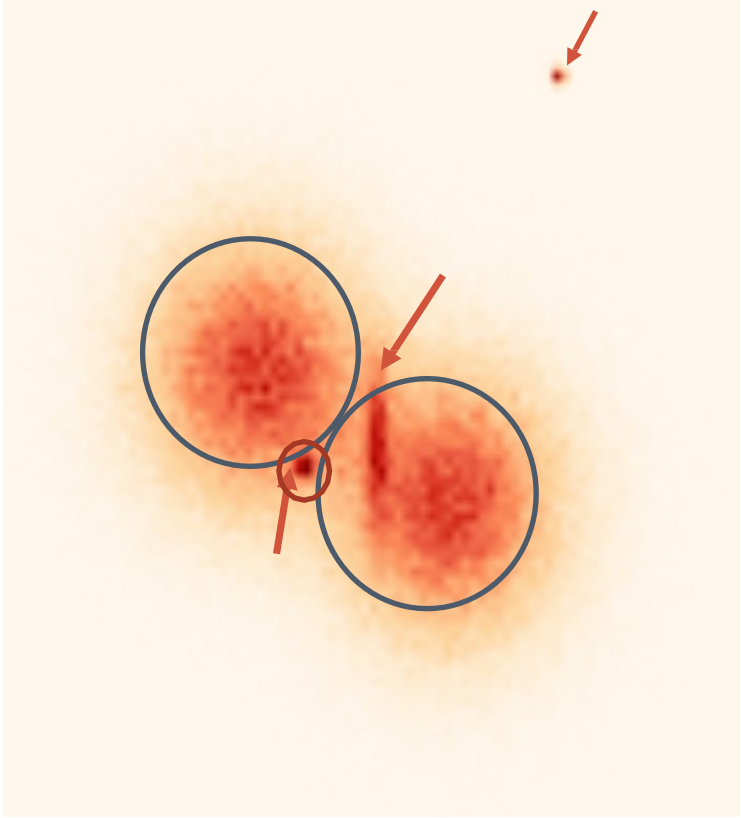
- Predecir Relaciones
- Predecir Atributos de Nodo
- Predecir Atributos de Relaciones
- Agrupamiento y detección de comunidad
- Detección de Fraude

Modelando el Comportamiento del Usuario



- Modelar usuarios normales y detectar anomalías son dos aspectos para comprender el comportamiento del usuario

Modelando el Comportamiento del Usuario



- Modelar usuarios normales y detectar anomalías son dos aspectos para comprender el comportamiento del usuario
- Los modelos complejos pueden capturar patrones normales y anormales

Dos lados de la misma Moneda

Recomendaciones

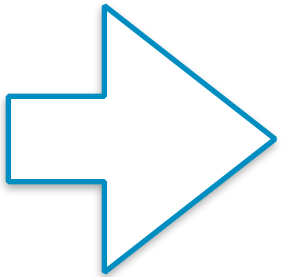
- Agregue la relación **que no** existe

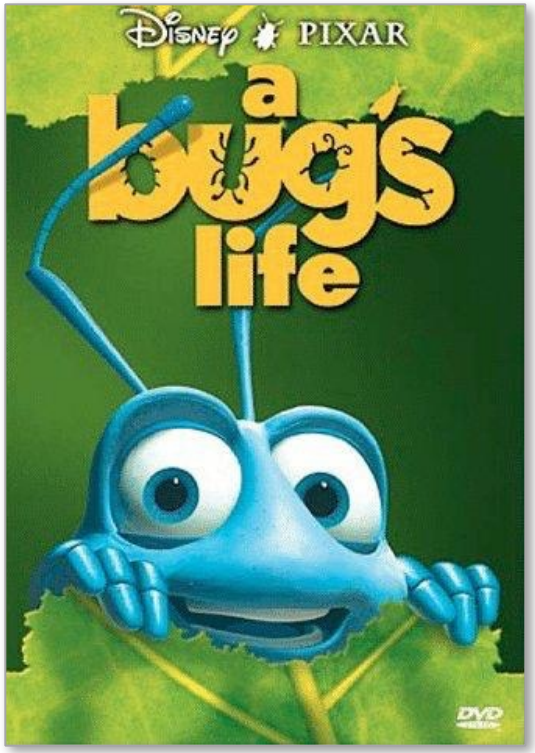
Detección de Fraude

- Encuentra las relaciones que **no deberían** existir

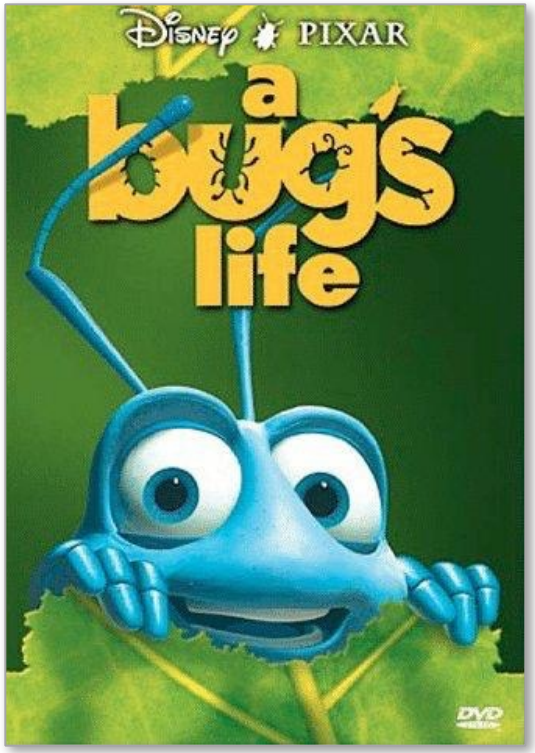
Sistemas de Recomendación

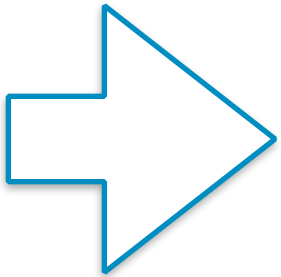
Recomendación Simple

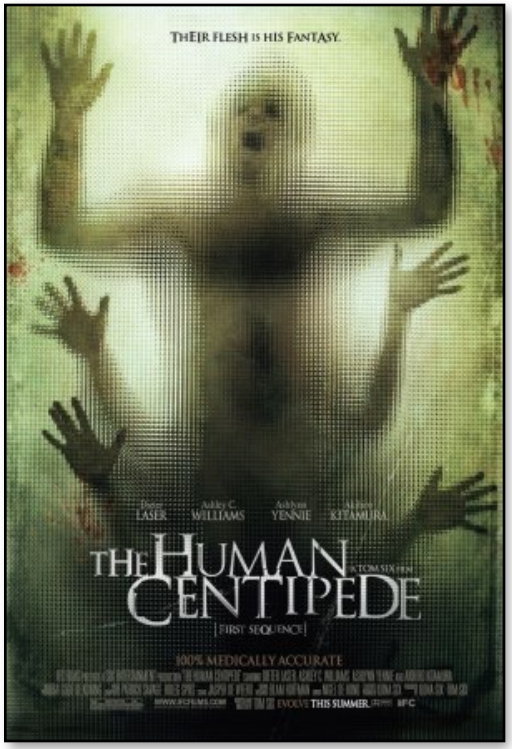




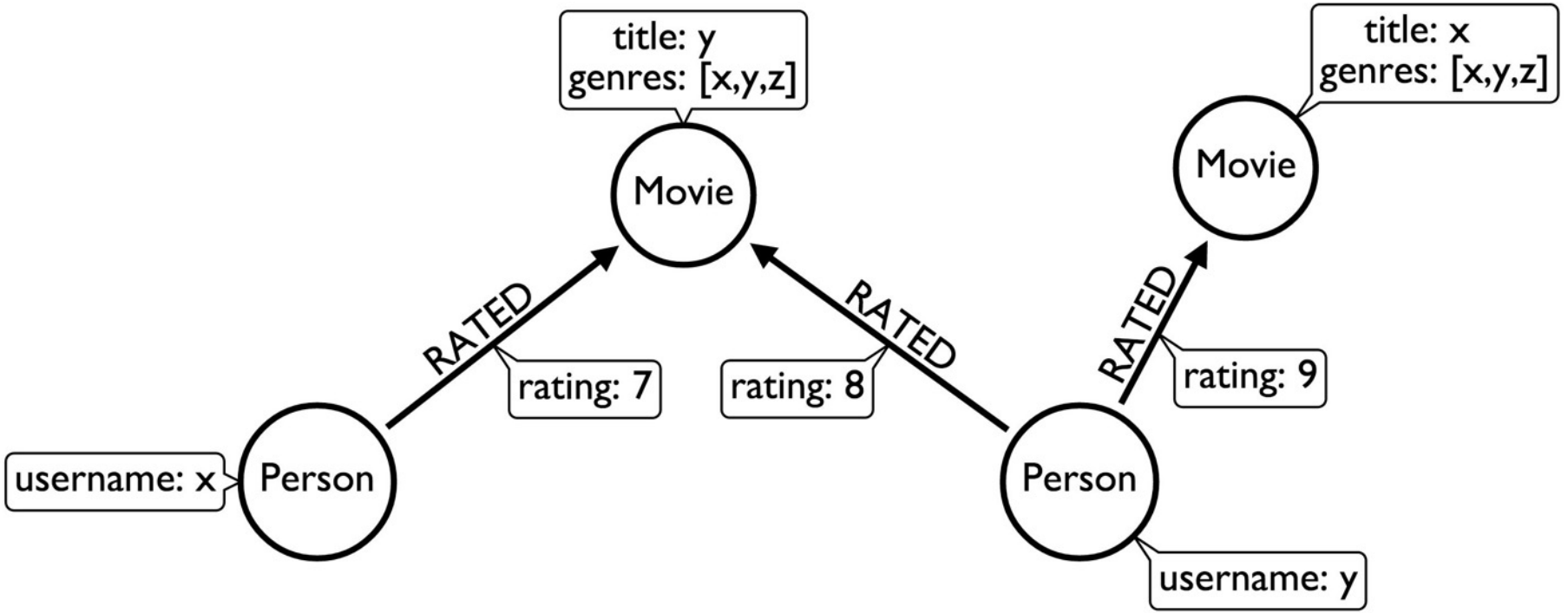
Recomendación Simple







Modelo de Datos de Películas



Recomendación de Película

¿Cuáles son las mejores 25 películas?

- que no he visto
- con los mismos géneros que Toy Story
- con altas calificaciones
- por personas que les gustó Toy Story

```
MATCH (viste:Pelicula {titulo:"Toy Story”}) <-[r1:VIO]- () -[r2:VIO]-> (no_viste:Pelicula)
WHERE r1.estrella > 7 AND r2.estrella > 7
AND viste.generos = no_viste.generos
AND NOT( (p:Persona) -[:VIO]-> (no_viste) )
AND p.usuario IN [“maxdemarzi”,”janedoe”,”jamesdean”]
RETURN no_viste.titulo, COUNT(*)
ORDER BY COUNT(*) DESC
LIMIT 25
```

Sistemas de Fraude

Fraude de Tarjeta de Credito

FEB 12 2014

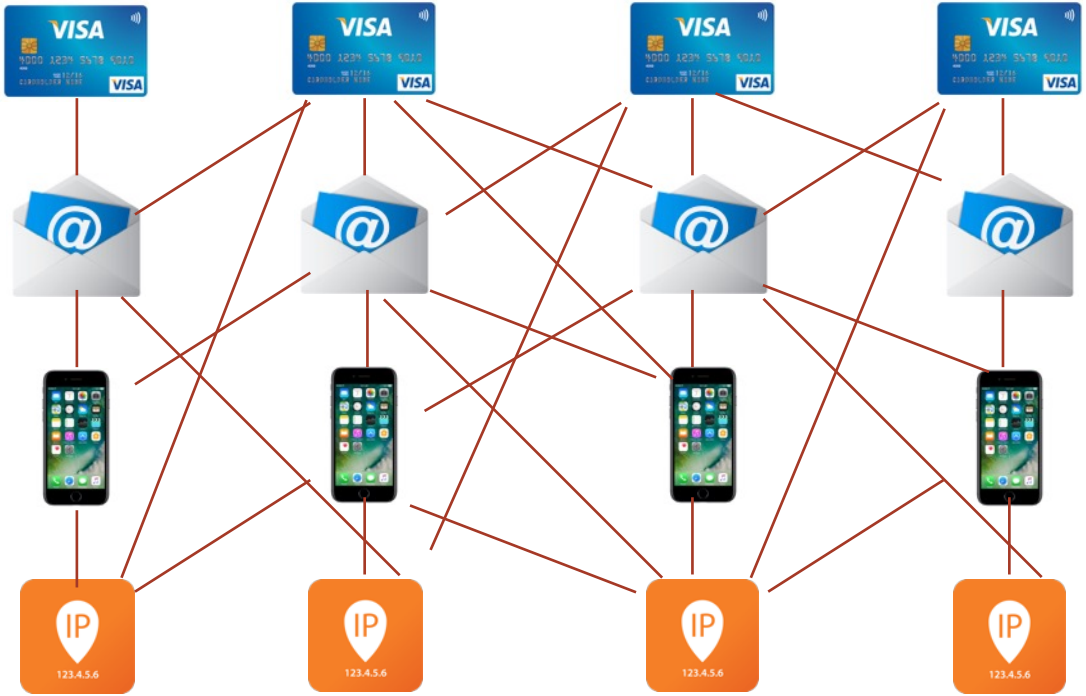
8 COMMENTS

JAVA, PROBLEMS, RANDOM

EDIT

ONLINE PAYMENT RISK MANAGEMENT WITH NEO4J

Referencia Cruzada



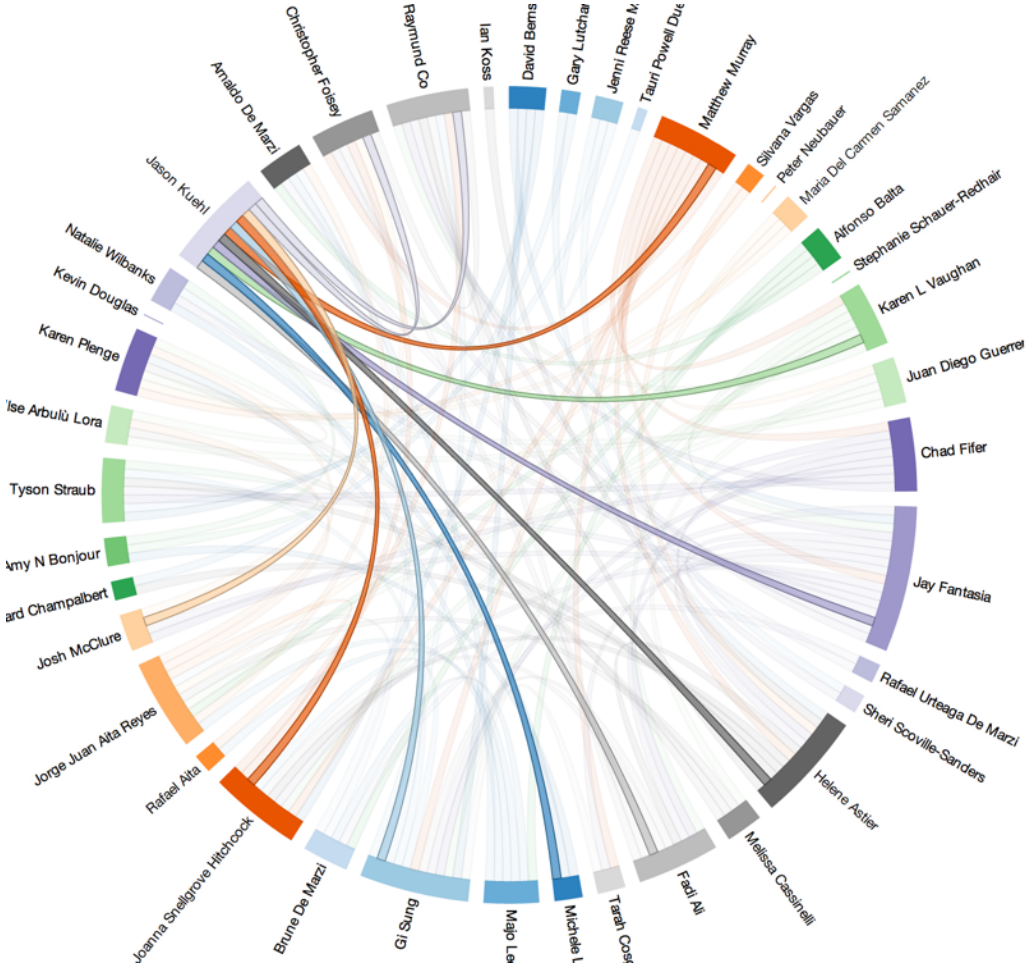
Una tarjeta

Un correo electrónico

Un telefono

Una dirección IP

Subgrafos

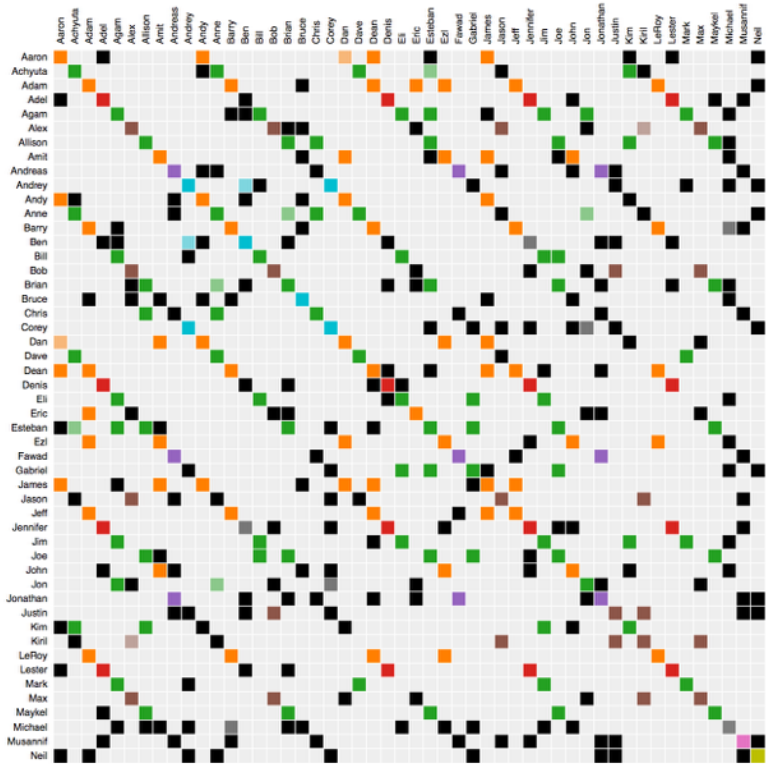


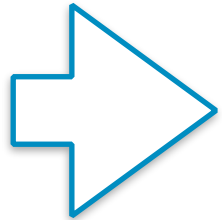
El vecindario de un Nodo

Gráfos como Matrices

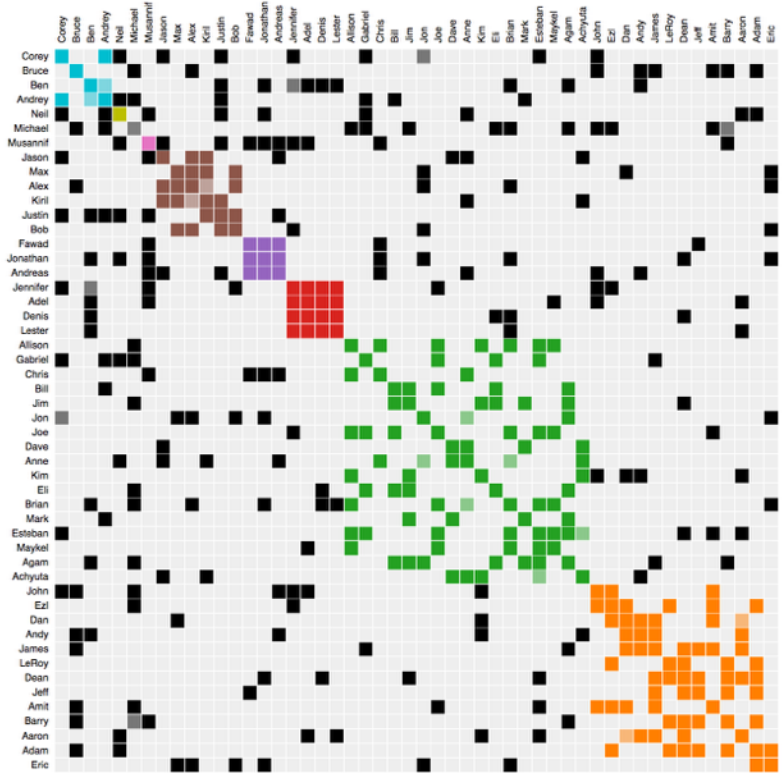
La Agrupación da Claridad

Encontrando pandillas de Nodos

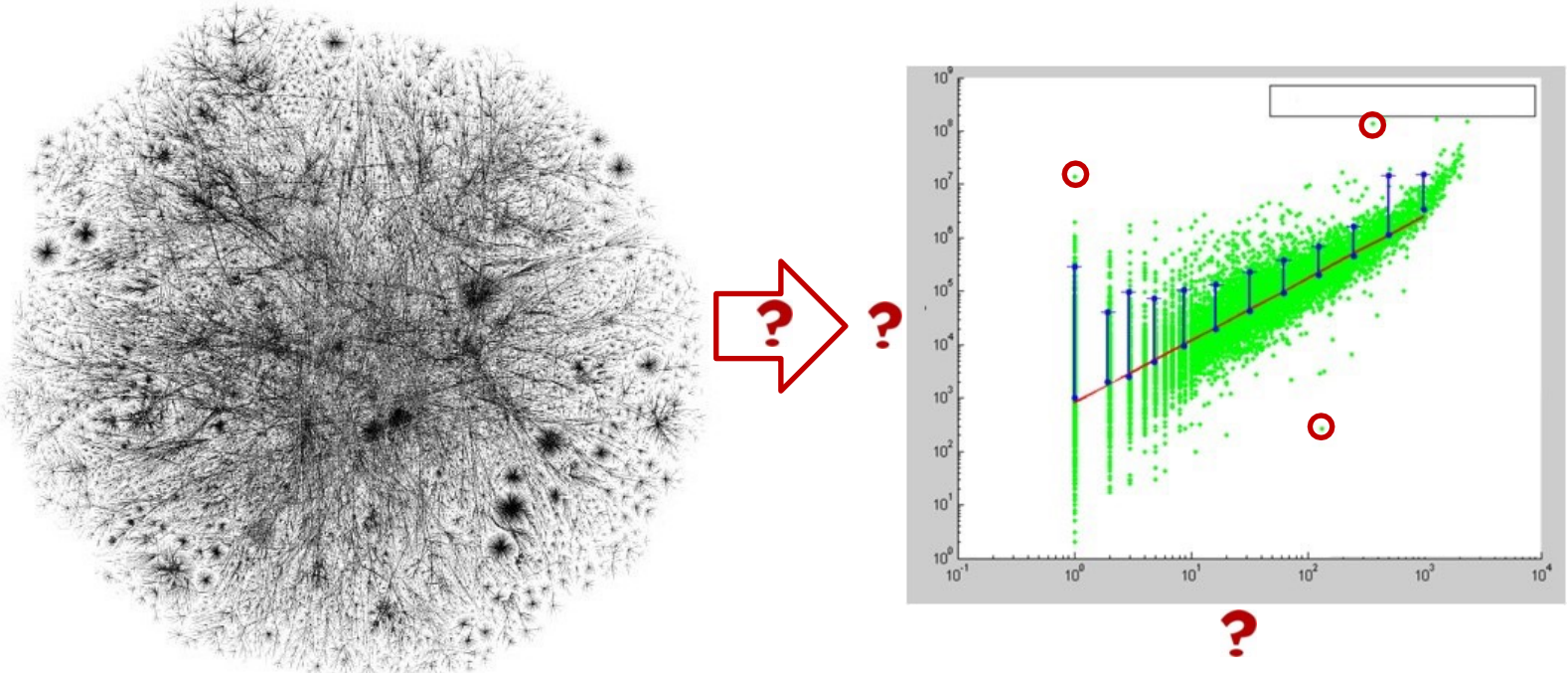




Link

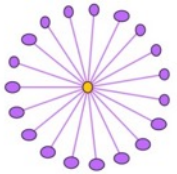


Patrones Egocéntricos

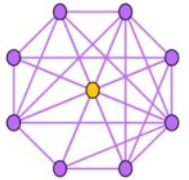


Patrones Egocéntricos

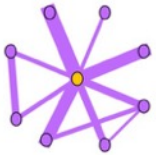
N_i: numero de vecinos de nodo i



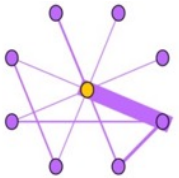
E_i: numero de relaciones entre los vecinos de i



W_i: peso total de las relaciones de i

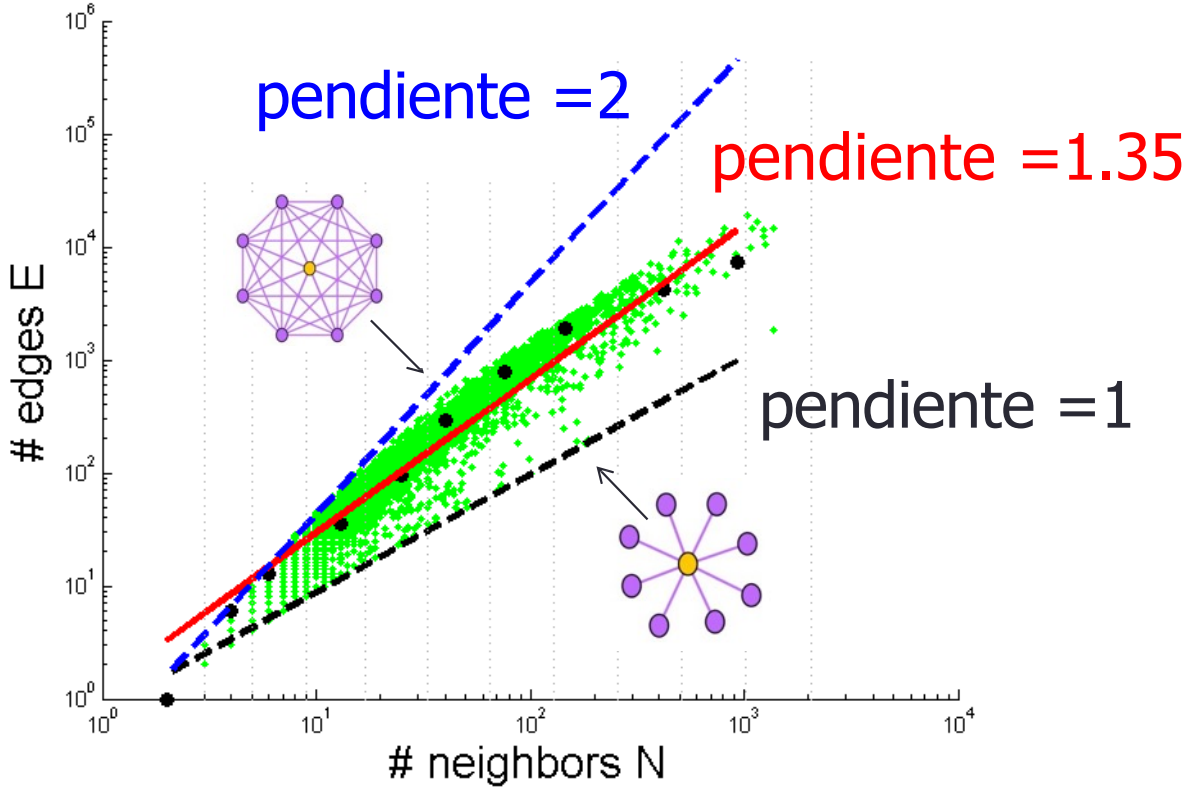


$\lambda_{w,i}$: valor propio principal de la matriz de adyacencia ponderada de i



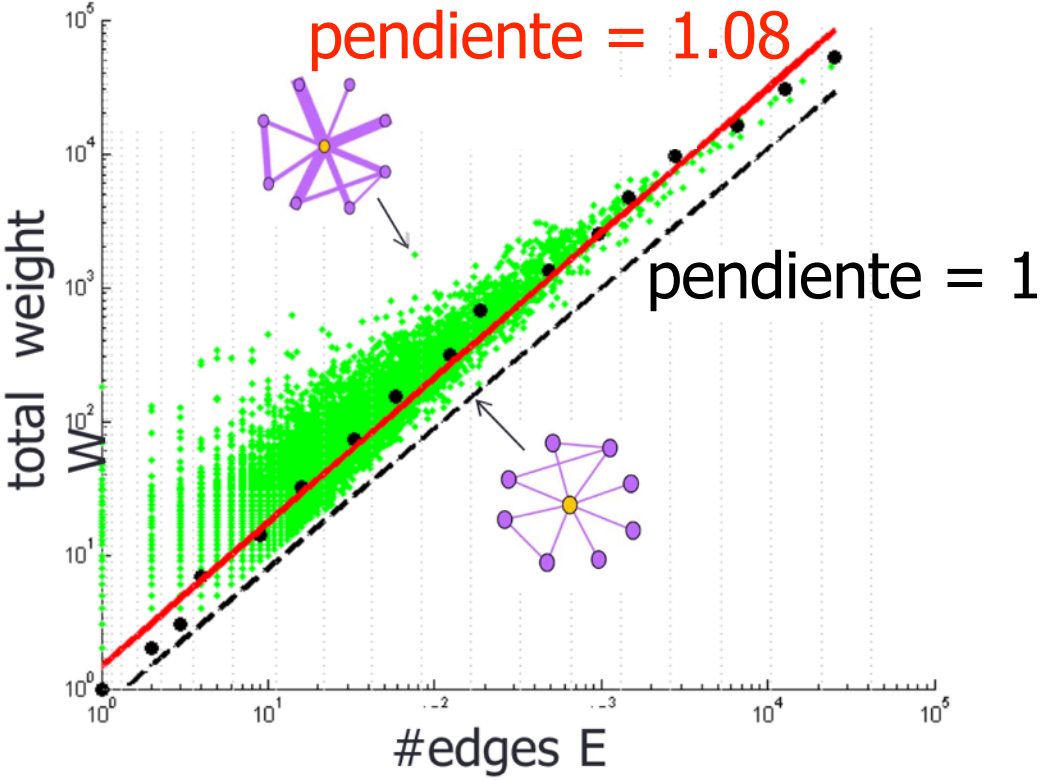
Relación mas fuerte

Densidad



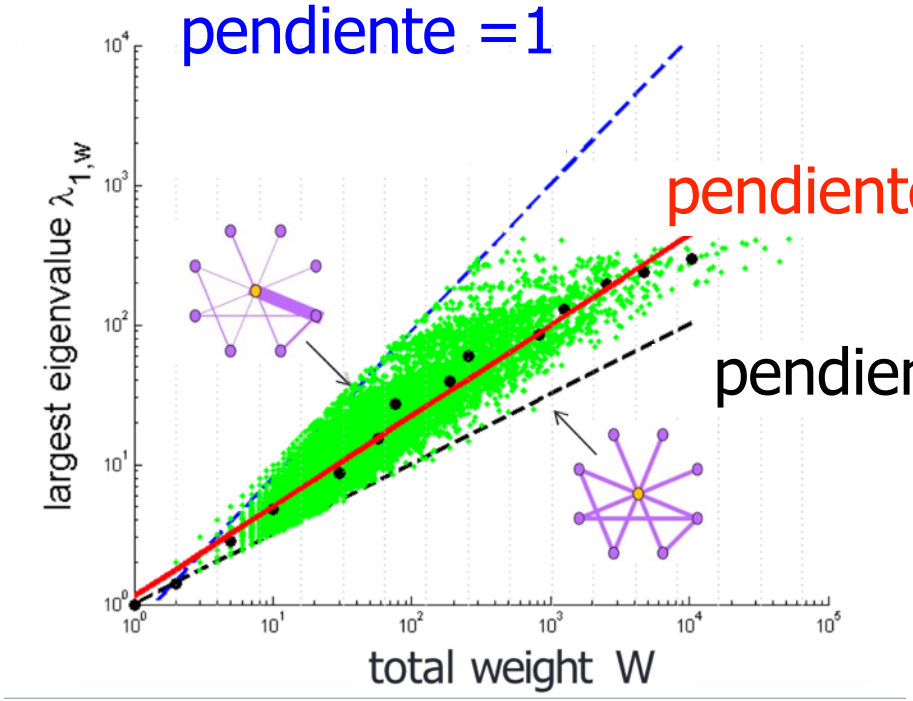
Numero de relaciones contra el numero de vecinos

Peso



Peso de los vecinos contra el numero de vecinos

Valor Propio (Eigenvalue)

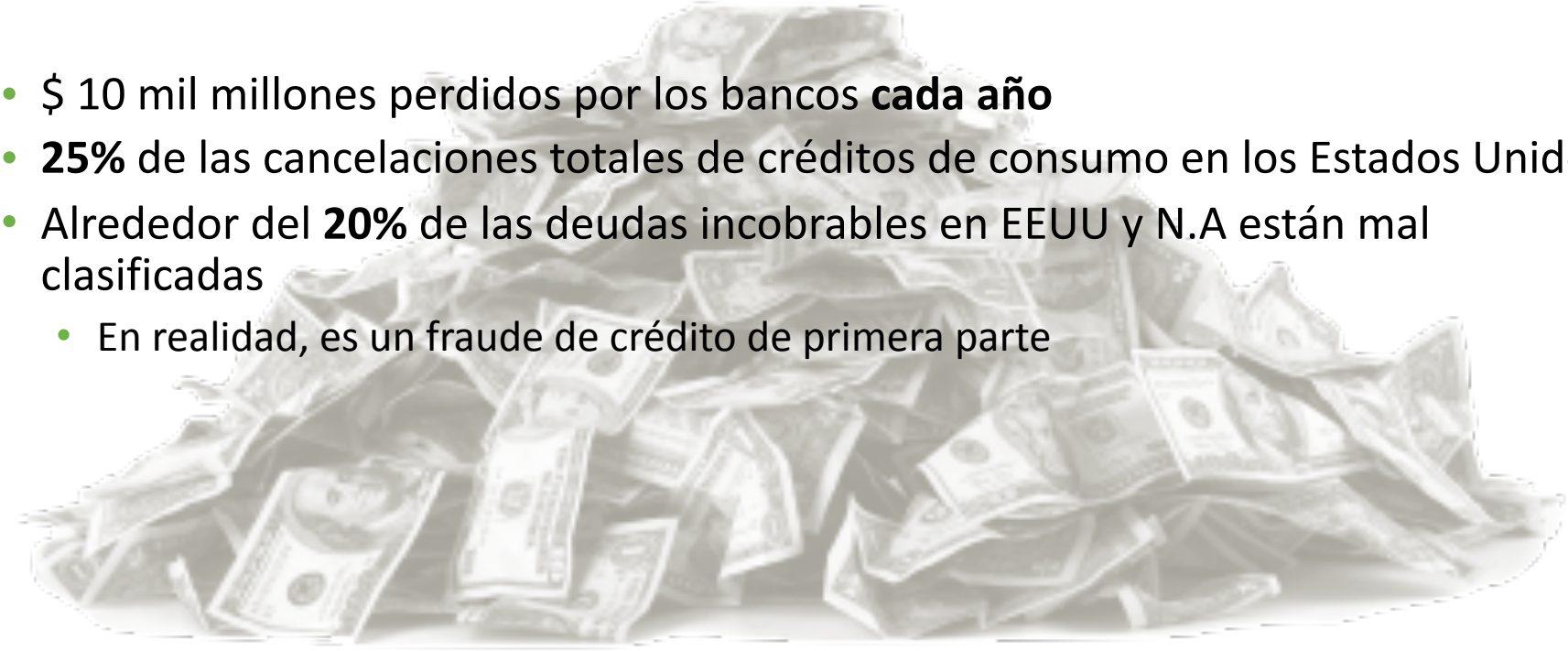


El peso mas fuerte contra peso total de los vecinos

Fraude de Credito

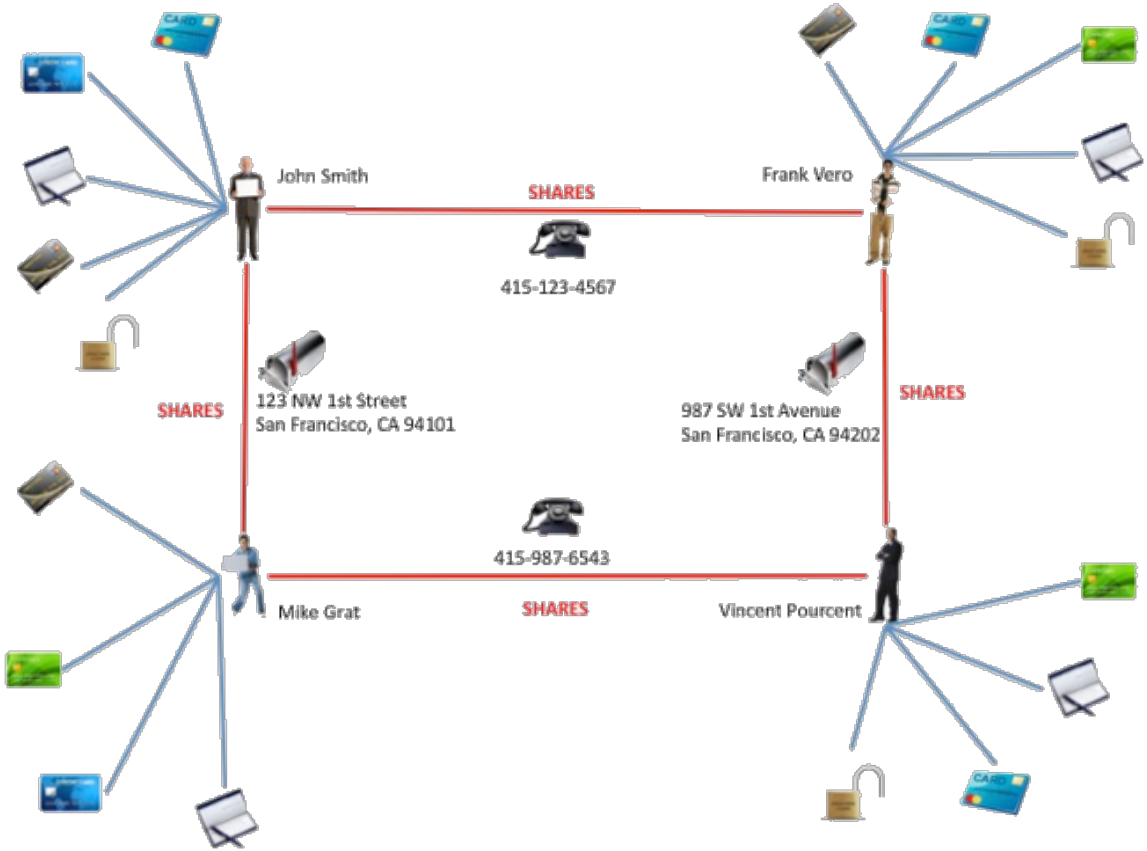
- Objetivo fraudulento: solicitar líneas de crédito, actuar normalmente, extender crédito, y luego ... salir corriendo con el dinero
- Fabricar una red de identificaciones sintéticas, agregar líneas de crédito más pequeñas en un valor sustancial
- Es un problema oculto ya que solo los bancos son afectados

El Problema



- $ 10 mil millones perdidos por los bancos **cada año**
- **25%** de las cancelaciones totales de créditos de consumo en los Estados Unidos
- Alrededor del **20%** de las deudas incobrables en EEUU y N.A están mal clasificadas
  - En realidad, es un fraude de crédito de primera parte

Anillo de Fraude



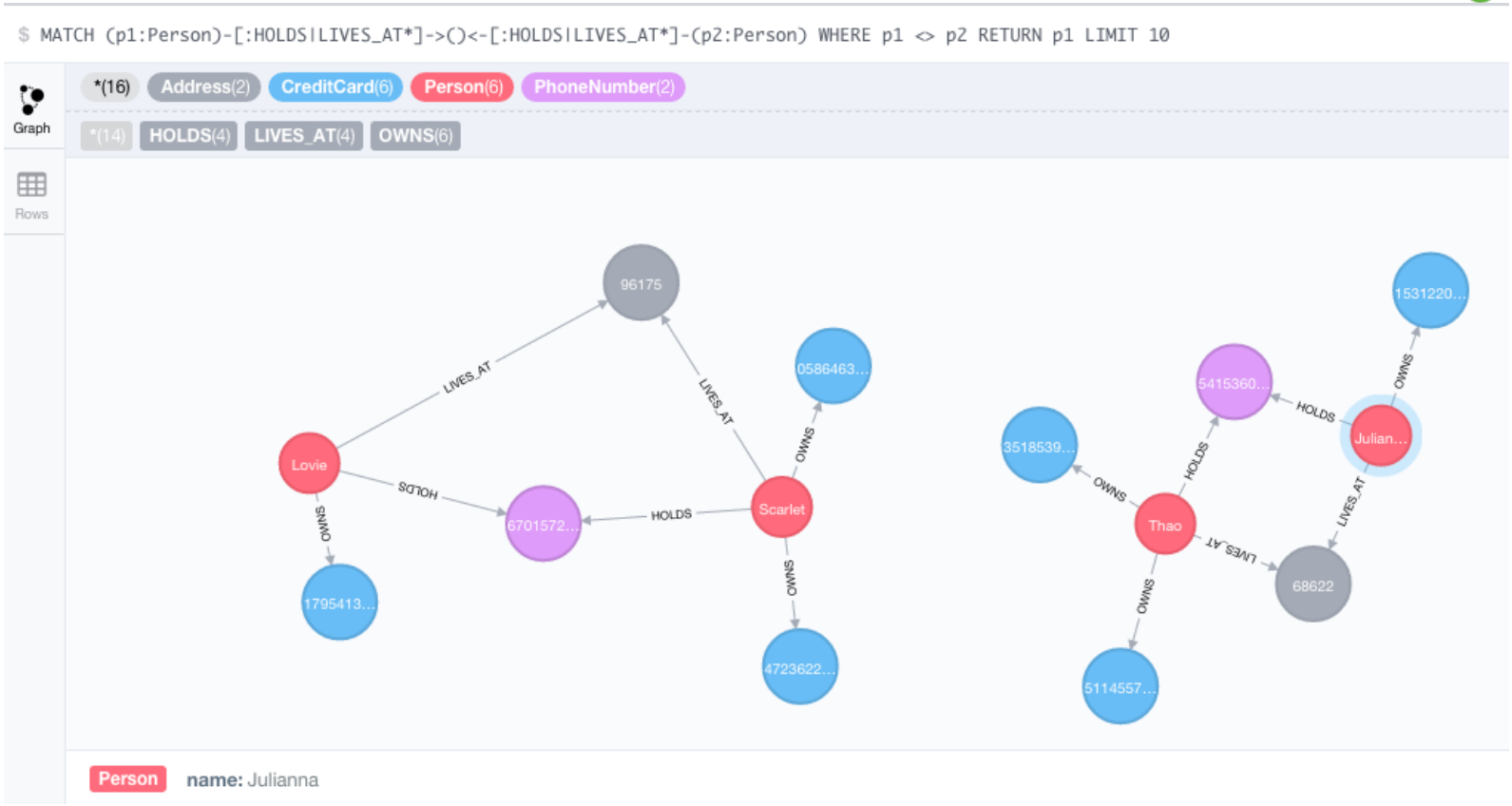
Entonces el Fraude ocurre…

- Estrategia de puertas giratorias
  - El dinero se mueve de una cuenta a otra para proporcionar un historial de transacciones legítimo
- Los bancos aumentan debidamente las líneas de crédito
  - Comportamiento de crédito responsable observado
- Los defraudadores agotan al máximo todas las líneas de crédito y luego desaparecen

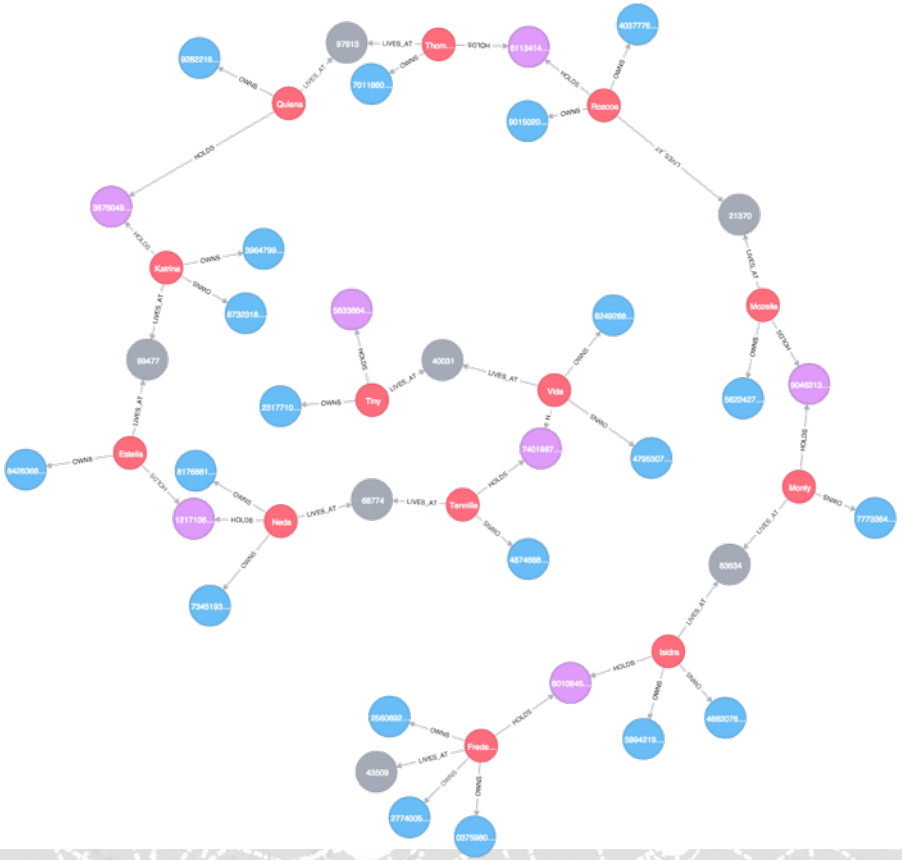
…y el Banco pierde

- El proceso de colecciones empieza
  - Direcciones reales son visitadas
  - Los estafadores niegan todo el conocimiento de las identificaciones sintéticas
  - Banco cancela deuda
- Dos estafadores pueden acumular fácilmente $ 80k
- Los anillos del crimen bien organizados pueden acumular cantidades muchas veces mayores a eso

Cohabitantes Probablemente no Fraudulentos

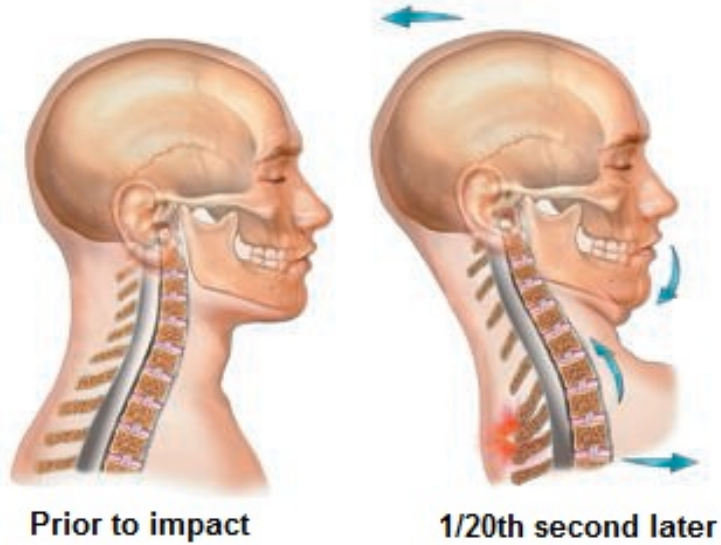


Cadena de Aspecto Sospechoso



Fraude de Autos

Dolor por Dinero



Prior to impact

1/20th second later

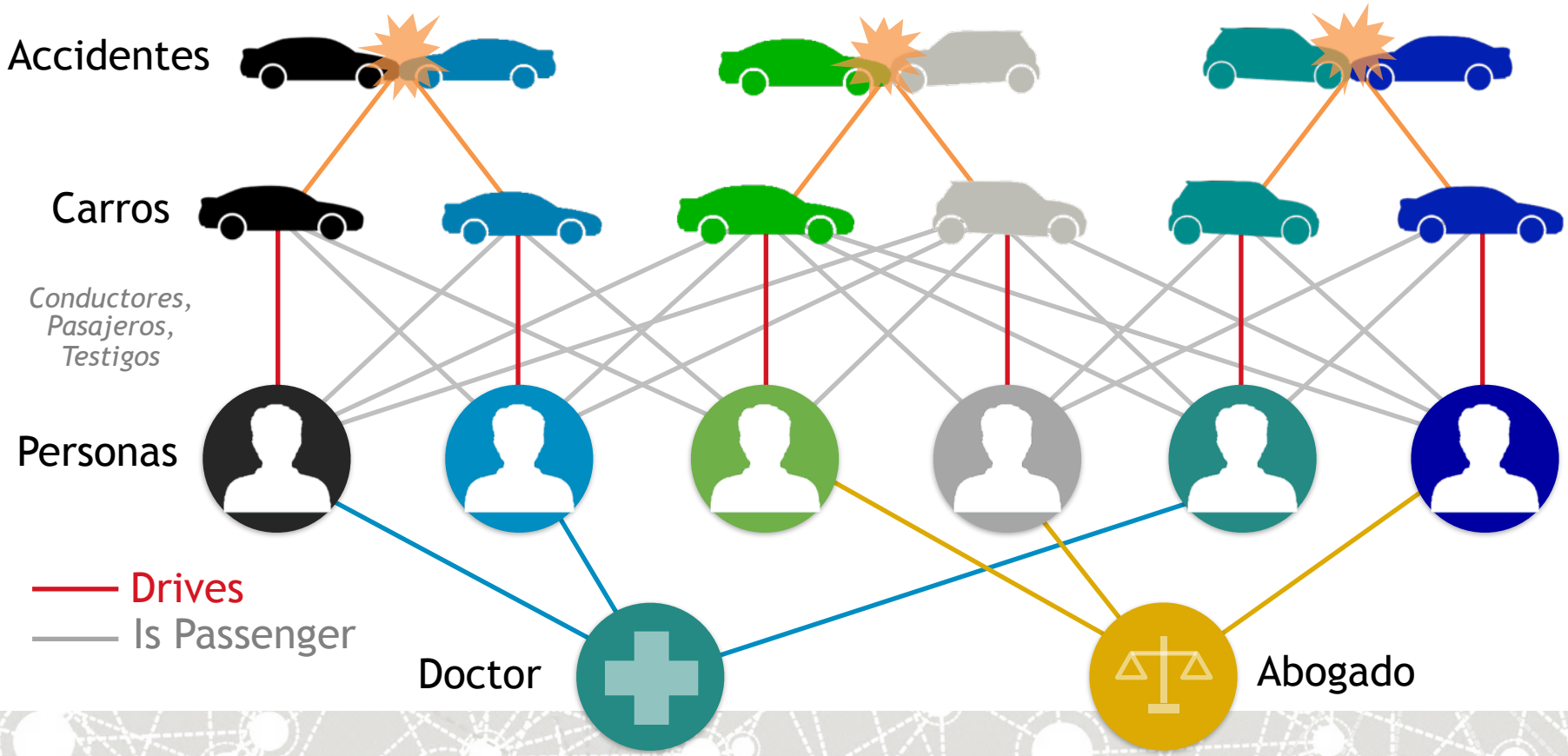
1/10th second later

6 months later

http://georgia-clinic.com/blog/wp-content/uploads/2013/10/whiplash.jpg

http://cdn2.holytaco.com/wp-content/uploads/2012/06/lottery-winner.jpg

Dolor por Dinero *Ejemplo*



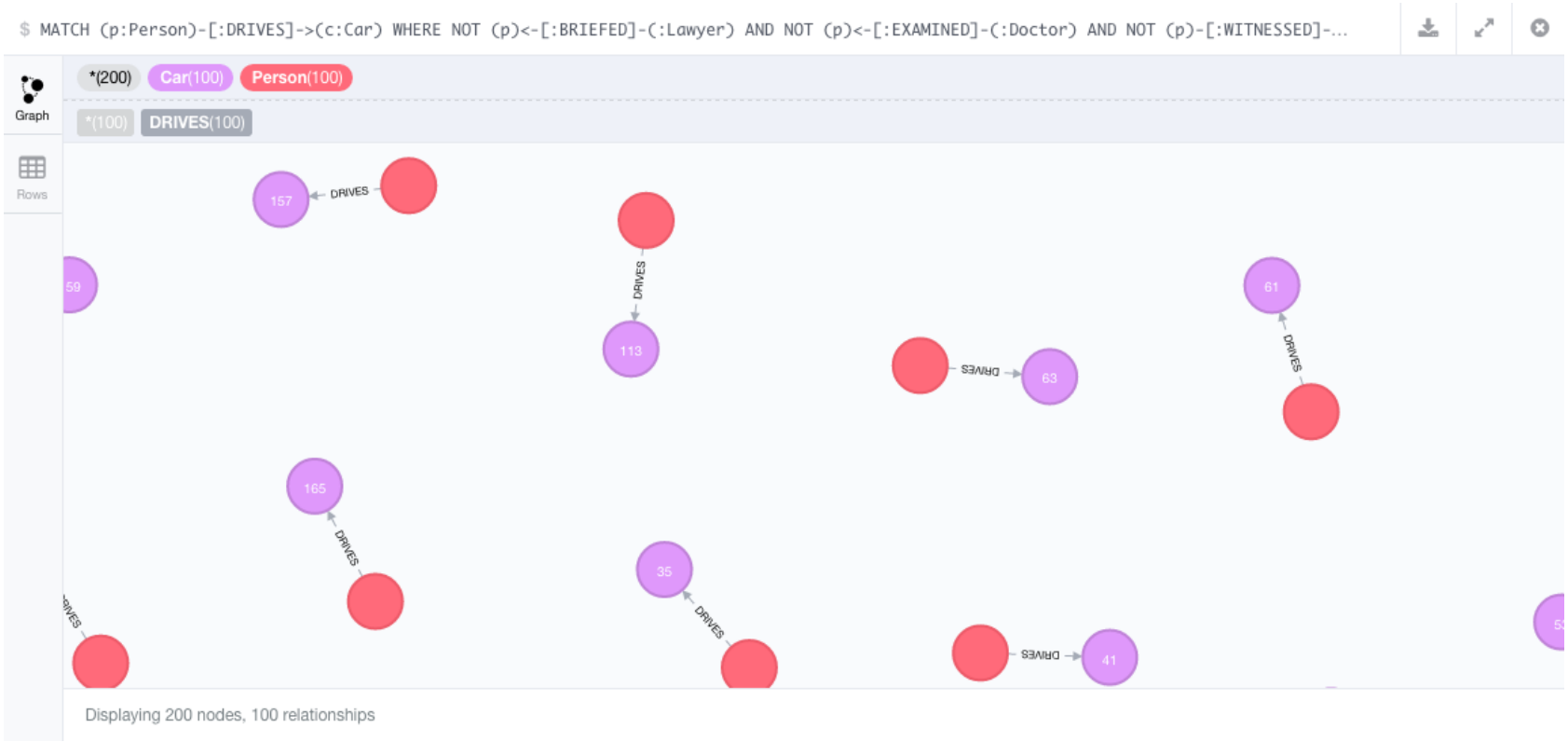
Fraude de Autos

Fraude Interno

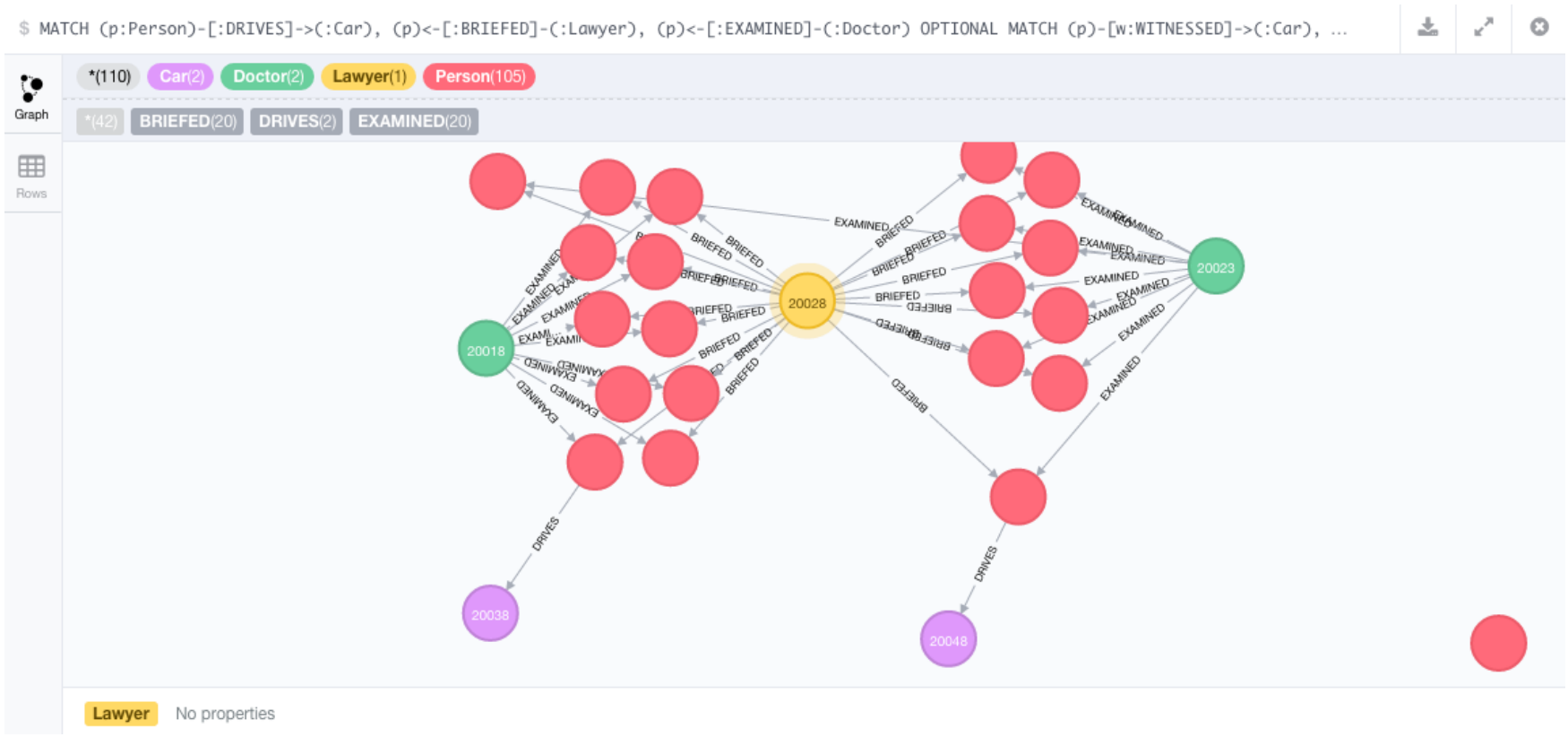
- Fraude de Seguros **en aumento** en el norte de México
- La cantidad de casos de fraude contra compañías de seguros en México ha crecido alrededor del **10%** durante el último período de cinco años (March 2016)
- De acuerdo con la información de OCRA, **los delincuentes** que cometen fraude **están conectados con los empleados** de las compañías de seguros **en casi todos los casos**

http://www.bnamericas.com/en/news/insurance/insurance-fraud-on-the-rise-in-northern-mexico/

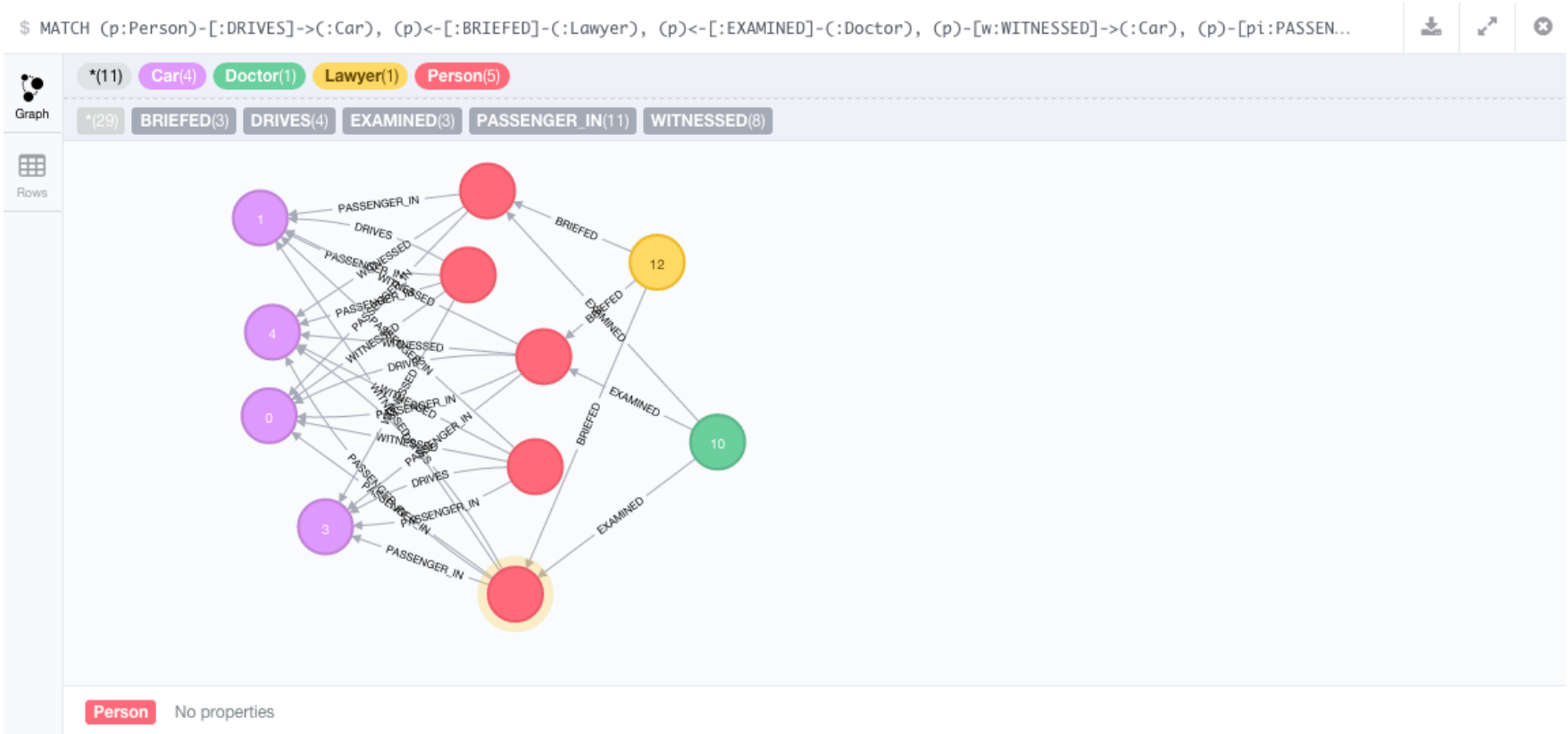
Conductores Regulares



Demandantes Genuinos

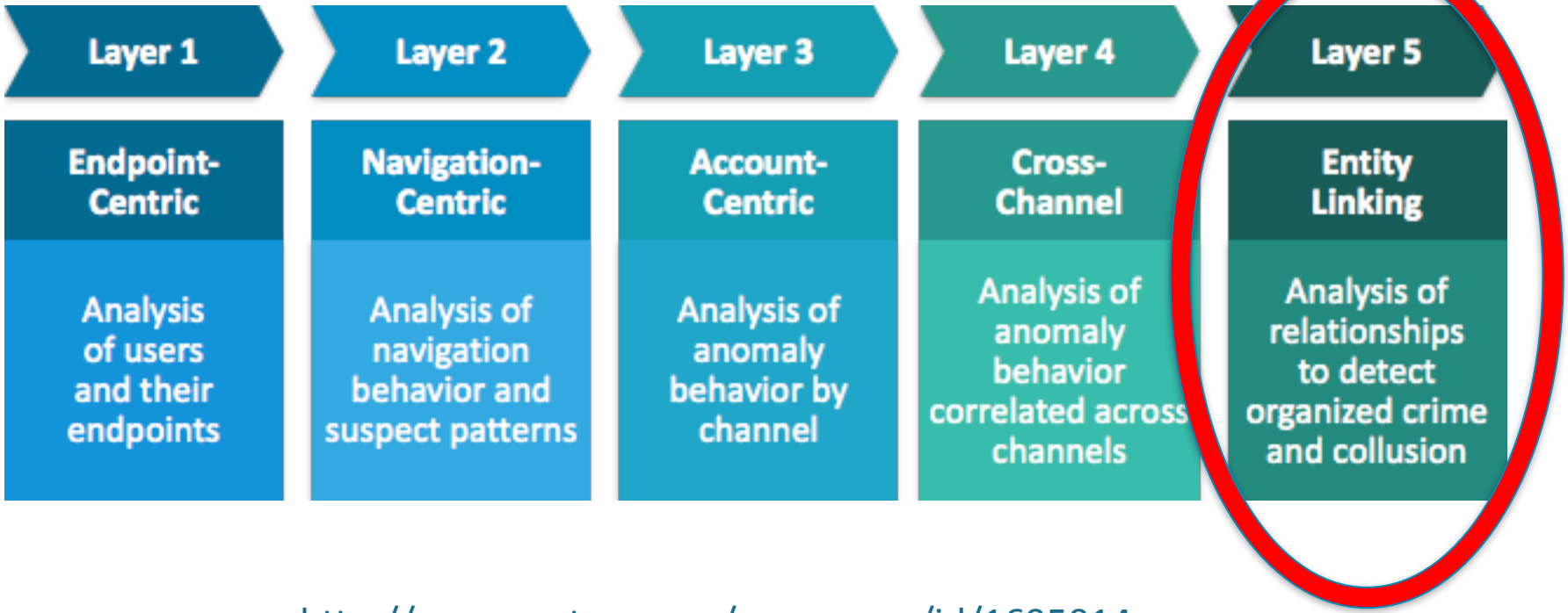


Estafadores



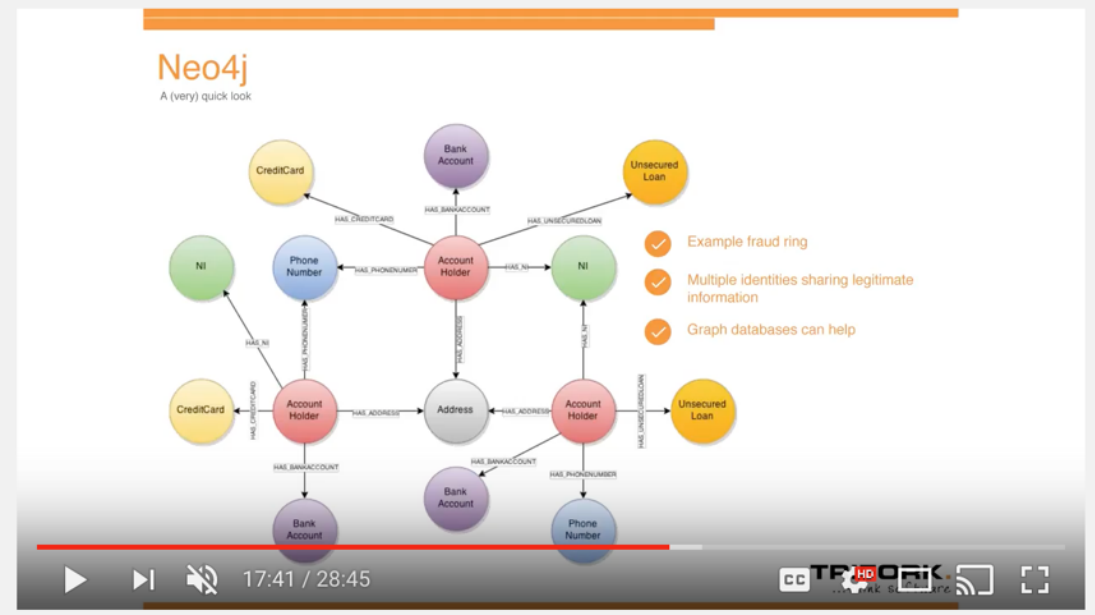
¿Cómo encaja Neo4j con la prevención tradicional de fraude?

Gartner's Layered Fraud Prevention Approach



http://www.gartner.com/newsroom/id/1695014

Red Neural Profunda(DNN) por Fraude Bancario



Usando Métricas Gráficas como Señales

https://www.youtube.com/watch?v=TAer-PeIypI

La detección de fraude comienza a mitad de camino (después de la introducción a redes neurales)

Gracias