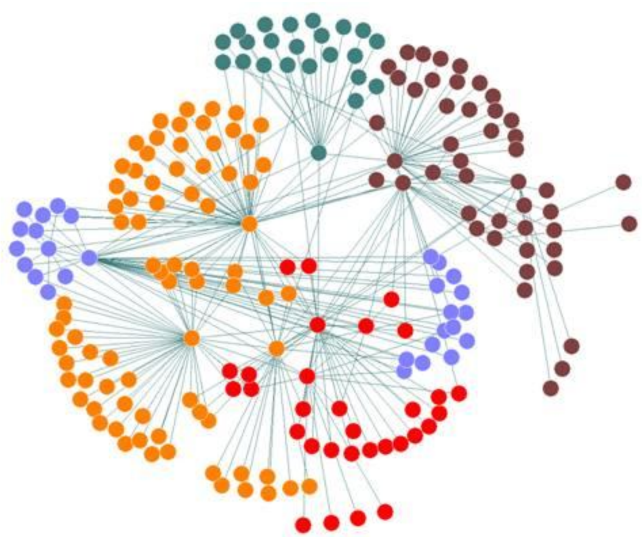
TEL AVIV UNIVERSITY

Algorithms and Applications in Social Networks



2025/2026, Semester A

Slava Novgorodov

Lesson #10

- Social Networks application examples:
  - Fraud
  - Crime
  - Terrorism
- Advices for in-practice social network analysis

Fraud detection and prevention

Motivation

- Fraud is everywhere:
  - Credit cards fraud
  - Taxes fraud
  - Fake companies fraud

- It costs our industry billions of dollars yearly

Based on: 1. GOTCHA! IMPROVING FRAUD DETECTION TECHNIQUES slides
2. Neo4j lectures about fraud detection: https://www.youtube.com/watch?v=AeNufTq1W5I

Fraud detection

Current <u>(non SNA)</u> methods:

- Machine learning algorithms that gives a score to each transaction (i.e. the probability to be fraud)
  - Improvement directions:
    - Better ML algorithms
    - More labeled data
- Rules based systems, which usually works as addition to ML techniques (usually written by experts)
  - Improvement directions:
    - Automatic rules generation
    - Better sharing of rules between experts

Example of fraud detection

Time	Amount	Transaction Type	Location	
18:02	107	Online, no CCV	Online Store	FRAUD
18:03	106	Online, no CCV	Online Store	FRAUD
18:04	112	Online, with CCV	Online Store	
19:08	114	Online, no CCV	Online Store	FRAUD
19:10	117	Online, with CCV	Online Store	
20:53	46	Offline, without PIN	GAS Station B	FRAUD
20:54	48	Offline, without PIN	GAS Station B	FRAUD
20:55	44	Offline, without PIN	GAS Station B	FRAUD
20:58	47	Offline, with PIN	Supermarket	
21:01	49	Offline, with PIN	GAS Station A	
:	:	:	:	:

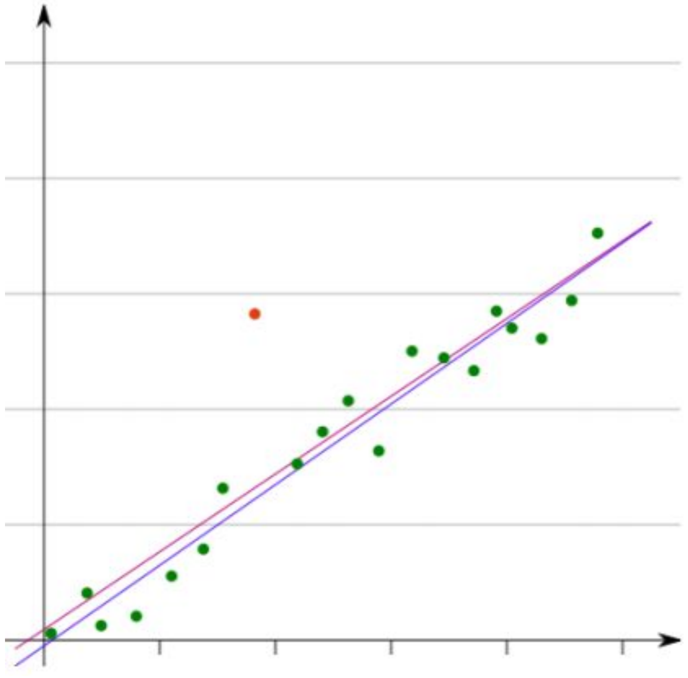
ML Score:
0.75
0.91
0.22
0.15
0.71
…

Rules:

1) Time $\in$ [18:00, 18:05] $\wedge$ Amt $\geq$ 110
2) Time $\in$ [18:55, 19:00] $\wedge$ Amt $\geq$ 110

Fraud detection

- Basic method: Anomalous behavior detection
  - Outlier detection: abnormal behavior and/or characteristics in a data set might often indicate that that person perpetrates suspicious activities.

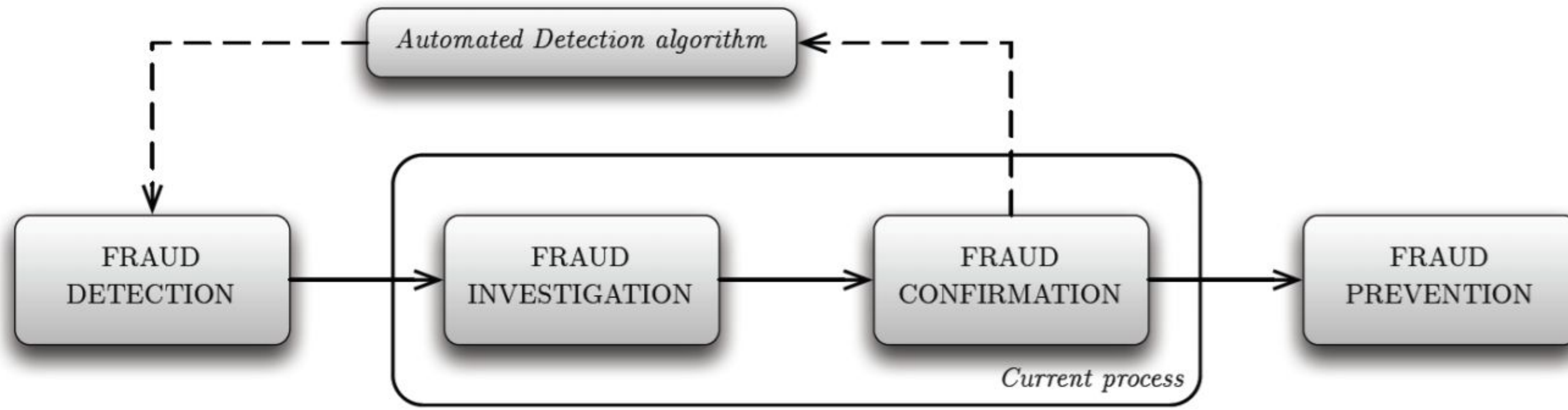


Fraud detection

- Basic method: Anomalous behavior detection
  - Pros: Very simple method
  - Cons: A lot of false positives and false negatives

Fraud detection

Current workflow:



Fraud detection

Main (not all) challenges with fraud detection:

- Unbalanced:
  - Extremely skewed class distribution
  - Big data, but only few fraudulent observations (often < 1%)
- Well-considered & Carefully organized:
  - Complex fraud structures are carefully planned
  - Outlier detection no longer sufficient: combination of patterns, preferably well-hidden
  - Relationships between fraudsters
- Imperceptibly concealed
  - Subtlety of fraud: imitating normal behavior, even in identify theft
  - Fraudsters are often first “sleeping”, pretending to be a good customer

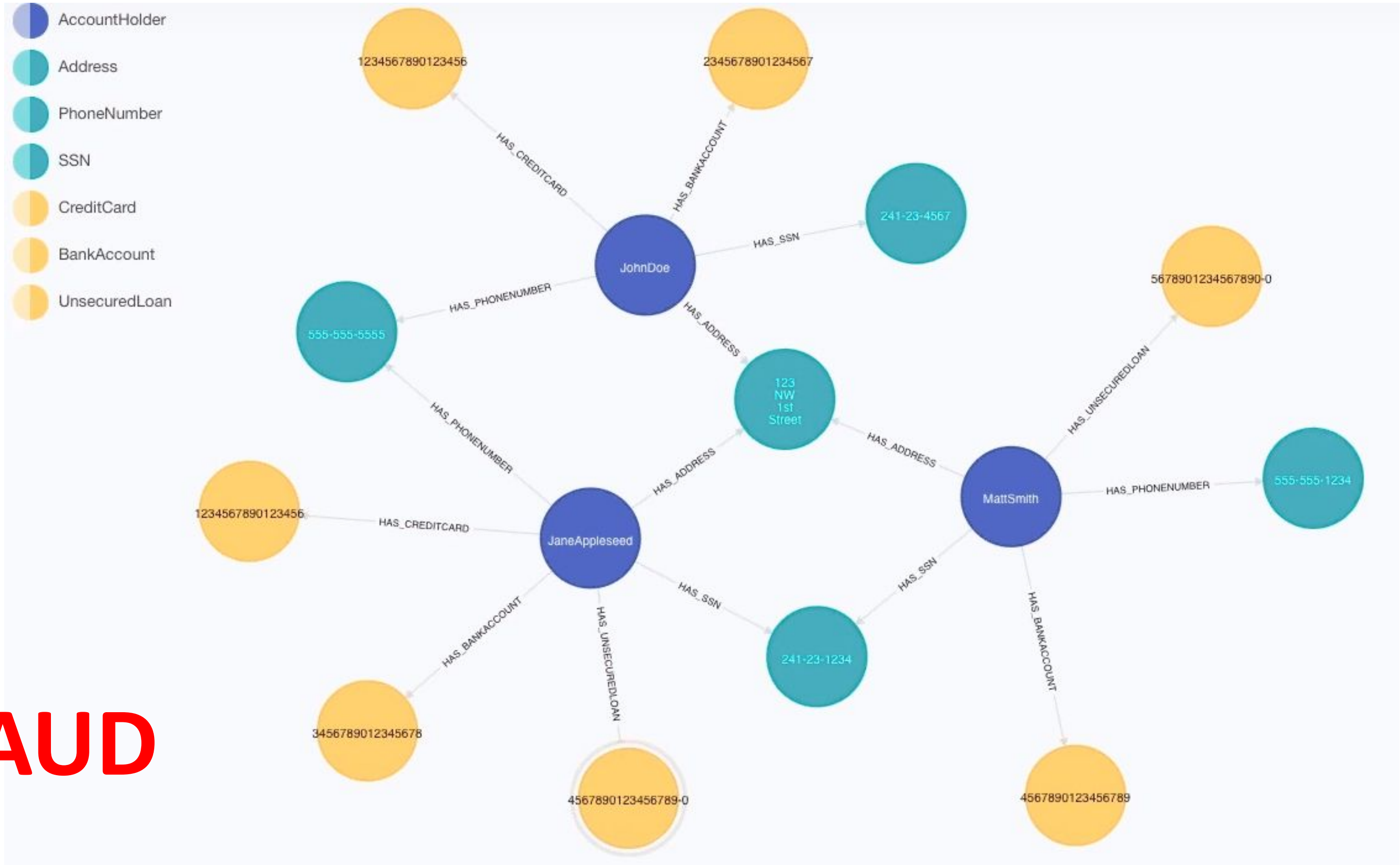
Social Networks Analysis for Fraud Detection

Model interactions as a network:

- Nodes:
  - People (Fraudsters/Victims)
  - Banks
  - Companies
  - Resources
  - ….
- Links:
  - Credit Card transactions
  - Loans
  - “belongs to” relation, “works at” relation …
  - …

Visualization can help!

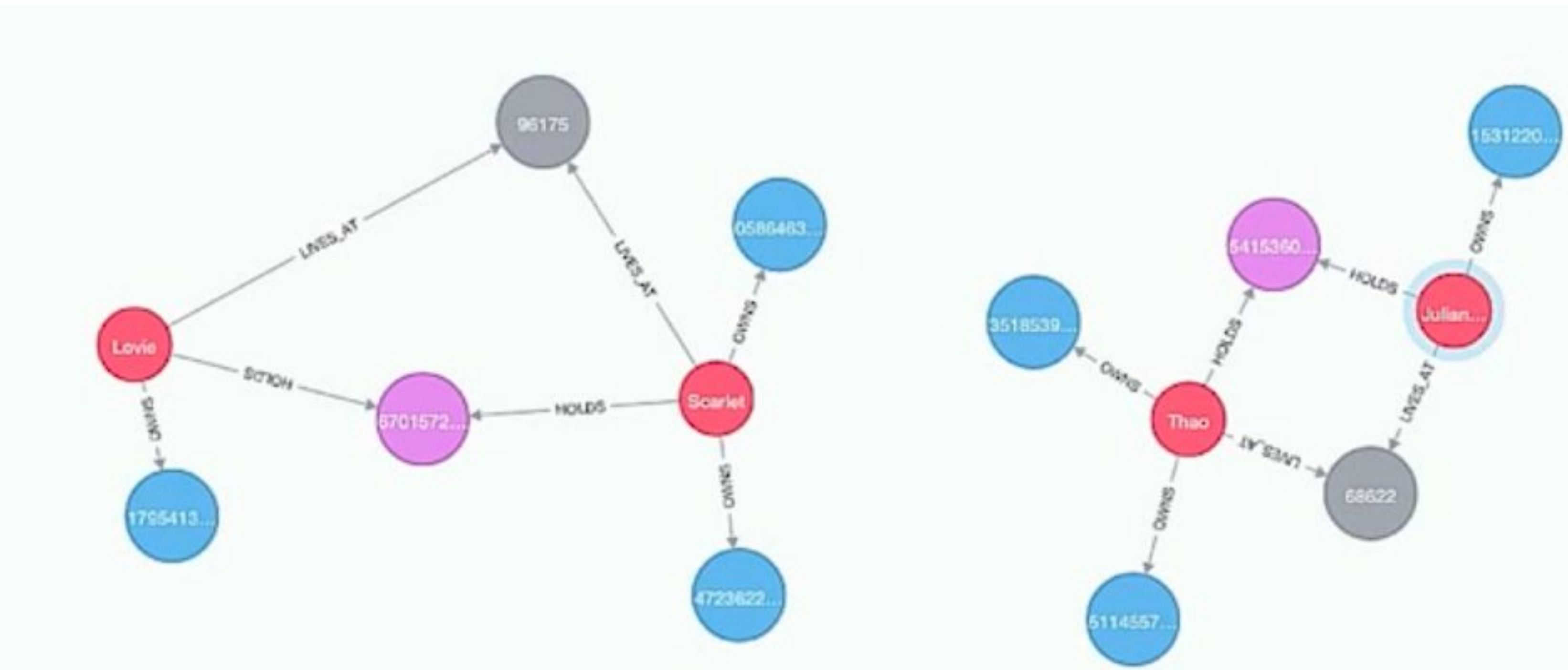
Modeling as a network can help even if you just visualize it…



FRAUD

Visualization can help!

Modeling as a network can help even if you just visualize it…

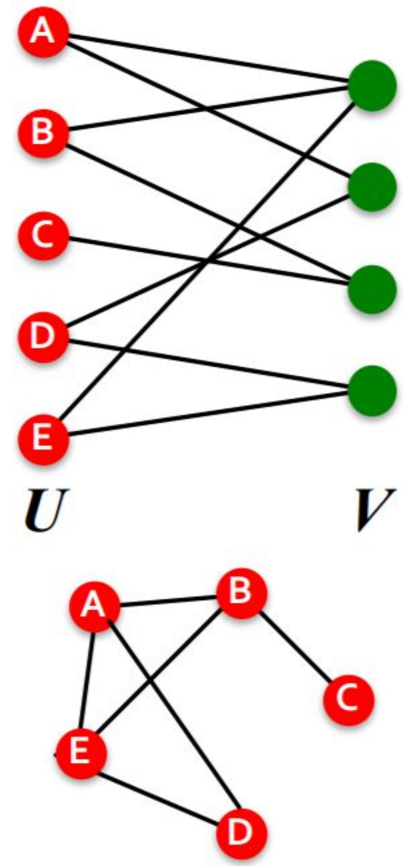


LEGITIMATE

Bipartite graphs folding

Folding:

Connect every red node to other red node if they are connected to same green node



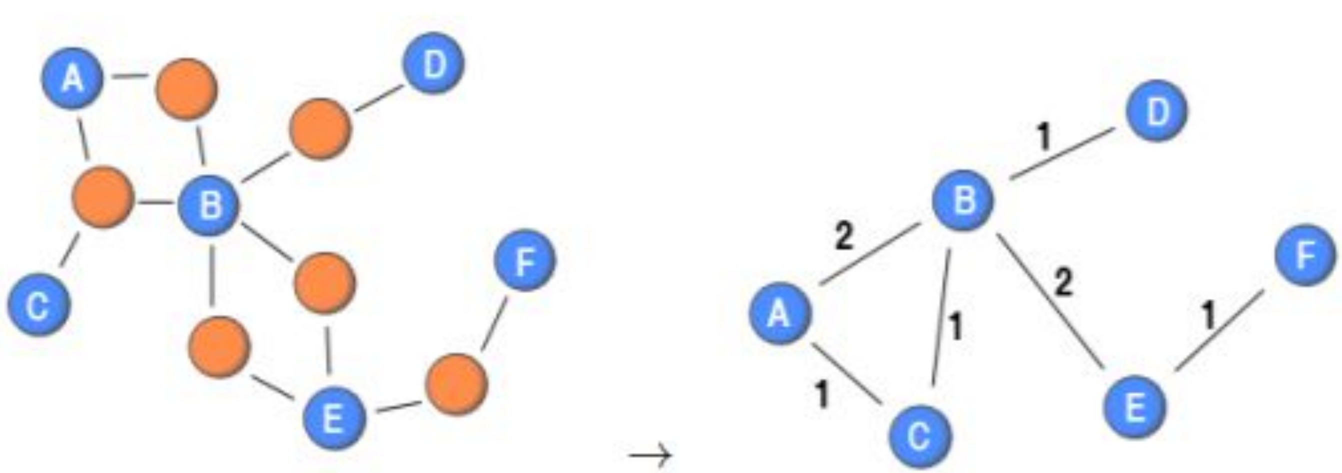
Folded version of the graph above

Bipartite graphs weighted folding

Folding:

Connect every blue node to other blue node if they are connected to same orange node.

If the node already exists, add 1 to its weight

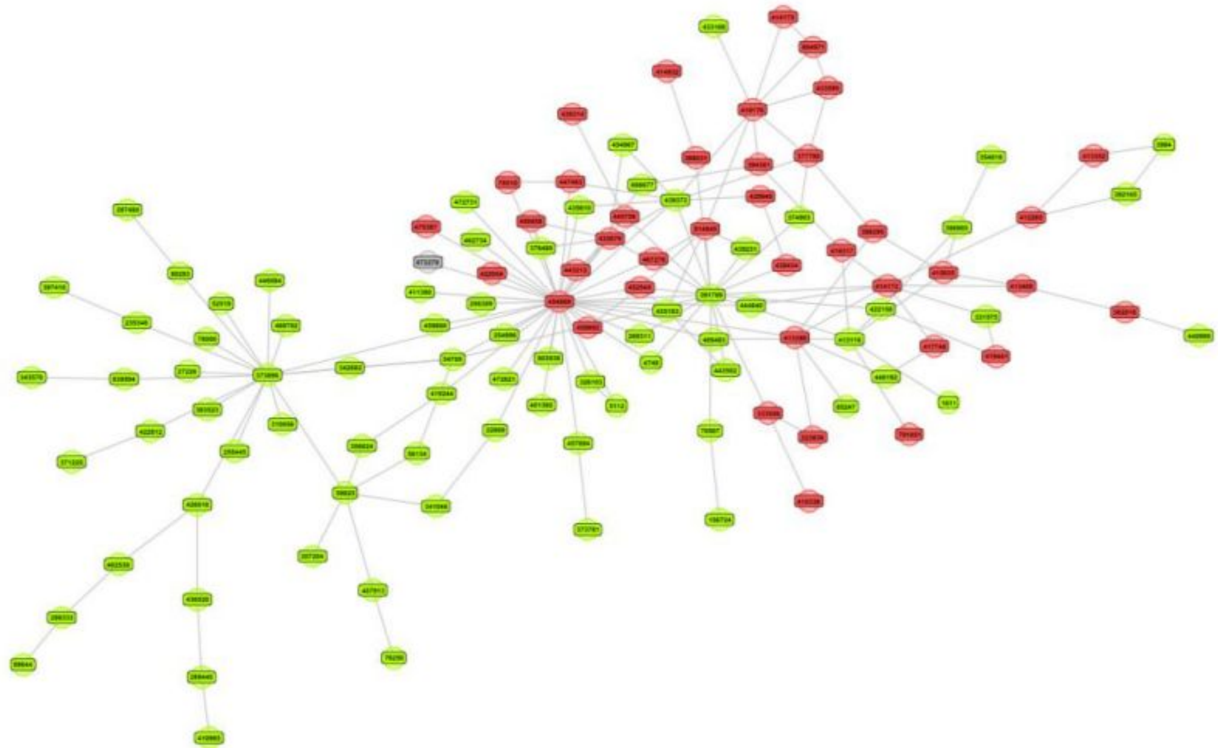


Fraud analysis “basic scheme”

1. Take the data and represent it as a network
2. Decide of the “sides” of the bipartite network
3. Fold it
4. Detect cliques, detect communities, measure centrality…

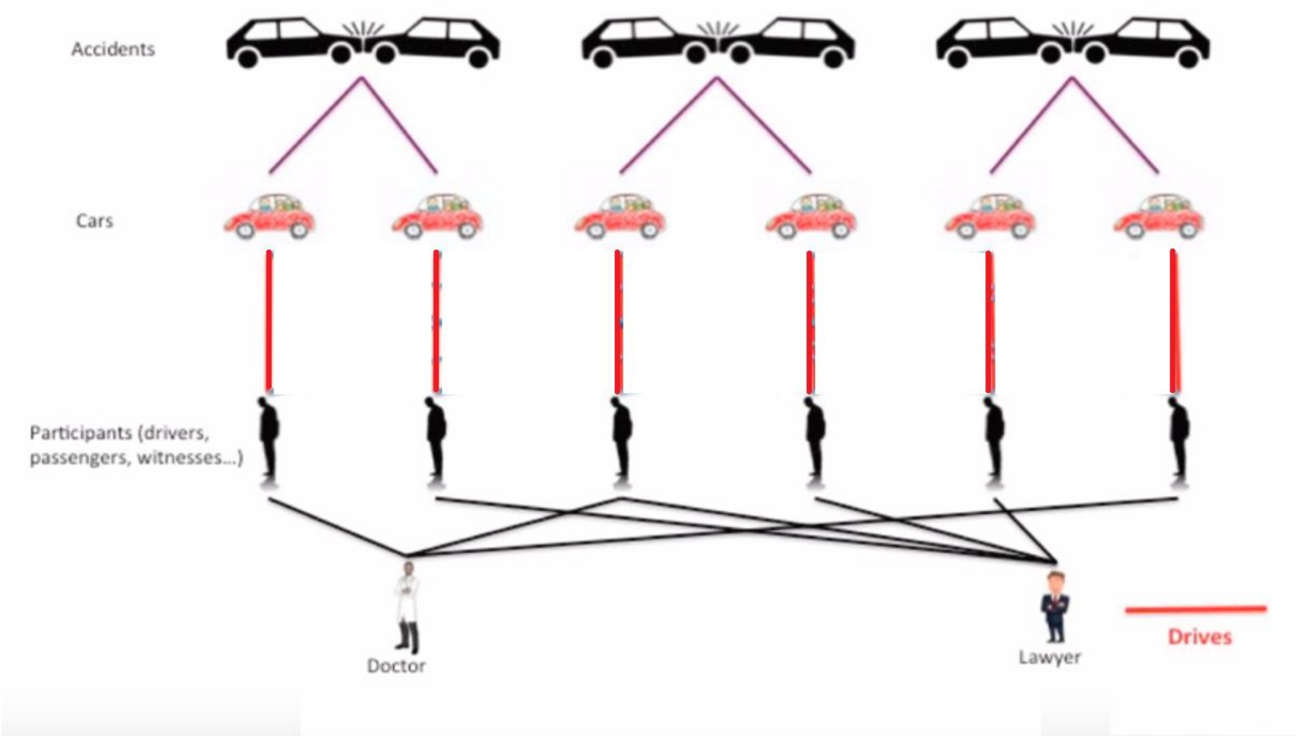
Homophily

- People tend to associate with other whom they perceive as being similar to themselves in some way. e.g.: same city, hobbies, interests…



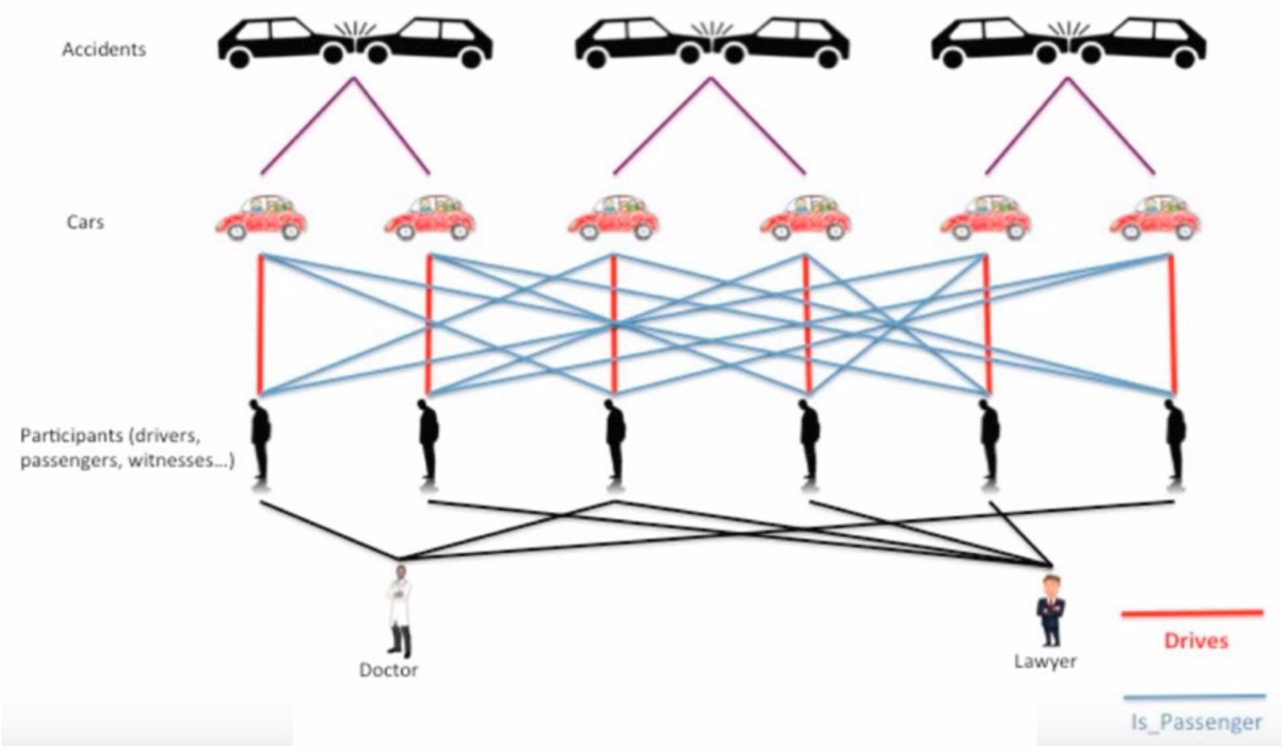
Insurance fraud

- Combining different types of links in one network can give much more information

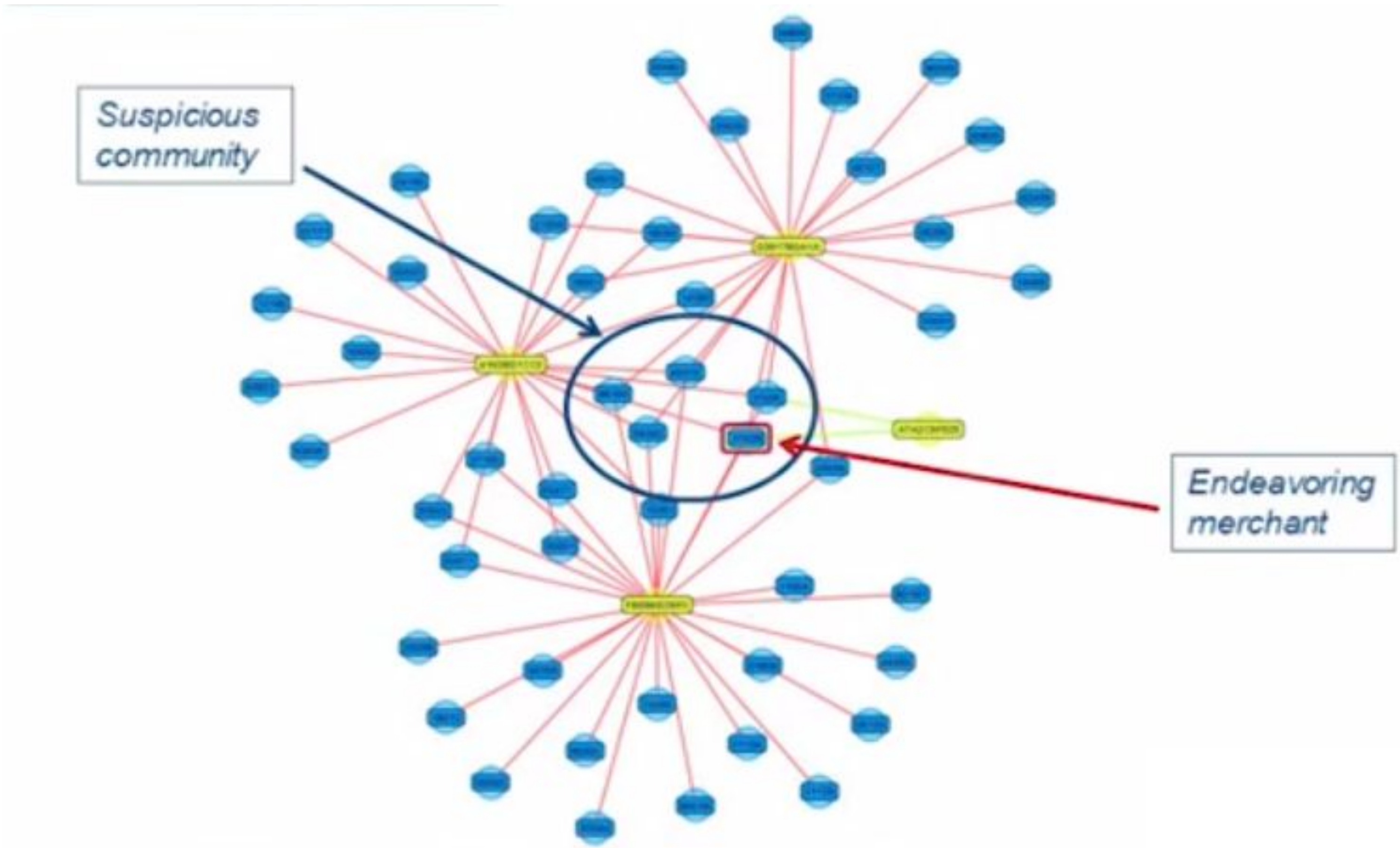


Insurance fraud

- Combining different types of links in one network can give much more information

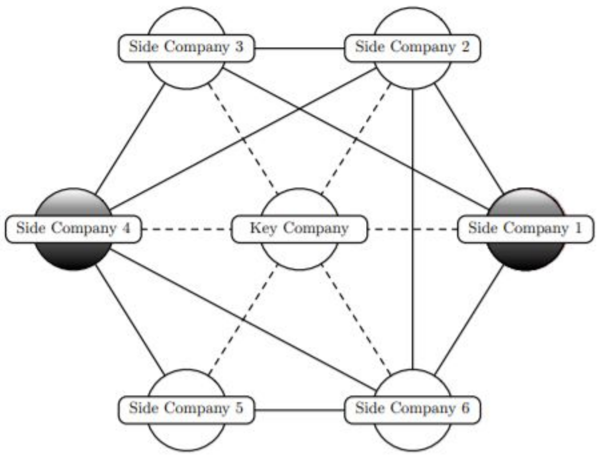


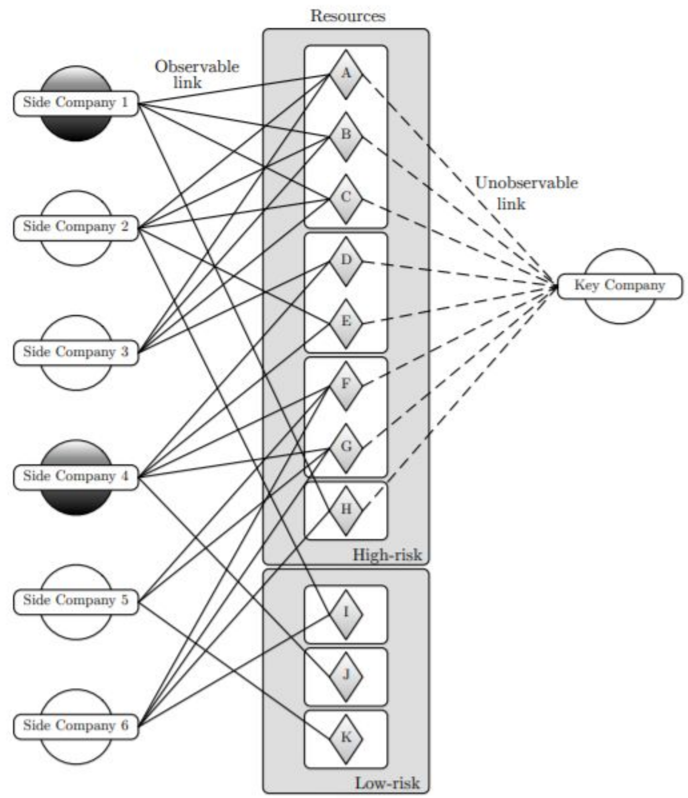
Credit Card Fraud



Very sensitive data

Taxes Fraud

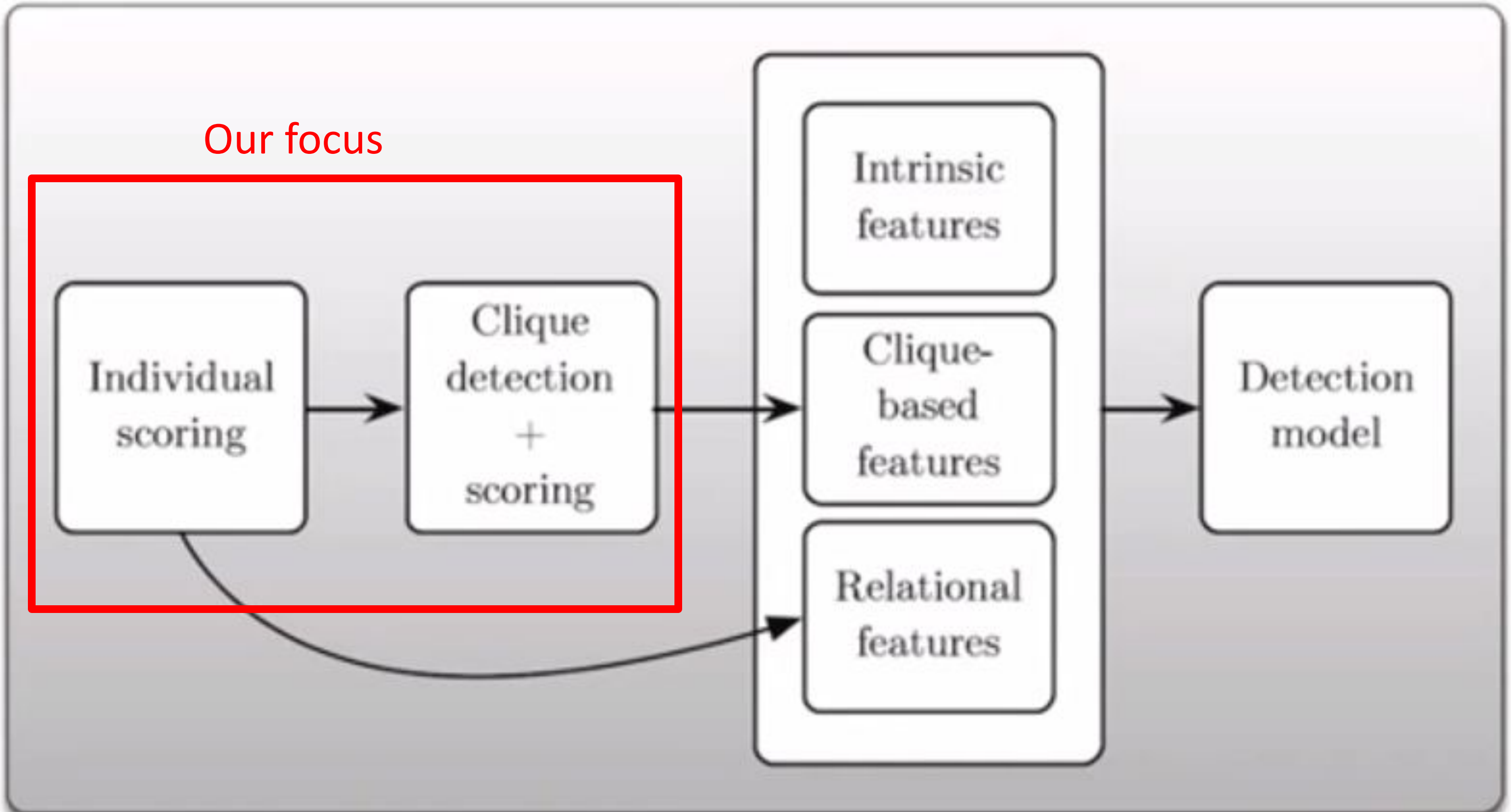




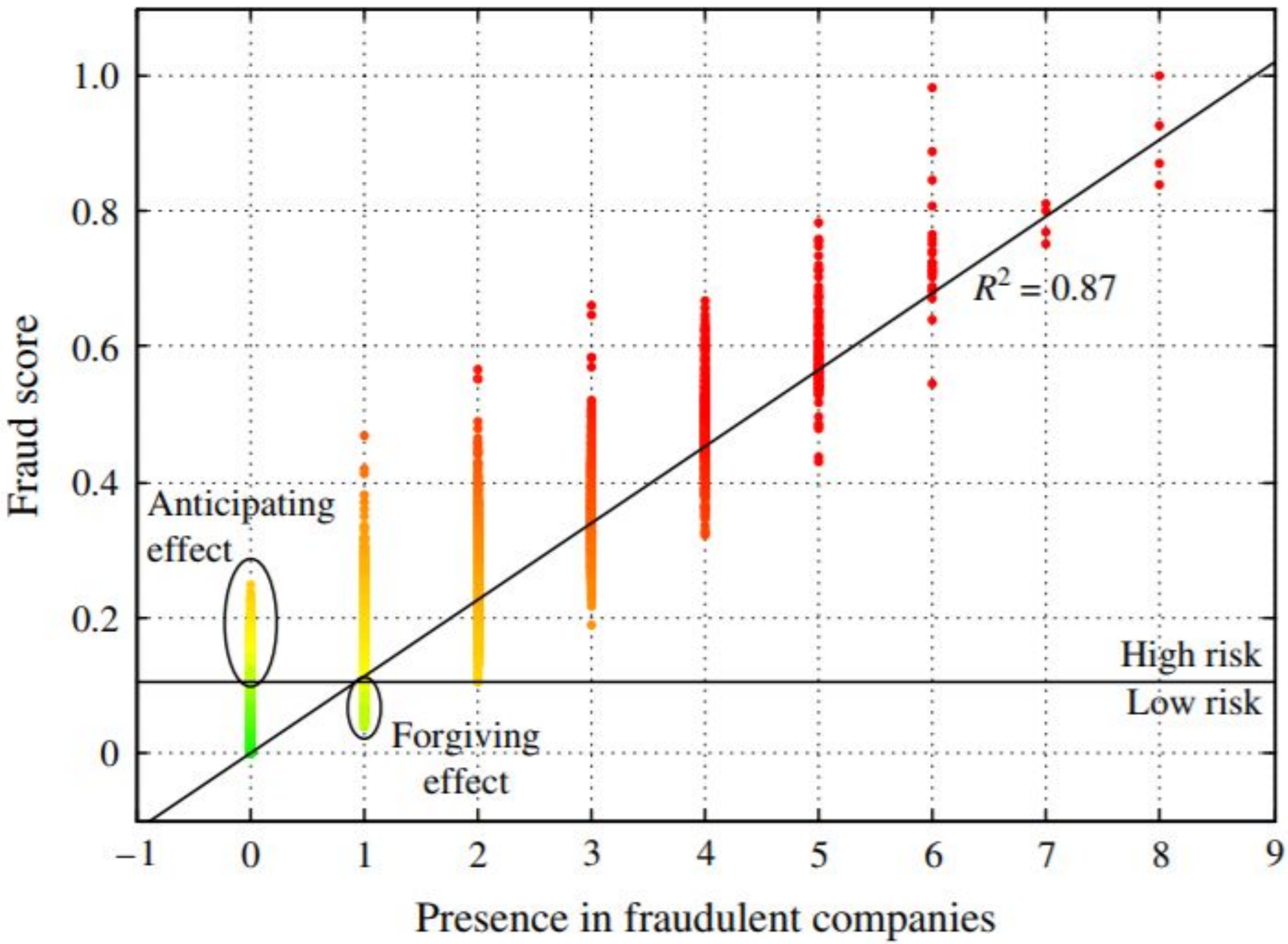
“Spider construction” fraud scheme – open a company, allocate resources, Bankrupt the company, move the resources…

The solution

- System called Gotcha! (Gotch'all):
  (by Van Vlasselaer et al.)



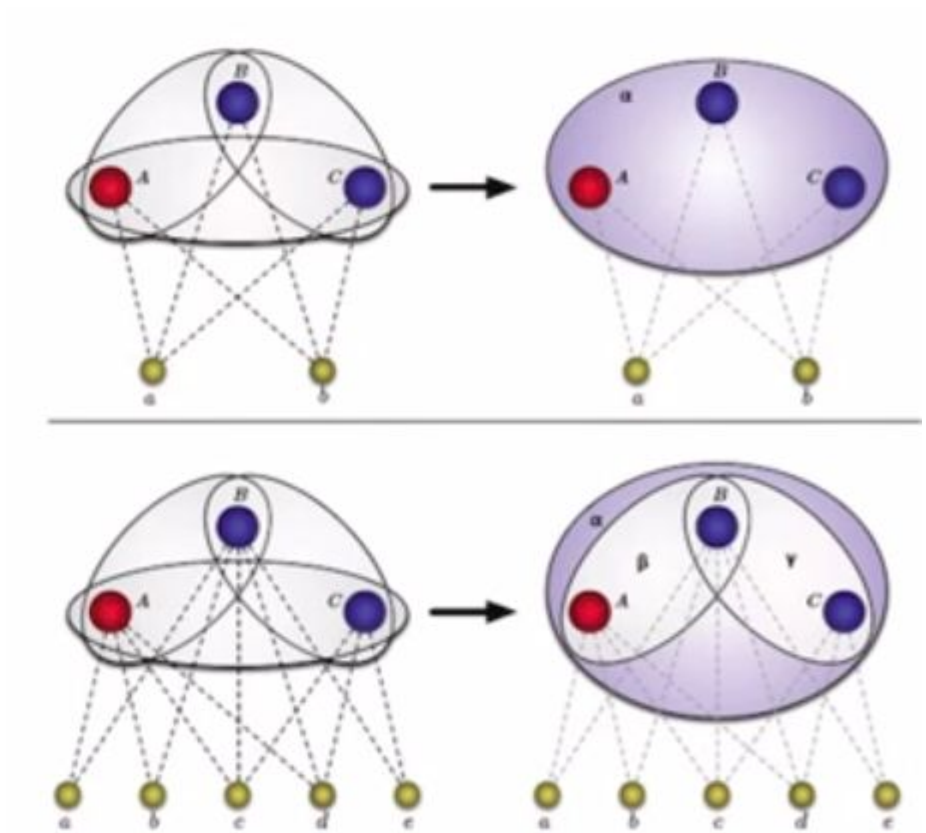
Individual Scoring



Clique detection

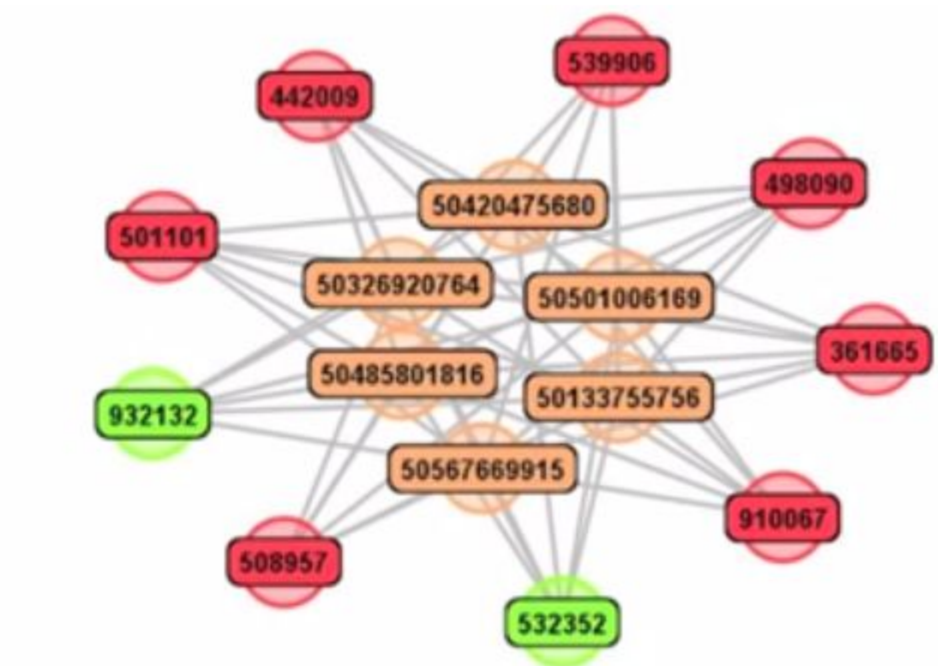
“Complete” clique

“Partial” clique



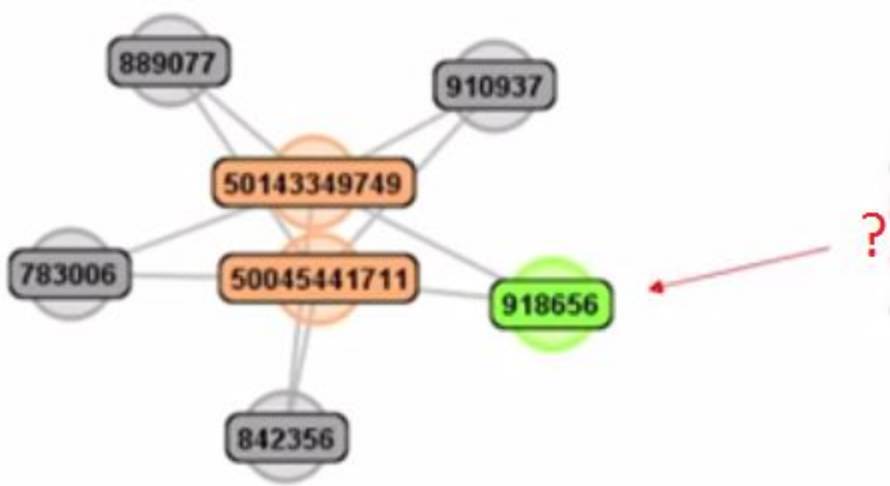
Clique scoring

Suspiciousness of the clique: How many bankrupts? How many frauds?



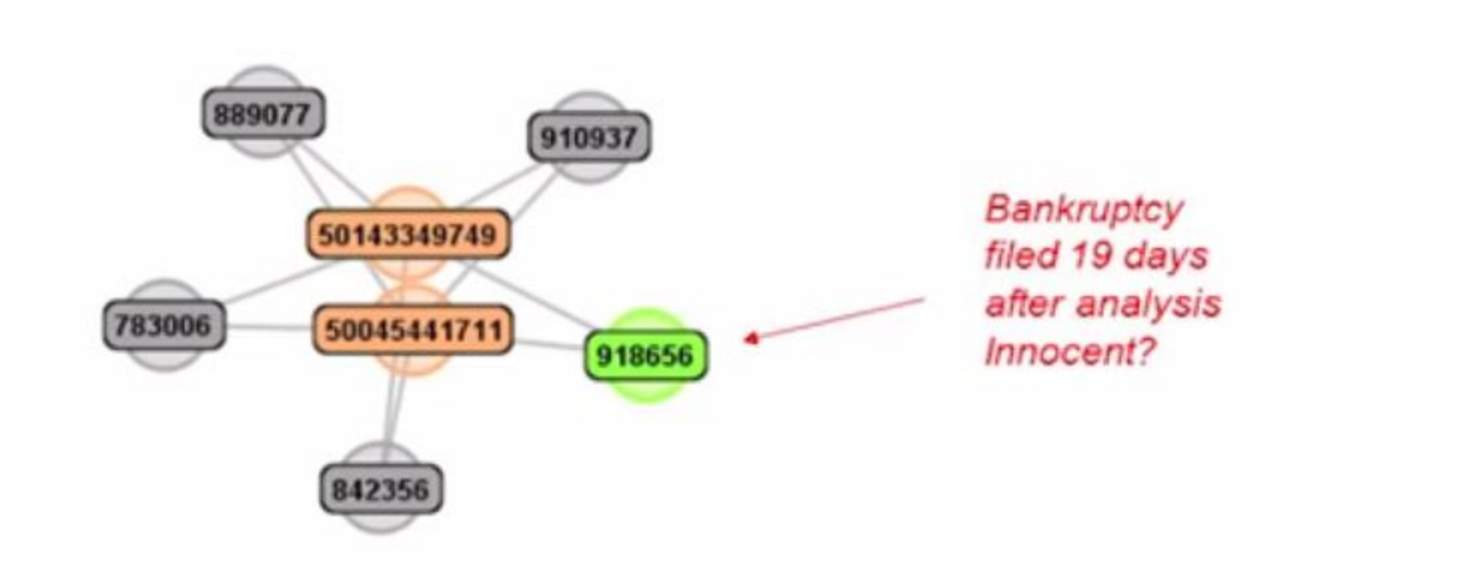
Empirical evaluation

- 5 companies, 2 resources
- 4 out of 5 companies are bankrupt
- What about the last company?



Empirical evaluation

- 5 companies, 2 resources
- 4 out of 5 companies are bankrupt
- What about the last company?



Crime detection

Motivation

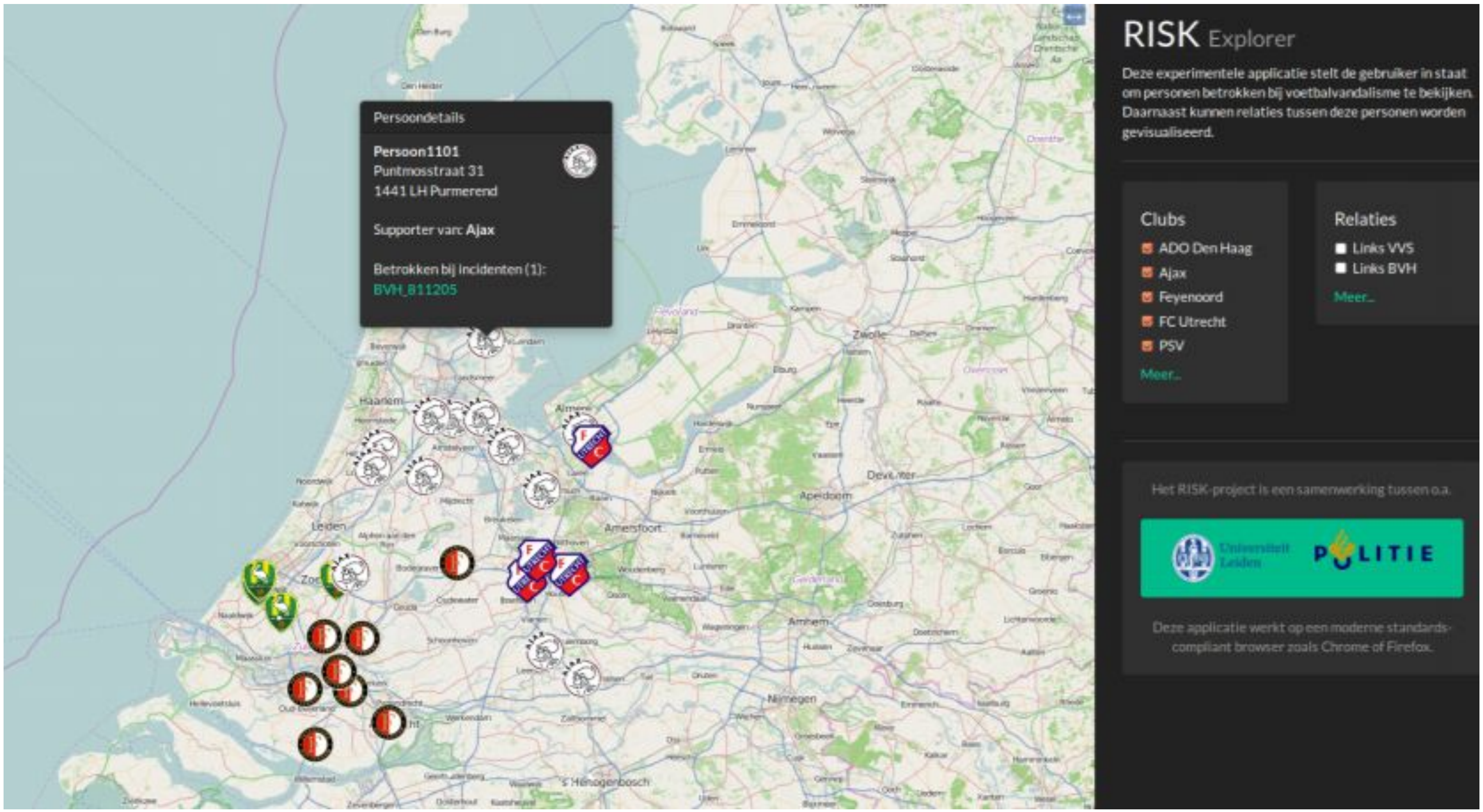
- Crime is often well organized, with individuals formed into groups/gangs, with structure and hierarchy.
- Crimes have a lot of “meta-data”, that can be better modeled as a network

Based on: 1. http://liacs.leidenuniv.nl/~takesfw/SNACS/lecture3.pdf

Dutch Police example

- Gain insight in social networks of soccer fans, group formation and organization
- Dataset: all entries in police systems of law violations of a particular group of people involved in soccer violence

Dutch Police example



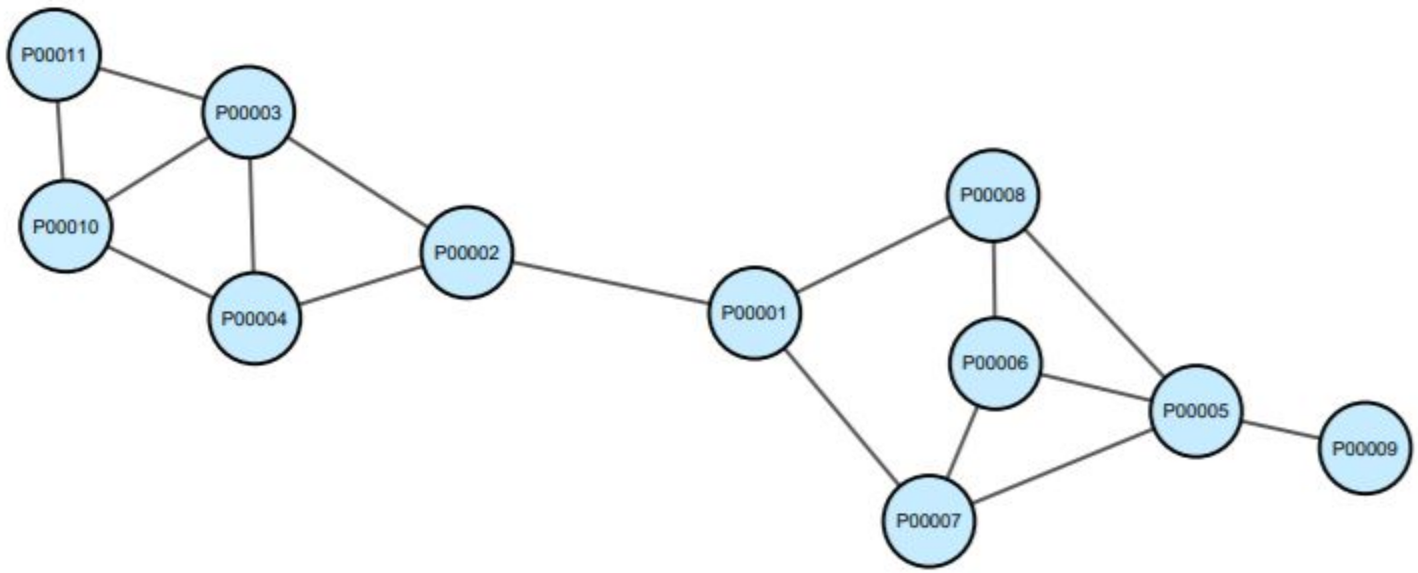
Dutch Police example - Dataset

Person ID	Incident ID	Incident Type
P000001	X00011	Straatroof/diefstal
P000001	X00014	Eenv. Mishandeling
P000002	X00011	Straatroof/diefstal
P000002	X00012	Eenv. Mishandeling
P000003	X00012	Eenv. Mishandeling
P000003	X00016	Bedreiging
P000004	X00012	Eenv. Mishandeling
P000004	X00017	Eenv. Mishandeling
P000005	X00013	Bedreiging
P000005	X00014	Eenv. Mishandeling
P000005	X00015	Straatroof/diefstal
P000006	X00013	Bedreiging
P000007	X00013	Bedreiging
P000008	X00013	Bedreiging
P000009	X00015	Straatroof/diefstal
P000010	X00016	Bedreiging
P000010	X00017	Eenv. Mishandeling
P000011	X00016	Bedreiging

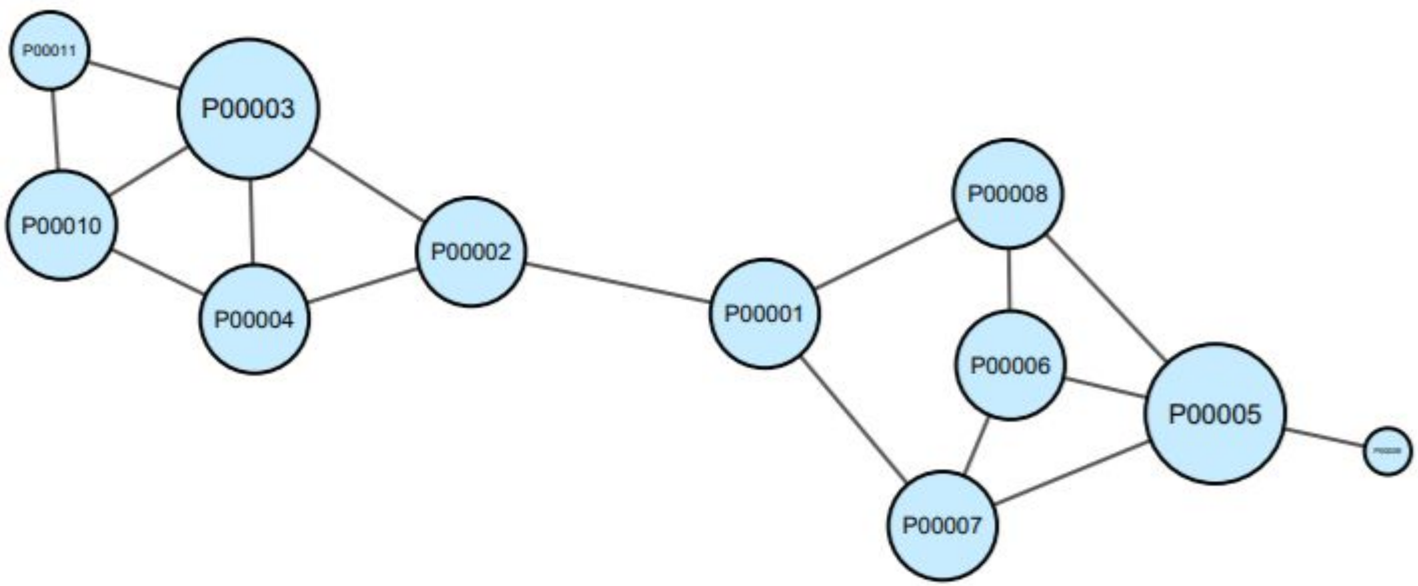
Dutch Police example - Dataset

Folded bipartite graph (people and incidents):

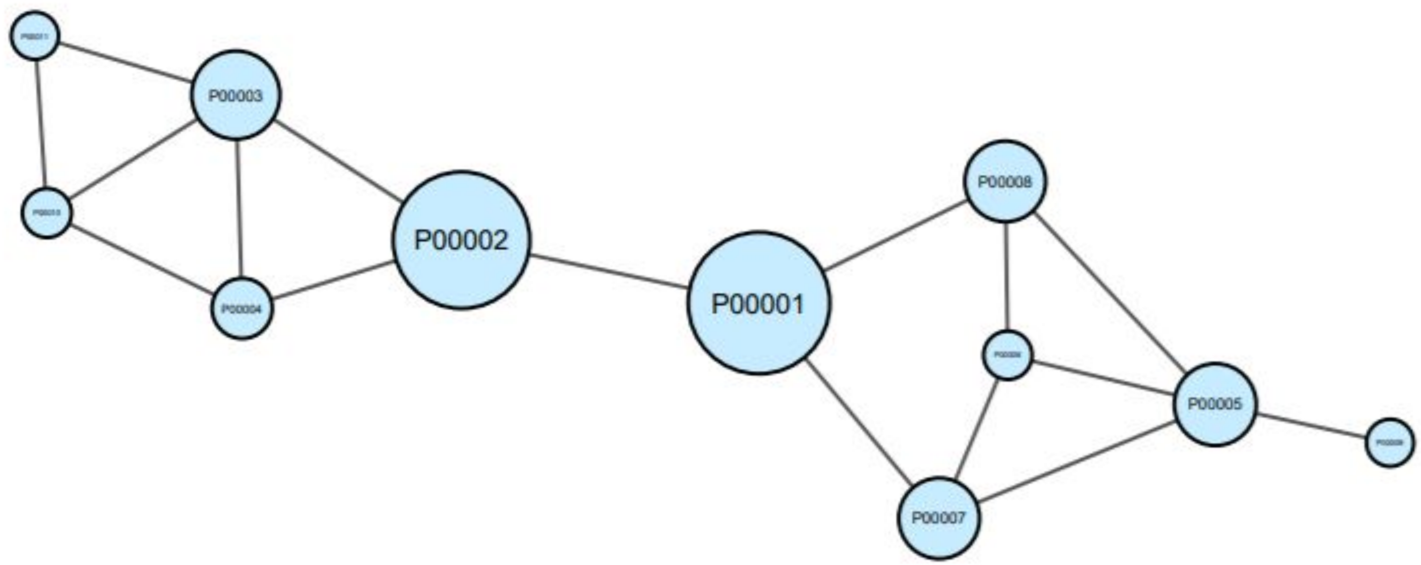
Dutch Police example - Visualization



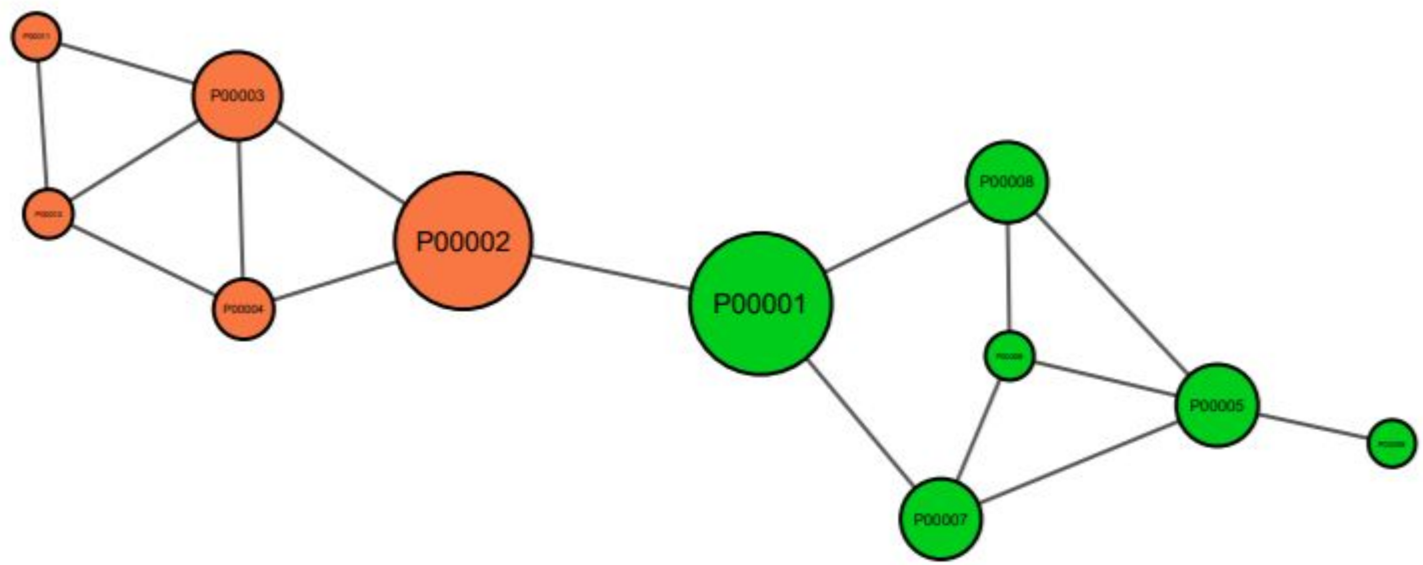
Dutch Police example - Centrality



Dutch Police example - Centrality

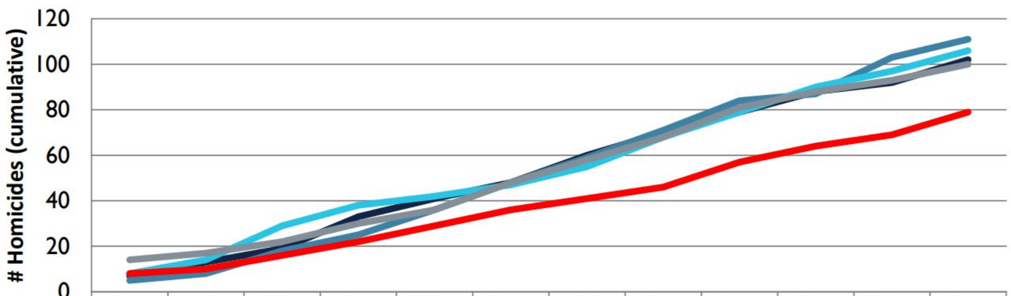


Dutch Police example - Communities



More examples from PD

- Kansas City crime – “Operation Clean Sweep” (2013):
  - Historically, one of the top 10 most violent cities in the US
  - Averages 106 homicides per year
  - Averages 3,484 aggravated assaults per year
- Results:



	Jan	Feb	Mar	Apr	May	June	Jul	Aug	Sep	Oct	Nov	Dec
2010	7	13	19	33	41	48	60	70	79	88	92	102
2011	5	8	18	25	36	48	59	71	84	87	103	111
2012	8	14	29	38	42	47	55	68	79	90	97	106
2013	14	17	22	30	36	48	58	68	81	88	93	100
2014	8	10	16	22	29	36	41	46	57	64	69	79

- Details: https://www.nationalpublicsafetypartnership.org/Documents/VRN%20Social%20Network%20Analysis%20Presentation%20July%2021%202015.pdf

Finding Terrorists Cells

9/11 Case Study

- Analyzing such networks is much easier in past, not in future. But still important for the prosecution and potentially detecting other members

- Based on Valdis E. Krebs analysis

http://insna.org/PDF/Connections/v24/2001_I-3-7.pdf

THE HIJACKERS ...

American Airlines 11
Crashed into WTC (north)


Mohamed Atta
(Egyptian)
Received pilot training


Waleed M. Alshehri
(Saudi)
Commercial pilot


Wail Alshahri
(Saudi)
Possible pilot training


Satam al-Suqami
(Nationality unknown)

No picture available
Abdulaziz Alomari*
(Saudi)
Possible pilot training

UnitedAirlines 175
Crashed into WTC (south)


Marwan al-Shehhi
(United Arab Emirates)
Received pilot training

No picture available
Fayez Ahmed
(Believed to be Saudi)


Ahmed Alghamdi
(Possibly Saudi)


Hamza Alghamdi
(Believed to be Saudi)
Possible pilot training


Mohald Alshehri
(Nationality unknown)
Possible pilot training

American Airlines 77
Crashed into Pentagon


Khalid al-Midhar
(Nationality unknown)
Received pilot training


Majed Moqed
(Nationality unknown)


Salem Alhamzi*
(Saudi)
Possible pilot training


Nawaf Alhamzi*
(Saudi)


Hani Hanjour
(Saudi)

United Airlines 93
Crashed in Pennsylvania


Ziad Jarrah
(Lebanese)
Received pilot training


Ahmed Alhaznawi
(Saudi)


Ahmed Alnami
(Nationality unknown)


Saeed Alghamdi*
(Seems to be Saudi)

*Disputed identity

AND HOW THEY WERE CONNECTED

Attended same technical college

Hamburg, Germany
Mohamed Atta
Marwan al-Shehhi
Ziad Jarrah

Took flight classes together

Pilot schools in Florida
Mohamed Atta
Marwan al-Shehhi

Pilot schools In San Diago
Khalid al-Midhar
Nawaf Alhamzi

Known to be together in week before attacks

Stayed together in a Florida motel
Mohamed Atta
Marwan al-Shehhi

Attended a gym in Maryland (Sept 2-6), also seen dining together
Khalid al-Midhar
Majed Moqed
Salem Alhamzi
Nawaf Alhamzi
Hani Hanjour

Bought flight tickets using same address

- Mohamed Atta*
  Marwan al-Shehhi
  Abdulaziz Alomari*

  * Also used same credit card
- Waleed M. Alshehri
  Wail Alshahri
- Fayez Ahmed
  Mohald Alshehri
- Ahmed Alghamdi
  Hamza Alghamdi

Bought flight tickets together

Mohamed Atta
Ziad Jarrah
Ahmed Alhaznawi

Picked up tickets bought earlier in Baltimore
Khalid al-Midhar
Majed Moqed

Bought from the same travel agent in Florida
Ahmed Alnami
Saeed Alghamdi

Last known address

Hollywood, Florida
Marwan al-Shehhi
Waleed M. Alshehri
Wail Alshahri
Ziad Jarrah
Hani Hanjour

Other cities in Florida
Mohamed Atta
Fayez Ahmed
Ahmed Alghamdi
Mohald Alshehri
Khalid al-Midhar
Ahmed Alhaznawi
Ahmed Alnami
Saeed Alghamdi

Outside Florida
Satam al-Suqami
Hamza Alghamdi
Abdulaziz Alomari
Majed Moqed
Salem Alhamzi
Nawaf Alhamzi

9/11 Case Study

- The beginning (January 2000):

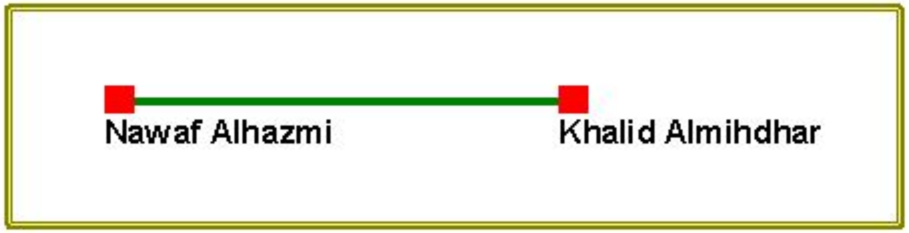


Figure 1 - Two known suspects in January 2000

9/11 Case Study

- USS Cole attack (October 2000)

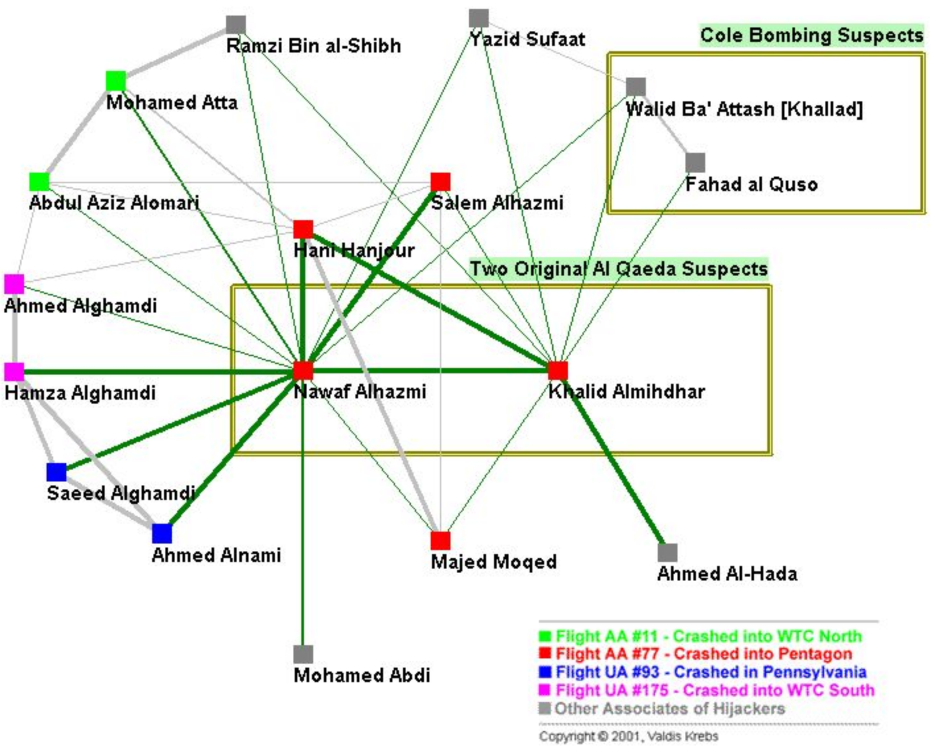
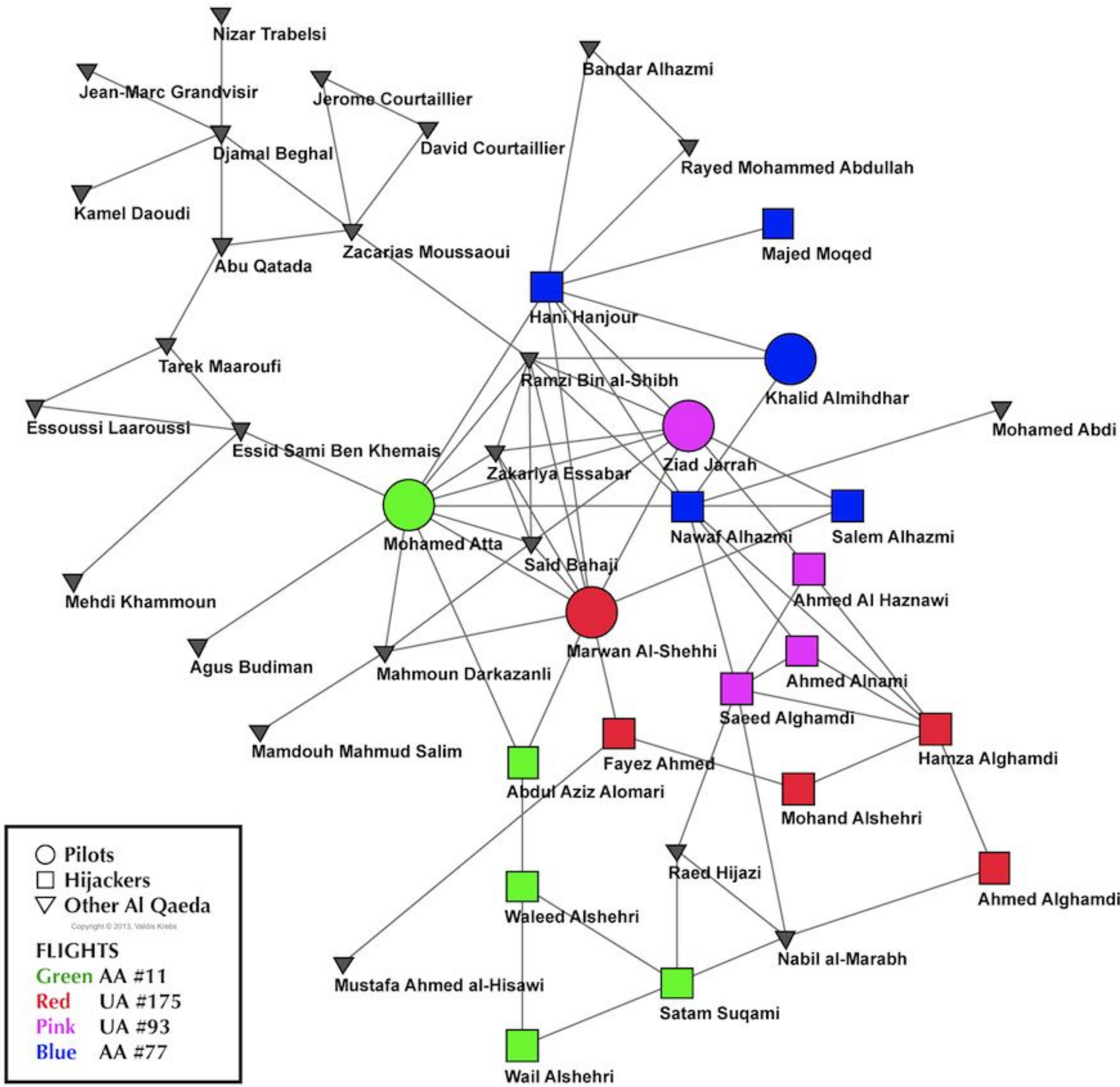


Figure 2 - All nodes within 1 step [direct link] of original suspects

9/11 Case Study

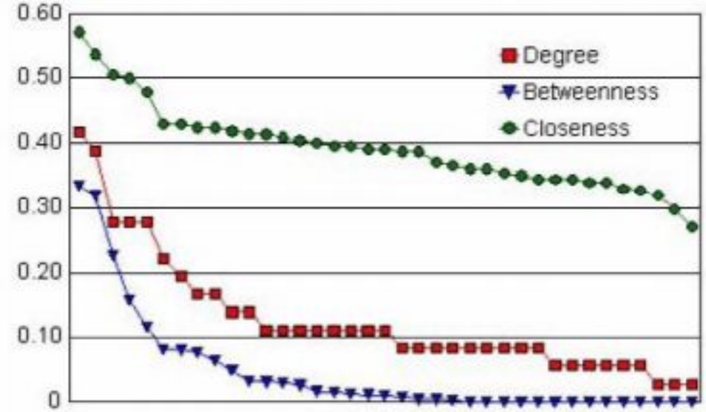


9/11 Case Study

Group Size	37
Potential Ties	1332
Actual Ties	170
Density	13%

Geodesics

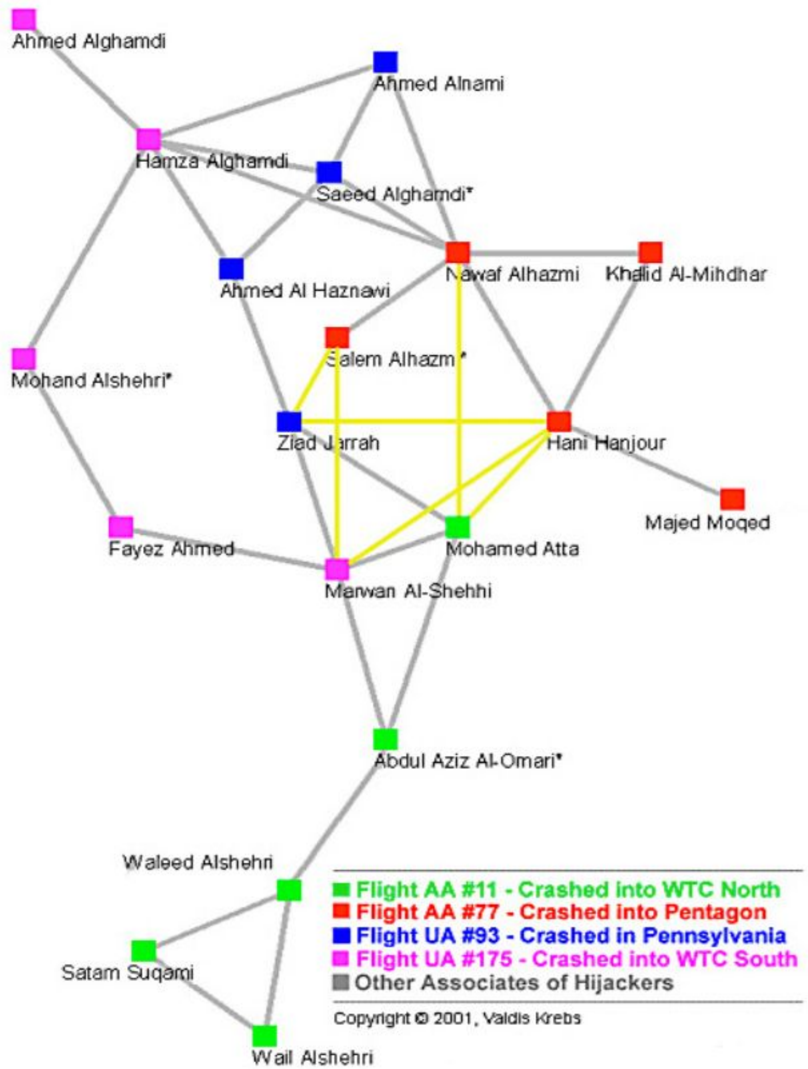
length	#
1	170
2	626
3	982
4	558
5	136
6	0

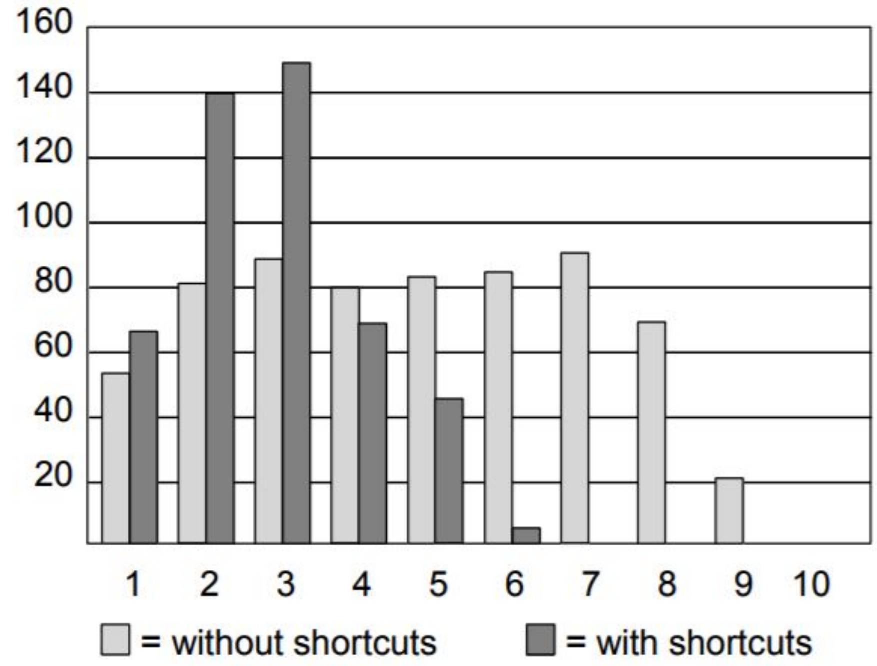


Degrees		Betweenness		Closeness	
0.417	Mohamed Atta	0.334	Nawaf Alhazmi	0.571	Mohamed Atta
0.389	Marwan Al-Shehhi	0.318	Mohamed Atta	0.537	Nawaf Alhazmi
0.278	Hani Hanjour	0.227	Hani Hanjour	0.507	Hani Hanjour
0.278	Nawaf Alhazmi	0.158	Marwan Al-Shehhi	0.500	Marwan Al-Shehhi
0.278	Ziad Jarrah	0.116	Saeed Alghamdi*	0.480	Ziad Jarrah
0.222	Ramzi Bin al-Shibh	0.081	Hamza Alghamdi	0.429	Mustafa al-Hisawi
0.194	Said Bahaji	0.080	Waleed Alshehri	0.429	Salem Alhazmi*
0.167	Hamza Alghamdi	0.076	Ziad Jarrah	0.424	Lotfi Raissi
0.167	Saeed Alghamdi*	0.064	Mustafa al-Hisawi	0.424	Saeed Alghamdi*
0.139	Lotfi Raissi	0.049	Abdul Aziz Al-Omari*	0.419	Abdul Aziz Al-Omari*
0.139	Zakariya Essabar	0.033	Satam Suqami	0.414	Hamza Alghamdi
0.111	Agus Budiman	0.031	Fayez Ahmed	0.414	Ramzi Bin al-Shibh
0.111	Khalid Al-Mihdhar	0.030	Ahmed Al Haznawi	0.409	Said Bahaji
0.111	Mounir El Motassadeq	0.026	Nabil al-Marabh	0.404	Ahmed Al Haznawi
0.111	Mustafa al-Hisawi	0.016	Raed Hijazi	0.400	Zakariya Essabar
0.111	Nabil al-Marabh	0.015	Lotfi Raissi	0.396	Agus Budiman
0.111	Rayed Abdullah	0.012	Mohand Alshehri*	0.396	Khalid Al-Mihdhar
0.111	Satam Suqami	0.011	Khalid Al-Mihdhar	0.391	Ahmed Alnami
0.111	Waleed Alshehri	0.010	Ramzi Bin al-Shibh	0.391	Mounir El Motassadeq

9/11 Case Study

Final meetings (shortcuts) in gold





9/11 Case Study

Data to build the network

Relationship / Network	**Data Sources**
1. Trust	Prior contacts in family, neighborhood, school, military, club or organization. Public and court records. Data may only be available in suspect's native country.
2. Task	Logs and records of phone calls, electronic mail, chat rooms, instant messages, web site visits. Travel records. Human intelligence – observation of meetings and attendance at common events.
3. Money & Resources	Bank account and money transfer records. Pattern and location of credit card use. Prior court records. Human intelligence – observation of visits to alternate banking resources such as Hawala.
4. Strategy & Goals	Web sites. Videos and encrypted disks delivered by courier. Travel records. Human intelligence – observation of meetings and attendance at common events

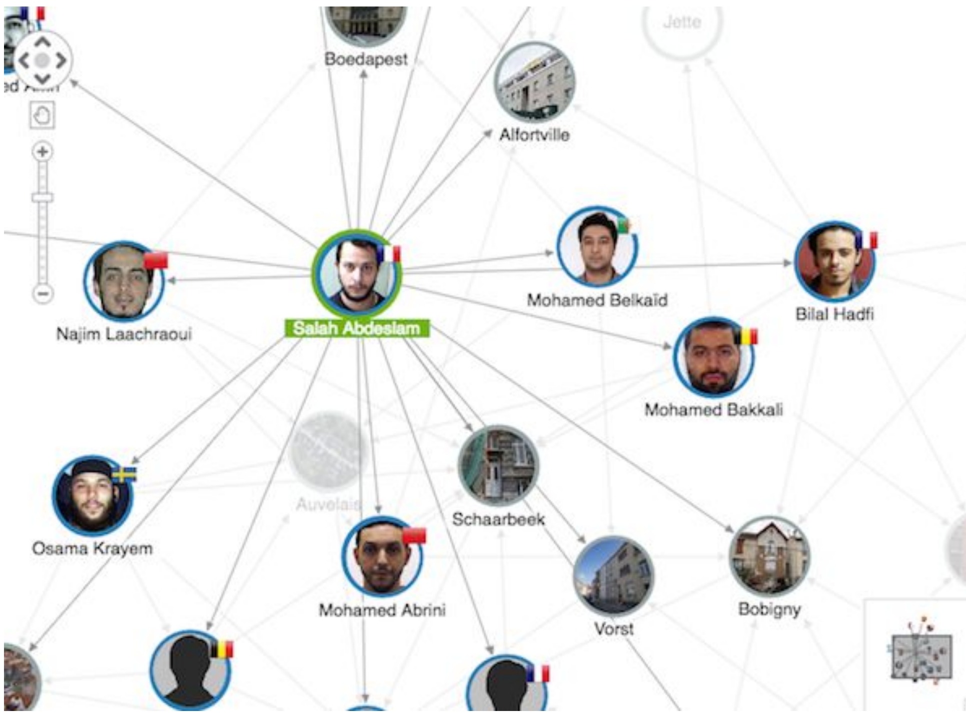
Technologies in practice

- Small networks or Initial/Partial analysis:
  - Python / NetworkX

- Huge networks
  - Graph databases, such as Neo4j
  - Distributed systems like Spark/Hadoop

Visualization, visualization, visualization…

- Very useful in Social Network analysis, helps faster identify patters and important details



Thank you!

Questions?