EXTENDED ABSTRACT

Title: A methodological approach for fraud detection with mathematical tools based on metric graphs.

Authors and e-mail of all: Calabuig Rodríguez, J.M. (*jmcalabu@mat.upv.es*)[1], Falciani, H. (*stella.ethical.buz@gmail.com*)[2] , A. Ferrer-Sapena (*anfersa@upv.es*)[3], Jiménez Fernández, E. (*jimeneze@uji.es*)[3], Sánchez Pérez, E.A. (*easanpe@mat.upv.es*)[4].

Department: 3: Department of Econonomía. 2: Taboow.org 1,4: Department of Applied Mathematics. 3: Department of Informatics.

University: 3: Jaume I de Castelló and 1,4 and 5: Politècnica de València.

Subject area: S10. Procesos espaciales a nivel de punto y redes espaciales.

Abstract

After giving an overview of the techniques for fighting against financial fraud, a new method for automatic fraud detection is shown. For the detection of cartels, we propose a topological graph-based approach, that is explained in this work.

1 Introduction

Based on some classical and new tools, fraud detection is becoming one of the main development themes of the financial world. Ten years after Lehman Brothers fall, entities involved in financial activities have meaningfully increased their control mechanisms, adopting more conservative policies in general, but also trying to identify potential partners who might be suspected of fraudulent practices. From the technological point of

view, a big effort has been made in recent years to increase the power of the mathematical and computational techniques for detecting and classifying possible fraudulent situations.

There are several aspect regarding fraudulent behavior that must be taken into account. First, there are a lot of different situations that can be considered as suspicious parctices and affect the proper functioning of economic activity. Let us present some situations that should be studied; we do not pretend to be exhaustive.

1) Classical fraudulent activities related to automatic financial processes. Fraud detection of the use of credit cards and other automatic financial procedures ?as improper use of bank accounts-, has been studied deeply from the technological/mathematical point of view. Based on well-known mathematical theories (as for example, outliers detection), it is a well established research framework, in which good and powerful results have been already obtained. Presumably, new contributions on the topic will be obtained in the next years, but no revolutionary methodological changes are expected.

2) Manipulation of contracts and falsification of documents. It seems clear that most of the tools to be used in this case should be done by analysts being experts on laws. However, some help from some computational tools related to semantic search, may be needed. In any case, it seems quite difficult to translate something that is deeply related to regulatory and legal norms into a mathematical problem. Automation does not seem to be foreseeable in this environment, at least in the next few years.

3) A technically more complex problem is the one that appears when one tries to analyze using computational tools the existing relations between the different actors that appear in a financial process. For example, how to discover that a group of companies acts forming a cartel in the context of public contracts? This is a difficult problem, and a mix of different techniques and interdisciplinary elements must be taken into account. However, it is a hot matter, both from the point of view of the private business world and the public affairs.

2 A graph-based topological tool for financial fraud detection

In this work, we are interested in explaining some recent developments related to the third point presented above. Let us explain first the main properties of the problem that must be taken into accout, for the aim of reducing the general problem to a mathematical algorithm. The general methodological context that we assume is supported by the so-called Fraud Triangle theory, that provides a sociological/psicologial explanation of why fraud appears. It has shown to be useful in applications (see for example [2, 4, 5, 10, 11]). However, our method is more formal, and goes directly to the creation of a general framework for a mathematical analysis of fraud.

Mathematical tools that can be taken into account for figthing against this particular class of fraud are already well develped. Statistics, game theory and graph theory are probably the better known ones; the reader can find information about in [1, 6, 8, 9, 12, 13]). The graph based analysis has lead, for example, to the creation of some commercial software, that is nowadays becoming the best analytical tool for fraud detection. However, this kind of tools still needs an analyst: no completely-automatic study of the graph is still possible.

We are developing a particular purpose for getting such an automatism. In order to do that, we have to center the attention on the construction of a topology on the graph. The idea is to use a quasi-pseudo-metrics (see for example [3, 7] for the definition).

Given a graph Ω, we consider a proximity function $\phi : \Omega \times \Omega \to \mathbb{R}^+$, in order to define define a quasi-pseudo-metric for Ω. This proximity function, that will be studied later, represents the mathematical formalization of all the aspects that must be taken into account for the construction of the quasi-pseudo-metric that models the problem. For example, if an analyst suspect that there is a cartel acting over a certain market, and the people that form the market are supposed to be related by family relationship, this must be used for defining ϕ.

The other mathematical element that gives a concrete presentation to the model is the

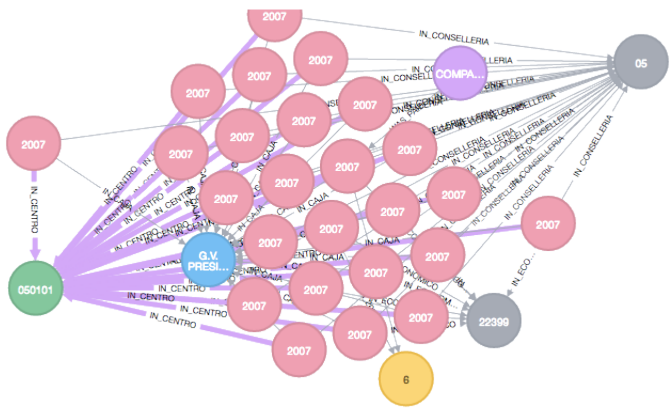

Figure 1: Grafical representation of a metric graph with Neo4j

notion of quasi-pseudo-metric. A quasi-pseudo-metric on Ω is a map $d : \Omega \times \Omega \to \mathbb{R}^+$ such that for $a, b, c \in \Omega$,

1. $d(a,b) = 0$ if $a = b$, and

2. $d(a,b) \leq d(a,c) + d(c,b)$.

Note that it is like a distance, but no symmetry in the definition is expected. This fits with the assumptions of the model. Relations among the entities that appear in a financial process are not symmetric. For example, if a transfers a certain quantity of money to b, the relation is clearly non-symmetric: a pays and b receives.

3 Proximity functions defined by means of correlation matrices: the standard model.

Let us explain the procedure of how to give a quasi-pseudo-metric to a graph for modelling a problem of fraud detection.

i) Take a set of N entities Ω that act in given financial process. Suppose that M is a class of properties of the elements of N that are relevant for the process. Assume that we can measure these properties by means of non-negative real numbers for each element of the graph $a \in \Omega$. Define a set of N vectors v_a with the values representing the properties of each $a \in \Omega$.

ii) Define a matrix C in which the rows are given by these vectors v_a given in i). They are suppose to have Euclidean norm equal to one. Otherwise, normalize them.

iii) Take the correlation matrix $A = C \cdot C^T$. We define the proximity matrix as $\Phi = \mathbb{I}_{N \times N} - A$, that is in this case symmetric. It can be used for defining the pseudo-metric d_ϕ using several techniques and adding other non-symmetric terms $g : \Omega \times \Omega \to \mathbb{R}^+$ if necessary. One of them might be given by an infimum, and so by the formula

$$d(a,b) = \inf \Big\{ \phi(a,c_1) + \phi(c_1,c_2) + ... + \phi(c_n,b) \Big\} \; + \; g(a,b), \quad a,b,c_1,...,c_n \in \Omega.$$

Using this kind of topological arguments, we can model the process of analysis of a possible financial fraud. The general procedure would be as follows.

- Consider a financial process in which fraud is suspected. Define the entities that must be considered as elements of the process: persons, entities, contracts,... Define the properties of the elements that could be relevant for the process. Define a graph which vertices are the entities and the axis the connections that are known among them. Model the way of weighting the relations that represent the axis, and the rest of the elements of the system.

- Define a proximity matrix from the information collected in the previous step. Apply the mathematical method explained above for constructing a quasi-pseudo-metric (often simply a metric) on the graph.

- Use the resulting topological space for analyzing the graph. For example, if an element $a_0 \in \Omega$ is suspicious of fraud, we must analyze *all the entities* that are "near" of a_0 in the graph, —that is, the quasi-pseudo-distance is less or equal to ε for a given $\varepsilon > 0$—.

Most of the steps of this scheme can be done automatically. Therefore, this model opens the door to the algorithm-based study of graphs (see Figures 1 and 2), that can be done without the intervention of an analyst. This would mean a revolutionary jump in the detection, for example, of cartels.

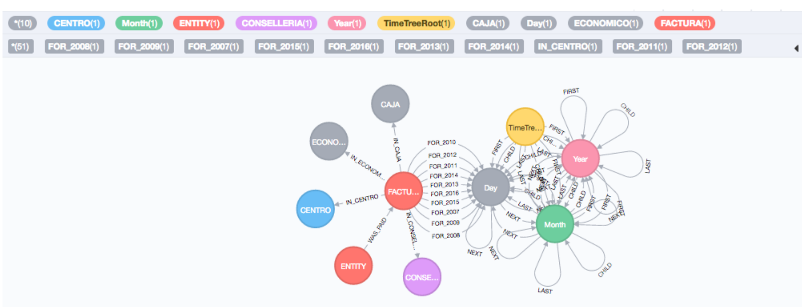

Figure 2: Metric graph representing relations among entities for fraud analysis

4 Conclusions

A general overlook on the financial fraud shows that there are some parts of its analysis that is far to be automatized, that is, analyst are required in most of the cases. We center the attention in the case of fraud that is commited by some ilegal association of persons or entities to violate the rules for clean financial processes. Using graph theory together with topology, an automatic system for fraud detection can be designed. Although in each particular case, the entities (vertices) and properties (weights of the edges) must be defined in a specific way, this research direction can lead to a new approach for the antifraud detection.

References

[1] Bolton R.J and and Hand D.J. *Unsupervised Profiling Methods for Fraud Detection.* (unpublished, available in Google Scholar).

[2] Jack Dorminey, A. Scott Fleming, Mary-Jo Kranacher,Richard A. Riley, Jr. *The Evolution of Fraud Theory.* Issues in Accounting Education: May (2012), Vol. 27, No. 2, pp. 555–579.

[3] Künzi, Hans-Peter A. *Quasi-uniform spaces?eleven years later.* In Topology Proc, vol. 18, pp. 143-171. 1993.

[4] Mansor, *Fraud Triangle Theory and Fraud Diamond Theory. Understanding the Convergent and Divergent For Future Research.*

[5] Theodore J. Mock, Rajendra P. Srivastava, and Arnold M. Wright. *Fraud Risk Assessment Using the Fraud Risk Model as a Decision Aid.* Journal of Emerging Technologies in Accounting. Spring (2017), Vol. 14, No. 1, pp. 37–56.

[6] Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., and Sun, X. *The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature.* Decision Support Systems, (2011) 50(3), 559-569. Some concrete techniques and applications of data mining procedures are shown.

[7] Reilly, Ivan L., P. V. Subrahmanyam, and M. K. Vamanamurthy. *Cauchy sequences in quasi-pseudo-metric spaces.* Monatshefte fr Mathematik 93.2 (1982): 127-140.

[8] Richhariya, P. and Singh P.K. *A Survey on Financial Fraud Detection Methodologies.* International Journal of Computer Applications (2012) 45, 22 pp.975 – 1007.

[9] Szárnyas, G., Koovár, Z., Salánki, A., Varró, D. *Towards the Characterization of Realistic Models: Evaluation of Multidisciplinary Graph Metrics.* In Proceedings of the ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems (2016), pp. 87–94.

[10] Gregory M. Trompeter, Tina D. Carpenter, Naman Desai, Keith L. Jones, and Richard A. Riley, Jr. *A Synthesis of Fraud-Related Research.* AUDITING: A Journal of Practice and Theory (2013), Vol. 32, No. Supplement 1, pp. 287–321.

[11] Gregory M. Trompeter, Tina D. Carpenter, Keith L. Jones, and Richard A. Riley, Jr. *Insights for Research and Practice: What We Learn about Fraud from Other Disciplines.* Accounting Horizons December 2014, Vol. 28, No. 4, pp. 769-804.

[12] T. Jeffrey Wilks and Mark F. Zimbelman. *Using Game Theory and Strategic Reasoning Concepts to Prevent and Detect Fraud.* Accounting Horizons (2004), Vol. 18, No. 3, pp. 173–184.

[13] Zhao, J., Lau, R. Y., Zhang, W., Zhang, K., Chen, X., Tang, D. *Extracting and reasoning about implicit behavioral evidences for detecting fraudulent online transactions in e-Commerce.* Decision Support Systems 86 (2016): 109–121.

keywords: Fraud detection, Graph, distance, quasi-pseudo-metric, Analytical Method.

JEL code: C60, C69, G19, Z0