



Tax and Customs Data Science

20 October 2026

AI-powered VAT Frauds Identification

Loris Cino
Data Scientist

Italian Revenue Agency

Outline



Introduction to Networks



VAT Frauds



Invoices as Transaction Networks



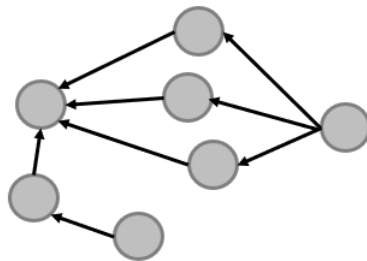
Creating the Graph



Identifying the Frauds

Networks and Graphs

A **graph** is an abstract mathematical object, whereas a **network** is its real-world counterpart



$$G = (V, E) = (\text{●}, \rightarrow)$$

Introduction to Networks

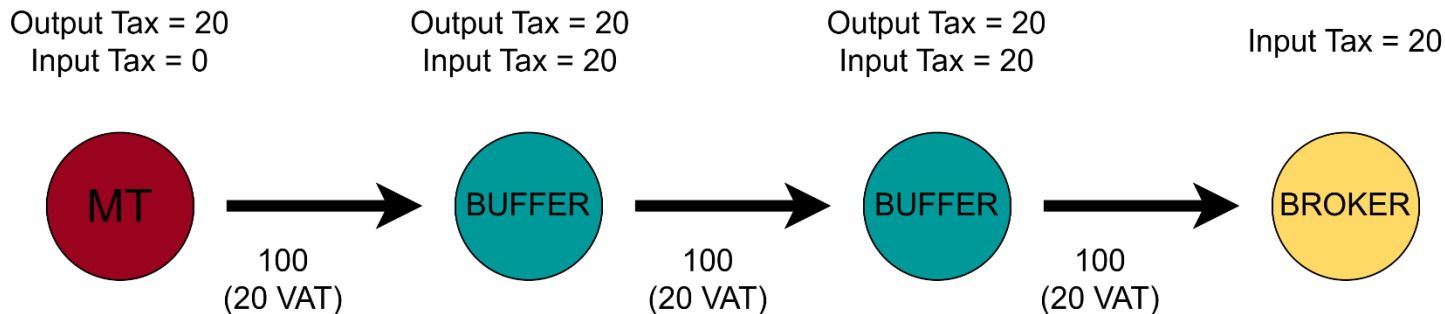
Many phenomena cannot be **understood** by examining their **individual components** in isolation. Their most important properties emerge from the **interactions** among the actors and from the **overall structure** of the system.

This is the fundamental idea behind **Network Science**.

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. SIAM Review, 45(2), 167-256.

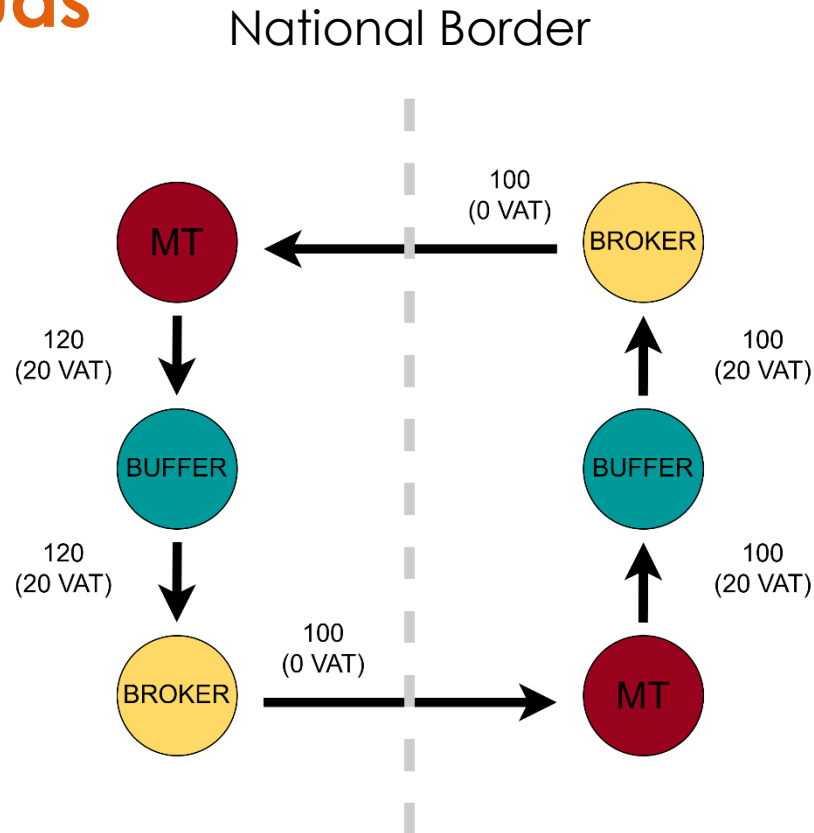
VAT Frauds

Fraud schemes often involve **multiple entities operating** in coordination. Traditional controls focusing on individual taxpayers may miss broader patterns.



Intra-European VAT Frauds

- ❑ One company collects VAT and disappears before remitting it to the tax authority ("missing trader").
- ❑ Other entities in the chain may claim fraudulent VAT refunds.
- ❑ The same goods may circulate repeatedly, creating a "carousel".



The main innovation of this study is: leveraging **e-invoicing data (networks)** as a ***risk factor*** for VAT fraud detection.

In two ways:

- reconstruct and analyze the fraud scheme
- capture a new form of risk

Invoice as Transaction Networks

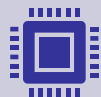
We understood how VAT fraud can be represented as a graph, but **identifying it remains difficult** due to the system in which they operate. Therefore, we need to **study the entire invoicing ecosystem**.

Watts, D. J. (2003). Six Degrees: The Science of a Connected Age. W. W. Norton & Company.

Studies on Transactions Networks



On the topologic structure of economic complex networks:
Empirical evidence from large scale payment network of
Estonia



Properties of B2B invoice graphs and detection of structures
(Italy)



Do the rich get richer? An empirical analysis of the BitCoin
transaction network

Transactions Networks

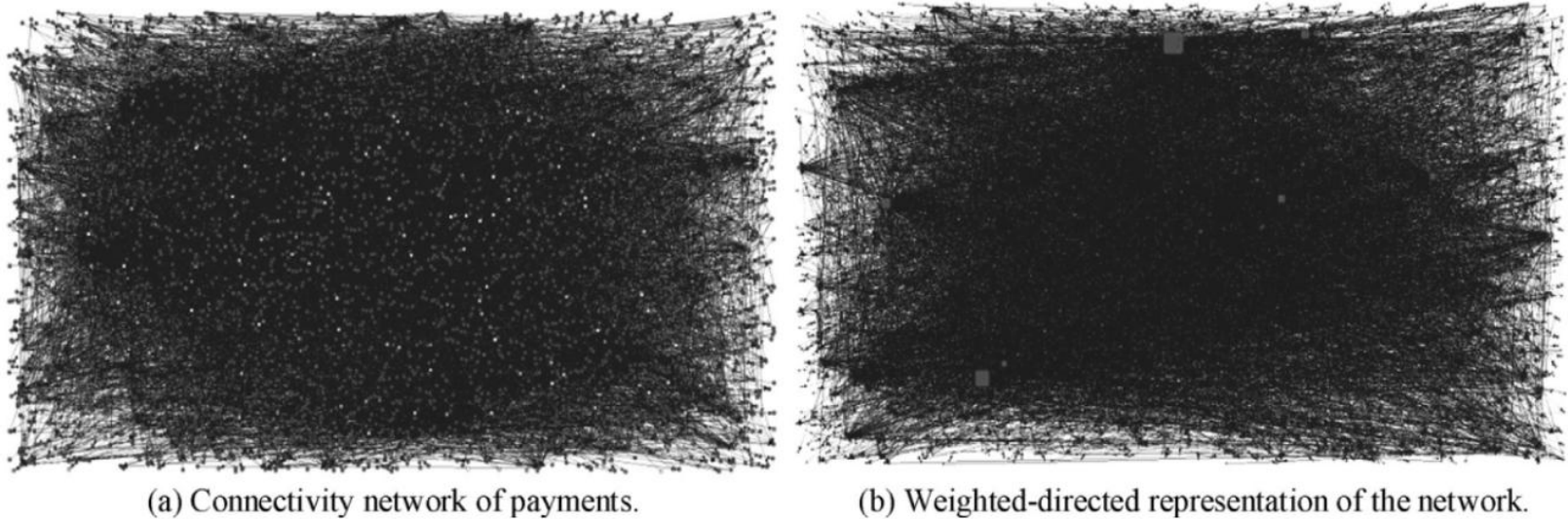
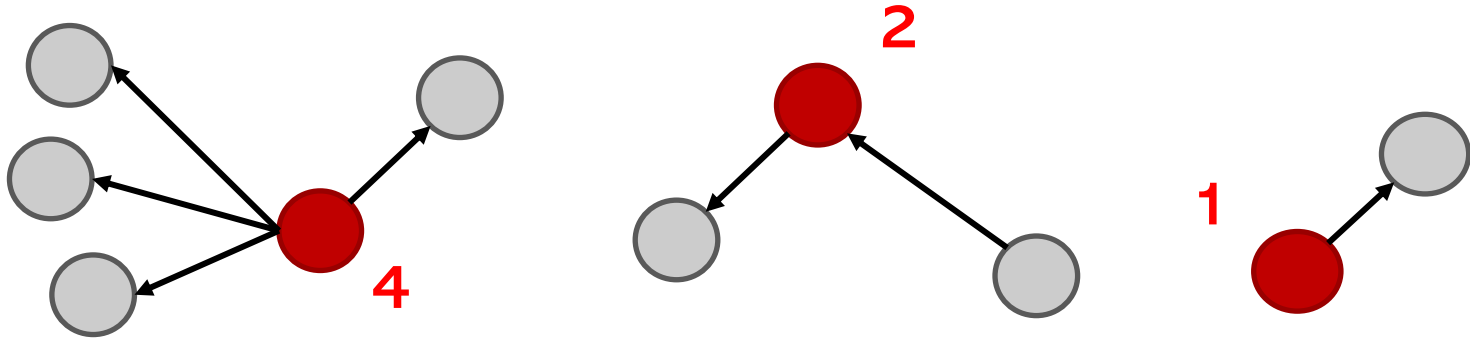


Fig. 2. (a) Connectivity network of payments. (b) Weighted-directed representation of the network.

Rendón de la Torre, S., Kalda, J., Kitt, R., & Engelbrecht, J. (2016). On the topologic structure of economic complex networks: Empirical evidence from large scale payment network of Estonia. *Chaos, Solitons & Fractals*, 90, 18–27.

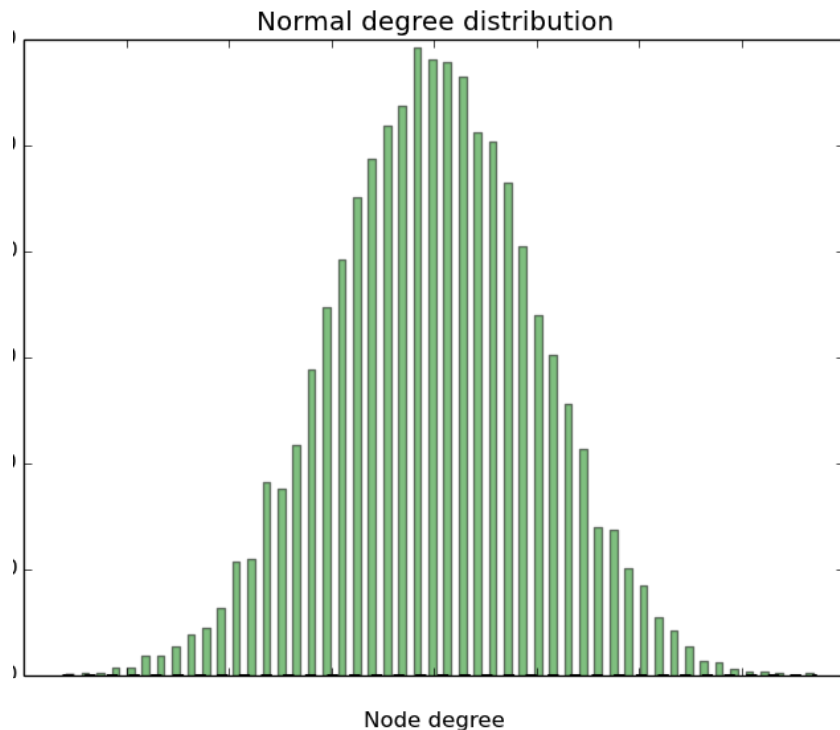
Node degree

The degree of a node is the total number of edges connected to that node.



Watts, D. J. (2003). Six Degrees: The Science of a Connected Age. W. W. Norton & Company.

Node Degree Distribution

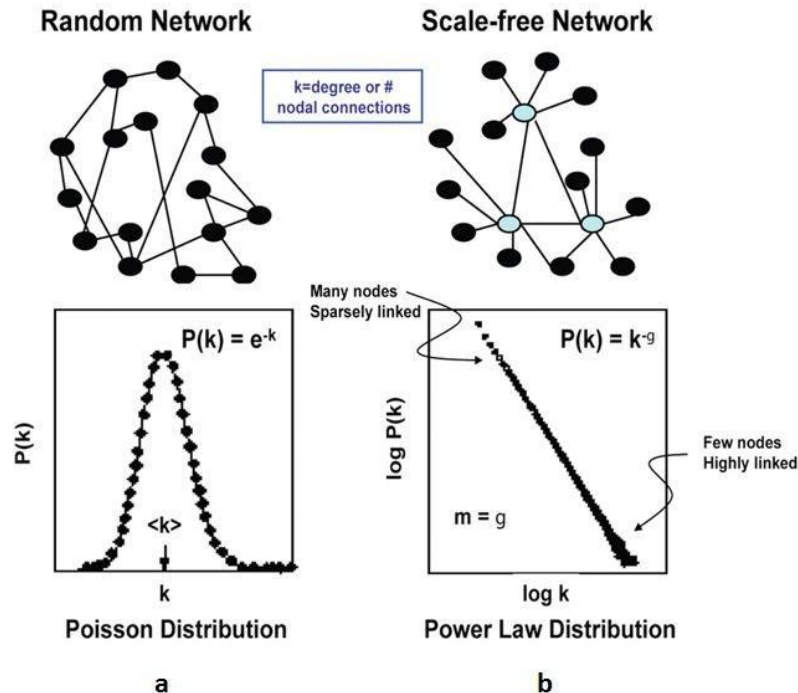


A common way to characterize a network is to plot its node degree distribution.

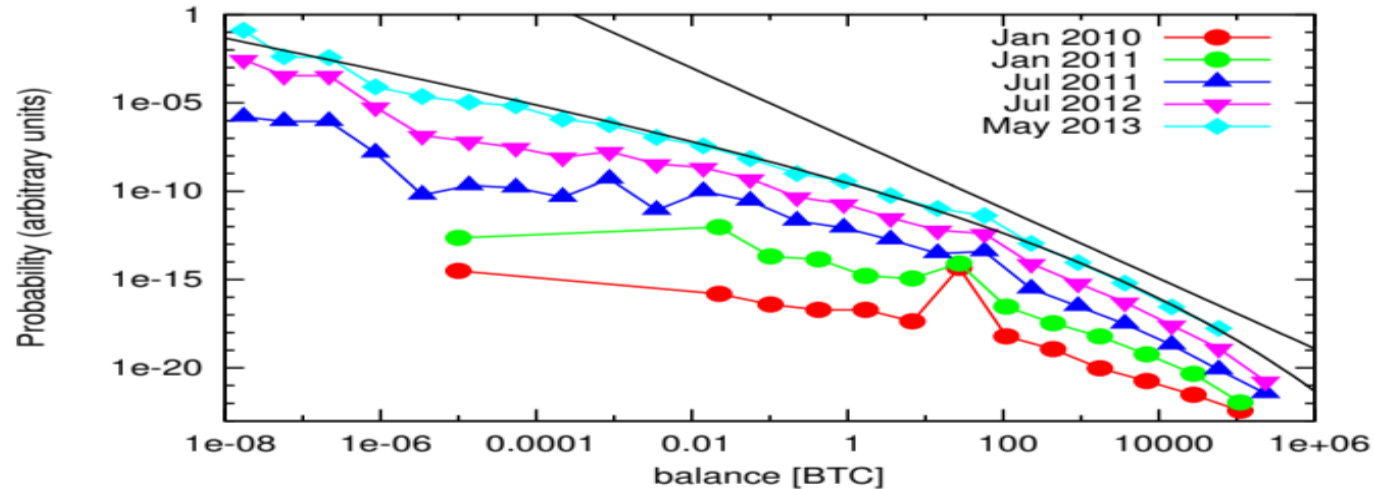
The **node degree distribution** is the probability distribution of the degrees of the nodes over the entire network.

Scale-free Networks

- **Scale-free:** Node degree follows a power law.
- **Hubs:** Few nodes have many connections, while many have few.
- **Asymmetric Robustness:** Resilient to random failures, fragile against targeted attacks.
- **Ultra-Small World**
- **Preferential Attachment:** Evolves through a "rich-get-richer" mechanism.



Transaction Network are Scale-Free



Kondor, D., Pósfai, M., Csabai, I., & Vattay, G. (2014). Do the rich get richer? An empirical analysis of the Bitcoin transaction network. PLoS ONE, 9(2), e86197.

Transaction Network are Scale-Free

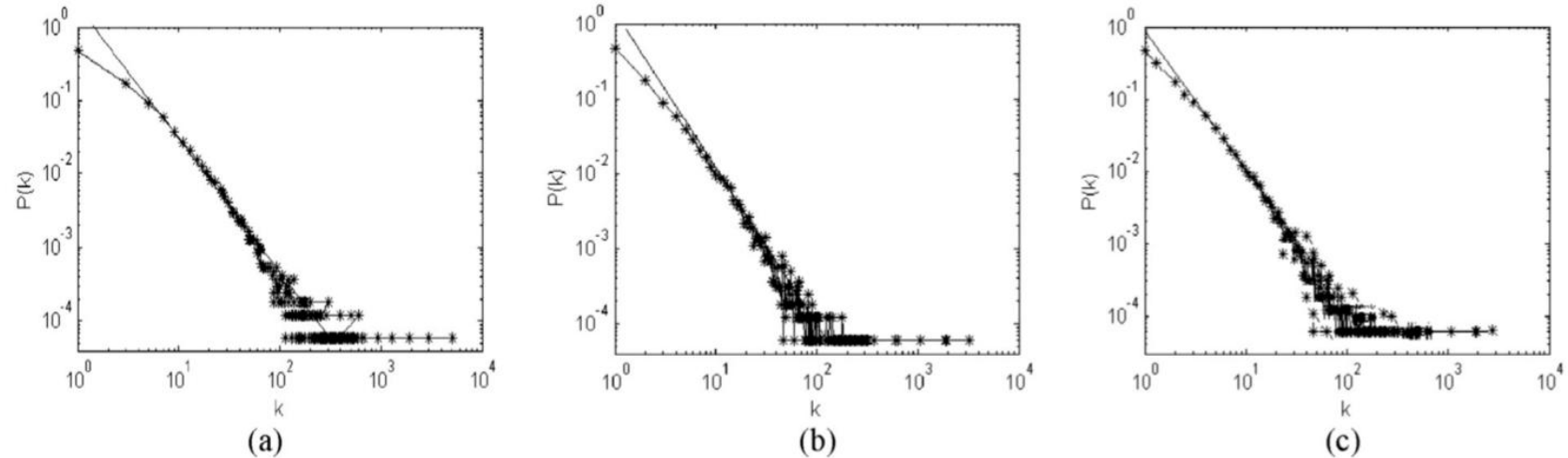


Fig. 3. X-axis corresponds to the number k of degrees and the Y-axis is $P(k)$. (a) Empirical degree distribution for the connectivity network. (b) Empirical in-degree distribution. (c) Empirical out-degree distribution. All the plots are log-log representations of histograms.

Rendón de la Torre, S., Kalda, J., Kitt, R., & Engelbrecht, J. (2016). On the topologic structure of economic complex networks: Empirical evidence from large scale payment network of Estonia. *Chaos, Solitons & Fractals*, 90, 18–27.

Assortativity: another similarity

Another network characteristics commonly studied is the **assortativity**.

This characteristics is the propension of nodes to connect with other nodes with similar degree (or other attribute)

More than scale-free property the transaction networks have a slightly **negative assortativity (disassortativity)**

What this means?

In disassortative network we have that hubs connect to leaf node (low degree) node.

This makes the network **more connected** but **less resilient** to targeted node removal.

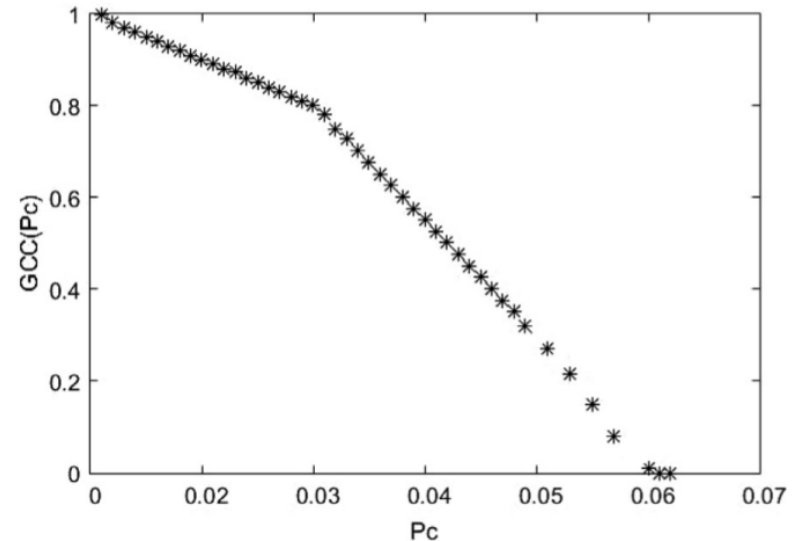


Fig. 8. GCC of the network of payments as a function of the percolation threshold p_c by using the collective influence algorithm.

Scale-Free networks

Lexical Networks

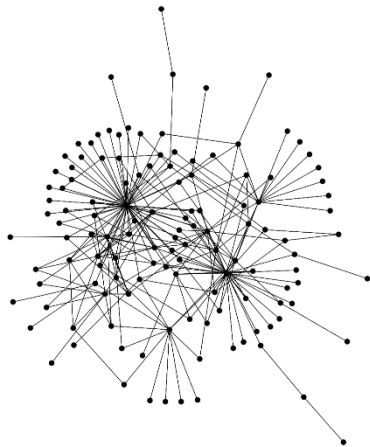


FIG. 5. The topology of Moby Dick. Two words appearing in Melville's book are linked if their mutual information is greater than a given threshold.

Ecological Networks

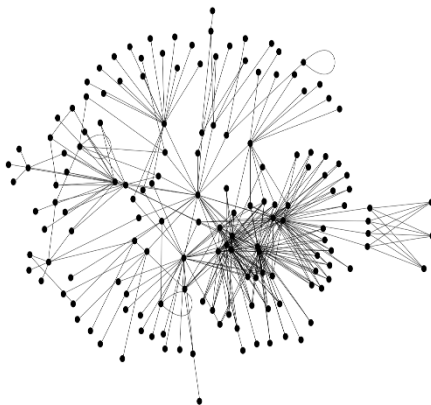


FIG. 4. The graph of species interactions of Silwood park web. Here each node represents one species. The central node in this representation is the Scotch broom *Cytisus scoparius*.

Software Networks

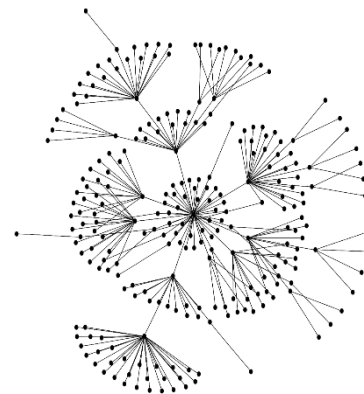
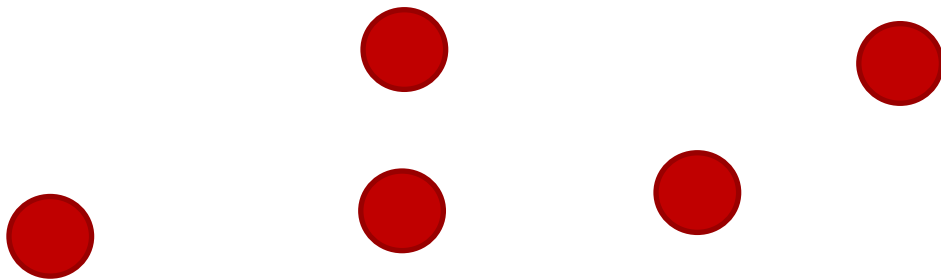


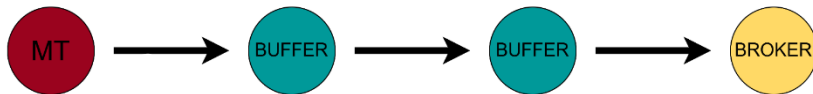
FIG. 7. An example of a medium-sized component of a software graph (extracted from the Java Development Kit 1.2 library). Here each node is a Java class and edges indicate relationships among classes.

Newman, M. E. J. (2003). The Structure and Function of Complex Networks. SIAM Review, 45(2), 167-256.

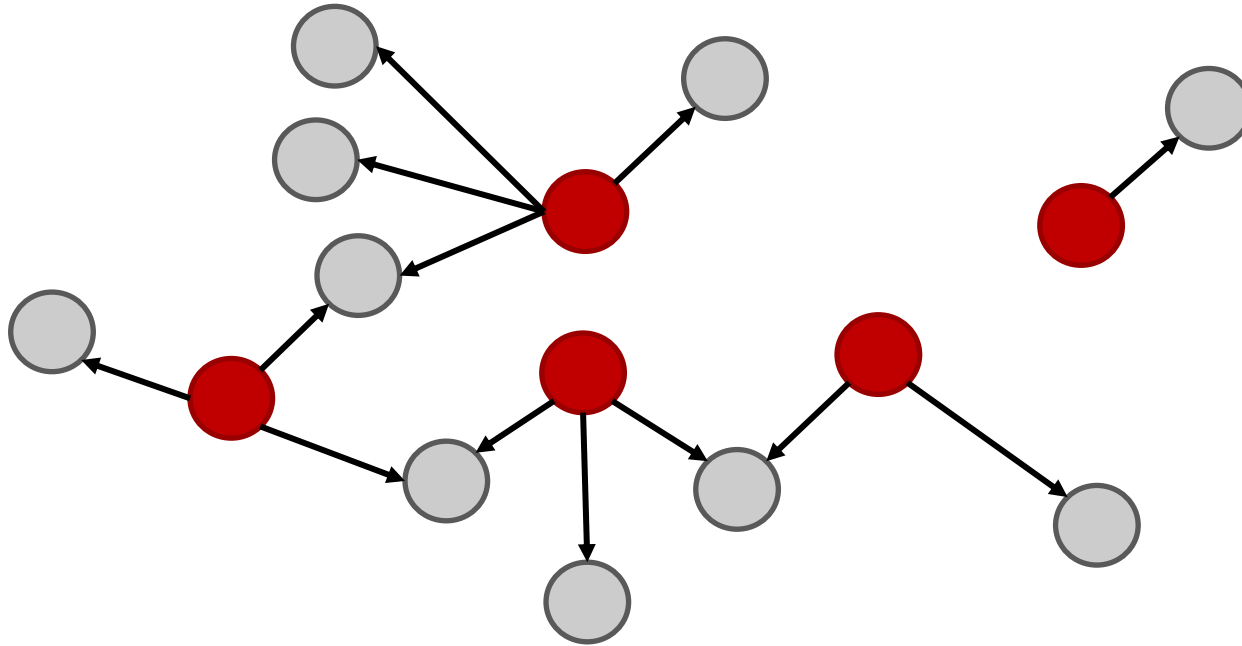
Creating The Graph (Network)



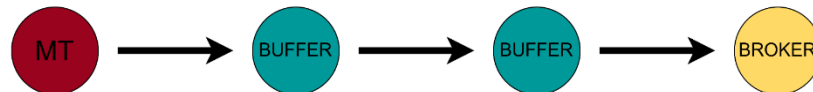
Identifying the taxpayers with a clear gap between e-invoice data and declarations/payments (likely missing trader – red dots)



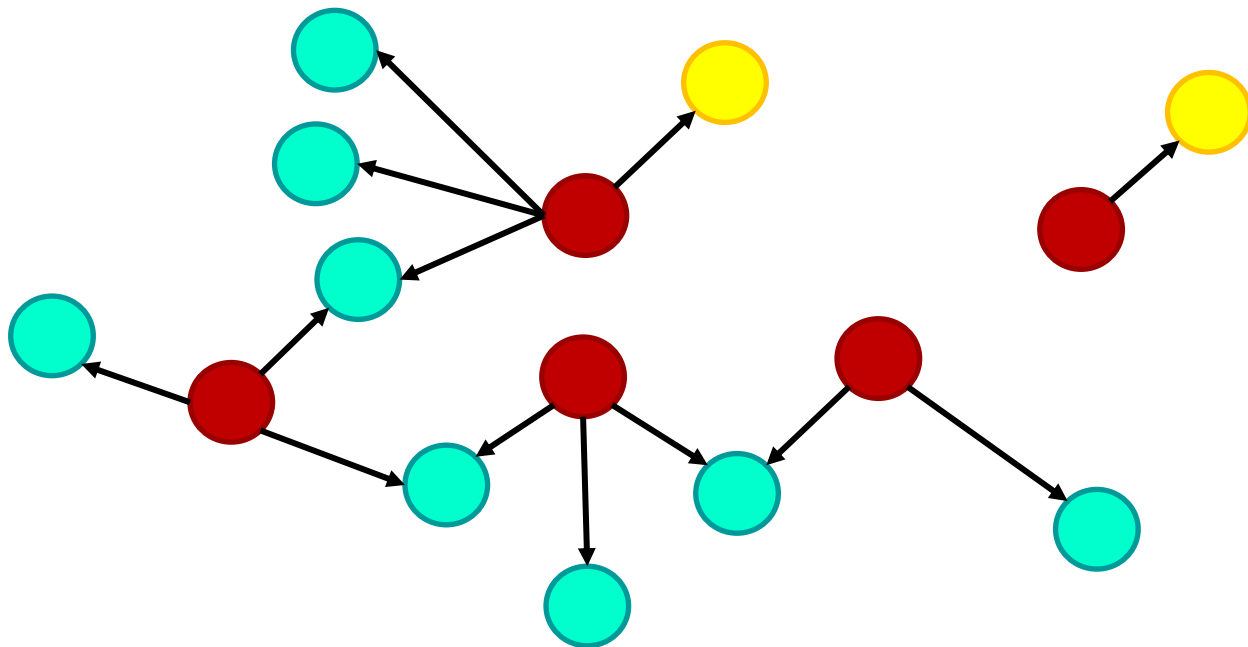
Creating The Graph (Network)



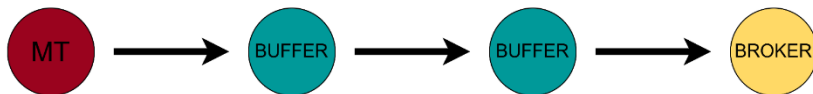
Adding the taxpayer that buy from missing traders (MT)



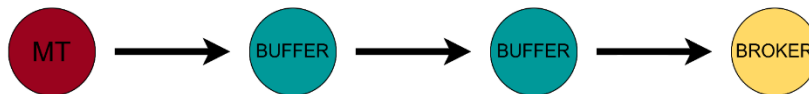
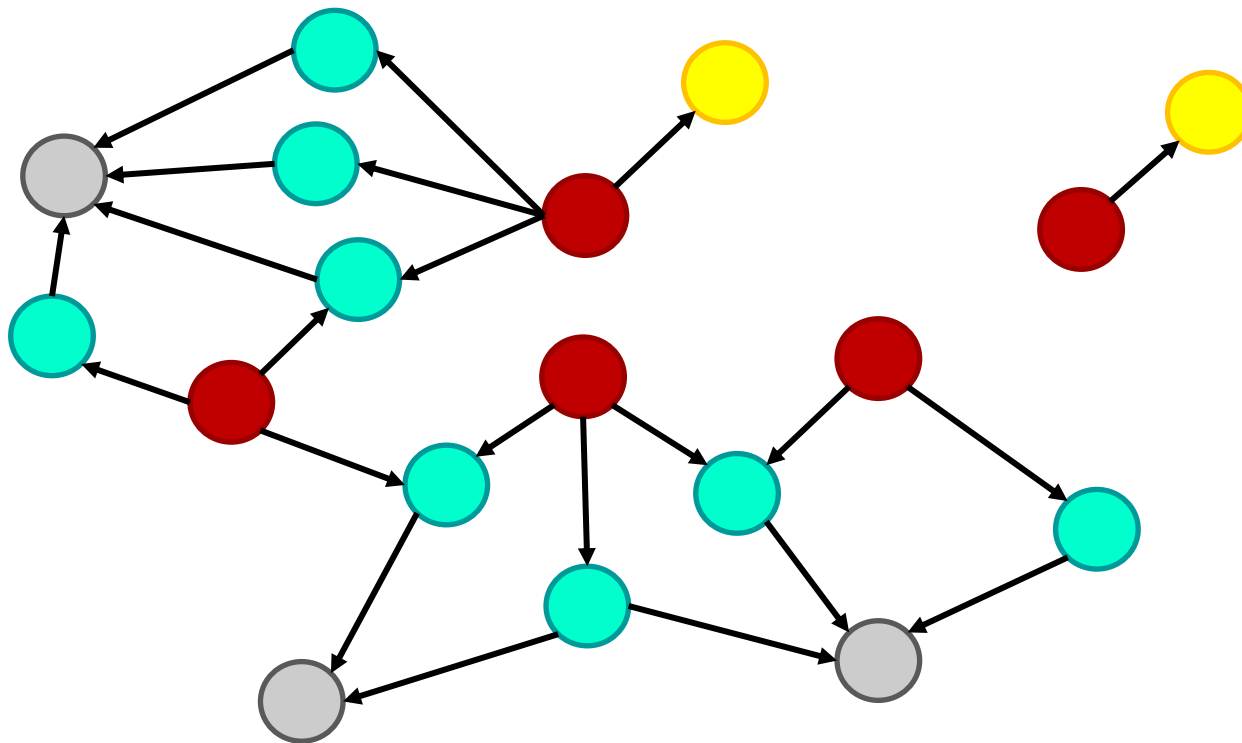
Creating The Graph (Network)



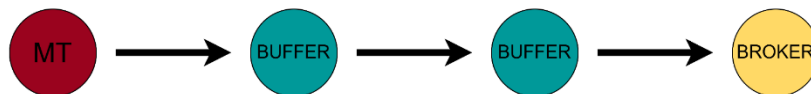
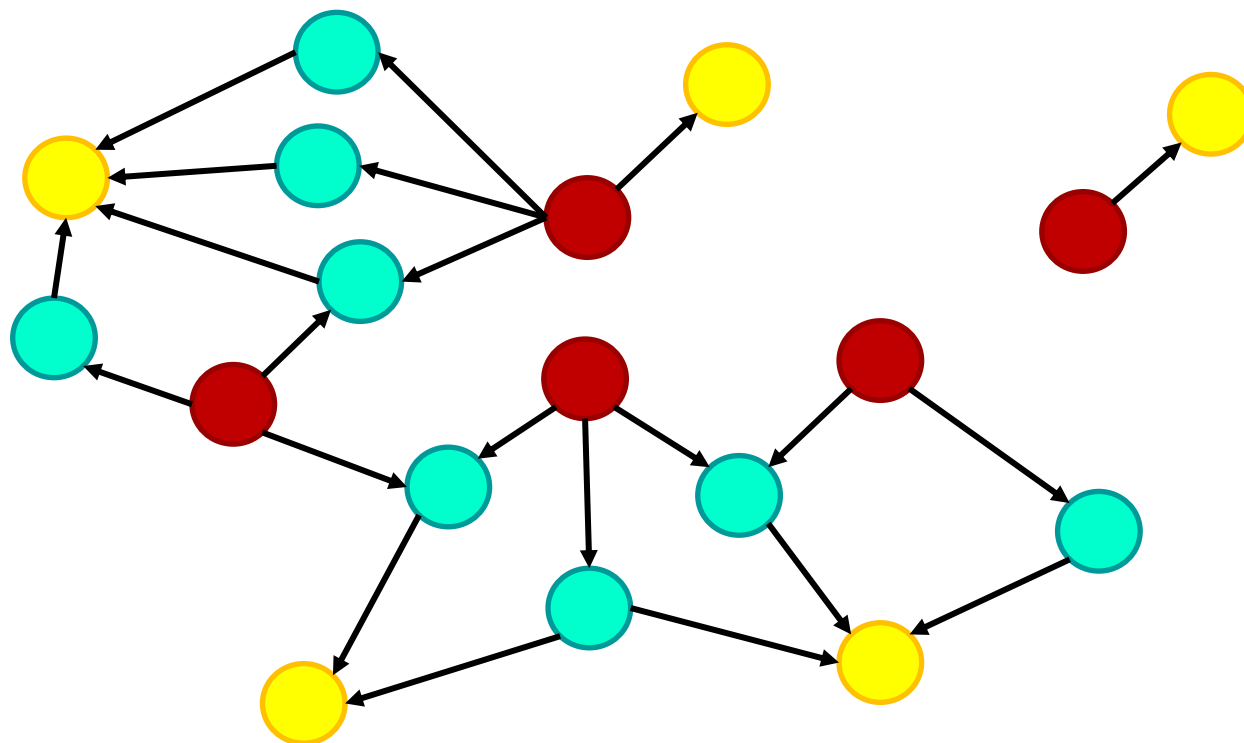
Classifying them as a possible buffer (light blue dots) or broker (yellow dots)



Creating The Graph (Network)



Creating The Graph (Network)



Community

A **community** is a group of nodes (like people or data points) that are strongly connected to each other, but only loosely connected to the rest of the network.

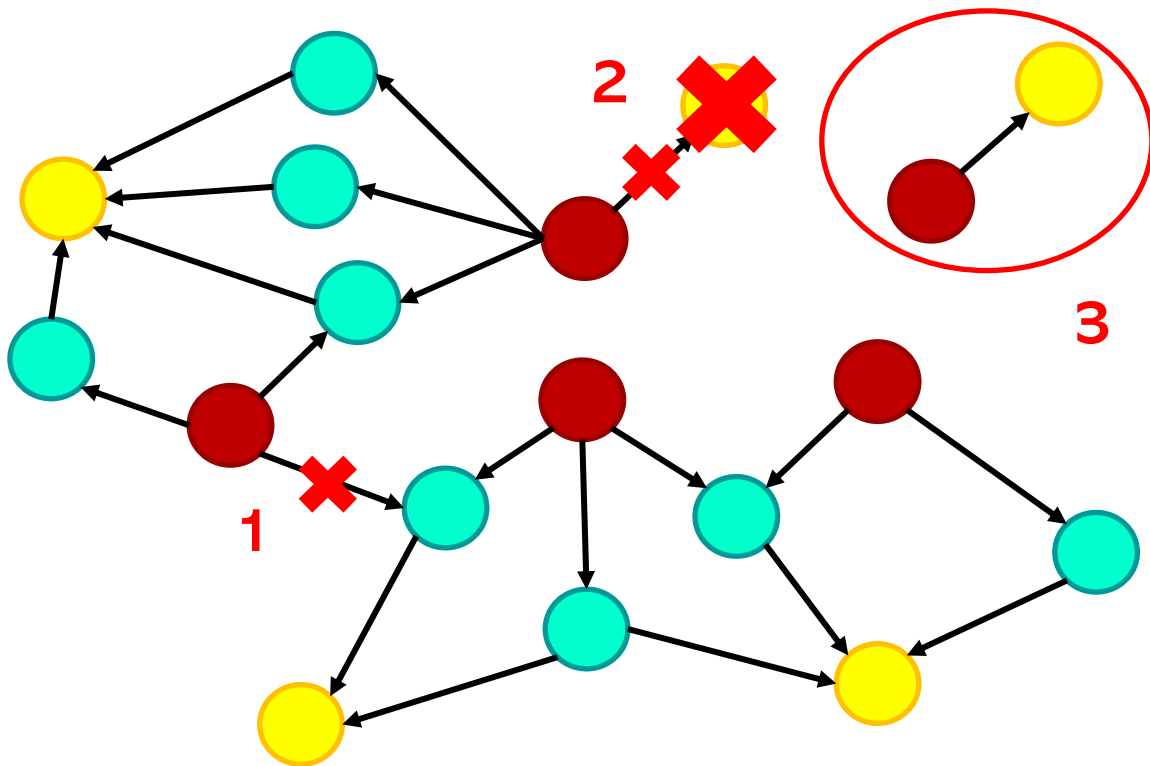
A **community detection algorithm** is the automated mathematical process that scans a massive network to discover and map these hidden groups.

The **Louvain algorithm** finds communities by trying to maximize a score called **modularity**.

Detecting communities

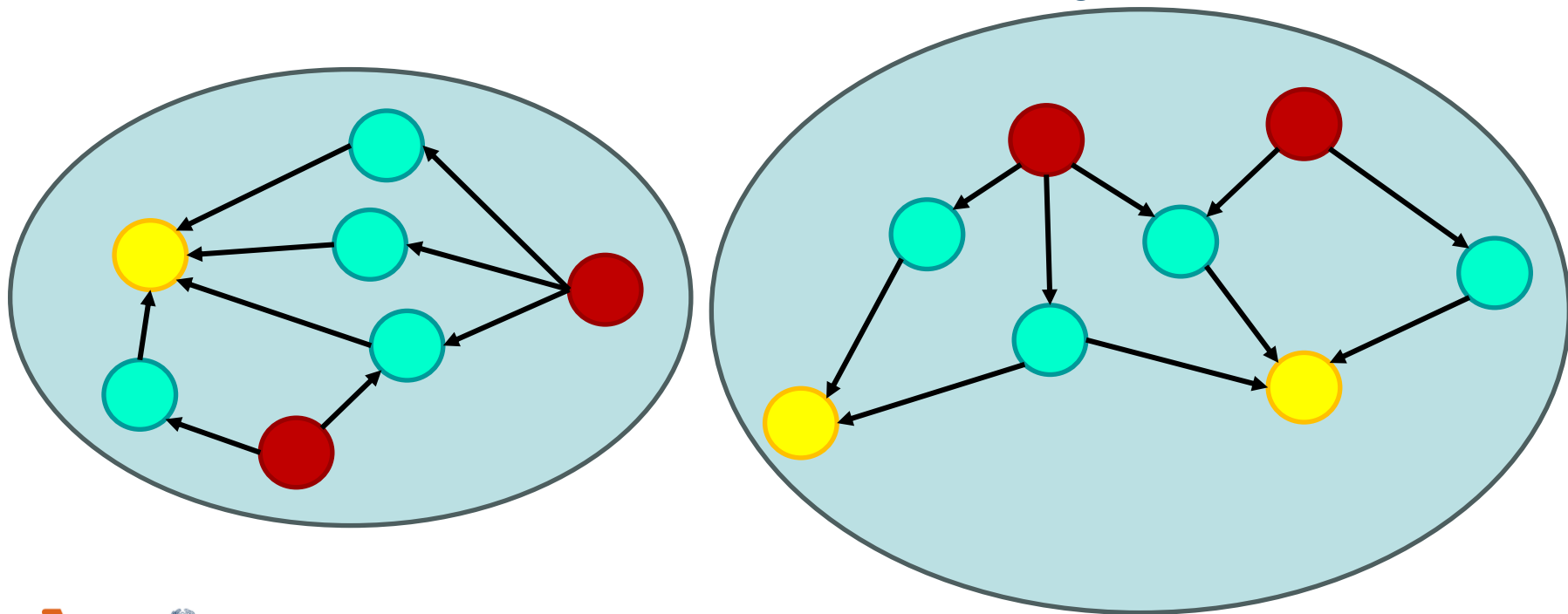
After applying Louvain algorithms identify three spurious relationship

1. Links (transactions) that **links two community.**
2. Links (and consequently nodes) that probably are not part of the frauds.
3. **Small community** that probably are not a fraud



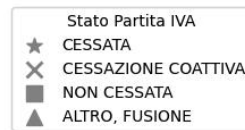
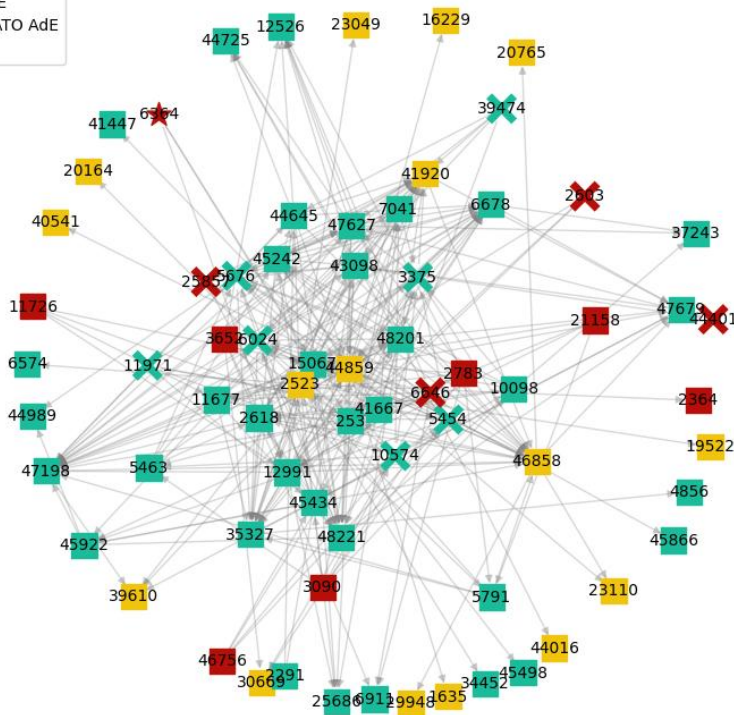
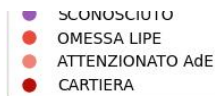
Detecting communities

Now we do have two well defined communities (possible frauds) that can be submitted to tax experts for further investigation.

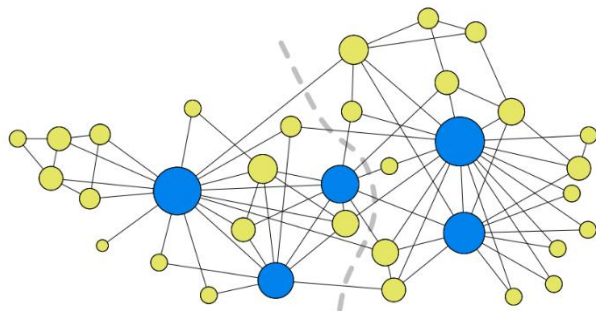


Results

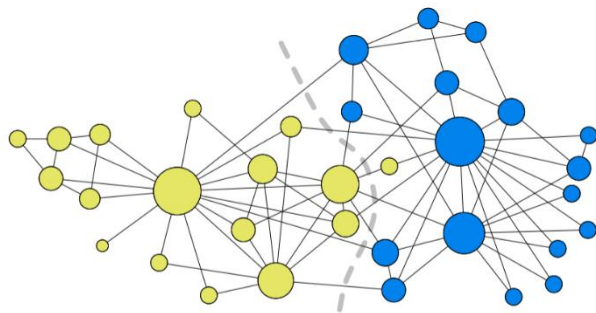
- The final outcome is a set of well-defined clusters (communities) sent to investigation.
- A real example is reported of the last year. The taxpayers reported with “X” are the one previously closed for fraudulent activities.
- Only in this cluster more than 10 are identified as frauds.



The graph structure matters



(a) Without degree correction



(b) With degree-correction

- Invoices that belong to frauds are still invoices. We found that they are “scale-free”
- “scale-free” properties has effect on the community detection algorithms

Improvements

- ❑ **Evaluate alternative community detection algorithms**
- ❑ **End-to-End Graph Machine Learning pipeline**
 - Graph embeddings
 - Graph Neural Networks (GNNs)
- ❑ **Temporal network analysis**
 - Detect frauds while they are emerging
 - Early warning signals

Conclusions

- ❑ VAT fraud is a network phenomenon, not an individual one
- ❑ Community detection improves fraud identification
- ❑ Network analysis provides risk indicators
- ❑ Transaction networks exhibit stable structural properties
 - Scale-free
 - Slightly disassortative
- ❑ Invoice data contains valuable network information
- ❑ International cooperation is essential for combating cross border VAT fraud