Emerging Technologies for mapping Risk (&) Stakeholders (using Artificial Intelligence (AI) and blockchain applications)

SAPIENZA
UNIVERSITÀ DI ROMA

Francesco Bellini

1st Workshop of the project “101101784 — 2022-IT-FRED2” Fraud Repression through EDucation2

Introduction to Rumsfeld's Theory: Unknown unknowns

“Reports that say that something hasn't happened are always interesting to me, because as we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns - the ones we don't know we don't know. And if one looks throughout the history of our country and other free countries, it is the latter category that tend to be the difficult ones”

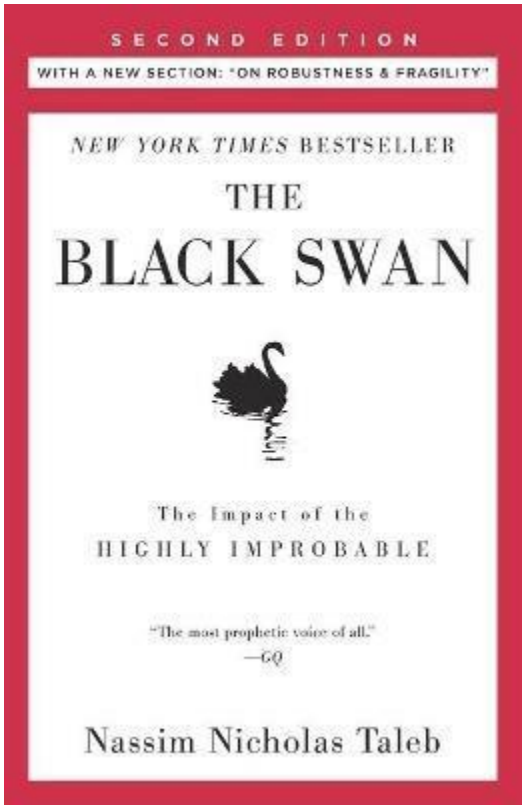
Donald Rumsfeld, United States Secretary of Defense, 2002

Black Swan Events and Modern Risk Analysis

The concept of Black Swan events as per Nassim Taleb

Attributes:

- Rarity
- Extreme Impact
- Retrospective Predictability



Frank Knight's Conceptual Distinction

Differentiating Risk and Uncertainty

- Risk: Known probability distribution
- Uncertainty: Unknown probability model

The Changing Landscape of Risk Post-9/11: how 9/11 changed the perception of risk and uncertainty?

Learning from Historical Disasters and how these events challenge traditional risk models

How psychological biases affect our understanding of rare events?

Approaches to Classifying Uncertainties

Different methods of categorizing uncertainties

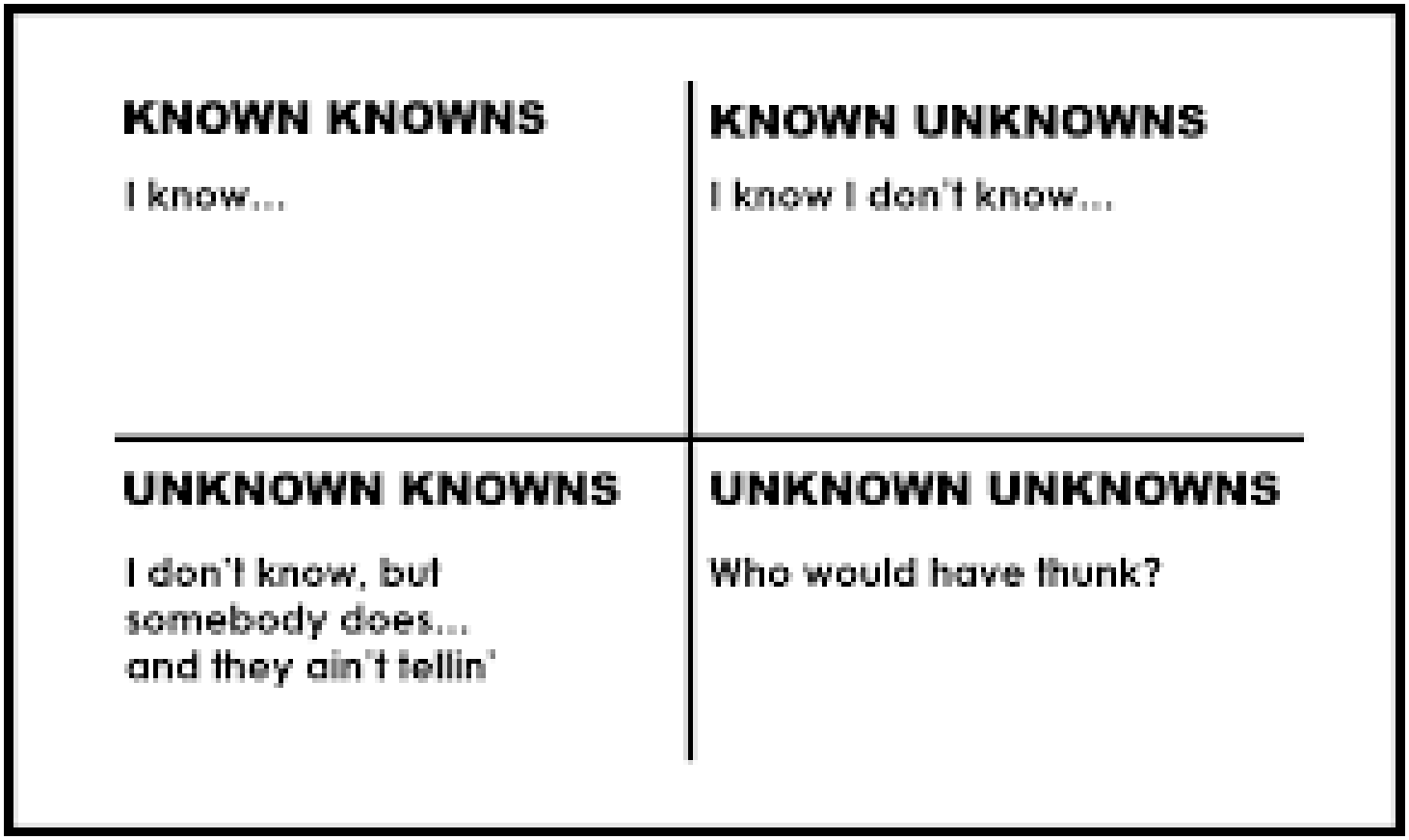
- subway uncertainty

- coconut uncertainty

Unknown unknowns

Event	Impact: Certain (Known)	Impact: Uncertain (Unknown)
Identified (Known)	Knowledge (Known knowns) Available data Predictability of future states	Identified risk (Known unknowns) Possible states identified
Unidentified (Unknown)	Unused knowledge (Unknown knowns) Searchable facts Unused resources	Inscrutable uncertainty (Unknown unknowns) Unknown relationships between key variables Unpredictable events

Unknown unknowns



Financial Frauds taxonomy

- **Application Fraud**: Involves submitting false or misleading information to obtain EU funds. This includes document forgery, overstatement of costs, or claims for non-existent projects.
- **Corruption and Collusion**: Entails using EU funds for corrupt activities, such as bribing officials to secure contracts or unduly influencing the fund allocation process.
- **Misuse of Funds**: Occurs when recipients use the funds for purposes other than those intended, often for personal enrichment or unauthorized activities.
- **Conflict of Interest**: Involves situations where individuals responsible for allocating or managing EU funds have a direct or indirect personal benefit from their use.
- **Money Laundering**: Some cases involve laundering money through EU funds to legitimize proceeds from criminal activities.
- **Irregularities in Management and Reporting**: This includes failing to comply with EU regulations on fund management and reporting, which can be either unintentional or deliberate.
- **Transnational Frauds**: Involve actors in multiple member states or third countries, often with complex fraudulent schemes exploiting different jurisdictions to evade controls.

Common points, workflow and deficits

- All these frauds are highly interconnected
- Identification of the real beneficiaries
- Limitations of the Information Systems
- Feedback based procedures
- Compliance procedures and costs
- False alerts
- Reliability of sources

What's next?

- Identification of financial frauds and criminal organisations
- From information systems to decision support systems
- A proposal for a multidisciplinary approach for the analysis of the big data streams

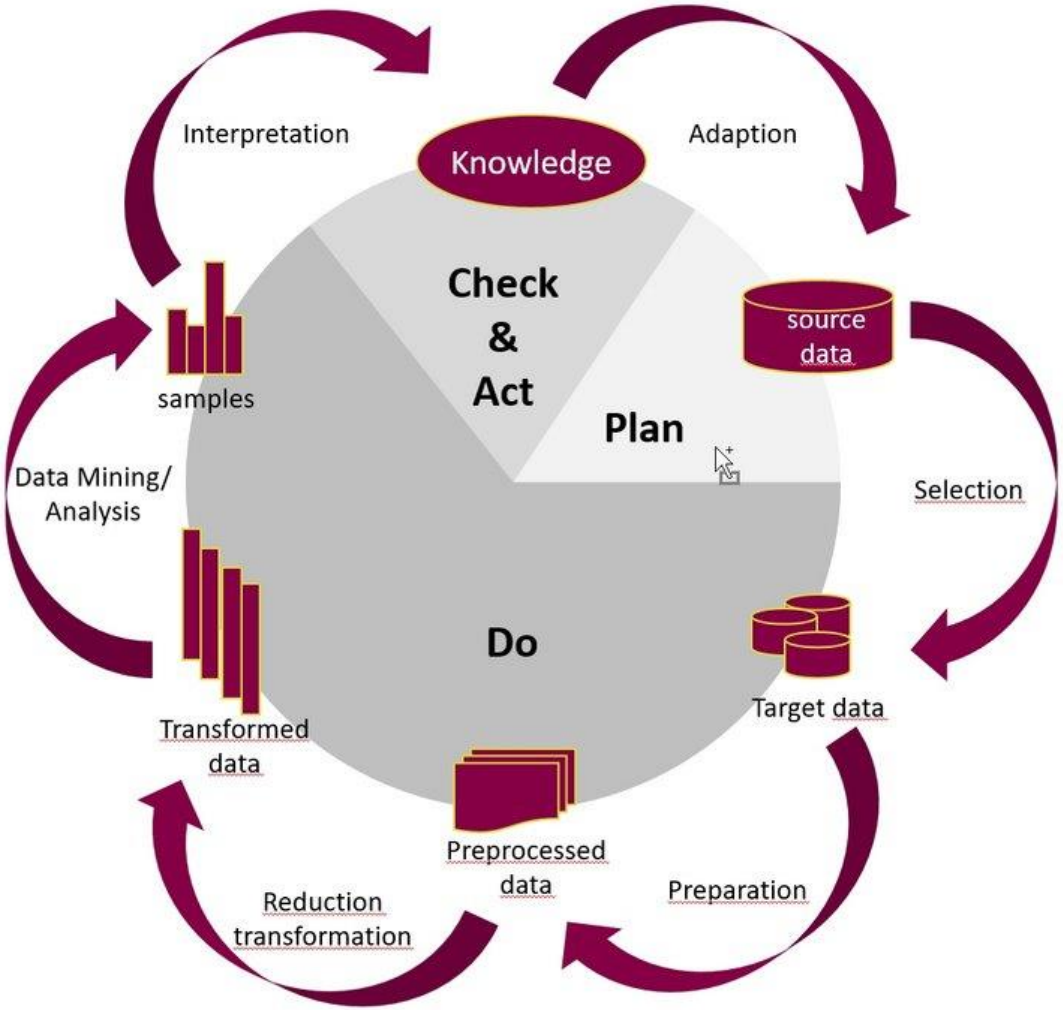
Techniques

- **Data Analytics and Big Data (DA & BD)**: Advanced data analytics tools are used to analyze large volumes of financial transactions and other relevant data. Big data technologies enable the processing of complex datasets to identify patterns indicative of fraudulent activities.
- **Machine Learning and AI (ML & AI)**: Machine learning algorithms and artificial intelligence are employed to learn from historical data, improving the ability to detect anomalies and potential fraud cases over time.
- **Blockchain (BC)**: In some instances, blockchain technology is explored for its potential in improving transparency and traceability of transactions, making it harder to commit fraud.
- **Network Analysis (NA)**: This involves mapping relationships and transactions between entities to identify unusual patterns or connections that might suggest fraudulent activities.
- **Forensic Accounting Tools (FAT)**: These tools are used for in-depth examination of financial records to uncover discrepancies and irregularities.
- **Risk Assessment Software (RAS)**: These applications assess the risk levels of various projects and beneficiaries, based on a range of factors, to flag high-risk cases for further investigation.

Data Analytics and Big Data

- Utilization of advanced algorithms to identify irregular patterns and anomalies in financial transactions across EU fund distributions.
- Integration of machine learning models to predict and detect fraudulent activities by analyzing historical data and recognizing suspicious behaviors.
- Implementation of natural language processing (NLP) to monitor and analyze unstructured data such as reports, contracts, and communications for signs of fraudulent language or hidden relationships.
- Use of predictive analytics to assign risk scores to various transactions and actors, enabling proactive measures and targeted audits on high-risk areas.
- Collaboration with EU member states' financial intelligence units, leveraging Big Data to streamline cross-border information sharing and enhance collective fraud detection capabilities.

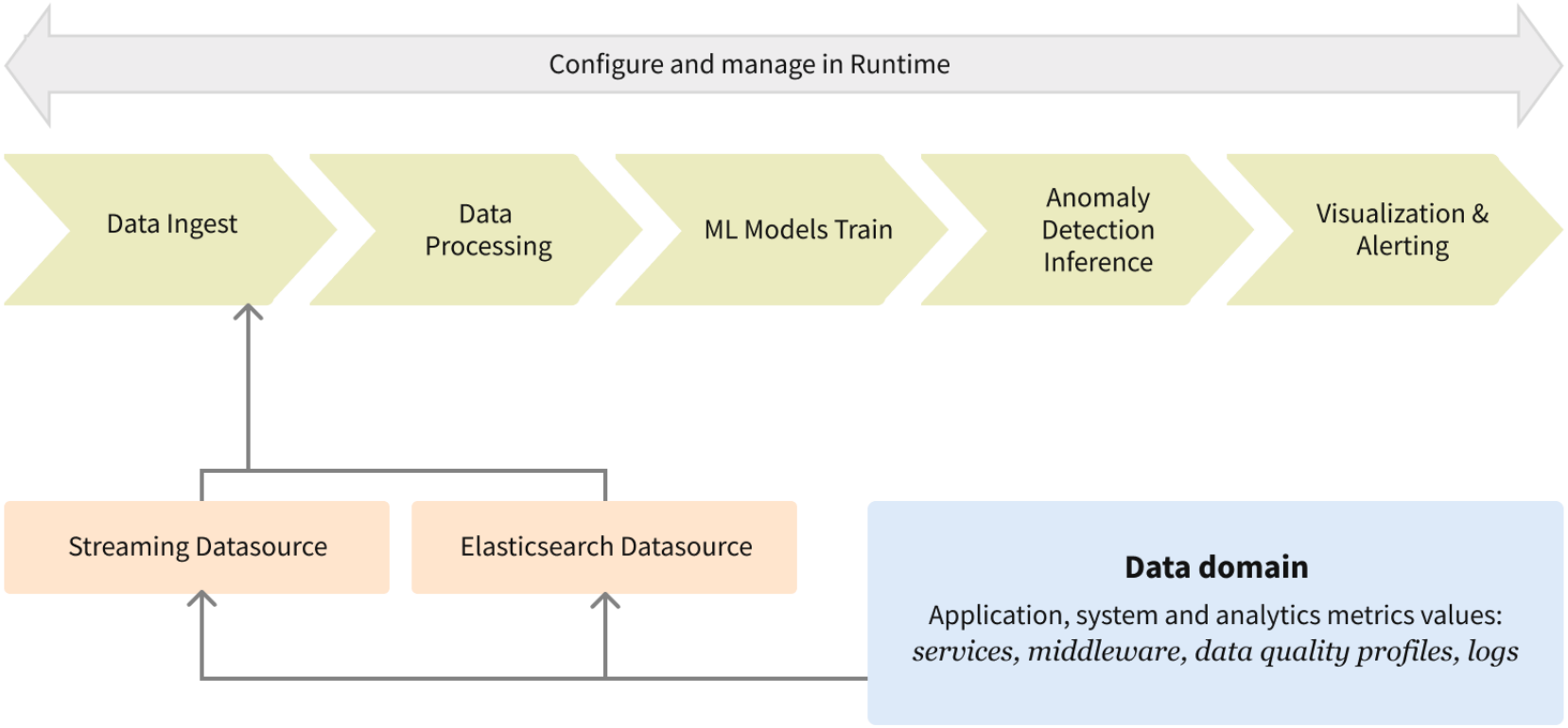
Data Analytics and Big Data



Machine Learning and Artificial Intelligence

- Advanced machine learning algorithms analyze vast datasets to spot deviations from normal expenditure patterns in EU funds.
- AI-driven anomaly detection systems flag unusual transactions in real-time, prompting immediate investigations to prevent misappropriation.
- Predictive modeling identifies potential fraud by correlating activities across multiple data points, including beneficiary history and network relationships.
- Natural Language Processing (NLP) tools sift through documentation and communications to detect linguistic indicators of fraud or collusion.
- Deep learning techniques cross-reference and learn from international databases, improving detection accuracy and reducing false positives.
- Continuous learning frameworks enable AI models to adapt to evolving fraudulent tactics, making the systems more resilient over time.

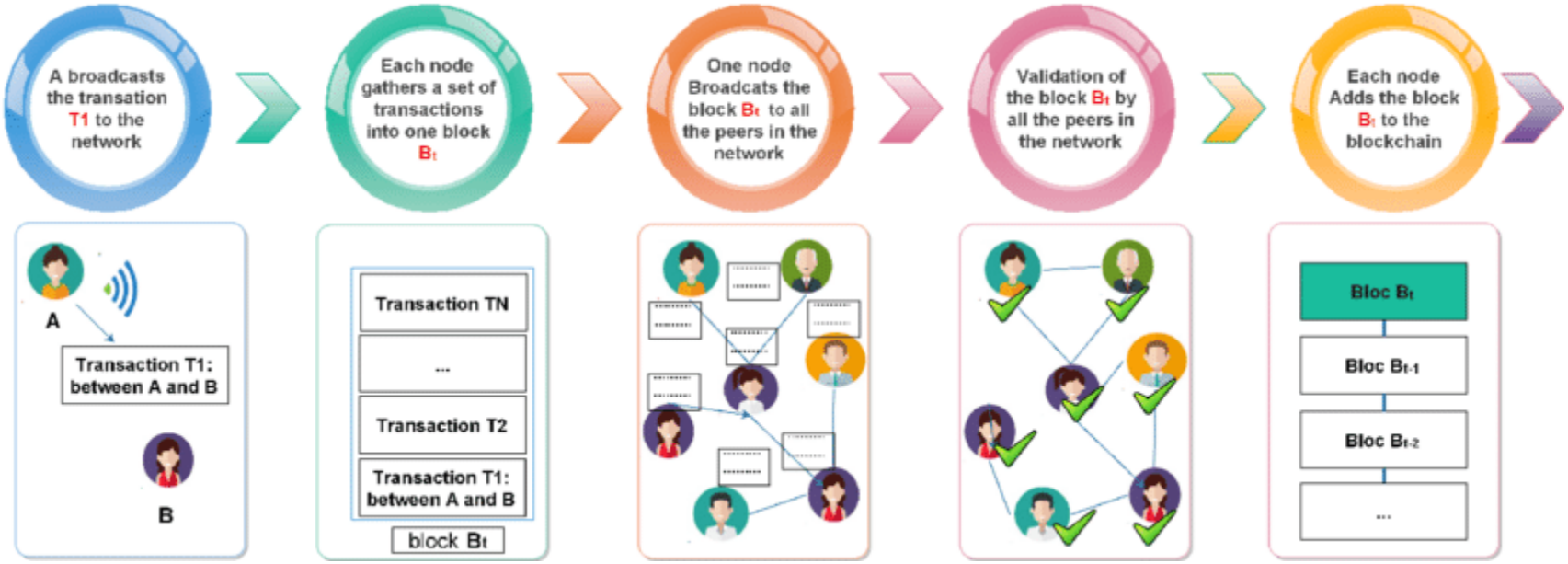
Machine Learning and Artificial Intelligence



Blockchain

- Blockchain's distributed ledger offers an immutable record of transactions, ensuring transparency and traceability in the disbursement of EU funds.
- Smart contracts on blockchain enable automated compliance checks against predefined rules, reducing the potential for fraudulent claims.
- Decentralization of data storage on the blockchain prevents tampering, making it easier to detect and trace irregular activities.
- Cross-referencing of transactions on the blockchain with AI analytics to pinpoint discrepancies and suspicious patterns indicative of fraud.
- Real-time auditing capabilities of blockchain systems provide ongoing oversight of financial flows, reducing the window of opportunity for fraud.
- Integration with existing financial systems to create a secure, transparent, and efficient infrastructure for managing and monitoring EU fund allocation.

Blockchain

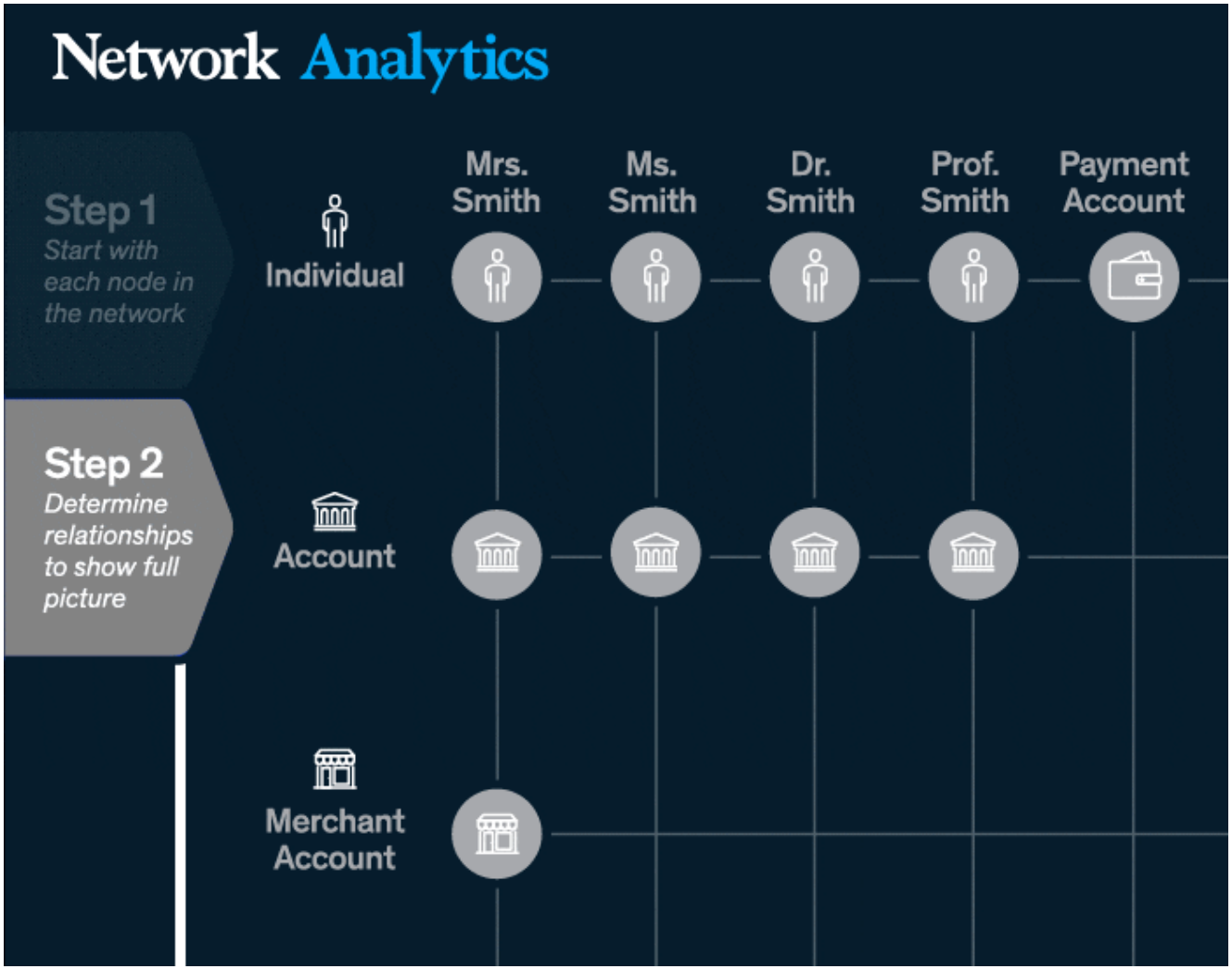


Empowering Financial Integrity with Blockchain for EU Funds

Network analysis

- Network analysis maps financial transactions and relationships, revealing complex patterns and connections that could indicate fraudulent activities.
- Identifies red flags such as unusual clusters of interactions or transactions that deviate from normal patterns within EU fund flows.
- Social network analysis (SNA) techniques detect indirect relationships and hidden networks among entities, helping to uncover collusion and kickback schemes.
- Integrates with machine learning to score and prioritize networks based on risk, focusing investigative resources on the most suspicious cases.
- Time-series analysis of transaction networks spots trends and sudden changes over time, which are often precursors to fraudulent behavior.
- Enhances collaboration between EU member states by sharing network insights, creating a unified front against cross-border financial crimes.

Network analysis



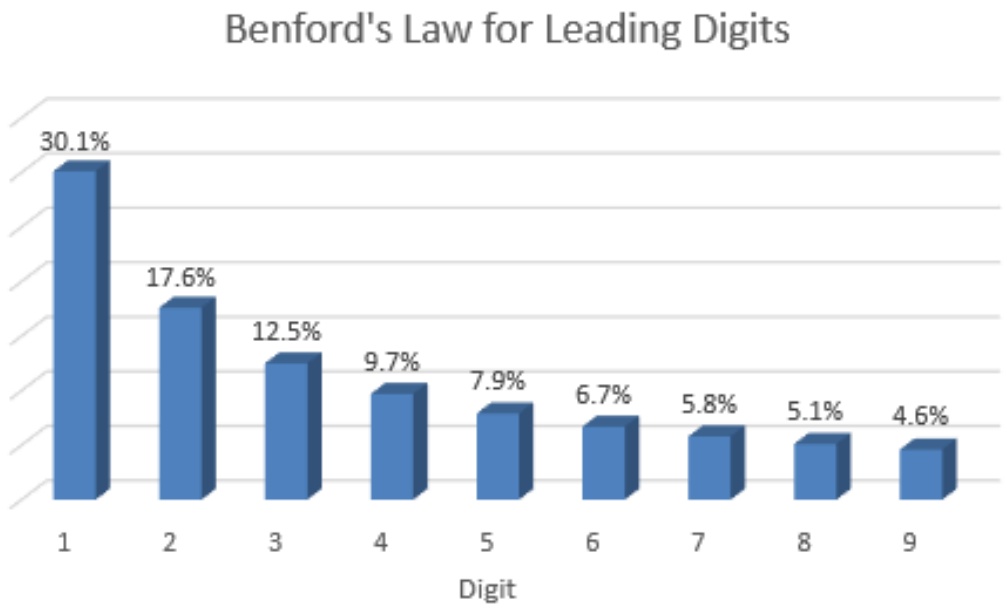
Forensic Accounting Tools

- Forensic accounting tools apply rigorous analytical methods to investigate EU fund disbursements for signs of misappropriation and embezzlement.
- Utilize advanced data-mining techniques to extract and analyze complex financial data, flagging anomalies indicative of fraudulent activities.
- Incorporate digital forensic practices to recover and examine electronic data, ensuring a complete audit trail for transactions and communications.
- Perform ratio analysis to compare financial data against established benchmarks, highlighting inconsistencies and outliers in fund usage.
- Implement Benford’s Law analysis for large datasets to identify unnatural patterns in financial figures that often signify fraud.
- Collaborate with law enforcement and regulatory bodies, providing detailed reports and evidence that can be used in legal proceedings against perpetrators of fraud.

Forensic Accounting Tools

$$NB = \log_{10}\left(1 + \frac{1}{d_1}\right)$$ for d1 = 1, . . . , 9.

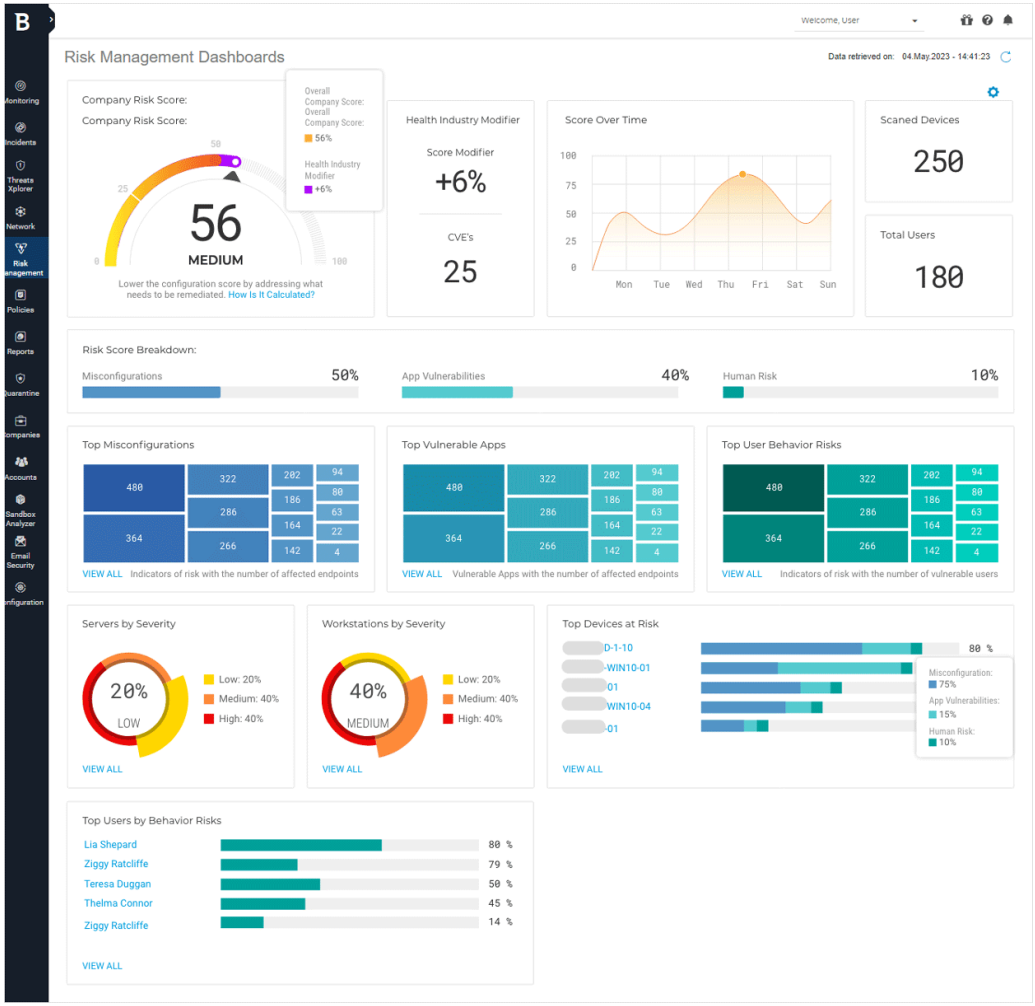
This probability is clearly decreasing with the value of d1, and it is higher than 30% for d1 = 1.



Risk Assessment Software

- Risk assessment software systematically evaluates transactional risks, identifying potential fraud in EU funding processes.
- Integrates historical data and current transactional behaviors to profile and predict risky activities, alerting to possible fraud.
- Uses weighted scoring systems to quantify risk levels, enabling prioritization of investigations based on severity and likelihood of fraud.
- Employs scenario analysis to simulate the impact of potential fraudulent activities, enhancing preparedness and response strategies.
- Supports continuous monitoring with dynamic risk models that adapt to new patterns of fraud as they emerge, maintaining robust defense mechanisms.
- Facilitates cross-referencing of financial operations across EU states, fostering an integrated approach to fraud detection and financial governance.

Risk Assessment Software

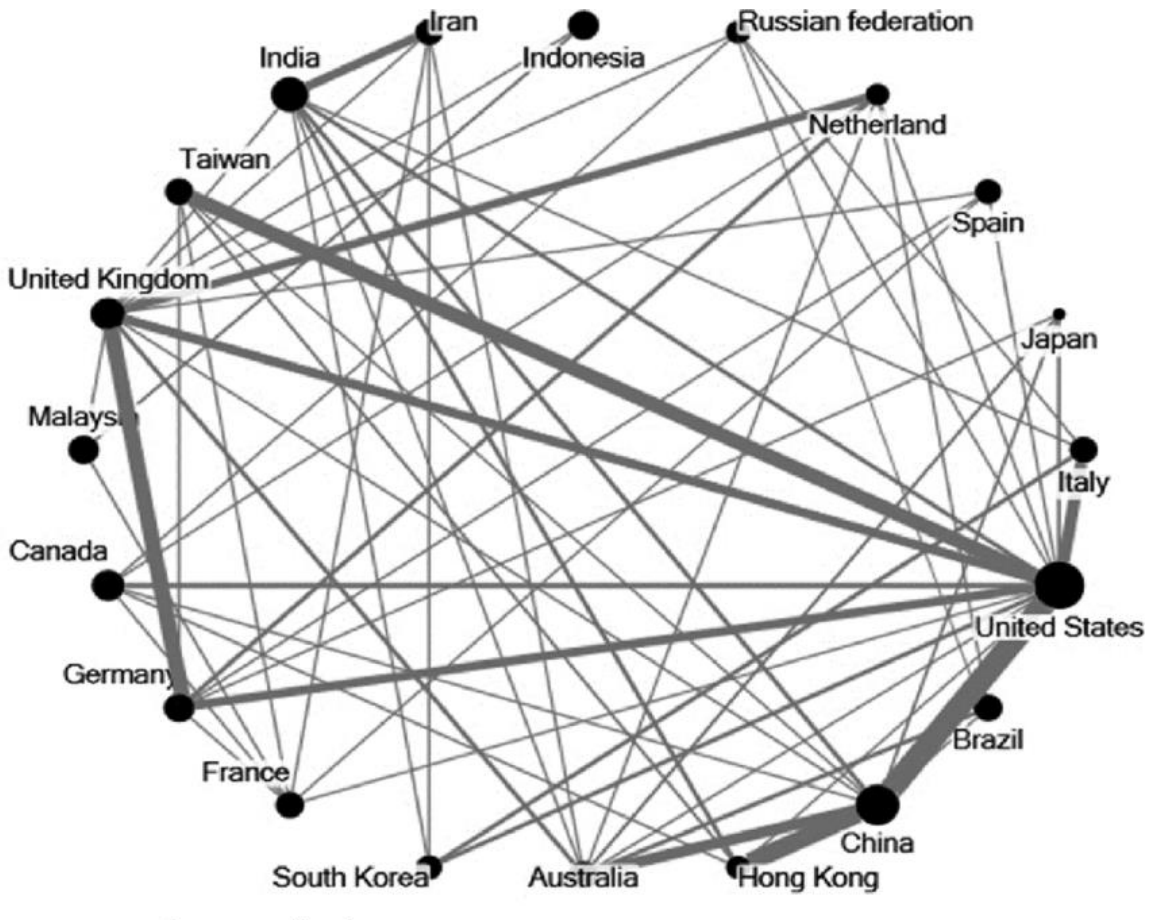


Proactive and Predictive: Risk Assessment Software at the Vanguard of Protecting EU Finances

Collaborative Platforms

- Collaborative platforms enable real-time sharing of data and intelligence across EU member states, enhancing collective fraud detection capabilities.
- Facilitate multi-agency cooperation, bringing together financial regulators, law enforcement, and financial institutions to combat fraud in EU funds.
- Foster a centralized repository for fraud indicators and patterns, streamlining the process of identifying and tracking suspicious activities.
- Support joint investigative teams with tools for collaboration on complex cases, ensuring that expertise and resources are optimally allocated.
- Integrate advanced communication systems to ensure secure, encrypted information exchange, protecting the integrity of shared data.
- Promote transparency and accountability by providing a platform for whistleblowers and auditors to report anomalies directly and anonymously if necessary.

Collaborative Platforms



Process: input channels

- Non-structured sources from Facebook, X,…
- Business registers (Open Data)
- Tax databases
- Banking system databases
- News providers
- Financial information providers
- ….

Process: Study of stylised facts (1/2)

In:

- stylized frauds, countries, sectors, persons, events
- time window
- order [relevance, topic, country, sector ...]

- Out:
- alert on potential frauds
- Beneficial Owners and Ownership Structure
- network of persons/companies
- tendency: % growing / falling
- Zooming in / out for Region
- Zooming in / out for time windows

Process: study of stylised facts (2/2)

Methods and tools applied :

- Cross lingual data extraction
- Multi lingual semantic for identification of potential frauds
- Supervised/knowledge based machine learning
- Network analysis
- Visualisation techniques

Process: Output and Visualisation

- Reports according the selected topic, sector, companies and individuals
- Visualizations according the defined parameters and scales
- Links to original sources and background material

Conclusions

- To go beyond formal controls
- Use the power of big data
- Increase the usage of AI empowered BI tool for identification of suspicious activities
- Multidisciplinary approach
- ….
- Discover unknown unknowns!

References

- AYACHE, Elie, 2010. *The Blank Swan: The End of Probability*. Chichester: Wiley.
- BREIMAN, L., 1961. Optimal gambling systems for favorable games. *In*: Jerzy NEYMAN, ed. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume I*. Berkeley: University of California Press, pp. 65–78.
- COSMIDES, Leda, and John TOOBY, 1996. Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition*, **58**(1), 1–73.
- de FINETTI, Bruno, 1937. La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, **7**(1), 1–68. Translated into English as 'Foresight: Its logical laws, its subjective sources' in H. E. Kyburg, Jr and H. E. Smokler, eds. *Studies in Subjective Probability*. New York: Wiley (1964), pp. 93–158.
- GIGERENZER, Gerd, and Ulrich HOFFRAGE, 1995. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, **102**(4), 684–704.
- GRAY, John, 2008. *Black Mass: Apocalyptic Religion and the Death of Utopia*. Penguin Books. First published by Allen Lane in 2007.
- KAHNEMAN, Daniel, and Amos TVERSKY, 1979. Prospect theory: An analysis of decision under risk. *Econometrica*, **47**(2), 263–292.
- KELLY, Jr, J. L., 1956. A new interpretation of information rate. *The Bell System Technical Journal*, **35**(4), 917–926.
- KEMENY, John G., 1955. Fair bets and inductive probabilities. *The Journal of Symbolic Logic*, **20**(3), 263–273.
- KNIGHT, Frank H., 1921. *Risk, Uncertainty and Profit*. Boston, MA: Hart, Schaffner & Marx; Houghton Mifflin Company.
- LEHMAN, R. Sherman, 1955. On confirmation and rational betting. *The Journal of Symbolic Logic*, **20**(3), 251–262.
- MOXON, Steve, 2010. Culture *is* biology: Why we cannot 'transcend' our genes—or ourselves. *Politics and Culture*, **1**.
- PINKER, Steven, 2011. *The Better Angels of Our Nature*. New York: Viking Books.
- RAMSEY, Frank Plumpton, 1926. Truth and probability. *In*: R. B. BRAITHWAITE, ed. *The Foundations of Mathematics and Other Logical Essays*. London: Kegan Paul, Trench, Trübner (1931), Chapter VII, pp. 156–198.
- SEWELL, Martin, 2012. The demarcation of science. Young Statisticians' Meeting, Cambridge, 2–3 April 2012.
- SHIMONY, Abner, 1955. Coherence and the axioms of confirmation. *The Journal of Symbolic Logic*, **20**(1), 1–28.
- TALEB, Nassim Nicholas, 2010. *The Black Swan: The Impact of the Highly Improbable*. Second ed. New York: Random House Trade Paperbacks.
- TVERSKY, Amos, and Daniel KAHNEMAN, 1973. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, **5**(2), 207–232.
- TVERSKY, Amos, and Daniel KAHNEMAN, 1992. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, **5**(4), 297–323.

SAPIENZA
UNIVERSITÀ DI ROMA

Francesco Bellini

francesco.bellini@uniroma1.it