A Framework for Occupational Fraud Detection by Social Network Analysis

Sanni Lookman, Selmin Nurcan

A Framework for Occupational Fraud Detection by Social Network Analysis

Sanni Lookman, Selmin Nurcan

Centre de Recherche en Informatique, Université Paris I, Panthéon-Sorbonne
lookman.sanni@malix.univ-paris1.fr, nurcan@univ-paris1.fr

Abstract. This paper explores issues related to occupational fraud detection. We observe over the past years, a broad use of network research across social and physical sciences including but not limited to social sharing and filtering, recommendation systems, marketing and customer intelligence, counter intelligence and law enforcement. However, the rate of social network analysis adoption in organizations by control professionals or even by academics for insider fraud detection purpose is still very low. This paper introduces the OFD – Occupational Fraud Detection framework, based on formal social network analysis and semantic reasoning principles by taking a design science research perspective.

Keywords: Design science, ontology, data mining, fraud detection, social network analysis, internal control, governance, risk, compliance.

1 Introduction

Frauds partly draw from human beings imaginative nature. Over the years, fraudster's attack methodologies have evolved from an opportunistic approach to some more sophisticated and traceless deception schemes and that, in a constantly yet automatizing but complexifying business environment. In recent years, several unethical behaviors within organizations have received significant attention. Celebrated cases range from financial scandals (Pechiney - 1988, Elf - 1994, Enron - 2001, Kerviel 2008) to data theft (WINDOWS 8 Beta - 2012, Korea Credit Bureau - 2014, SONY - 2014) and have proven that fraud is likely to happen at any level of an organization. The Association of Certified Fraud Examiners, in his 2014 report to the nations on occupational fraud and abuse [ACFE, 2014], estimates a global loss of 5% of revenues to fraud (3.7 trillion dollars if applied to the 2013 Gross World Product). They additionally reported that fraud cases were mostly uncovered by tips or chance (40%). That is an anonymous fraud hotline would even anticipate a lot of fraud damage and yet, knowledge discovery and data mining techniques are teeming.

Detection innovations include automated rules, watch lists matching, supervised and unsupervised classification, data fusion and link analysis. Such techniques have received increased industry specific interests for external frauds (i.e. committed by people outside of the organization) detection. Those would include cybercrimes by

computer or network intrusion, credit card, insurance, telecommunication and credit application frauds [Phua et al., 2004], [Yufeng et al., 2004], [Cox et al., 1997], [Wheeler et al., 2000]. In the meantime, internal or occupational fraud, defined by the ACFE as the use of one's occupation for personal enrichment through the deliberate misuse or misapplication of the employing organization's resources or assets, has proved to be more prevalent than external fraud. PriceWaterhouseCoopers' 2014 Global Economic Crime Survey reports in France an average of 56% of internal fraud [PWC, 2014].

This paper elicits problems faced by investigators in the process of occupational fraud detection and comes up with a solution which contributes to solving these problems. Following the design science research paradigm [Wieringa, 2009], formal social network analysis and semantic modeling concepts have been reused to suggest a new perspective on the architecture of an effective fraud detection system.

The remainder of this paper is organized as follows. Section 2 introduces properties, formal analysis of social networks and motivation for their use to address fraud detection issues. Section 3 then demonstrates the design of the OFD framework and the validation of its design within a context of fraud detection from journal entries. In the last section, related works, concluding remarks and their implication for further research on social network analysis were taken up.

2 Social Networks

2.1 What is Social Network Analysis (SNA)?

A social network is a concept referring to a structure made of social actors sharing interests, activities, etc… Joseph Moreno is cited by most research papers on the topic of social network analysis as being the first to introduce methods and tools for a formal analysis. In the 1930s, he was the first one to use all four properties that characterizes SNA at the same time in a study aiming at explaining a spate of girls' runaways: (1) the intuition that links among social actors are important. (2) It is based on data that record social relations that link actors. (3)It draws heavily on graphic imagery to reveal and display the patterning of those links. (4) It develops mathematical and computational models to describe and explain those patterns [Freeman, 2011]. Basically, SNA aims at understanding relationships between the network participants, by means of mapping and measuring. SNA has received increased attention from organizations seeking to understand connection between patterns of interactions. It applies to a wide range of business problems including collaboration in workplaces, team building in post merger configuration, employee's engagement measurement, online reputation, customer intelligence, business strategy, disease contagion, counter terrorism, etc. It was a SNA which led US military to the capture of Saddam Hussein in December 2003 [PSU, 2007]. The tool Inflow for example is credited with contribution to the analysis of terrorist networks surrounding the September 11th events and contact tracing for HIV transmission in a state prison [INFLOW, 2010].

2.2 Formal Social Network Analysis

Whether used for infectious disease spread modeling, professional relations analysis, concentration of resources or power identification, SNA would follow two different approaches. Researchers distinguish between egocentric and socio-centric analysis of networks [Chung et al., 2006]. In the former type of analysis, the focus is made on local structure of networks, i.e. the network around a given node while the latter considers the network as a whole, looking at interactions patterns and the overall network structure by quantifying relationships between people. This distinction would impact the SNA process during data collection and graph visualization.

SNA provides both a visual and a mathematical analysis of relationships between the entities participating to the network. From the visual perspective, social networks are represented as "sociogram" [Scott et al., 2011] or graphs showing actors as nodes that are tied by one or many types of interdependency (values, ideas, visions, sex, friendship, kinship, collaboration, trade, antagonism, etc…). From a mathematical perspective, the social relations datasets translate into a matrix, underlining the visualized graph. This perspective serves at uncovering the graph's theoretic properties (e.g.: number of edges, number of vertices, degree, multiplexity, centrality, density, closeness, betweenness, etc…), supported by metrics computed from the matrix, that help characterizing and even querying the network at hand.

2.3 Why Social Network Analysis for Addressing Insider Fraud Detection?

Social Network Analysis can bring value to occupational fraud detection in at least three ways. First, nowadays organizations are networked (staff, management, customers, suppliers, etc…) and fraud can originate from any part of the network. SNA brings the ability to analyze behaviors and reveal hidden connections that would have not been seen in raw text format.

Secondly, the dynamic nature of fraud makes detection challenging for the traditional rule based algorithms. Fraudsters are constantly adapting to circumvent the existing controls and any new pattern would not be covered by such static algorithms. As people excel at detecting patterns and their judgment when reviewing anomalous activities or transactions very valuable, we believe combining this human ability to computer's capability to iteratively and tirelessly search for defined instances would improve the overall detection process.

Thirdly, SNA can help saving time during manual investigation, which is a necessary step for validating any potential fraud case uncovered by a tool. Traditional computer-aided audit tools are transaction oriented [ACL, 1987], output rows of incriminated transactions without the view of other related transactions performed by the same entities and thus make the manual investigation process labor intensive. With SNA the involved entities and their overall activities is readily available in a graph view for fraud examiners, who in turn are able to quickly visualize false positives and can focus on more risky cases.

3 The Occupational Fraud Detection Framework - OFD

3.1 Framework design

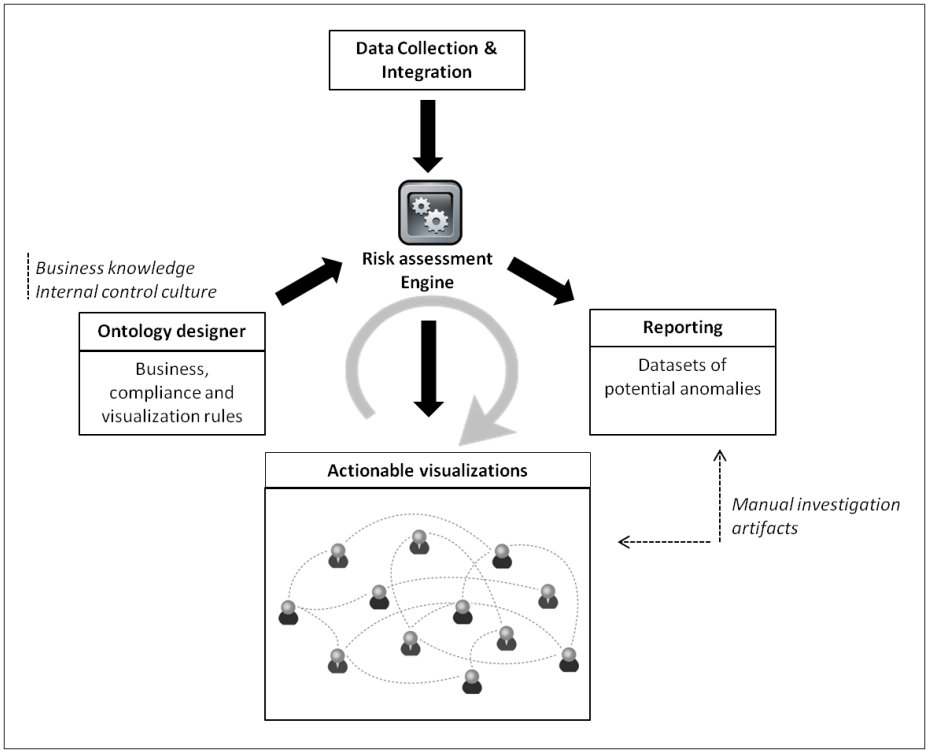

Fig. 1. Overview of the OFD framework

The OFD framework as envisioned in this paper, starts with the selection of a process well integrated with IT, from which historical facts (e.g.: sales, journal entries, purchases, etc…) can be extracted from. OFD is built around four main components:

- The ontology designer, which is the specification component. As fraud examiners are barely proficient in data sciences to maintain robust systems their own and fraud schemes always evolving, the need of a layer of semantic in fraud detection systems architecture is mission critical. Actor types, interaction types and their respective characteristics are to be specified via this component. The way in which they are represented in the final sociogram (shape, color, etc…) is essential and to be defined here as well. Compliance related rules, conflicting interactions, antecedence or association rules are also regarded. The idea behind this component is its complete flexibility so as to enable a fraud examiner to not only control

the rationale behind how fraudulent cases are uncovered, but also the display of the different visualizations available.

- The risk assessment engine is made of a data parser, for ensuring proper integration of raw data collected and a semantic reasoning system. The reasoner is meant to infer logical consequences from the rules specified in the ontology designer. It would analyze the parsed data with both socio-centric and ego-centric perspectives. At one hand, ego-centric analysis will highlight individual interactions which violate the set of rules specified, while at the other hand, socio-centric analysis will enable the identification of internal control deficiencies (e.g.: no segregation of duties) and the detection of fraud not pertaining to a specific transaction, or entity (e.g.: conflicts of interest, management frauds, etc…).
- The reporting component, like what exist today in the industry, would report on cases of violation of the specified rules, by outputting rows of potentially fraudulent transactions.
- The visualization component with its set of actionable sociograms includes a multidimensional social network view, showing several interaction types in the same network, what goes along with the socio-centric perspective mentioned earlier. Drill down and rollup capabilities would help zooming into transactions pertaining to a specific interaction type, or a specific actor of the network (ego-centric analysis). On reduced set of interactions, the time dimension would also be viewable. This component is critical to the overall detection process as through it, fraud examiners would uncover new unforeseen patterns to be specified in the ontology editor, thus paving the way for a continuously improving fraud detection engine.

3.2 OFD framework evaluation by early prototyping

Before jumping to the development of a generic, sound and theoretically grounded tool for supporting the framework introduced earlier, a review of the design has been performed. The aim of such evaluation was threefold:

a. Assess the extent to which graphs can fairly and faithfully represent the diversity of interactions happening between actors of the same or different organizations.
b. Measure the expressiveness of a social network in terms of red flagging of fraudulent interactions or transactions.
c. Gain insights on the perceived complexity by fraud examiners in the use of such visualizations to support fraud detection.

To this end, we ran a case study using accounting journal entries extractions as input for two different organizations of different size. The case study was conducted in collaboration with a population of internal auditors, who have been surveyed on various multidimensional social networks generated from the accounting journal entries (actionable visualization component). The number of auditors involved in the evaluation cannot be revealed in the presence of non disclosure agreement with the

cooperating organization. The R project and the network analysis package "IGRAPH 0.7.1" [Csárdi et al., 2006] were used for scripting raw data parsing, business and design logic. Figure 2 illustrates the overall multidimensional network for one of the entities studied.

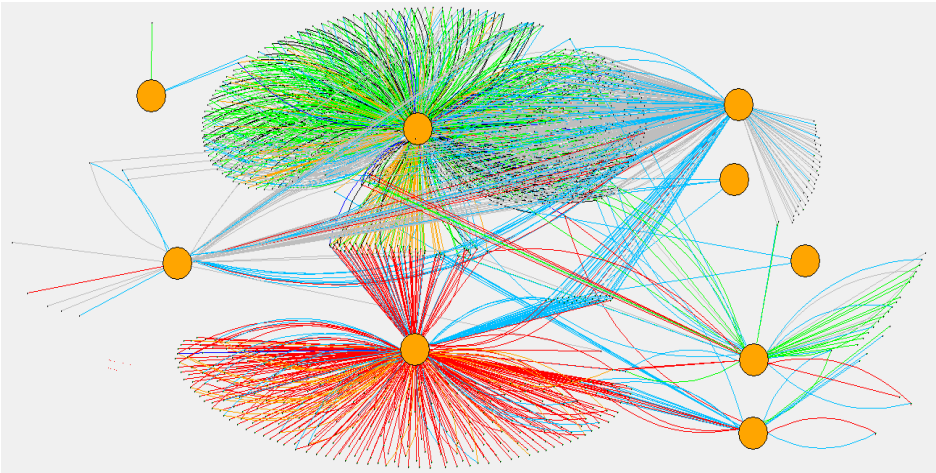

Fig. 1. Global multidimensional social network of accounting journal entries

Each edge in the graph above corresponds to a type of interaction happening between an employee (orange nodes) and a third party (other nodes - customer or supplier in this case). Red edges correspond to outgoing payments, orange ones being purchase invoices, etc… Different drills down or subsets of what is shown in figure 2 have been submitted to the auditors, like the one in figure 3, illustrating supplier only related interactions for the same entity as above.

The key takeaways from this evaluation exercise are as follows:

- Not all journal entries involve a third party (customer, supplier, etc…), what could be perceived as a threat to the validity of our social network oriented approach. Fortunately, such entries (depreciation, amortization, miscellaneous incomes, etc…) are usually subjected to rules which can be reasoned by the risk assessment engine and atypical entries solely highlighted in the reporting engine.
- The manipulation of the proposed visualizations is not that intuitive for auditors, even with a help document attached. Training should not be neglected as 20% of the surveyed auditors perceived the visualization as being too complex, embedding too much information at once. They actually did not provide any further answer to the questionnaire.
- The remaining participants' high level observations or socio-centric conclusions were identical (e.g. non effectiveness of segregation of duties), denoting the good expressiveness of graphs for serving such purpose.
- At the other hand, the ego centric findings were diverse and varied from an auditor to another one, but not contradictory. The variability in the red flags of interest might be explained by the difference in the past experiences of each one of the au-

ditors. They tend to focus their testing procedures on the types of anomalies they expect to come across (what is quite aligned with traditional rule based static detection algorithms); the visualization can help then going beyond that, by expanding the range of possibilities and suggesting further investigation axes.

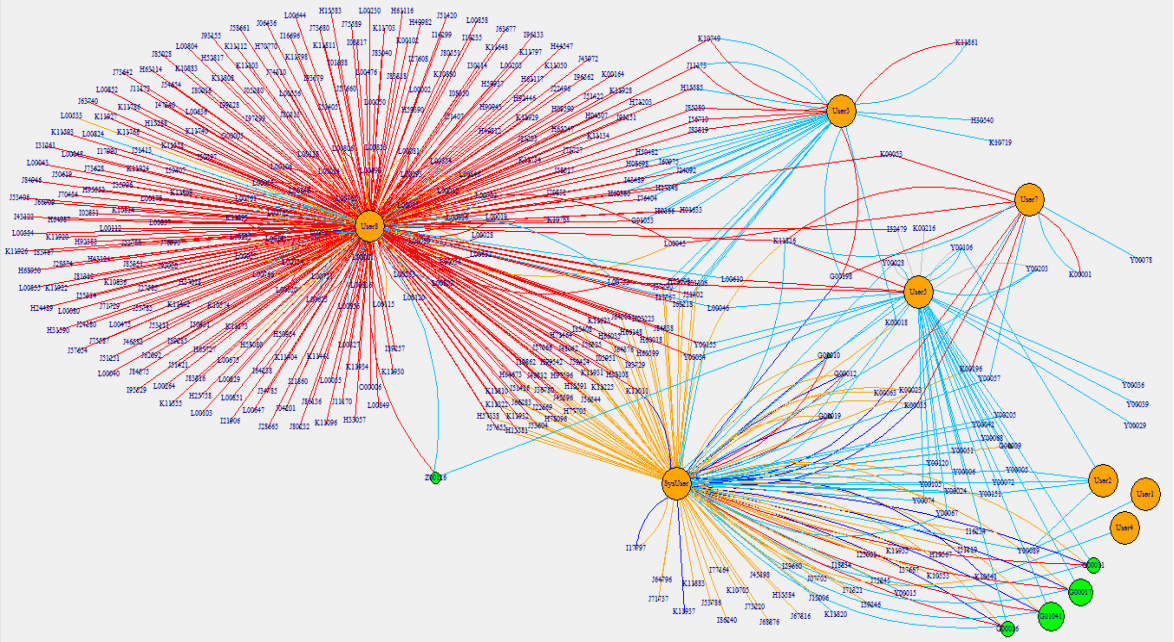

Fig. 2. Drill of the overall social network down to suppliers only related interactions

4 Conclusion and future works

To the best of our knowledge, only few research papers tackled issues in occupational fraud and even fewer integrated visual analytic concepts to their approach. Those last include unsupervised approaches like graph pattern matching techniques [Eberle et al, 2011], with strong focus on structural anomalies identification, but unfortunately forgoing real world business specificities and rules, what leads to a high rate of false positives and complex maintenance by end users. Other approaches like [Luell, 2010] or [Argyriou et al, 2013], rely on innovative but tailor made visualizations which cannot be applied to other business processes. The framework presented in this paper extends existing data mining techniques used for occupational fraud detection by offering not only visualizations to be used by auditors to uncover new fraud patterns, but also semantic reasoning capabilities for integrating those new patterns to the fraud detection engine. The targeted architecture is then scalable and extensible provided the only maintenance of specified ontologies. Our assessment of the serviceability of the sociograms on accounting journal entries delivered promising

results and future directions for this research will be towards the design and the evaluation of a full prototype for supporting the framework. The generic nature of the framework presented herein and its network oriented approach also open perspectives for investigation beyond the scope of occupational fraud detection. Cyber criminality in an environment where information systems are more and more interoperable may also be investigated following a likely approach.

References

- [ACFE, 2014] ACFE 2014 Report to the nations on occupational fraud and abuse http://www.acfe.com/uploadedFiles/ACFE_Website/Content/documents/2004RttN.pdf
- [ACL, 1987] www.acl.com
- [Argyriou et al, 2013] Evmorfia N. Argyriou, Aikaterini A. Sotiraki, Antonios Symvonis. Occupational Fraud Detection Through Visualization, In Proc. of the 11th IEEE Intelligence and Security Informatics (ISI 2013), pages 4-7, 2013.
- [Chung et al., 2006] Kenneth K Chung, Liquat Hossain, Joseph Davis. Exploring sociocentric and egocentric approaches for social network analysis. KMAP 2005: Second International Conference on Knowledge Management in Asia Pacific (pp. 1-8). New Zealand: Victoria University of Wellington.
- [Cox et al., 1997] Kenneth C. Cox, Stephen G. Eick, Graham J. Wills, Ronald J. Brachman. Visual data mining: Recognizing telephone calling fraud. Data Mining and Knowledge Discovery 1997, Volume 1, Issue 2, pp 225-231.
- [Eberle et al, 2011] William Eberle - PhD, Jeffrey Graves. Insider Threat Detection Using a Graph-Based Approach, Journal of Applied Security Research, 6:32–81, 2011
- [Freeman, 2011] Linton C. Freeman. The development of social network analysis - with an emphasis on recent events. The SAGE Handbook of Social Network Analysis, SAGE Publications Ltd.
- [Csárdi et al., 2006] Gábor Csárdi, Tamás Nepusz: The igraph software package for complex network research. InterJournal Complex Systems, 1695, 2006.
- [INFLOW, 2010] http://www.orgnet.com/cases.html
- [Luell, 2010] "Employee fraud detection under real world conditions," Ph.D. dissertation, 2010. [Online]. Available: http://www.zora.uzh.ch/44863/
- [Phua et al., 2004] Clifton Phua, Vincent Lee, Kate Smith & Ross Gayler. A comprehensive Survey of Data Mining-based Fraud Detection Research. Arxiv preprint arXiv: 1009.6119.
- [PSU, 2007] https://courseware.e-education.psu.edu/courses/bootcamp/lo09/08.html
- [PWC, 2014] PriceWaterCoopers. 2014 Global economic crime survey. La fraude continue à être une vraie menace pour les entreprises.
- [Scott et al., 2011] John Scott, Peter J. Carrington. The SAGE handbook of social network analysis. SAGE Publications Ltd.
- [Wheeler et al, 2000] Richard Wheeler, Stuart Aitken. Multiple algorithms for fraud detection. Knowledge-Based Systems 13(3): 93-99.
- [Wieringa, 2009] Roel Wieringa. Design science as nested problem solving. Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology (DESRIST '09), Philadelphia, Pennsylvania, USA.
- [Yufeng et al., 2004] Yufeng Kou, Chang-Tien Lu, Sirirat Sirwongwattana, Yo-Ping Huang. Survey of fraud detection techniques. Proceedings of the 2004 IEEE. International Conference of Networking, Sensing & Control. Taipei, Taiwan, March 21-23, 2004.