

# Analisi di Sentenze Penali

Giovanni Giuffrida, Giuseppe Giura, Carlo Pennisi, Calogero Zarba  
Dipartimento di Sociologia e Metodi delle Scienze Sociali  
Università di Catania - Italy

Giuseppe Pigola  
Dipartimento di Matematica e Informatica  
Università di Catania – Italy

# Introduzione

1. Circa 900 sentenze di mafia emesse dal 2000 al 2006 dalle Direzioni Distrettuali Antimafia siciliane (Catania, Caltanissetta, Messina, Palermo) divenute irrevocabili in ogni ordine e grado;
2. Sentenze in formato cartaceo;
3. Estrarre “Entità” in modo automatico al fine di popolare un Codebook:
  1. Tipo di sentenza e dati identificativi;
  2. Autorità emittente;
  3. Giudici;
  4. Luogo;
  5. Imputati (dati anagrafici);
  6. Avvocati difensori;
  7. Pubblici ministeri;
  8. Imputazioni;
  9. Parti civili;
  10. Esito del processo;
  11. Etc. etc.

# Introduzione

N. 896/94 R.G.N.R.  
N. 2232/94 R.G.G.I.P.  
N. 93/02 R. Sent.



1

TRIBUNALE DI CALTANISSETTA  
Ufficio del Giudice per le indagini preliminari

\*\*\*

**Repubblica Italiana**  
*In nome del popolo italiano*

Il Giudice Dr. Francesco Antoni ha pronunciato la seguente sentenza  
nel procedimento nei confronti di:

# Introduzione

***imputato***

*del reato p. e p. dagli art. 81 cpv., 110 e 416-bis C.P., perché, con più azioni esecutive di un medesimo disegno criminoso, poste in essere in tempi diversi, nella qualità di magistrato componente della Sezione Misure di Prevenzione della Corte di Appello di Palermo – Sezione Sesta, nell’ambito della quale svolgeva le funzioni di consigliere, pur non appartenendo formalmente all’associazione mafiosa denominata “Cosa Nostra” operante nella provincia palermitana – che si avvale della forza di intimidazione del vincolo associativo e della condizione di assoggettamento e di omertà che ne deriva per commettere delitti quali omicidi, traffico di sostanze*



# Schema

## PREPROCESSING

- Parserizzare le sentenze con una procedura automatica identificando entità e relazioni tra queste;
- Memorizzazione dei dati in un DB;

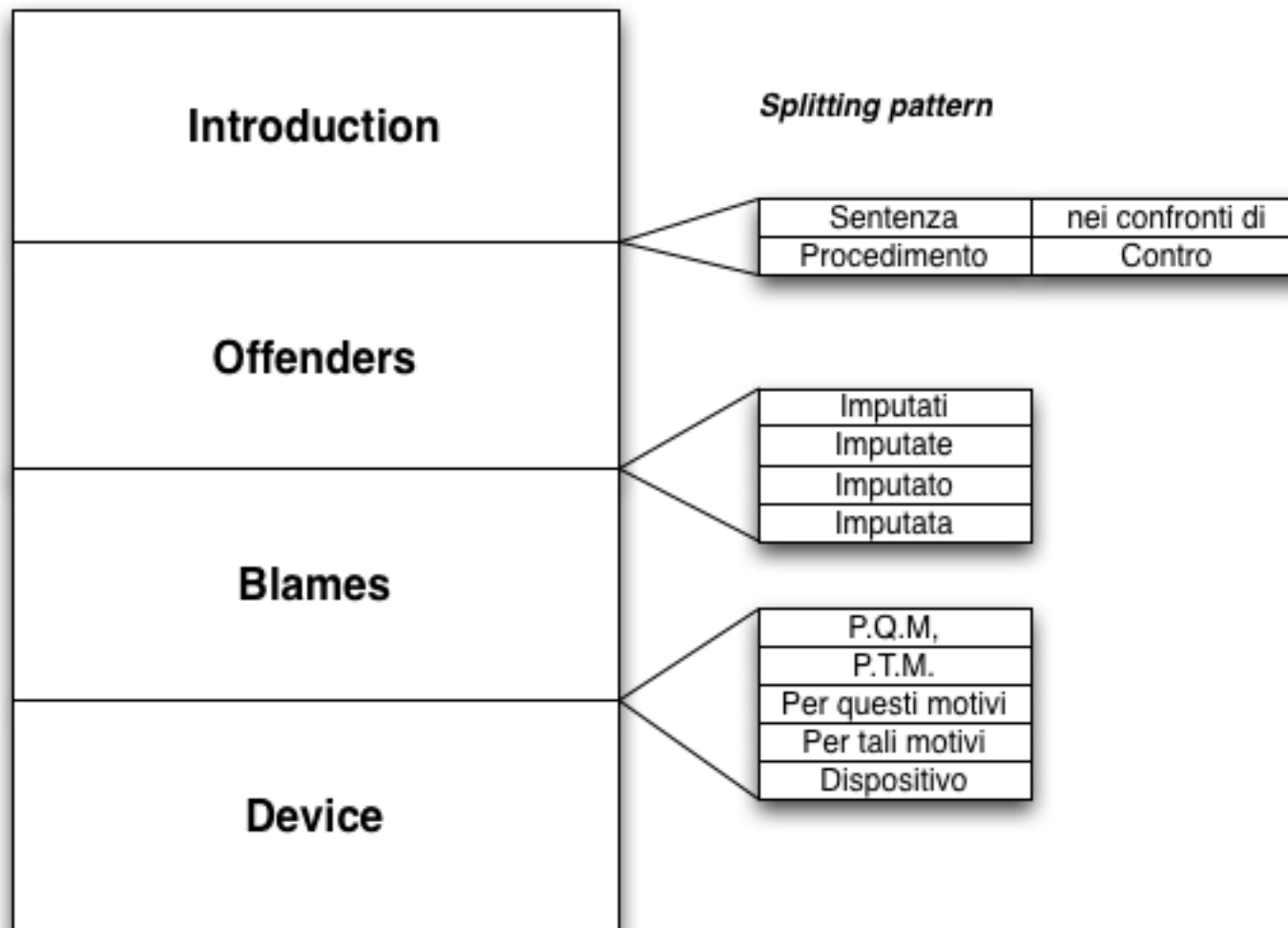
## ANALYSIS

- Statistiche;
- Interrogazione;
- Social Network Analysis;

# Parsing: Individuare Entità

- E' uno dei passi più complessi:
  - **Splitting**
    - La sentenza è suddivisa in sezioni: Introduzione, Imputati, Accusa, Dispositivo;
  - **Parsing**
    - Estrazione delle entità in base alla loro posizione nelle varie sezioni: Vengono costruiti degli automi specializzati nella ricerca di determinate entità;
  - **Relationships**
    - Le entità sono associate tra di loro in base a relazioni intrinseche e dedotte.
- Entità dedotte:
  - Number\_judgement
  - Date\_judgement
  - Lex\_Crime
  - Lex\_Mode
  - Lex\_Request
  - Lex\_Result
  - Place\_judgement\_authority
  - Authority\_judgement
  - Place\_birthcity
  - Person\_offender
  - Person\_lawyer
  - Person\_other
  - Person\_pm
  - Person\_judge\_president
  - Person\_judge\_consigliere
  - Person\_judge\_generic
  - Person\_clerk
  - Person\_collaborator

# Splitting



# Parsing

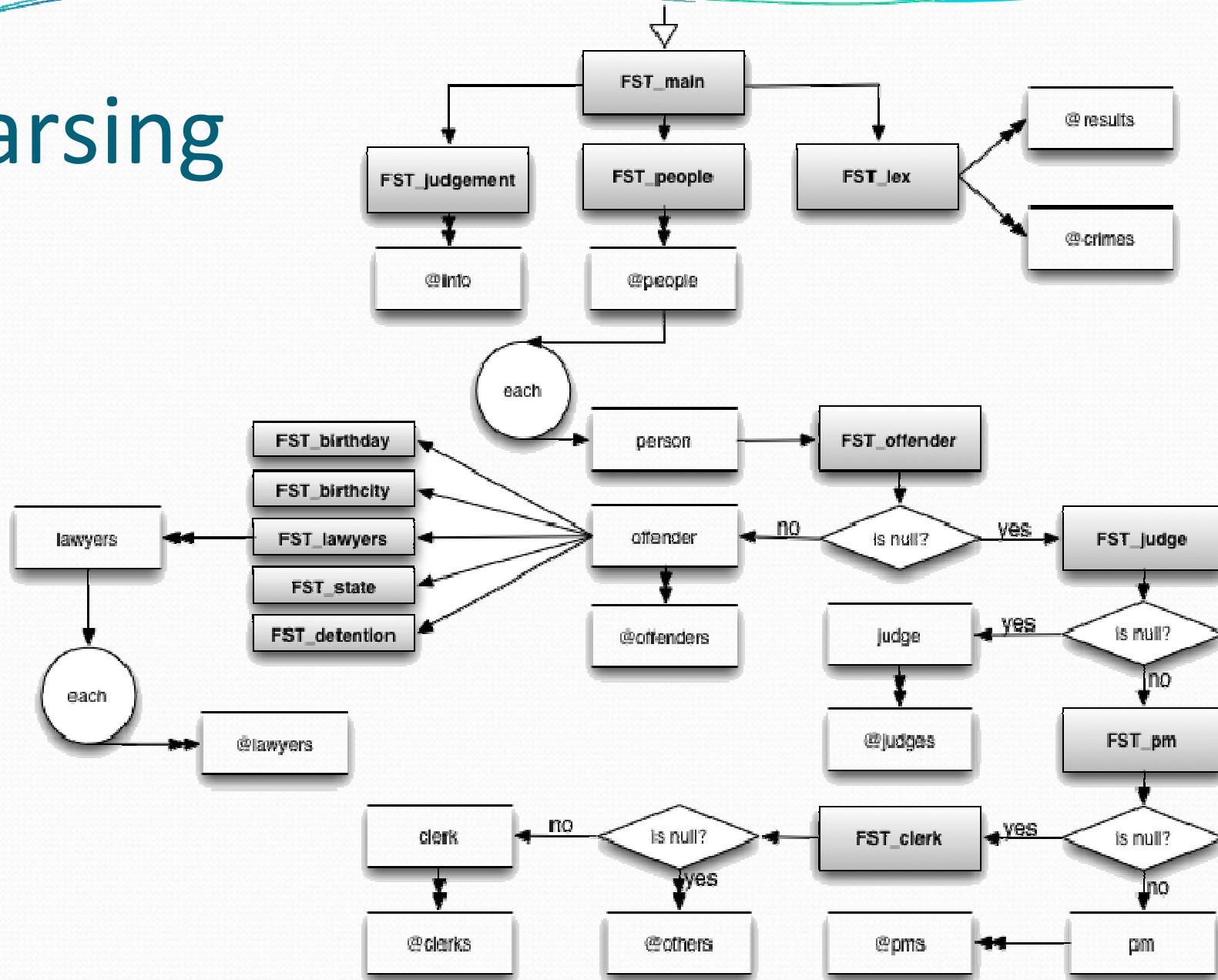
- Traduttori a stati finiti (FST):
- Viene effettuata una analisi morfologica di una parola in input e viene restituita la sua struttura:

CATS → CAT NOME PLURALE

- PRO: Sono abbastanza efficienti. CONTRO: E' laborioso costruire tutte le regole che permettono di analizzare e classificare un testo;
- Database di Nomi e Città per identificare Persone e Luoghi;
- Uso intensivo di Espressioni regolari;



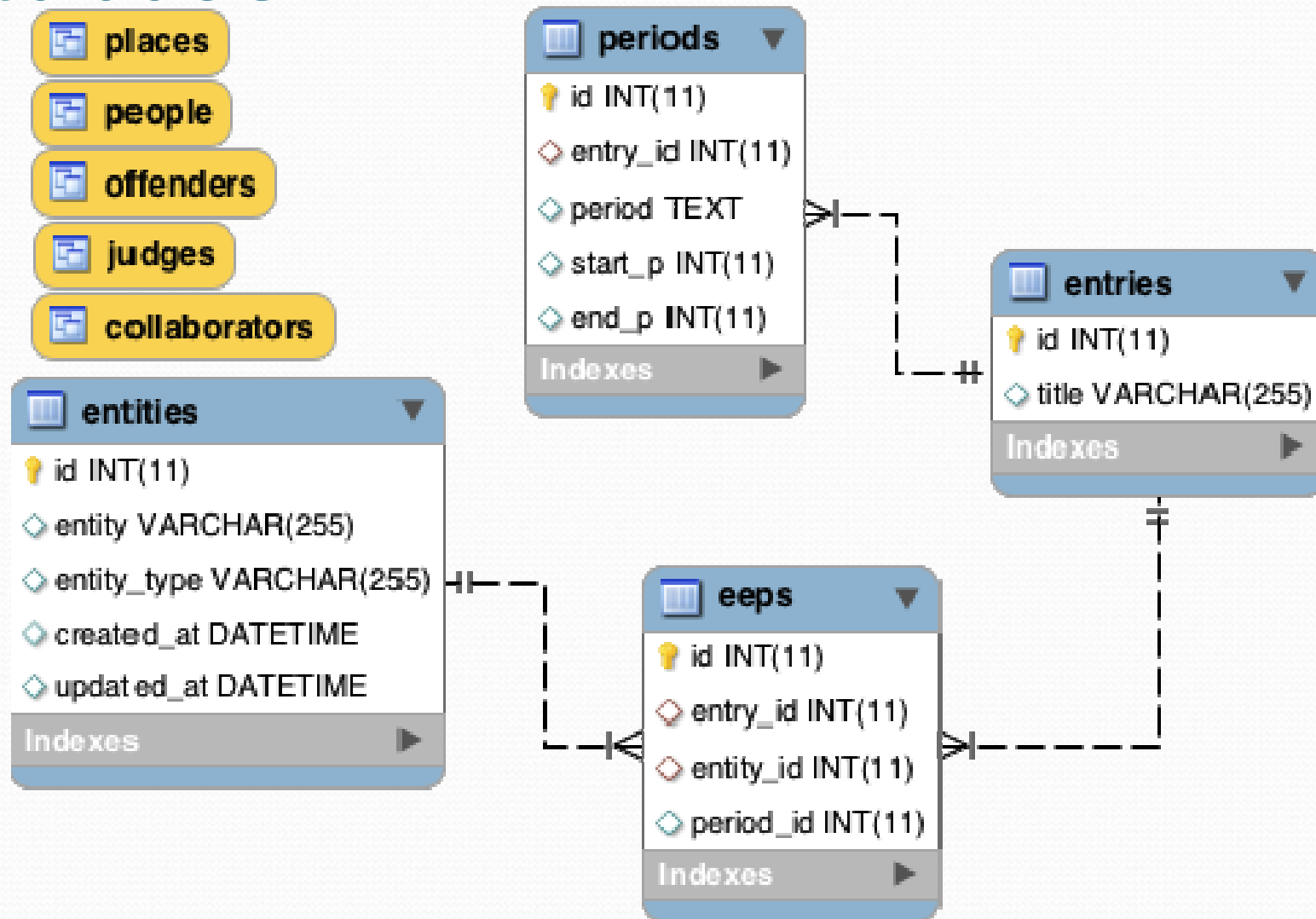
# Parsing



# Relazioni

- **Associazioni intrinseche**
  - Entità che appaiono nella stessa sentenza;
  - Relazioni tra giudici, imputati;
  - Giudici e sentenza;
  - Avvocati e imputati;
  - Imputati e loro data e città di nascita;
  
- **Associazioni dedotte**
  - Entità che appaiono nello stesso periodo (Relazioni Pesate);
  - .....

# Database



# Statistiche: Codebook

- Sentenze in base al numero di imputati;
  - 1 – Un imputato;
  - 2 – Da 2 a 5 imputati;
  - 3 – Da 6 a 10 imputati;
  - 4 – Da 11 a 15 imputati;
  - 6 – Da 16 a 20 imputati;
  - 7 – Più di 20 imputati;
- Sentenze in base al luogo (Catania, Messina, Palermo, Caltanissetta);
- Autorità emittente (GIP, Tribunale, Corte di Appello, Corte di Assise, Corte di Assise di Appello);
- Richieste di Pena;
- Giudici (Nomi e qualifiche);
- Avvocati (Nomi e qualifiche);
- Imputati (Anagrafica + informazioni aggiuntive (Presente, Assente, Contumace, etc. etc.));
- Altre Persone;
  
- Grafici statistici possono essere costruiti a partire dal DB.

# Query by Entity

- Mostra le sentenze in cui uno o più avvocati (AND/OR) sono stati difensori di imputati poi assolti;
- Mostra le sentenze in cui gli imputati sono stati assolti per insufficienza di prove e era presente uno o più Giudici, Avvocati, PM, etc.;
- Mostra tutte le sentenze aventi in comune il determinato giudice;
- Mostra le sentenze in base al crimine commesso;
- Mostra le sentenze in base al loro esito;

Per ogni entità individuata da una query è possibile visionare il periodo in cui essa appare

333	NR0350_S186_04_TRIB_MARSALA.txt	170818	170863	GASPARE PIZZO, GIGI ADAMO, LUIGI SCOMA, FRANCO
333	NR0350_S186_04_TRIB_MARSALA.txt	233222	233440	In tale contesto, lo stesso MARCECA informava il CONCETTO di essere andato la domenica precedente a parlare della vicenda con GIGI (da identificarsi nell'ADAMO) e che questi si era impegnato ad indagare su tale vicenda.
333	NR0350_S186_04_TRIB_MARSALA.txt	233441	233774	Il MARCECA continuava il racconto dicendo di avere ricevuto, la sera successiva, la visita di VITO e CARLUCCIO ( TITONE) "tutti cacaf i quali avevano già ricevuto "la risposta che GIGI (ADAMO) aveva portato a FRANCO ( BONVENTRE)" e che l'immediatezza di tale richiesta l'aveva sorpreso ( "perche questa risposta con tanta fretta?").
333	NR0350_S186_04_TRIB_MARSALA.txt	234449	234512	di avere saputo il giorno precedente da GIGI della vicenda e che

# Query by Judgment

- Dato una o più sentenze, mostra specifiche entità coinvolte (e possibilmente filtra per un tipo di entità);
- Selezionando più sentenze possiamo evidenziare eventuali relazioni tra esse (dedotte dalla presenza della stessa entità);

aldo ercolano	person_other	NR1154_S19_05_CAA_CT.txt	Risulta, così, dalle dichiarazioni del Di Raimondo (vedi sopra, sub IV-1) che il Di Fazio e il Di Grazia furono dall' Ercolano - subito dopo la decisione, dallo stesso adottata, di uccidere il Celano e il Marsengo - incaricati, unitamente al Branciforte, di prendere accordi con il Nardo per la esecuzione del delitto.
filippo branciforte	person_other	NR1154_S19_05_CAA_CT.txt	DETENUTO PER ALTRO - PRESENTE 2) BRANCIFORTE FILIPPO nato a Catania 1120.1.1964
filippo branciforte	person_other	NR1154_S19_05_CAA_CT.txt	* APPELLANTI NARDO SEBASTIANO e BRANCIFORTE FILIPPO avverso la sentenza della Corte di Assise di Catania del 7.10.2003; nonché il P.M. di Catania contro DI GRAZIA FRANCESCO e DI FAZIO UMBERTO, con la quale veniva deciso come segue:
filippo branciforte	person_other	NR1154_S19_05_CAA_CT.txt	Dichiara NARDO Sebastiano e BRANCIFORTE Filippo colpevoli dei reati:



# Social Network Analysis

- E' possibile usare un grafo per rappresentare **social networks** dove gli individui sono in relazione tra loro in base a determinate caratteristiche (amicizia, condivisione di passioni comuni, appartenenza allo stesso gruppo di discussione etc.);
- **Il ruolo di un "Attore" della rete è definito dalla sua posizione nella network;**
- L'importanza dei ruoli di un individuo nella network può essere quantificato considerando il numero di contatti del nodo corrispondente (**degree centrality**), la velocità con cui le informazioni vengono trasferite (**closeness centrality**-minimizza la somma delle distanze da ogni altro nodo nella network) e la sua potenza di controllo del flusso di informazioni (**betweenness centrality**- la frazione di cammini minimi che passano per esso);
- Tecniche di clustering possono essere applicate al fine di individuare "**Nodi altamente connessi**" (probabilmente coinvolti nello stesso ruolo o funzione);

# DEGREE

Dato un vertice  $v$ , la misura più semplice di centralità è data dal degree:

$\text{DEGREE}(v)$

$\text{IN-DEGREE}(v)$

$\text{OUT-DEGREE}(v)$

- È utilizzata per network statiche.
- E' una misura locale perché dipende solo dai nodi vicini.



# ECCENTRICITY

Dato un grafo  $G=(V,E)$  connesso e indiretto. Definiamo  $d(u,v)$  come la lunghezza dello shortest path tra  $u$  e  $v$ .

Definiamo eccentricità per un nodo  $u$ :

$$e(u) = \max\{ d(u, v) : v \in V \}$$

La misura di centralità indotta è:

$$C_E(u) = \frac{1}{e(u)} = \frac{1}{\max\{ d(u, v) : v \in V \}}$$

Tale funzione cresce se la distanza massima di  $u$  decresce.

# CLOSENESS

- Supponiamo di voler trovare la posizione ottimale per un centro commerciale tale che la distanza totale da tutti i clienti è minimizzata.
- Questo ottimizzerebbe lo spostamento per la maggior parte dei clienti.
- Denotiamo la somma delle distanze da un vertice  $u$  ad ogni altro come:

$$\sum_{v \in V} d(u, v)$$

- La centralità indotta da tale funzione sarà (cresce se la distanza totale di  $u$  decresce):

$$C_C(u) = \frac{1}{\sum_{v \in V} d(u, v)}$$

# BETWEENNESS CENTRALITY

Consideriamo:

$$\delta_{st}(v) = \frac{\sigma_{st}(v)}{\sigma_{st}}$$

dove  $\sigma_{st}$  rappresenta il numero di shortest paths tra  $s$  e  $t$ .

Tale rapporto rappresenta la probabilità che il nodo  $v$  sia coinvolto in una “comunicazione” tra  $s$  e  $t$ .

Possiamo definire Betweenness Centrality come:

$$c_B(v) = \sum_{s \neq v \in V} \sum_{t \neq v \in V} \delta_{st}(v)$$

# IDENTIFICARE CLUSTER

Individuare sottografi altamente connessi.

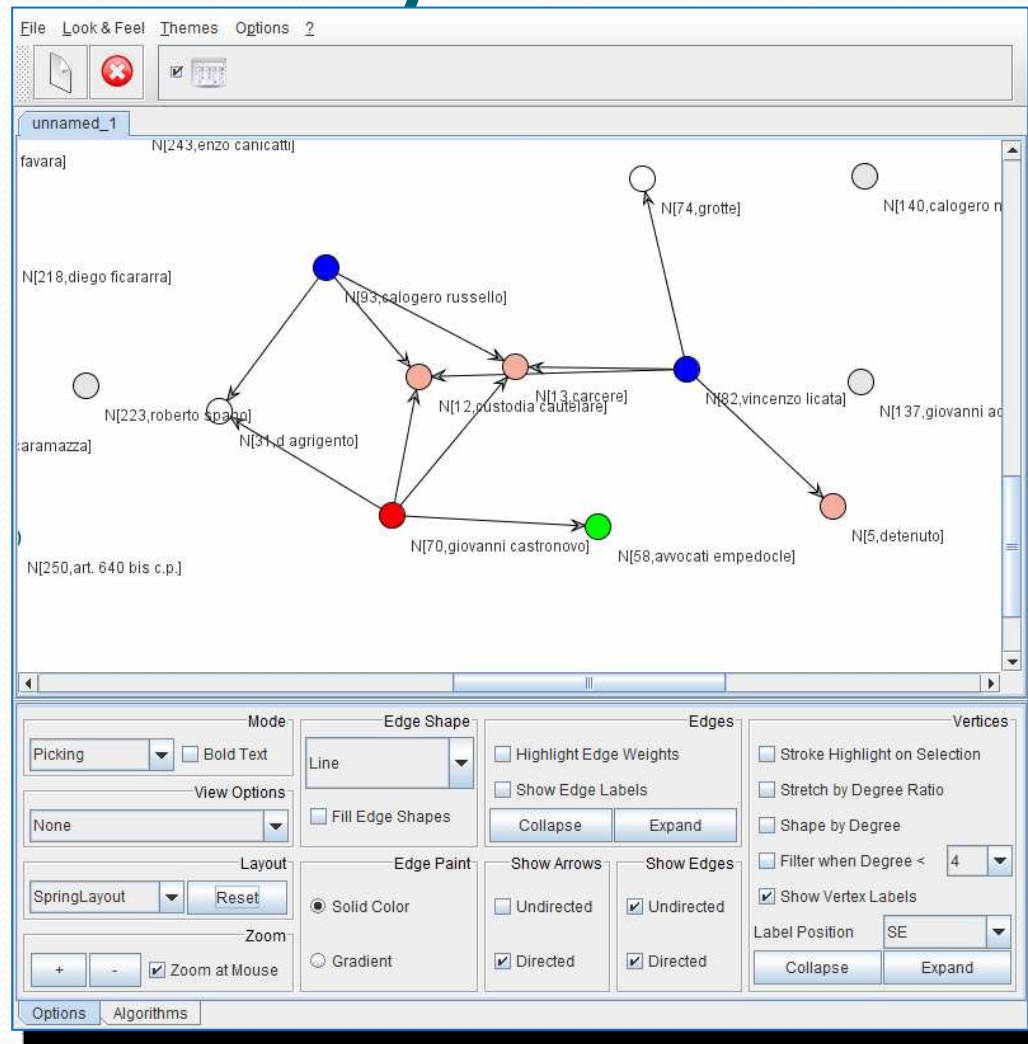
- Clique massimali (o eventualmente quasi-clique). Approccio Incrementale:
  - Per trovare una clique di dimensione  $n$  considerare solo quelle di dimensione  $n-1$ ;
  - Se il grafo è sparso può essere molto veloce;
- Gli algoritmi esatti sono esponenziali. Approcci approssimati sono necessari per grafi di grandi dimensioni;

# IDENTIFICARE SOTTOSTRUTTURE

- Ricercare le occorrenze di un grafo query all'interno di una network:
  - Equivale a ricercare un determinato insieme di relazioni tra individui del social network;
- Algoritmi di matching;

# Social Network Analysis

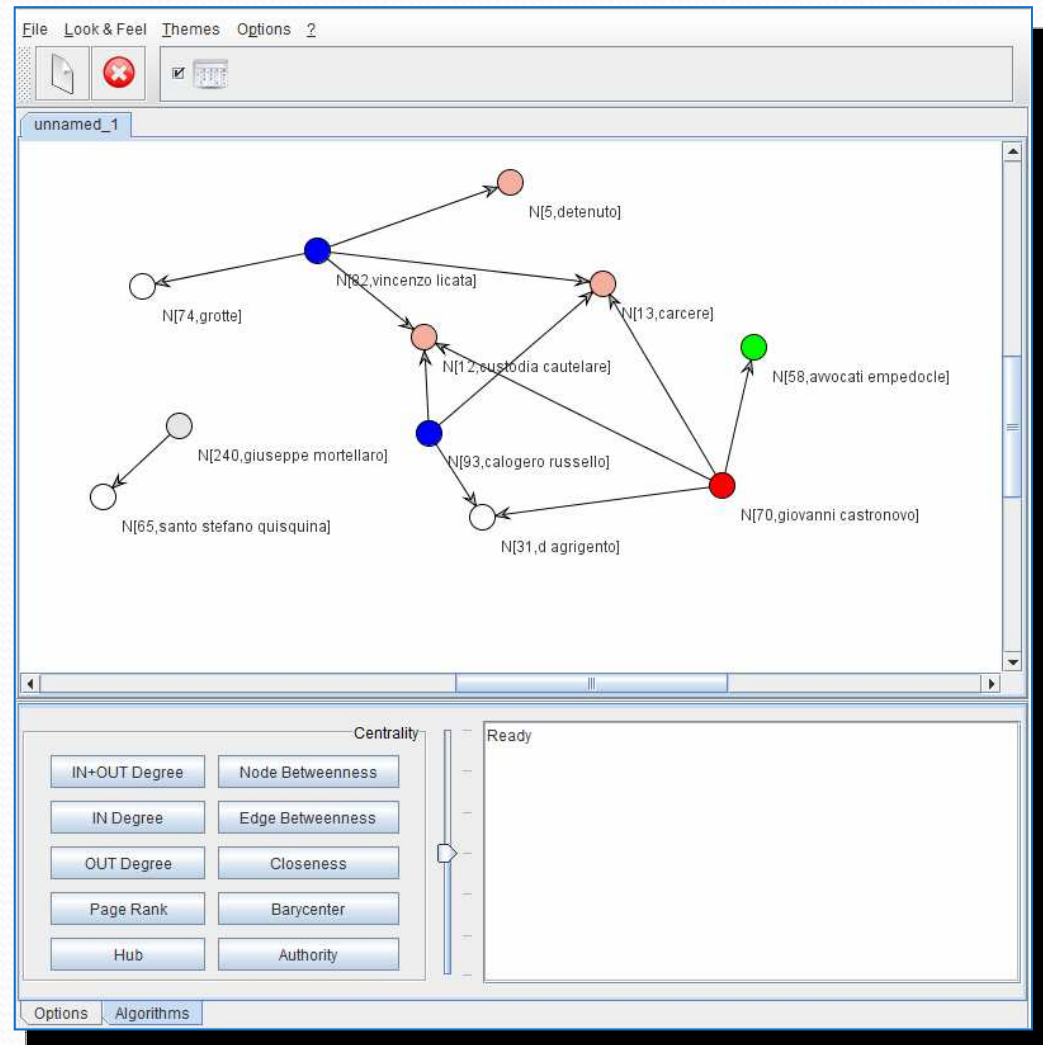
- Data una o più sentenze, possiamo provare a costruire automaticamente un social network in cui i nodi sono le entità (imputati, avvocati, giudici, crimini, esiti del processo, etc) e gli archi rappresentano le relazioni tra di esse;



# Social Network Analysis

## OBIETTIVI FUTURI:

Identificare relazioni nascoste tra entità anche di sentenze diverse;



# Social Network Analysis

## OBIETTIVI FUTURI:

Analizzare le proprietà topologiche e dinamiche delle relazioni;

- Nodi “**Centrali**” hanno probabilmente un ruolo importante (degree, closeness, betweenness, Page Rank, etc.);
- Nodi “**Altamente connessi**” sono probabilmente coinvolti nello stesso ruolo (Clustering);
- Ricercare sottostrutture;

