Association for Information Systems
AIS Electronic Library (AISeL)

MCIS 2014 Proceedings

Mediterranean Conference on Information Systems (MCIS)

Summer 9-4-2014

BIG DATA ANALYTICS FOR FINANCIAL FRAUDS DETECTION

Francesco Bellini
University La Sapienza, francesco.bellini@uniroma1.it

BIG DATA ANALYTICS FOR FINANCIAL FRAUDS DETECTION

Research in Progress

Bellini, Francesco, University La Sapienza – Dept. of Management/Eurokleis s.r.l., Roma, IT,
francesco.bellini@uniroma1.it/francesco.bellini@eurokleis.com

Abstract

Criminals and criminal organizations often make use of companies and other corporate entities to hide their identity, conceal illicit flows of money, launder funds, finance terrorist organizations, evade taxes, create and hide slash funds, commit bribery, corruption, accounting frauds and other financial crimes. These legal entities are frequently organized into complex ownership schemes set up in different countries, and with a "Chinese boxes" structure, in order to make it harder to determine who ultimately controls them and benefits of the illegal conduct.

Currently there are a lot of competitors in the market of Financial Fraud Detection but the software that they propose are mainly oriented to supervise and manage the institutions' internal compliance processes such as the management and transmission of Suspicious Activity Reports (SAR) instead of providing intelligence tools for proactively discovery potential threats and identify the final beneficiaries of illegal operations. Consequently there is a potential for Financial Fraud Detection focused on the on-line, real-time statistical analysis of transactions, operators behaviour, price movements and the use of data mining algorithms that work on heterogeneous sources of big data. After having described the schemes used for executing the three most relevant financial frauds this research proposes a novel approach for the detection of illicit behaviours and suspect transactions. The approach benefits of a multidisciplinary approach for the analysis of the big data streams coming heterogeneous sources such as TV stream, social media and public (official and unofficial) data bases.

Keywords: financial frauds, big data, business intelligence, real-time analytics.

1 Introduction to financial frauds and Chinese boxes

The well known Chinese boxes scheme mentioned is common to at least three of the most relevant crimes related to financial activities:

- **Market abuse** includes the insider trading (average insider purchases are more than $56 million a day, or nearly $1.2 billion a month; average insider sales: more than $600 million a day, or nearly $13 billion a month – US SEC), the market manipulation that can be defined as distorting the price-setting mechanism of financial instruments (trade-based manipulation) or as disseminating false or misleading information (information-based manipulation). Such a 'broad' understanding of market abuse has been adopted by the legal system of the European Union. Technological innovations and significant changes to the structure of the European market landscape have contributed to a rapidly increasing presence of High Frequency Trading (HFT) that, on one hand, optimally executing certain complex trading strategies but, on the other hand, presents opportunities for abuse and market manipulation**.** There are many market abuse schemes (i.e. pools, churning, stock bashing, pump and dump, bear ride etc.) that can be monitored and stylized in order to train the detection engine.

- **Tax fraud** is a huge source of public imbalances: VAT evasion 106 billion + 94 other taxes in EU 27 (out of 700 billion/year) needed to face the crisis up to 2020). Value Added Tax (VAT) has a key strength: the government effectively collects a proportion of the tax at each stage in the sales chain, rather than concentrating the potential revenue, and the risk of evasion, in the final seller. This "fractionated" nature is also VAT's principal weakness: many more transactions have to be logged and reported by traders, raising both the regulatory burden and the difficulty of monitoring compliance. In addition, the seller of an item is collecting the tax on behalf of the revenue authority, rather than the revenue authority collecting it directly from the taxpayer. An example of carousel fraud is shown in the following figure

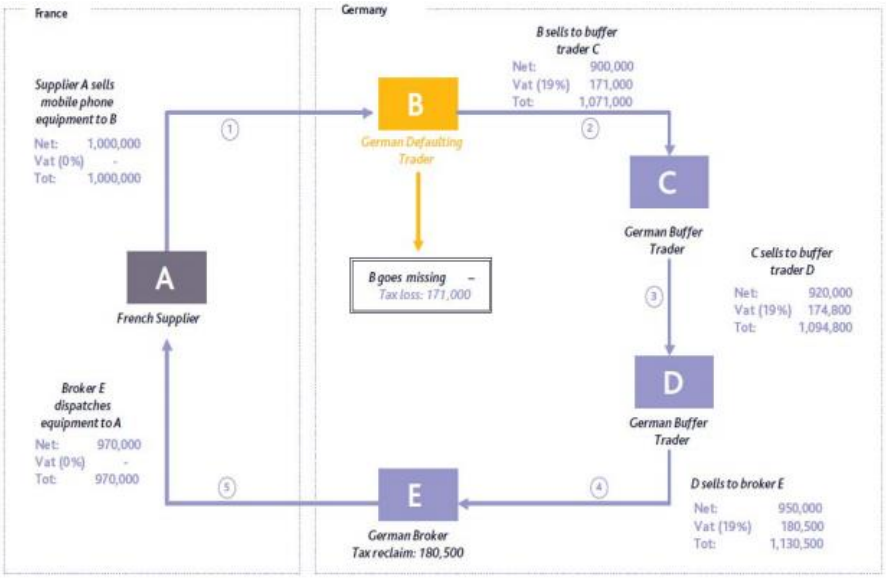

Figure 1. VAT carousel fraud - Source: International VAT Association (2007).

A German missing trader (B) buys goods from a French company (A) without VAT (destination principle) and sells them at a VAT-inclusive price to (C), without remitting the tax to the authorities. After several transactions, the same goods are exported to (A) and the carousel starts again. At each turn, the German broker (E) claims back the input VAT paid (€180,500). The net loss for the German government is €171,000 (VAT collected by (B) and not remitted to the authorities). This is the basic scheme but the reality proposes more complex situations with dozens of players involved.

- **Money laundering** remains a hot topic in the FS sector, where it is almost five times more likely to occur than in other industries. The report of United Nations Office on Drugs and Crime (World Bank and UNODC, 2011) estimates that criminal proceeds amounted to 3.6% of global GDP with up to 2,8 $ trillion being laundered. FS organizations fear the fallout of being caught up in money laundering – almost 30% believed that the most severe impact is reputational. The latest statistics available from Eurostat show that in 2010 266.388 Suspicious Activity Reports (SAR)[1], 118.559 Unusual Transaction Reports (UTR)[2] and 126.116 Suspicious Transaction Reports (STR)[3].

  In the initial - or placement - stage of money laundering, the launderer introduces his illegal profits into the financial system. This might be done by breaking up large amounts of cash into less conspicuous smaller sums that are then deposited directly into a bank account, or by purchasing a series of monetary instruments (cheques, money orders, etc.) that are then collected and deposited into accounts at another location.

It is also important to underline that **all these frauds are strictly inter-connected** since the illegal profits coming market abuses and tax frauds need to be "laundered" as well as the VAT carousel fraud is itself a form of money laundering.

2 Current workflow for frauds detection and deficits

2.1 Market abuse schemes

If we refer to market abuse there are three broad groups of detection methods:

- Simple detection procedure that deals with raw market data.
- Procedures that utilize some statistical market models to forecast the market.
- Non-parametric methods and numerical algorithms.

[1] Suspicious Activity Report (SAR) is a disclosure made to an FIU by a professional having an obligation to disclose based on any suspicious activity of money laundering or terrorist financing. The main difference with STR is the fact that the SAR scope is broader as it may not include a transaction.

[2] Unusual transaction report (UTR) is a disclosure made to an FIU by a professional with an obligation to disclose, based on unusual behaviour in a client's profile. The main distinction between an STR and UTR is the higher standards and quality expected of STRs.

[3] Suspicious transaction report (STR) is a disclosure made to an FIU by a party having an obligation to disclose based on any type of suspicion of money laundering or terrorist financing which are required by regulations, which may include unusual behaviour. Suspicious transactions are handled to the appropriate law enforcement units for investigation. The counting unit was specified as the initial STR received for each case opened by the FIU from each category of obliged entity.

The main disadvantages of these approaches are the need to constantly and precisely calibrate the algorithm parameters and a potential bias towards ambiguous signals of the system. Numerical algorithms need to be tested thoroughly before one can judge their effectiveness and put them into practice. Moreover these methods are based on the market data analysis without exploring the outer dimension coming from news, blogs, TV, radio and social channels.

2.2 VAT Frauds and money laundering

Feedback is an important technique in improving intelligence and maximising the impact of the effort that goes into reporting suspicious activity but large financial institutions can generate thousands (and in some cases, hundreds of thousands) of SARs every year, so AML compliance costs can be substantial. Many scenarios will generate superfluous or unproductive alerts, that is, alerts where suspicious activity is anticipated but none is found. More unproductive alerts mean higher compliance costs. Unstructured data can be a fundamental problem. In some instances, SAR databases are made up of many thousands of narrative reports containing unstructured data that can be difficult to analyse. Unstructured data is information presented as text – for example, a written paragraph describing the nature of the suspicious activity – which is challenging for analytics to parse. As of yet, there is no easy and reliable way to perform quantitative analysis on the informational components of a written paragraph. However, augmenting the SARs with descriptive (yet structured) information will help support the use of analytics and data mining tools.

After having identified the illegal transaction the following action is to identify the Beneficial Owner and the actual Ownership Structure. The BOWNET project carried out a survey among the EU competent authorities asking what data/information is used to reconstruct the OS and identify the BO of suspicious corporate entities

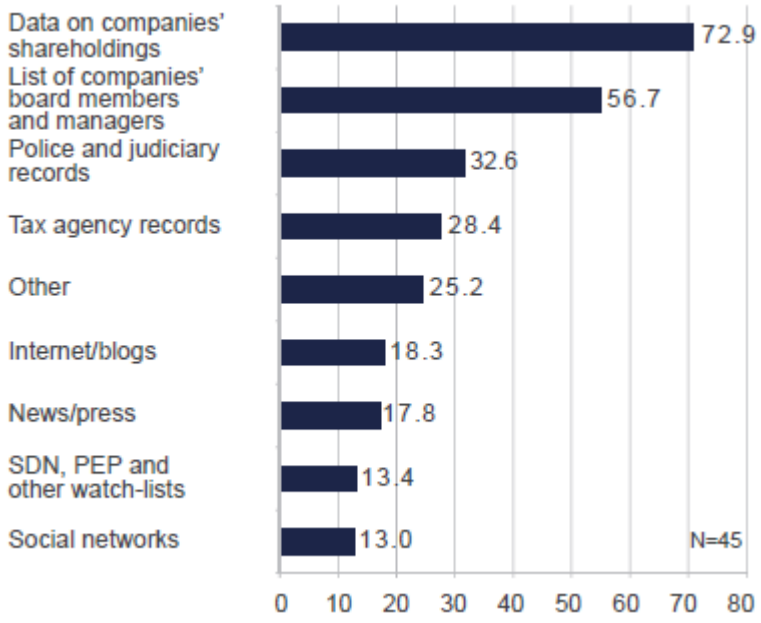

Figure 2. Data sources currently used for reconstructing the OS and identify the BO of suspicious corporate entities (source BOWNET project).

It is consequently not surprising to find that the data sources most frequently used in investigations of this kind are **business registers**, which usually provide, among other data, also information on the shareholders and directors of the legal persons registered. But if most SMEs, after all, are straightforward structures where the shareholders are the same as the directors and they hold the shares in their own names for some, particular those that want to obfuscate this information, there are many opportunities to do so these days – through nominee directors and shareholders, offshore entities, and complex networks of obscure company structures the rules

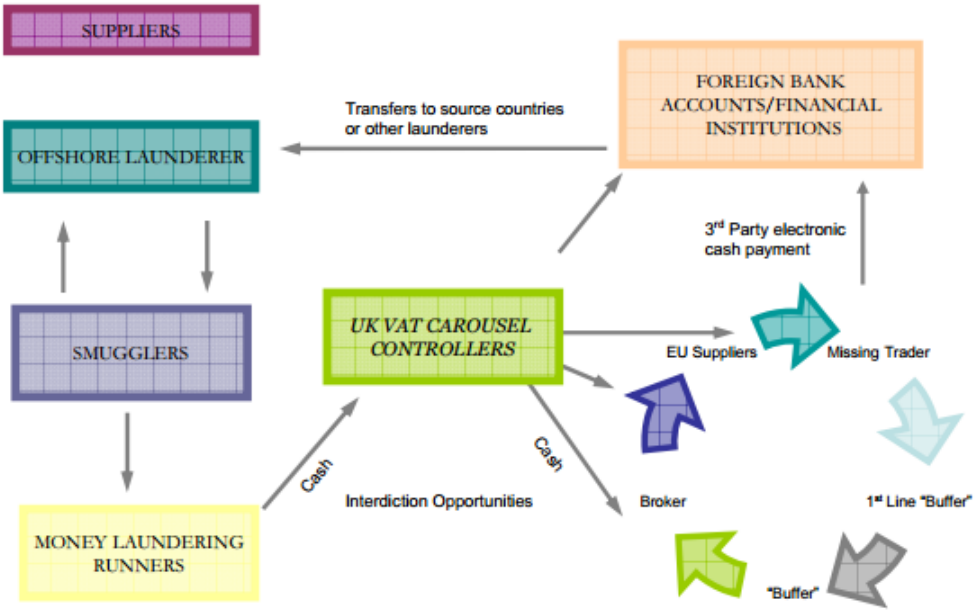

Figure 3. Financial links between VAT carousel and drugs street cash (Source FATF).

For all these frauds becomes fundamental to identify the signals of a suspicious transaction and trace the connections that may lead to the Beneficial Owner (BO) i.e. the actual responsible of the criminal offense. The Financial Action Task Force (FATF, 2012) defines Beneficial Owner as: "the natural person(s) who ultimately owns or controls a customer and/or the natural person on whose behalf a transaction is being conducted. It also includes those persons who exercise ultimate effective control over a legal person or arrangement".

Directive 2005/60/EC (Third AML Directive or 3AMLD), integrated by implementing measures (Directive 2006/70/EC, see below), which comprises the highest number of provisions regarding BO identification and the disclosure of information on BOs and Ownership Structure OS. The 3AMLD addressed gaps highlighted in previous scoping studies and impact assessments, such as Transcrime in 2007 and Howell e Van Reenen in 2005, in particular referring to the lack of regulation requiring disclosure of full information on BOs. In fact, the directive, drawing on from previous FATF recommendations and adopting an intermediary-based approach, requires the following categories: credit institutions, financial institutions, auditors, external accountants and tax advisors, notaries, and other independent legal professionals, trust or company service providers and real estate agents. Identifying then the money flow after it enters the financial system is therefore crucial to identifying the associated money laundering, and Suspicious Transaction Reports (STRs) play a crucial role in this task. Therefore, whistleblowing mechanisms appear to be more prevalent than before, however doubts remain over their effectiveness. Indeed, a recent survey of FTMT-GAFI among countries representatives only 2 (out of 7) EU respondents were able to say how many intelligence reports

concerning VAT carousel fraud were disseminated to other countries and 25% of respondents failed to conduct annual fraud risk assessments.

3 Features for a new model of Financial Frauds Detection

Technological innovation, combined with the increase in the number of trading venues, makes it necessary for the Financial Institutions to use an electronic detection and analysis system to identify suspicious transactions, Beneficial Owners and Ownerships Structures. This enables to trace certain types of frauds abuse in a targeted way. In the future, the anti-fraud detection will focus more on the statistical analysis of transactions, price movements and use the use of algorithms to conduct searches of reported transactions (data mining) on big data[4] rather than use single reports (i.e. SAR) as a starting point of the analysis. Our system for Finance Frauds Detection (FFD) aims to provide a new generation of services to better support the identification of the financial frauds in the previously described areas. The system will offer the financial institutions support for the whole workflow from the discovery of suspect behaviours to the identification of the Beneficial Owner and the actual Ownership Structure.

Functions	**Parameters**
Selection of input channels	
from all over the world from public Web Sites e.g. from online newspapers or press agencies, Social Networks, TV channels, Tax databases, banking system databases, company registries. They could select the sources, which are most relevant for the target users, e.g. on the basis of received input, news agencies or Web access statistics. The end consumers can be located in the region but also at any place in the world and are interested in some specific domains	In: • list of Channels with Type Details related to channel type: • Web Sites: list of Pages social networks: topics, ... • TV Streams Time Schedule • company registries • market price time series
Identification of stylized facts representing a signal of possible fraud, such as:	
• For market abuse o a trader places both buy and sell orders at about the same price o savvy online message board posters (a.k.a. "Bashers") who make up false and/or misleading information	In: • stylized frauds, countries, sectors, persons, events time window

[4] According to Gartner's definition "Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."

about the target company in an attempt to get shares for a cheaper price o a group of traders create activity or rumours in order to drive the price of a security up o selling and repurchasing the same or substantially the same security for the purpose of generating activity and increasing the price. • For VAT and money laundering o Business has changed commodities and sectors quickly. o Turnover of business grows substantially in short time period o Unsolicited approaches from organisations with little history in the market offering a guaranteed profit on high-value deals. o Repeat deals at the same or a lower prices and small or consistent profit, e.g. €1 per item. o Instructions to pay less than the full price to the supplier. o Using goods that are of high value and low volume and attract a high tax rate such as computer parts or mobile phones. o The total amount of money channelled through an account is considerable, although the balance is usually very low. o Not publishing the companies' annual records, pursuing activities that are not part of the corporate goals. o Obligatory elements of invoices, such as VAT number, date etc. are missing. o Foreign nationals in charge of companies, who have often never been a director of a company in the jurisdiction and may not have an address in the jurisdiction. o Invoices for services not usually associated with the business. o Export of goods do not match normal market rate**.**	• order [relevance, topic, country, sector ...] Out: • alert on potential frauds • Beneficial Owners • Ownership Structure • network of persons/companies • tendency: % growing / falling • Zooming in / out for Region • Zooming in / out for time windows The methods and tools applied are: • Cross lingual data extraction • Multi lingual semantic for identification of potential frauds • Supervised/knowledge based machine learning • Network analysis • Visualisation techniques
Visualization	
to get an overview over most important new trends during the last hours or day or highlight specific developments. Users in charge of the control will utilize the visualization of transactions, BOs and OS from the observed channels to update and adopt the needed actions including the association of the sentiment to markets behavior and the face recognition of the suspects.	In: • topics, countries, persons, • time window • type of visualization [diagram, curbe, table…] Out: visualization figures

	variable formats
Exposé generation	
to get an ad hoc and intraday report: users can select domains where they want to get information from in order to collect input materials for their own control activity. In-depth list of potential threats which have to be updated several times per day. Include new sources that comply with the flexibly ordered digest of different new topics and support the FFD.	In: topics, countries, sectors, persons • level of detail • time window • type of materials Out: links to background materials • type of material • links to big data sources, aggregated items relevance values

Table 1. Functions and Parameters of the Financial Frauds Detection system.

Input Channels

Input channels are social media (from Twitter to specialized blogs through Facebook), news channels (text and voice), financial information providers, information gathered by State bodies such as business registers, tax offices and made available under the Open Data scheme. Indeed, there is growing trend to make available to a wider public the information because it is essential to give the greatest publicity to the affairs of companies, which everyone may know on what grounds he is dealing[5]. These sources use text, pictures and video in the several languages that are used for business conversations for finance and trading:

Internet, blogs, social media:

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

Output

Analysis results for FFD system will be shown through a dashboard, which content can be exported to file reports in PDF format for in-house use. Different aggregates of the information shown in the

[5] http://registries.opencorporates.com/

dashboard could be exported including entities (company names, frauds, locations etc) mentions, trends related to a given entities and/or concepts appearing along with selected entities**.**

The output of the FFD services will be delivered according the selected time windows and can be transferred in-house to another department or forwarded to external customers:

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

4 Conclusions

We believe that the proposed approach goes beyond the current state of the art (Ngai et al., 2011) by applying monitoring of social networks, registers and web sources to extract cases of Market abuse, VAT frauds and Money laundering. Indeed the control chain of Financial Frauds can be only reconstructed by merging the heterogeneous information from the business registers and other private business information providers, and by collecting information from open sources such as LEA reports, the press, social networks.

The electronic detection and analysis to identify suspicious transactions, Beneficial Owners and Ownerships Structures is not widely exploited to trace certain types of frauds abuse in a targeted way. There is a potential for an anti-fraud detection which focus more on the statistical analysis of transactions, price movements and use the use of algorithms to conduct searches of reported transactions (data mining) on big data rather than use single reports (i.e. SAR) as a starting point of the analysis. The minor role played by open sources (e.g. news, press archives) is mostly due to problems of reliability and organisation of the information but also to the fact that **almost half of the respondents (46.7%) did not make use of software** (Transcrime, 2013) and IT tools to analyse such information.

Here comes the role of our FFD system, indeed, if the non-use of software may not necessarily have consequences in terms of either the effectiveness of investigations or of their outcomes, it surely implies more time-consuming research or a loss of potential investigative leads. But at the present stage the business intelligence solutions are only used by a small number of public investigators in fraud identification activities due mainly to a lack of software able to collect, merge and process information coming from different sources different from business registers.

References

Cholewiński, R., (2009), Real-Time Market Abuse Detection with a Stochastic Parameter Model, vol. 284, 2009, pp. 261-284.

FATF (2010), Best Practices Confiscation (Recommendations 3 and 38), http://www.fatf-afi.org/media/fatf/documents/recommendations/Best%20Practices%20Confication%20R3%20and%20R38%202012.pdf

FATF (2012), The Forty Recommendations http://www.fatfgafi.org/media/fatf/documents/recommendations/pdfs/FATF%20Recommendations%20%28approved%20February%202012%29%20reprint%20May%202012%20web%20version.pdf

Howell J. and Van Reenen P. (2005), Scoping Study: Cost Benefit Analysis of Transparency Requirements in the Company/Corporate Field, Contract No. DG JAI/D2/2004/02, http://ec.europa.eu/homeaffairs/ doc_centre/crime/docs/study_cost_benefit_ transparency_en.pdf

Keen M. and Smith S. (2007), VAT Fraud and Evasion: What Do We Know, and What Can Be Done?, IMF Working Papers (2007), WP/07/31

Minenna, M. (2003), The detection of market abuse on financial markets: a quantitative approach, Quaderni di Finanza, CONSOB 2003.

Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559-569.

Öğüt H. Doğanay M. Mete, and Aktaş R., (2009) Detecting stock-price manipulation in an emerging market: The case of Turkey," Expert Systems with Applications, vol. 36, Nov. 2009, pp. 11944-11949.

Transcrime (2007), Cost Benefit Analysis of Transparency Requirements in the Company/Corporate Field and Banking Sector Relevant for the Fight Against Money Laundering and Other Financial Crime, http://transcrime.cs.unitn.it/tc/851.php

Transcrime (2013), Final Report of Project BOWNET - Identifying the beneficial owner of legal entities in the fight against money laundering

World Bank and UNODC (2011), The Puppet Masters: How the Corrupt Use Legal Structures to Hide Stolen Assets and What to Do About It, http://www1.worldbank.org/ finance/starsite/documents/Puppet%20Masters%20 Report.pdf