GRAPH-BASED ANOMALY DETECTION

SHIVA SHADROOH

NTNU

sfi Centre for Research-based Innovation

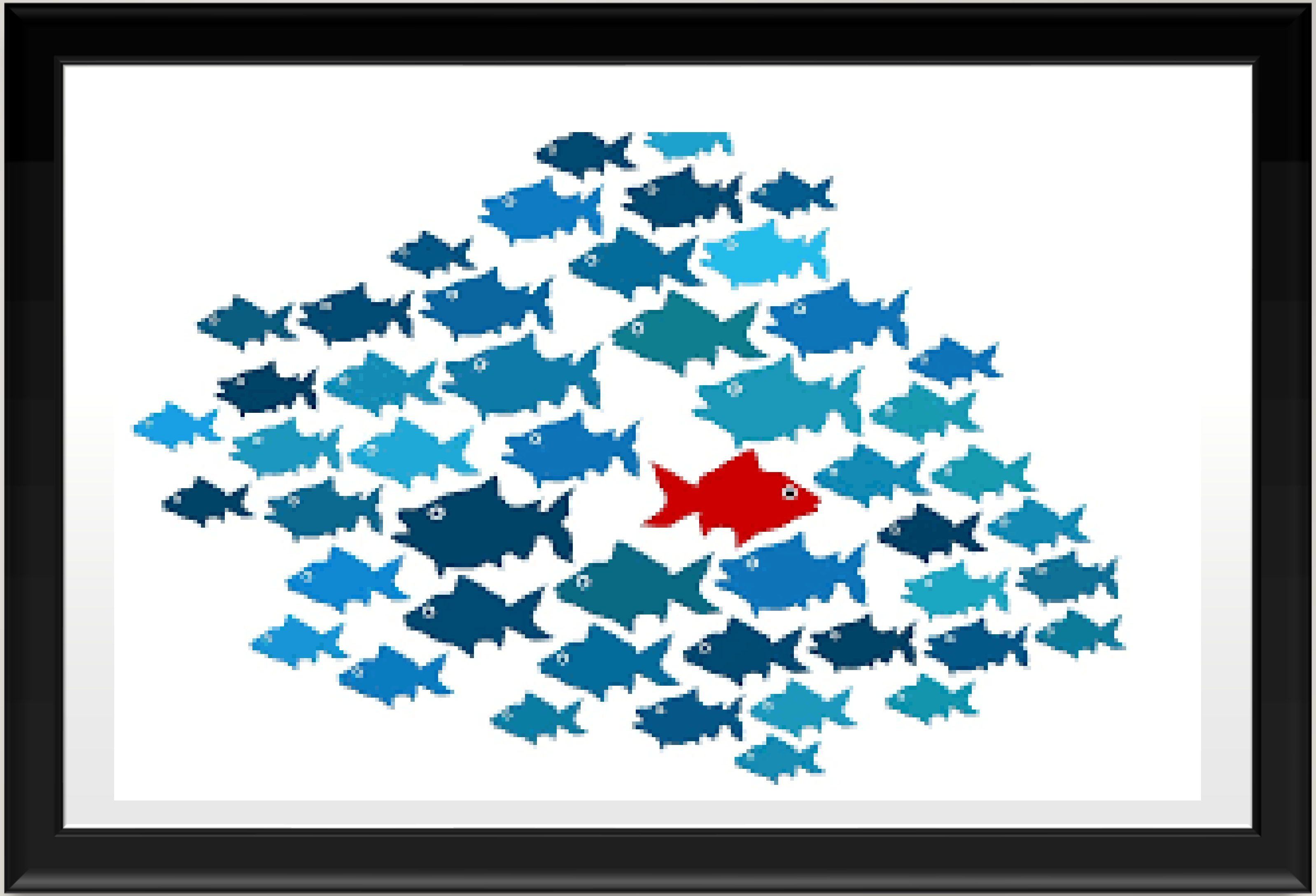
The Research Council of Norway

ROADMAP

- Introduction
- Formalization of anomalies
- Graphs anomalies and applications
- Anomaly detection in dynamic financial network

ANOMALY: THAT STANDS OUT

observations in collections of data that stand out, typically those which not obey the norms of the data.



ANOMALY DETECTION: USE-CASES

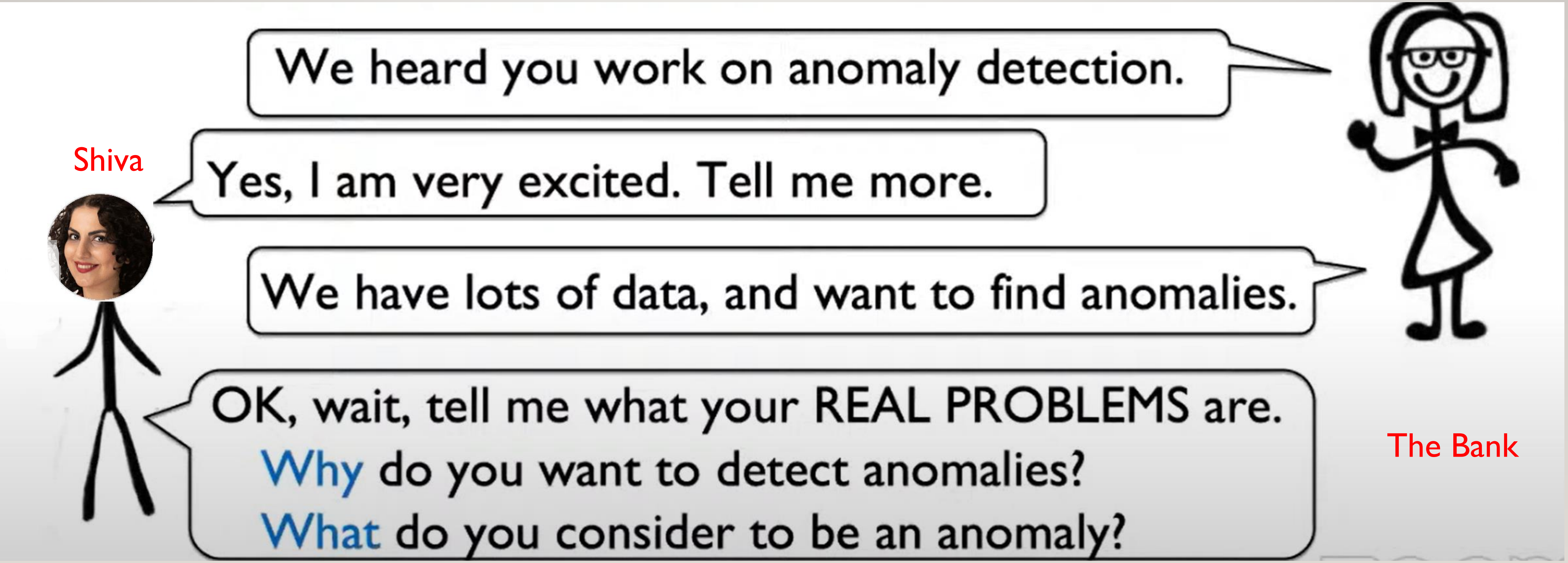
- Cyber security (cyber terrorism)
- Surveillance
- Fraud detection (insurance, credit card)
- Money laundering
- Advertisement fraud
- Fake identity or opinion fraud in social networks
- Insider trading

FORMALIZING ANOMALY DETECTION

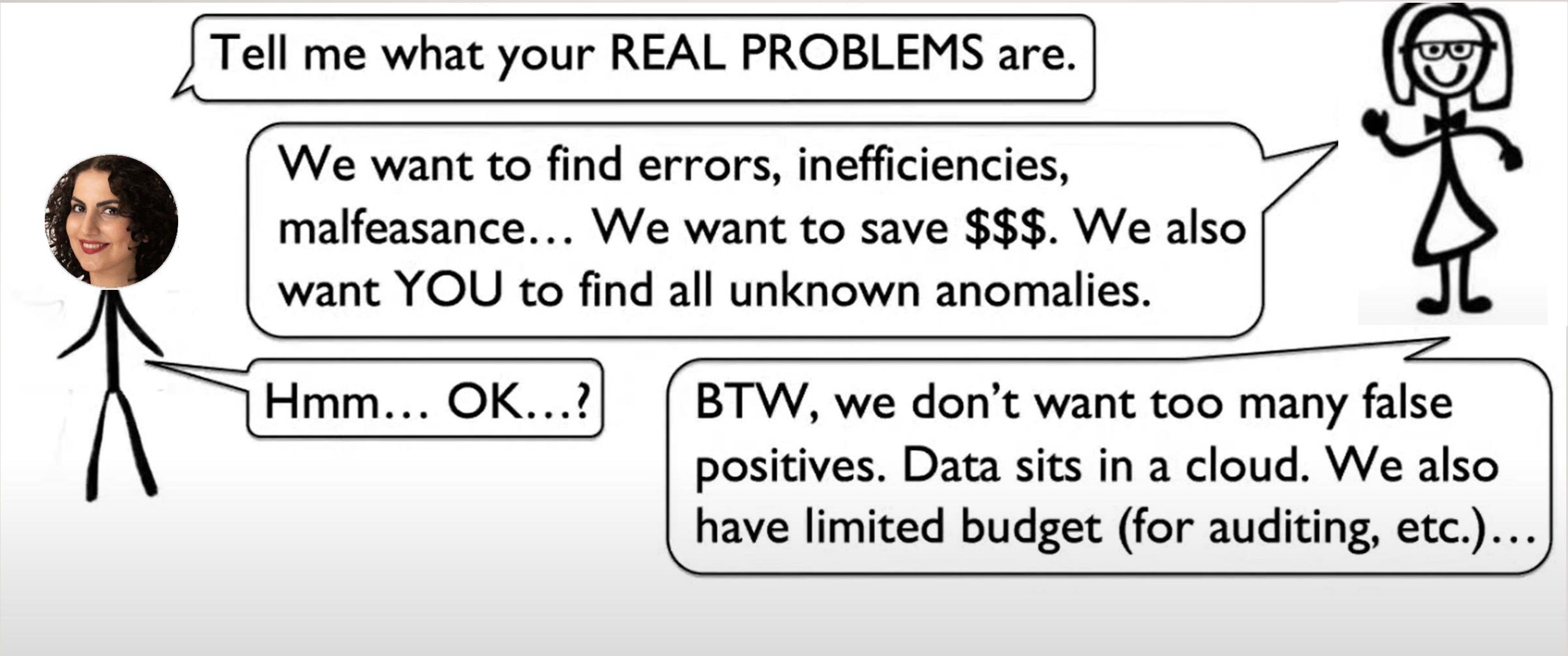
- Concrete problem settings.
- Three types of anomalies:
  - Global outliers
  - Local outliers
  - Collective outliers
- Real world? A bit more complex

FORMALIZING ANOMALY DETECTION (REAL WORLD)

- Given <DATA>, Find <ANOMALIES> : e.g. Given million transactions, find abnormalities



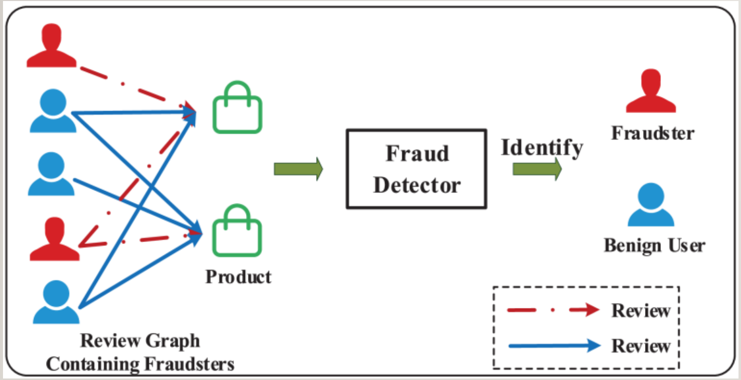
FORMALIZING ANOMALY DETECTION



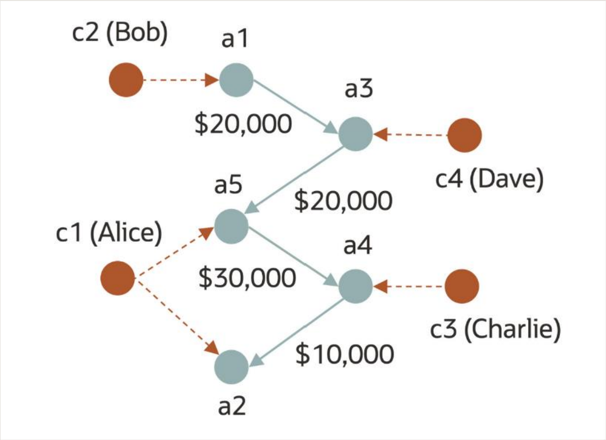
Given <DATA>, Find <ANOMALIES> s.t. <CONSTRAINT>

WHAT IS THE TYPE OF OUR DATA ?

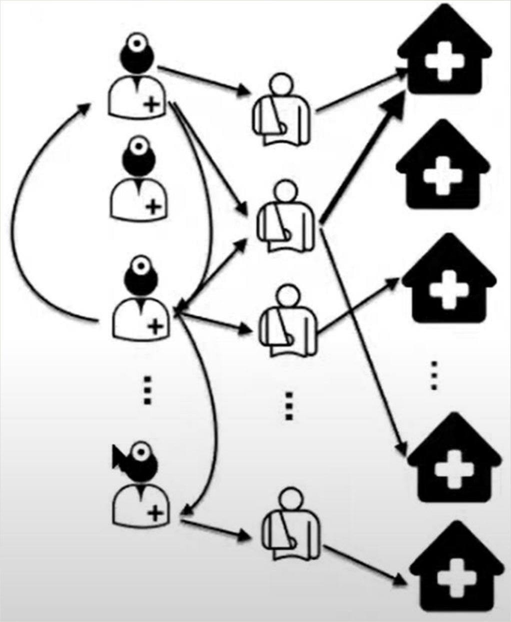
- In many many scenarios underlying data is relational. So we can actually build a graph to study the problem from graph point of view.



User-business review



Financial transactions



Physician-patient provider

CHALLENGES

Given <DATA>, Find <ANOMALIES> s.t. <CONSTRAINT>

Data

- Graph heterogeneity (node/edge labels, attributes, multi edges, edge weights, edge timestamps, etc.) (how to fold meta-data into a graph)

Anomalies

- Definition/Formalization of anomalies (group anomalies vs. anomalous groups)

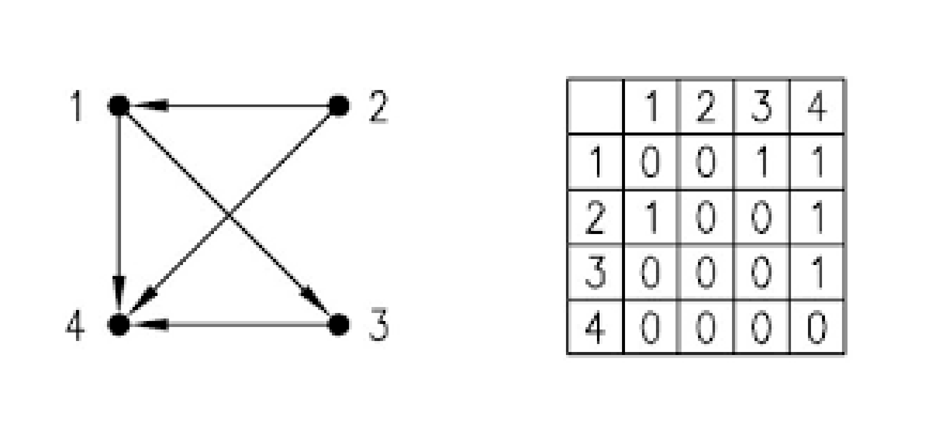
Constraint

- System/Application requirements, e.g. distributed/streaming/massive data, identify attributes on the system(who), explainability (why claiming this to be anomaly)

GRAPH TYPES?

STATIC

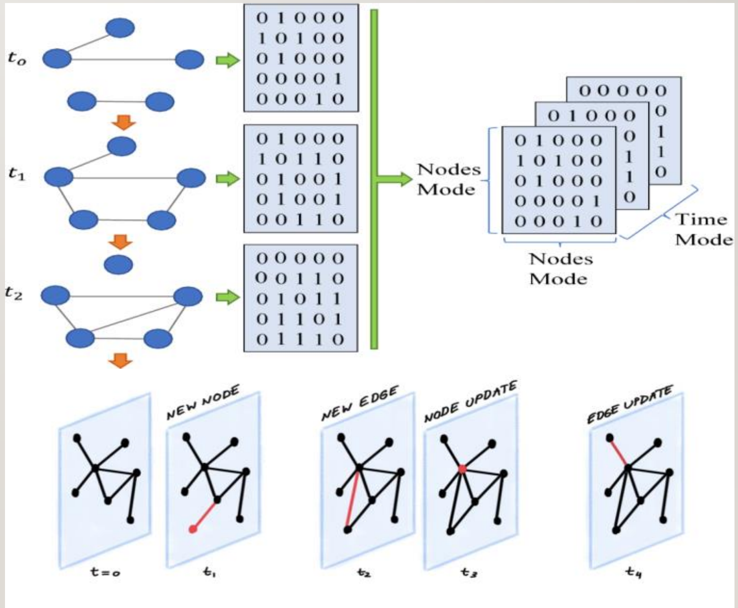
- Nodes and edges are fixed.



	1	2	3	4
1	0	0	1	1
2	1	0	0	1
3	0	0	0	1
4	0	0	0	0

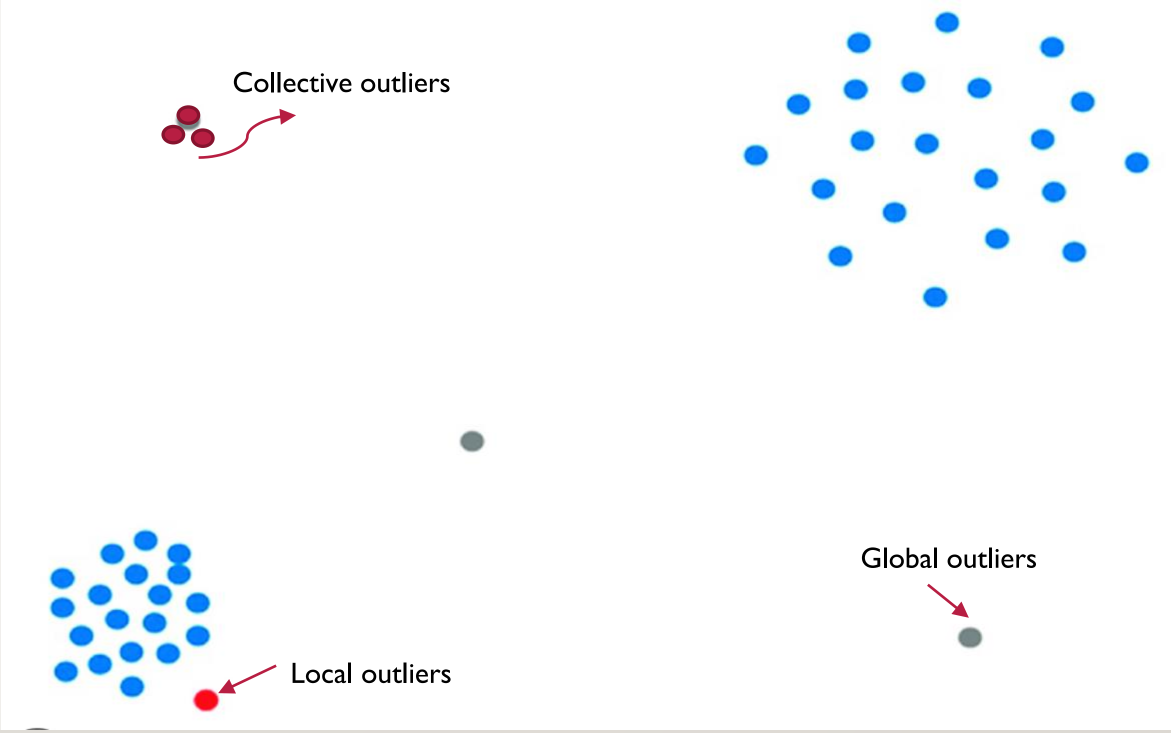
DYNAMIC (REALTIME APPLICATIONS)

- Addition and deletion of nodes and edges are allowed throughout time.



GRAPH BASED ANOMALY DETECTION

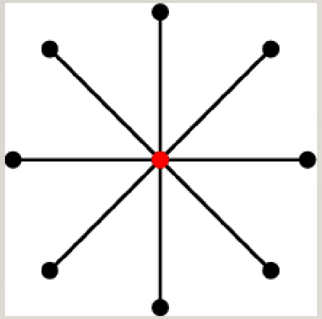
- General purpose (single graph)
  - Global- anomalous nodes
  - Local- group anomalies
  - Collective- anomalous group

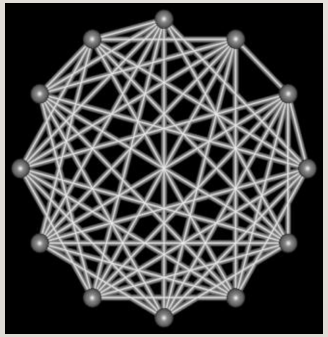


GLOBAL- ANOMALOUS NODES

ANOMALOUS VERTICES(STATIC)

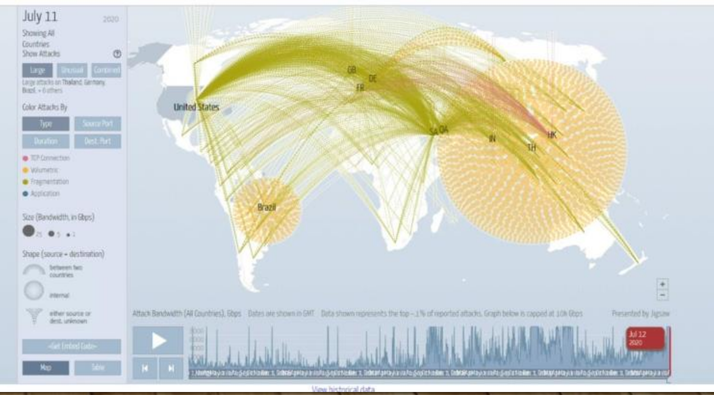
- Anomalous nodes with many edges
- Telemarketer, spammer, port scanner, popularity contests





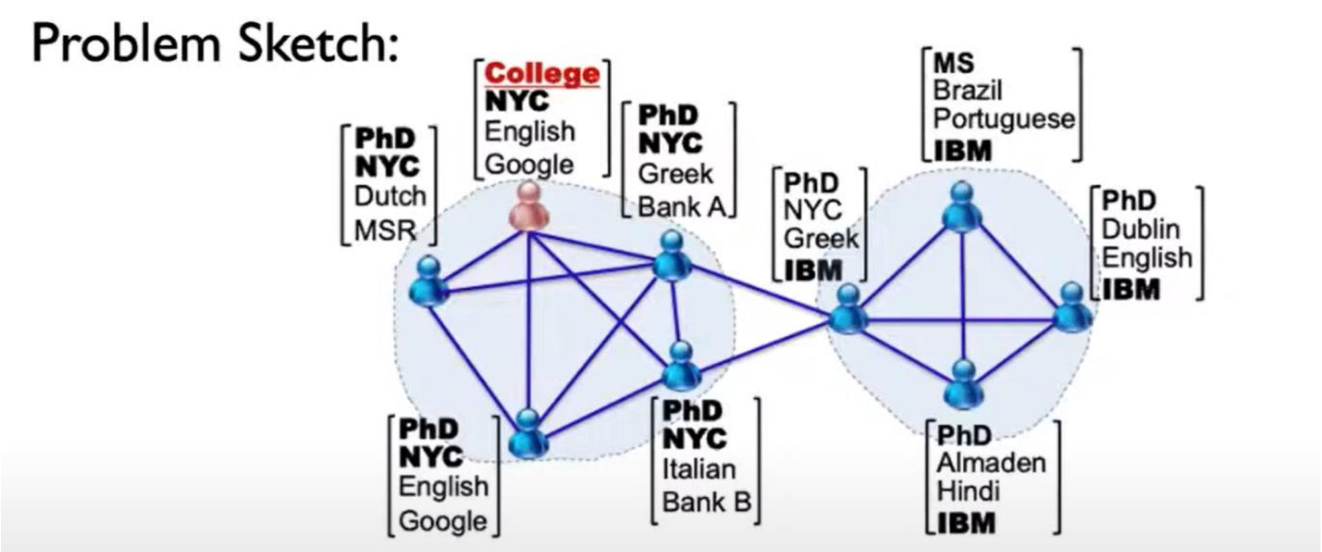
ANOMALOUS VERTICES(DYNAMIC)

- Anomalous nodes receives or lose many edges in a short period (social media fan)
- DDoS attack to a node in network, Realtime congestion in a road network

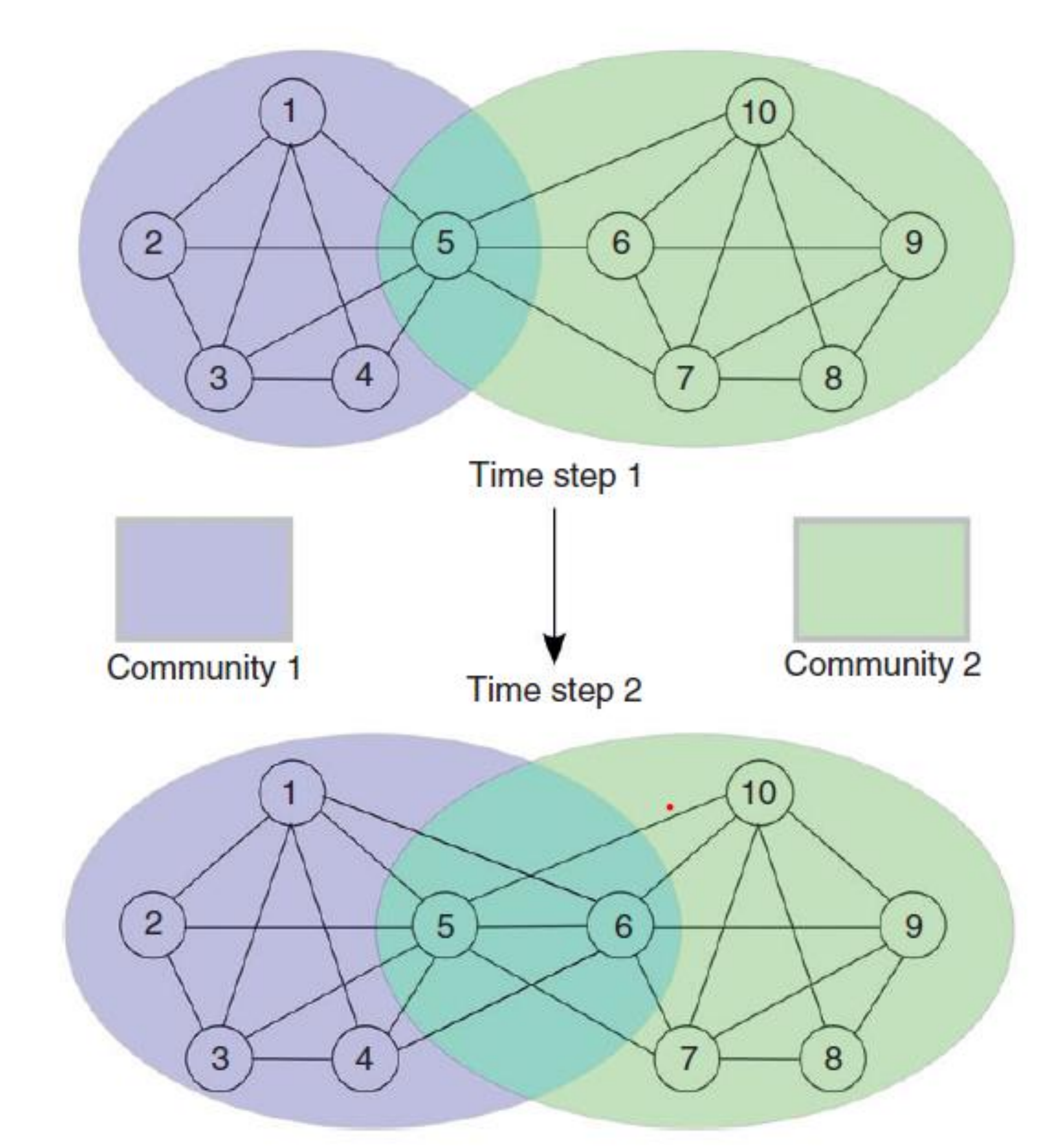


LOCAL- GROUP ANOMALIES

Left community:
{degree, location}
Right community:
{work}



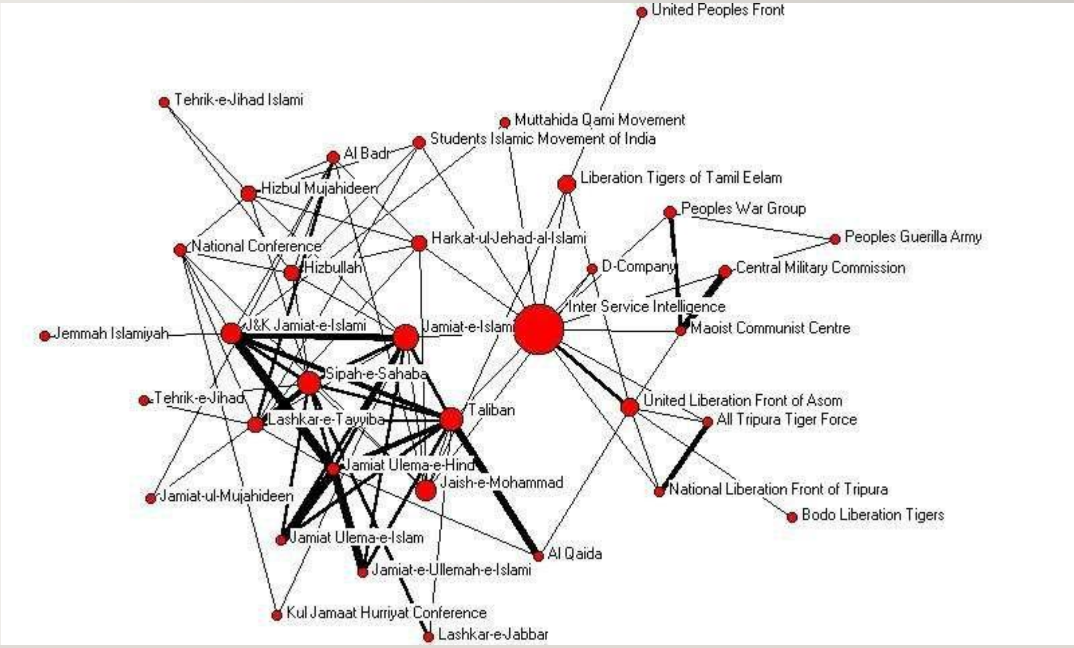
LOCAL- GROUP ANOMALIES (DYNAMIC)



- Node 6 make a local group anomaly in time step 2, since it belongs to both communities.

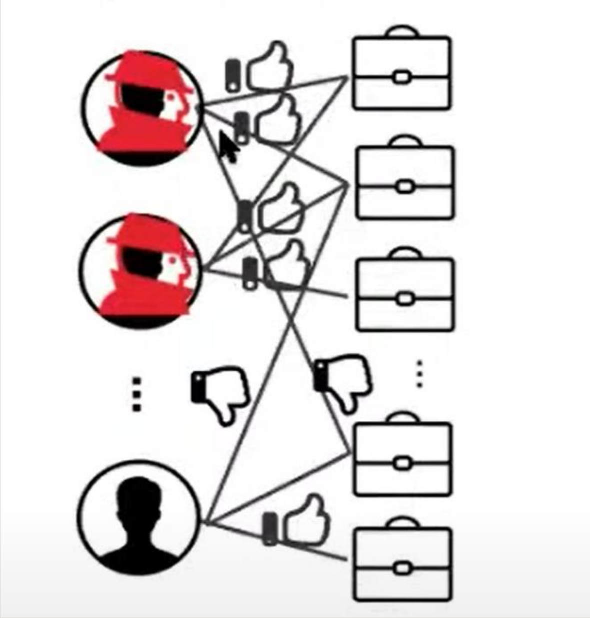
COLLECTIVE- ANOMALOUS GROUP (STATIC)

- Too densely connected groups may be indicative of fraud.
- 9/11 hijackers were densely connected via:
  - Kinship
  - School, training
  - Travel
  - Meetings

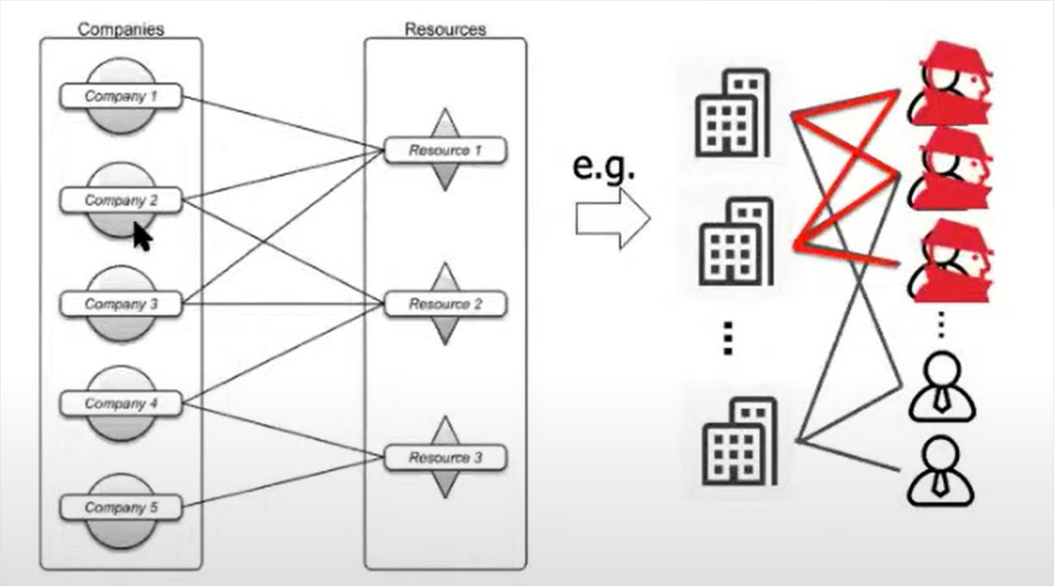


COLLECTIVE- ANOMALOUS GROUP (STATIC)

Opinion fraud: Groups of users promoting/demoting businesses

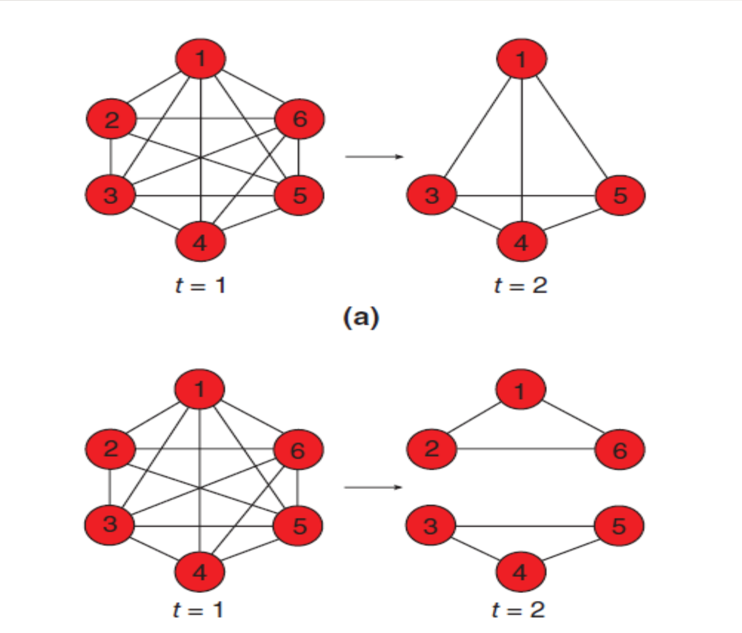


Social securities tax fraud: Groups of resources transferred between "shadow" companies

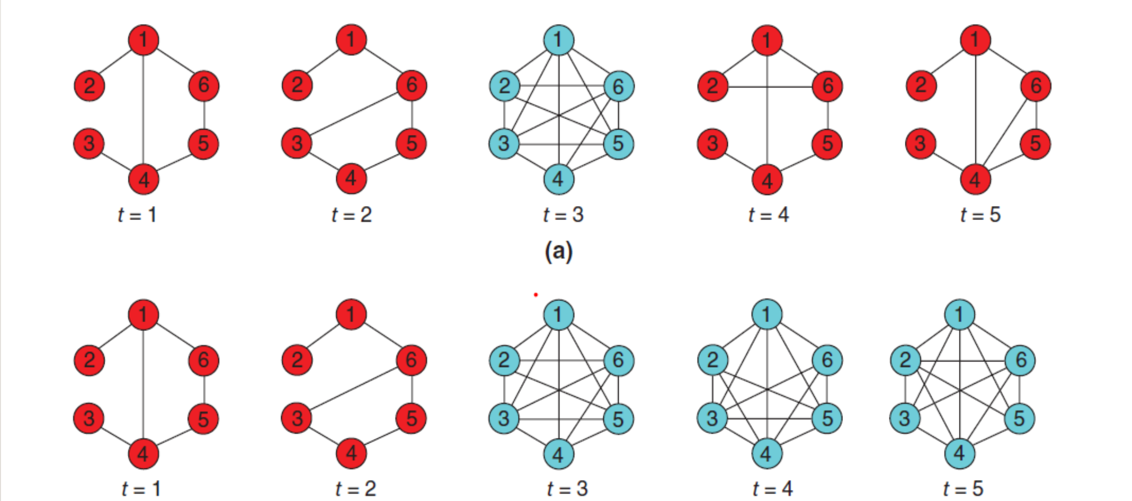


COLLECTIVE- ANOMALOUS GROUP (DYNAMIC)

ANOMALOUS SUBGRAPHS



EVENT AND CHANGE DETECTION



MY THESIS (ANOMALY DETECTION IN DYNAMIC FINANCIAL NETWORK)

DATA CHALLENGES

- Real datasets are not clean!
- Anonymizations are mostly reversable!
- Synthetic datasets are also reversable (no benchmark exist)!
- No annotated data (you should inject)!

ANOMALIES CHALLENGES

- Collective outliers (Anomalous group)
- Find a pattern anomaly which is most common but also it is so similar to a normal pattern.
- Find a pattern that might occur in other areas so that you can use other public data
- Same anomaly pattern have different behaviour in different domains (still you should make a general algorithm)

Given <DATA>, Find <ANOMALIES> s.t. <CONSTRAINT>

MY THESIS (ANOMALY DETECTION IN DYNAMIC FINANCIAL NETWORK)

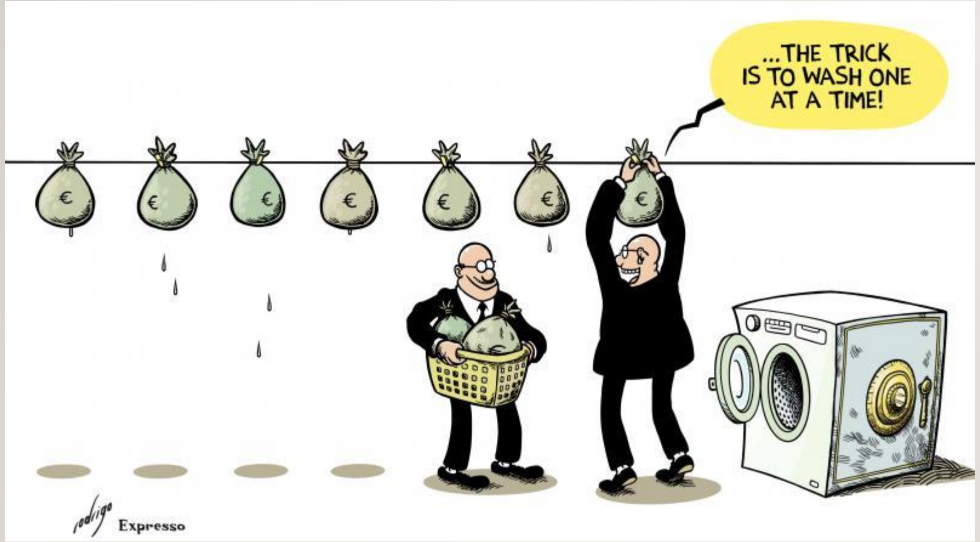
CONSTRAINT CHALLENGES

- Secure data challenges!(Bank can not disclose any data)
- We are using Neo4g graph database, so make algorithm work good on it
- Your methods shall be scalable, we have massive datasets ☺
- Your algorithms shall be easy to understand ☺
- Realtime detection of patterns ☺

Given <DATA>, Find <ANOMALIES> s.t. <CONSTRAINT>

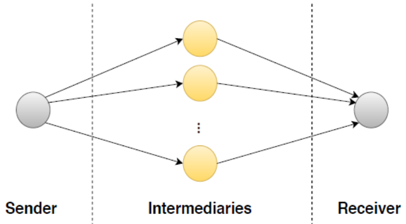
WHY ME?

REAL WORLD MONEY LAUNDERING EXAMPLES



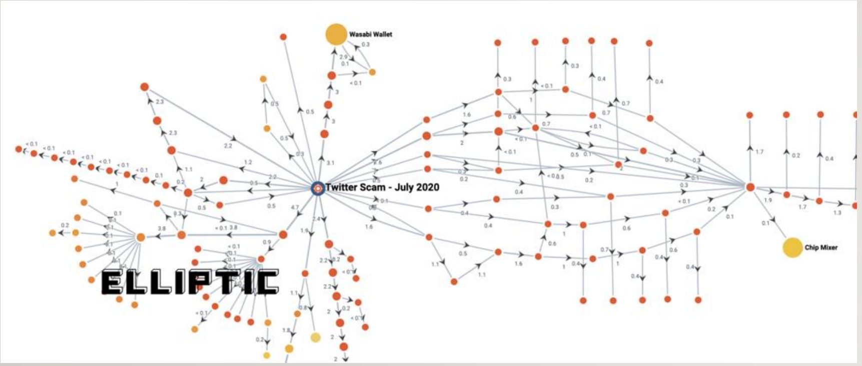
- **Structuring and smurfing**: spreading many small cash deposits to accounts call as smurfs, to avoid anti-money laundering report requirements.
- **Bulk cash smuggling**: smuggling cash to offshore financial institutions.
- **Cash intensive business**: resaurants, casinos, etc.
- **Round tripping**: money deposited offshore, brought back as investment to avoid taxation.

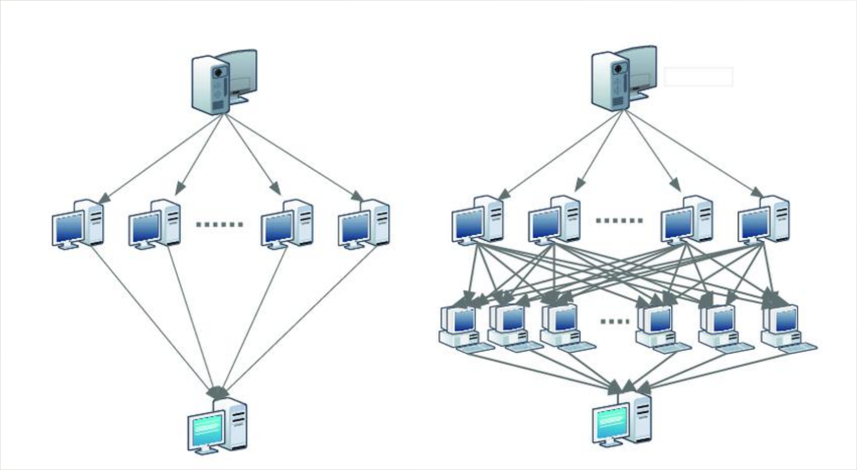
SMURFING



“Smurfing”: One of the most frequent types of money laundering

- Smurfing is **a money-laundering technique involving the structuring of large amounts of cash into multiple small transactions**. Smurfs often spread these small transactions over many different accounts, to keep them under regulatory reporting limits and avoid detection.
- What makes smurfing financial deposits so complicated, is that in many situations, the same behavior used to smurf accounts is considered just making a legal bank deposits.





WHAT IS MY RESEARCH ABOUT?

- Given a dynamic graph of banking transactions we would like to retrieve top-k smurfing patterns and the period within this pattern is valid.
- Given a dynamic graph of banking transactions which factors can help to distinguish between a smurfing fraud and a random pattern?
- Given a dynamic graph of banking transactions how can we spot smurfing patterns efficiently and scalably near realtime?
-

REFERENCES

- NJIT Data Science Seminar: Leman Akoglu, Carnegie Mellon University
- Anomaly detection in dynamic networks: a survey
- https://datascience.njit.edu/

THANK YOU FOR YOUR ATTENTION!