




# Data visualization in Network Economics

乔卓彪

[jqiaozb@sjtu.edu.cn](mailto:jqiaozb@sjtu.edu.cn)



# CONTENTS


- 1 Introduction
  - 2 Knowledge graph
  - 3 Current accomplishment
  - 4 Future work
- 



# PART 01

# Introduction

This project is put forward by police which wants to establish a criminal information library and make a Relation map between criminal members.





01

## Introduction

Objective:

Fight against selling counterfeit cigarettes through Internet



### Background

At present, police eliminate some criminal gang, they have collected a lot of information stored in their mobile phone and they have gathered and put them in order with some commercial information, they want to select from data to establish a criminal map just like Acemap in order to show out the relationship between members and gangs.





01

## Introduction

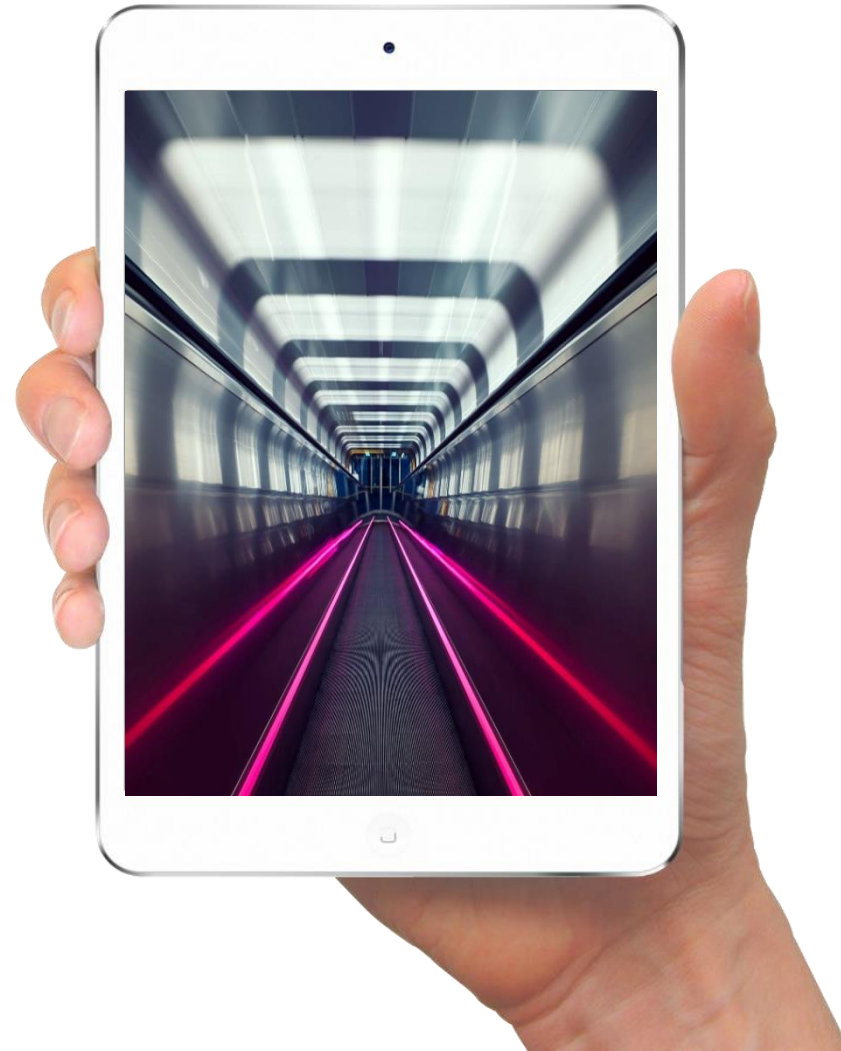
### Data Crawling

Extract data from HTML files and XML file.



### Web Framework

Create a web framework for collecting data.



### Data Masking

Design a series of algorithm to mask data in order to cover the sensitive information.



### Data Visualization

Draw a information network to show out the hidden relation





**PART  
02**

# **Knowledge graph**

It's basic theory of our task.





Social Network :

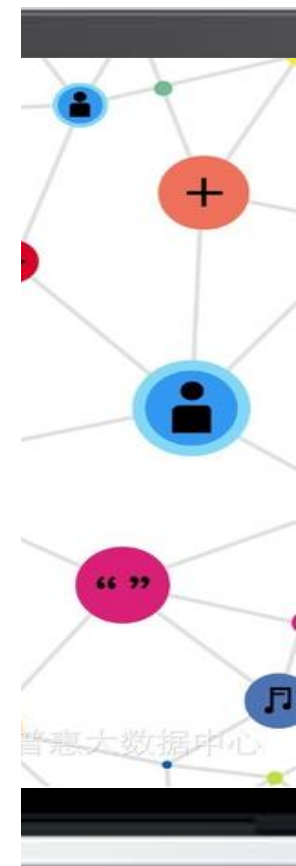
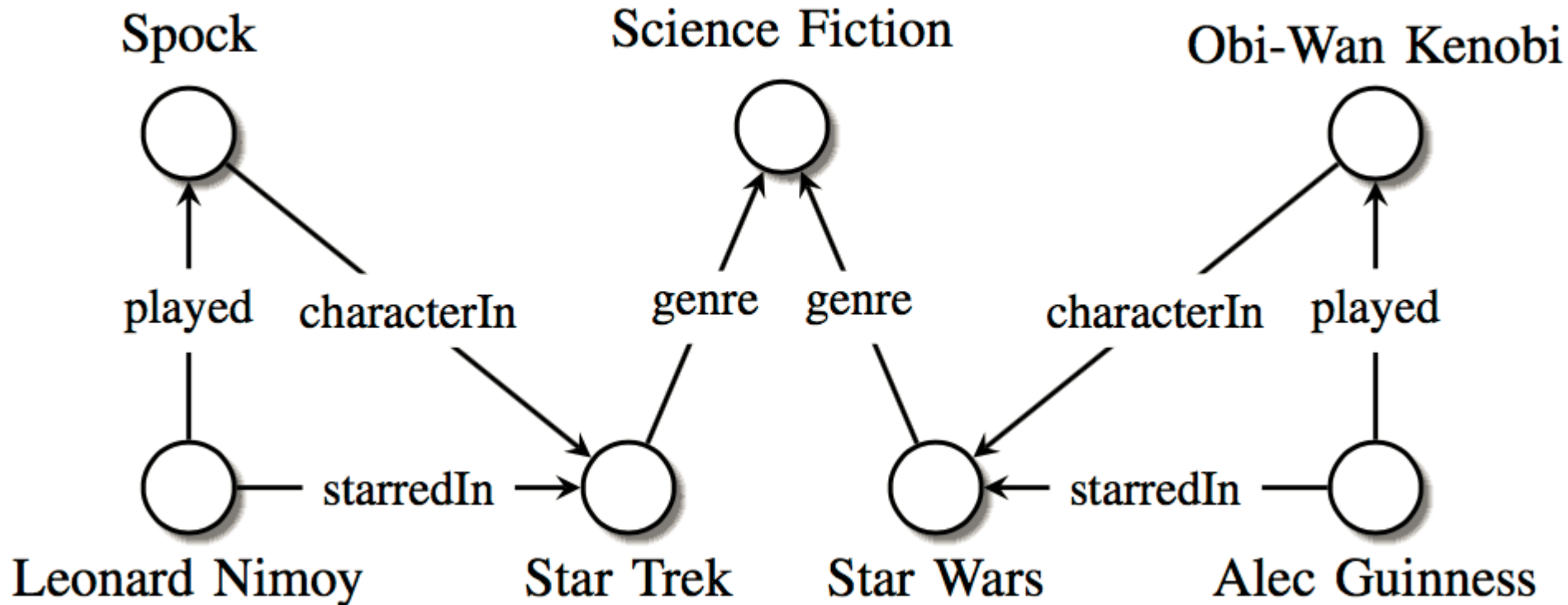
Based on people



Knowledge graph :

Based on everything









## Knowledge representation

W3C RDF standard

(subject, predicate, object) (SPO) triples



## Open vs. closed world assumption:

- Closed world assumption (CWA) :

Non-existing triples indicate false relationships

- Open world assumption (OWA):

A non-existing triple is interpreted as unknown



*subject*

*(LeonardNimoy,*  
*(LeonardNimoy,*  
*(LeonardNimoy,*  
*(Spock,*  
*(StarTrek,*

*predicate*

*profession,*  
*starredIn,*  
*played,*  
*characterIn,*  
*genre,*

*object*

*Actor)*  
*StarTrek)*  
*Spock)*  
*StarTrek)*  
*ScienceFiction)*

Leonard Nimoy was an actor who played the character Spock in the science-fiction movie Star Trek





		<i>subject</i>	<i>predicate</i>	<i>object</i>
Subject	Property	Object		
1903_Tennessee_Volunteers_football_team	opponent	1903_Vanderbilt_Commodores_football_team		ion, Actor)
Reseburg_(community),_Wisconsin	pushpinLabel	Reseburg		In, StarTrek)
A_Boy_Without_a_Girl	thisSingle	¶Khomustakh,_Batagaysky_Rural_Okrug,_Ust-Aldansky_District,...		Spock)
Derry_City_F.C.	leagueChampions...	3		StarTrek)
2009_Eusébio_Cup	referee	João Ferreira		erIn, ScienceFiction)
(subject) Nagaland_(Lok_Sabha_constituency)	votes	52785		
Norman_Saint	bowlAvg	47.05		
SZD-20X_Wampir_II	sinkRateNote	at		
1992-93_Indiana_Hoosiers_men's_basketb...	score	81		
New_York's_23rd_congressional_district_s...	candidate	Bill Owens		
Sengoku_Basara:_Samurai_Kings	originalairdate	2010-09-12		
Spotted_Island_Air_Station	latDeg	53		no played the character
Grade_II*_listed_buildings_in_Pembrokeshire	location	Tregwynt, Granston/Treopert, Pencaer		vie Star Trek
Ashley_Mallett	lasttestdate	--08-28		
Non-exists On_an_Island_with_You	caption	Theatrical release poster		
2013-14_Hazfi_Cup	goals	Hossein Zarei		
● Open 2012_SANFL_Grand_Final	homeQtr	1.1		
Pascal_Zuberbühler	totalcaps	446		
A non-exists Chills_(album)	recorded	2008		
List_of_20_Dakika_episodes	episodenummer	24		



Open vs

- Close

Non-exists

- Open

A non-exists





## 02 Uses of knowledge graphs

- *Smarter Search Engine*  
Google, Bing, Baidu
- *Semantically aware question answering services.*  
IBM's question answering system Watson  
Siri, Cortana, or Google Now.
- *Integrate multiple sources of biomedical information.*  
Bio2RDF, Neurocommons, and LinkedLifeData





## Main tasks in knowledge graph

- *Link prediction*  
*Predict the existence (or probability of correctness) of (typed) edges in the graph*
- *Entity resolution*  
*Identify which objects in relational data refer to the same underlying entities.*
- *Link-based clustering.*  
*Entities are not only grouped by the similarity of their features but also by the similarity of their links.*





**PART  
03**

**Current  
accomplishment**







# Achievement



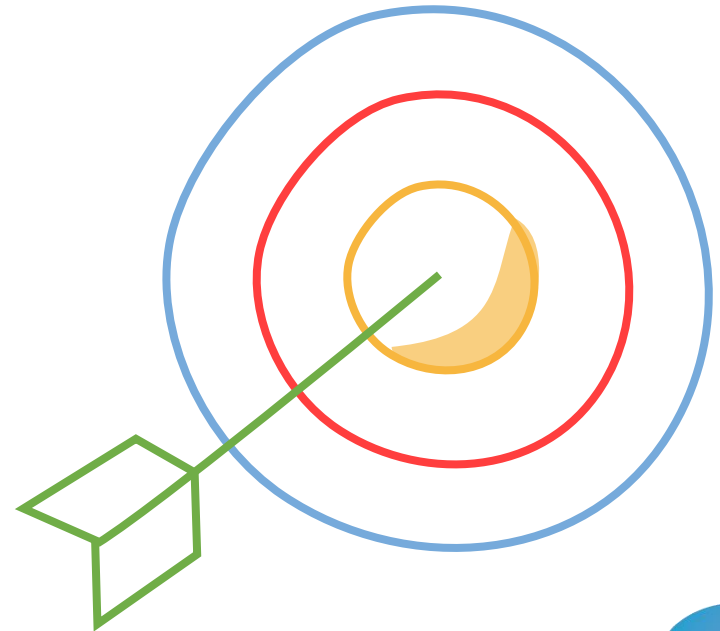
- *Extract information from HTML file*



- *Extract information from XML file*



- *Establish a Web Framework*





# HTML and XML

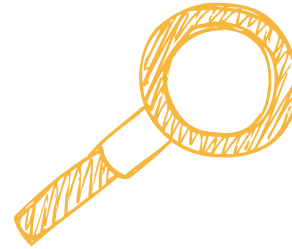
---

## Algorithm 1 Crawling Algorithm for callLog

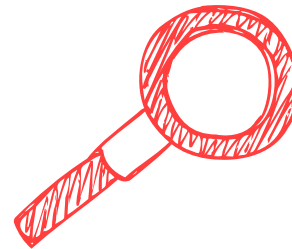
---

```
1: procedure MYPROCEDURE
2:   initial:
3:     url ← find global address of content0.html
4:     response ← scrapyRequest(url)
5:   selection:
6:     calladdress ← select address for callLog
7:     cut the key segments base on the key words
8:     terminal pages ← key segments cut
9:   parsing:
10:    establish .csv file in specific output path
11:    open .csv file in specific output path
12:    select segment to write according to its format
13: end procedure
```

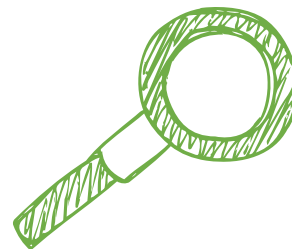
---



Call Log



Address Book

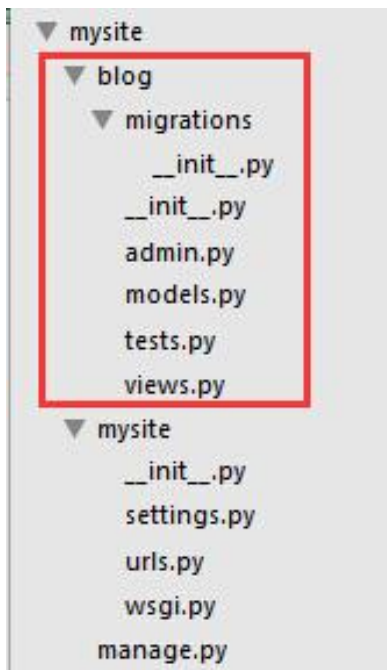


Message





# Web framework\*



```
Username (leave blank to use 'fnngj'):  
Email address: fnngj@126.com      邮箱地址  
Password:      密码  
Password (again):      重复密码  
Superuser created successfully.
```

Django administration

Username:

Password:

## Start app

Create in mysite

## Initialize database

Support different database

## Set admin app

<http://127.0.0.1:8000/admin>



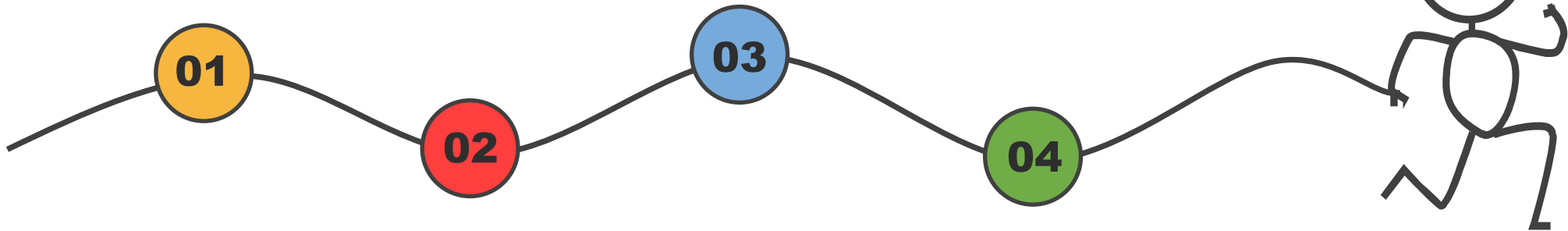
# Web framework\*

## Design table

Design model  
Initial database again

## Create component

Template    View    URL



## Design webpage

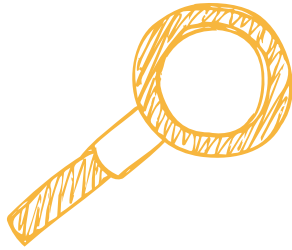
Establish contents and attributes

## Add pattern

Upload and show

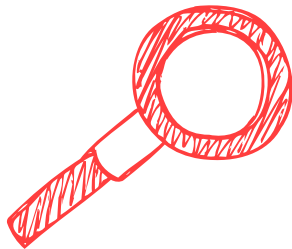


# Data Masking



## Data Mask

- *Algorithm:*
  - DES
  - RSA



## Data Creation

- *Tools:*
  - *Python MySQLdb*



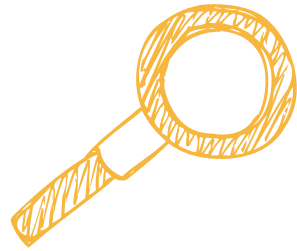


# Data Masking

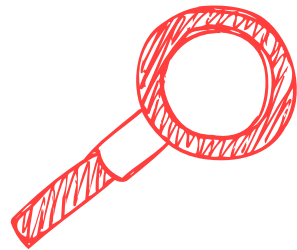
id	user_id	vote_id	group_id
8800410	3vfyTDjwuIVmR5H44ZzK	702	74
8800409	4oCIzyUGAGvm5arzmWtl	981	36
8800408	JeQkXC4ns1BMXhlHlr3L	313	76
8800407	s3KpFUo53Uc7RK1GdSxW	791	58
8800406	C6rMrFTHyKXlJqOBq9jz	670	86
8800405	qNyc92BHwofSvL63MDff	948	46
8800404	SiJytvX5sPBoYyEscqnr	746	90
8800403	QncGHg2g3iaLbuIWq79g	562	48
8800402	UZ38mf1bHNHV03Tb4FXC	887	6
8800401	7fif0UkNOw2rVaREvqrO	387	98



# Data Visualization



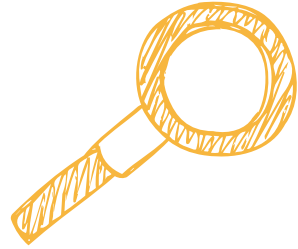
Sigmajs



PRA



# Data Visualization



## Sigmajs

**Sigma** is a **JavaScript library** dedicated to **graph drawing**. It makes easy to publish networks on Web pages, and allows developers to integrate network exploration in rich Web applications.

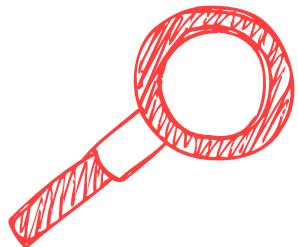


# Data Visualization





# Data Visualization



PRA

Extend the idea of using random walks of bounded lengths for **predicting links** in multi-relational knowledge graphs.





# Data Visualization

Let  $\pi_L(i, j, k, t)$  denote a path of length  $L$  of the form

$$e_i \xrightarrow{r_1} e_2 \xrightarrow{r_2} e_3 \cdots \xrightarrow{r_L} e_j$$

$$\Pi_L(i, j, k) \quad \phi_{ijk}^{\text{PRA}} = [P(\pi) : \pi \in \Pi_L(i, j, k)]$$

set of all such paths of length  $L$ , ranging over path types  $t$ .



# Data Visualization

$P(\pi_L(i, j, k, t))$  computed recursively by a sampling procedure, similar to PageRank

$$f_{ijk}^{\text{PRA}} := \mathbf{w}_k^\top \phi_{ijk}^{\text{PRA}}$$

Logistic regression

The key idea in PRA is to use these path probabilities as features for predicting the probability of missing edges.



# Data Visualization

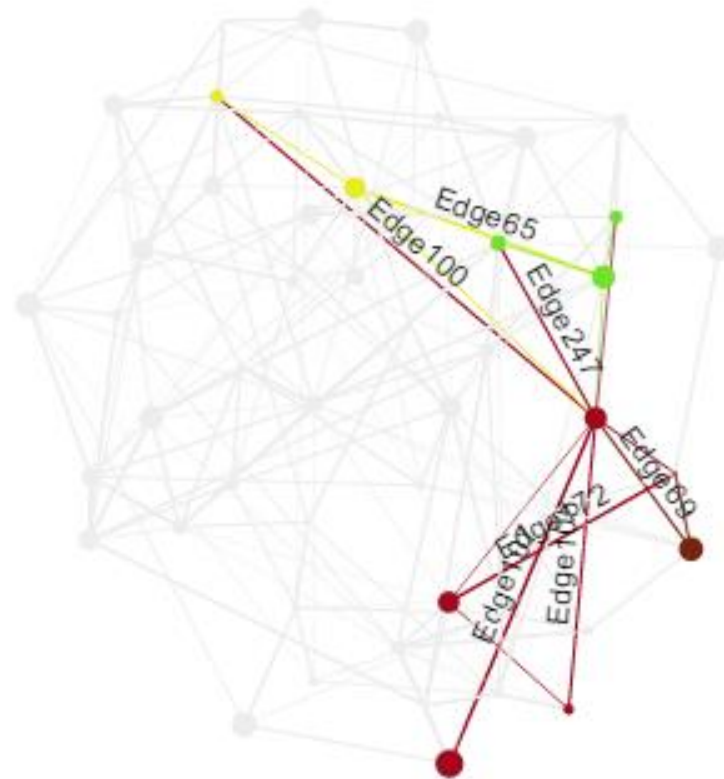
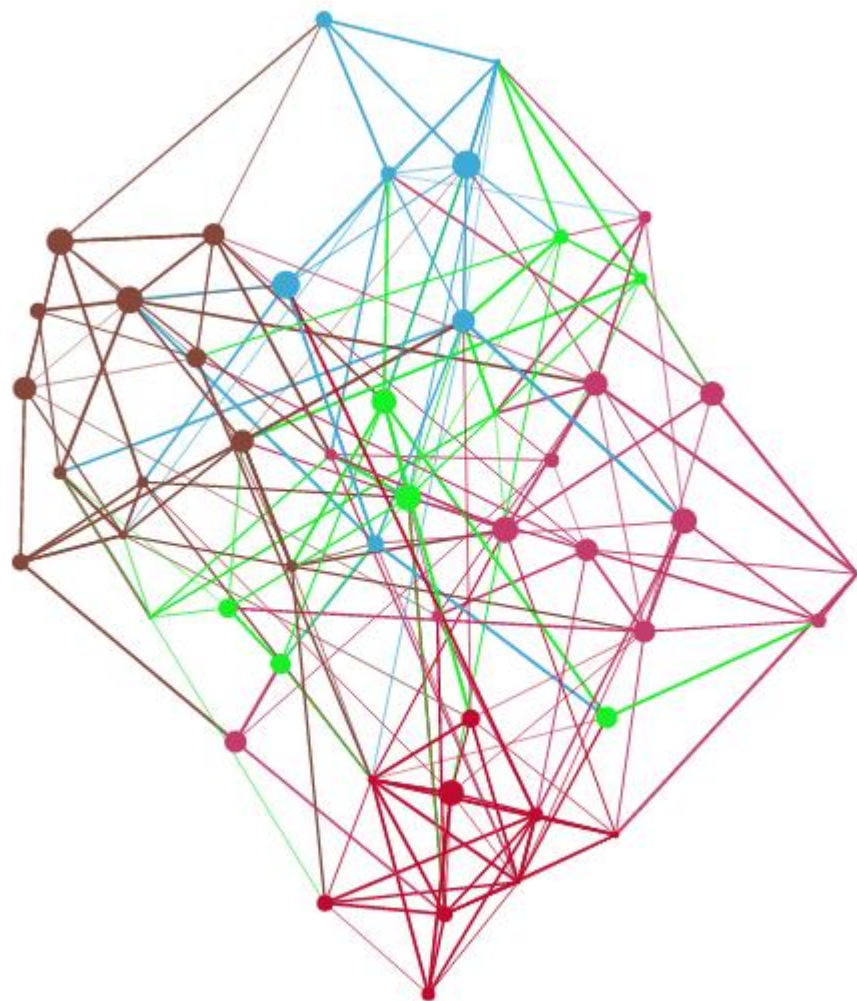
Relation Path	F1	Prec	Rec	Weight
<i>(draftedBy, school)</i>	0.03	1.0	0.01	2.62
<i>(sibling(s), sibling, education, institution)</i>	0.05	0.55	0.02	1.88
<i>(spouse(s), spouse, education, institution)</i>	0.06	0.41	0.02	1.87
<i>(parents, education, institution)</i>	0.04	0.29	0.02	1.37
<i>(children, education, institution)</i>	0.05	0.21	0.02	1.85
<i>(placeOfBirth, peopleBornHere, education)</i>	0.13	0.1	0.58	6.4
<i>(type, instance, education, institution)</i>	0.05	0.04	0.34	1.74
<i>(profession, peopleWithProf., edu., inst.)</i>	0.04	0.03	0.33	2.19

By using a sparsity promoting prior on  $w_k$ , we can perform feature selection, which is equivalent to rule learning.

$$(p, \text{college}, c) - (p, \text{draftedBy}, t) \wedge (t, \text{school}, c) .$$



# Data Visualization



## CURRENT INFORMATION

### NODE INFORMATION:

ID:n31

### EDGE INFORMATION:



**PART  
04**

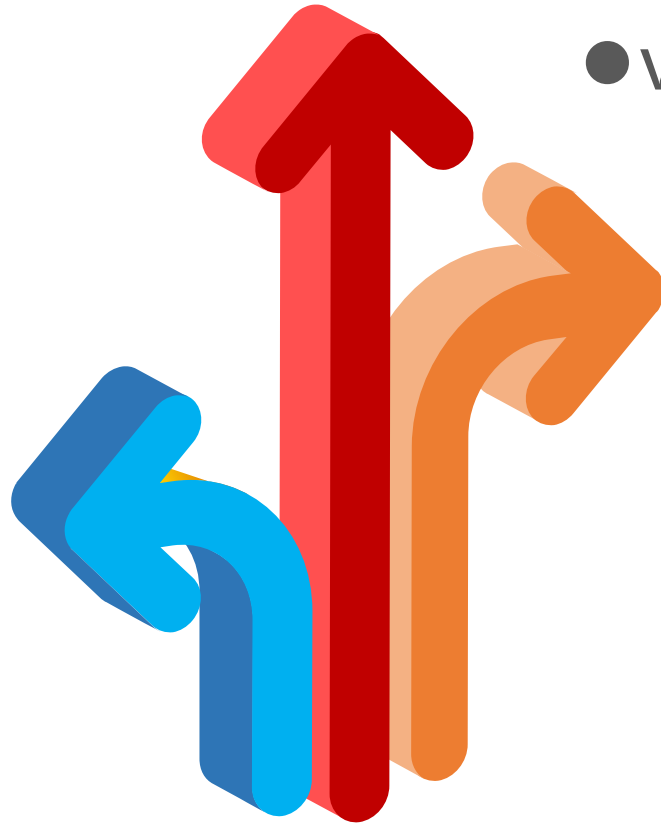
**Future work**





## Data Visualization

- Link prediction
- Feature selection



## Data collection

- Depends on larger dataset from police.
- verifiable relationships

## NLP

- Police Search
- Meaning of Secret



The final goal is to generate ten billion class database that contains more than 10000 verifiable relationships and implements associations between different types of data, as well as providing a learning sample for a more complex machine learning model.





