Network Analysis for Financial Fraud Detection

Roger Almeida Leite[1], Theresia Gschwandtner[1], Silvia Miksch[1], Erich Gstrein[2] & Johannes Kuntner[2]

[1]Vienna University of Technology (TU Wien), Austria
[2]Erste Group IT International, Austria

Abstract

Security and quality are main concerns for private and public financial institutions. Data mining techniques based on the profiles of customers of a financial institution are commonly used to avoid fraud and financial damage. However, these approaches often are limited to the analysis of individual customers which hinders the detection of fraudulent networks. We propose a Visual Analytics approach for supporting and fine-tuning customers' network analysis, thus, reducing false-negative alarms of frauds.

Categories and Subject Descriptors (according to ACM CCS): Human-Centered Computing, Visual Analytics, Information Visualization, Time Series Data, Business and Finance Visualization, Financial Fraud Detection, Financial Fraud Analysis.

1. Introduction

Detecting fraudulent events is an important task in several domains such as insurance companies, credit card companies, public sector, and banks. Automatic approaches for fraudulent event detection are often used in order to reduce the amount of false-positive and false-negative alarms. However, this type of automatic system needs to be constantly administrated and updated to ensure good detection rates and quality. In this work we focus on financial fraud detection (FFD) for bank transaction data (unauthorized transactions, money laundering, and others). This data contains time-oriented and multivariate features, which are of complex nature [AMST11] and demand appropriate visualization and exploration means.

Artificial Intelligence (AI) techniques and fraud detection metrics commonly look for patterns and/or outliers in the financial transaction domain. One of the main drawbacks of applying AI techniques only is that constantly changing strategies and behavior adaptation of the fraud creators might not be detected. To cope with such situations, it is important to (1) always balance and adapt the parameters of the algorithms in order to identify frauds but not to overload too many alarms and (2) constantly reason about the algorithms' results. In real world, false-positive alarms might lead to the accusation of innocent people. On the other hand, false-negative alarms mean that a fraudster succeeds. However, the fine-tuning of both is usually coupled. More sensitive algorithms lead to the detection of more frauds but also more false-positive alarms will be generated. To this end, the calibration of their sensitiveness is essential.

Another challenge on FFD is detecting fraudulent events through a network. It is known that analyzing all individuals' relations using all possible levels will result in an increase of the algorithm's complexity, which is not always traceable. From our close collaboration with a financial company and in a literature study, we could identify important tasks within this research field. Our contributions are: (1) enumerate the challenges of fraud detection focusing on customer network analysis, (2) the integration of a Visual Analytics (VA) loop into the network analysis process, and (3) the prototypical implementation of a VA approach for the investigation of suspicious behaviour and fine-tuning of automatic alert systems.

2. Related Work

One of the first contributions combining fraud detection and visual analysis was investigated by Kirkland, et al [KSH*99]. They propose the combination of AI, visualization, pattern recognition, and data mining to support alerts (pattern detection) and exploration. WireVis [CGK*07] was the first approach, which explores FFD and network analysis. In the WireVis's approach big amounts of transaction data are visually explored using a multiple-coordinated view visualizations to identify fraudulent cases through transaction keywords' investigations within transactions. This approach aims to depict relationships among accounts and keywords over time. Huang, et al. [HLN09] presented a VA framework for stock market security. One of their main goals was reductions of false-positive alarms by applying traditional AI techniques. From their visual design, they combined a 3D tree map with a node-link diagram. In EVA [LGM*18] we presented the integration of a VA step to the current "detection and decision" workflow. EVA combined automatic methods with well-known visualization techniques, which our domain experts are mostly familiar with.

According to our close collaboration with FFD experts and our literature study the following challenges can be derived: (a) false-positive and false-negative alarm reduction; (b) development of a comprehensive VA design for network analysis; (c) enhancement

© 2018 The Author(s)
Eurographics Proceedings © 2018 The Eurographics Association.

DOI: 10.2312/eurp.20181120

delivered by EUROGRAPHICS DIGITAL LIBRARY www.eg.org diglib.eg.org

of the scalability for network monitoring; (d) knowledge base construction and customer transactions behavior classification in order to support further fraud identifications; and (e) support for different types of frauds.

3. Conceptual Design

We design our interactive VA approach with respect to the data, users, and tasks [MA14].

Data: Financial transaction events.
Users: Analysts from financial institutions that monitor, investigate, and validate transactions and alert systems.
Tasks: The overall task is fraud detection by means of network analysis based on a profile scoring system. This task includes fine-tuning of automatic alert algorithms as well as managing the trade-off between the sensitivity of the approach and the reduction of false-negative and false-positive alarms.

The most common approach for AI techniques in this domain is to create, update, and explore account profiles when a new transaction happens. The transaction is compared to known profiles and afterwards categorized as suspicious or not [CCM*14]. Many parameters have an influence on the suspectedness score. Constantly fine-tuning the algorithms is one of the main challenges that, when done wrong, can result in opportunities for fraudsters.

Financial frauds can be divided into different classes [LGM*18]. Examples of such classes are 'unauthorized transaction', 'money laundry', and 'fake user'. The detection of different types of frauds requires different measures, metrics, and parameters. Available solutions are often focused on one of the fraud types only. More extensive amount of works exists on the topic of 'unauthorized transaction'. However, just a few approaches cover fraud types that involve network analysis (e.g., 'money laundry' and 'straw person') due to its analysis scaling and complexity challenges. Even without considering loops, finding all paths between two nodes would have a computational complexity of O(v+e) [RR96].

We used a state of the art method for fraudulent event detection based on scores [LGM*18, CCM*14], to flag accounts as 'fraudulent receivers' and 'fraudulent senders'. Based on this initial classification, we created four classifications in order to clarify network behaviors. **Couples**, when a fraudulent sender has only a fraudulent receiver and the fraudulent receiver has non-alarmed connections besides the target sender. **Sender Flowers** and **Receiver Flowers**, when one sender/receiver has more than one receiver/sender but does not present fraudulent accounts in the second relationship layer. **Complex**, when a receiver or sender node present two or more relationship layers.

With these network behavior classification sets, we analyzed an anonymized real world dataset with approx. 20.000 accounts (provided by our collaborating FFD experts) in order to query for patterns. We plotted the amount of accounts that matched each category in a horizontal bar chart aiming for a better comparison between accounts that are detected to be suspicious by automatic means and non-suspicious accounts (see Figure 1 A). In our plots, we do not plot non-suspicious accounts that match the 'complex' classification due to its massive number and low risk. In Figure 1 B, we represent the processed accounts in an interactive node-linked diagram that allows brush selection, drag, zoom-in and zoom-out. By having both visual representations we could find some interesting insights. (i) We could observe non-suspicious accounts 'in between' two suspicious accounts, which is a strong indicator for further investigation on that account. This could be a 'money laundry' scheme. (ii) We could identify the behavior of potential 'fake account creations'. This could fit in a 'straw person scheme'. (iii) Couples have a higher chance of being fraudulent as well as Receiver Flowers. (iv) Sender Flowers as well as Complex have a higher chance of not being fraudulent. We propose the usage of well established FFD solutions for supporting fraudster network detection. Information such as (iii) and (iv), when confirmed, could be used for the fine-tuning process of the detection algorithm.

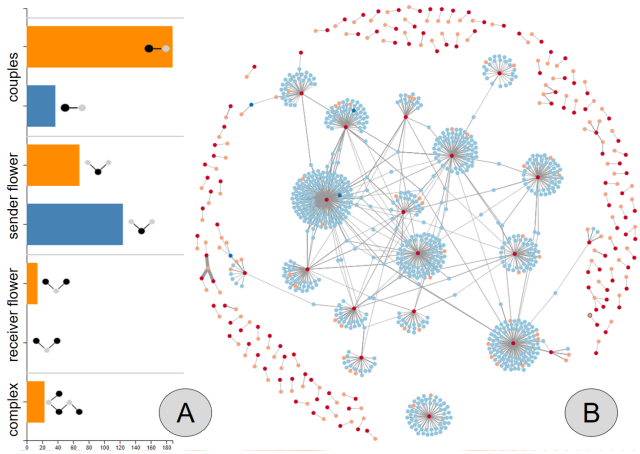

Figure 1: *(A) horizontal bar chart shows in orange the amount of suspicious accounts and the variance of network behavior. In (B) we represent the network by a interactive node-linked diagram where orange nodes represent suspicious receivers accounts, red nodes represent suspicious senders accounts, light blue nodes initially non-suspicious receivers, and dark blue nodes initially non-suspicious senders.*

4. Conclusion and Further Work

In this work we compile the main challenges of the detection of fraudulent networks. Based on the found challenges we propose a score based VA approach to support intra-network relationship analysis. Aiming to support decision making, we integrate this approach into the fraud detection workflow used by our collaborators, combining automatic techniques and visual reasoning. Using anonymized real data, we could already find insights from fraudster cases and map them to the proposed behavior classification. We believe that similar domains, such as malware detection or tax usage analysis, can also benefit from applying our approach.

5. Acknowledgements

This work was supported by the Austrian Federal Ministry of Science, Research, and Economy via CVAST, a Laura Bassi Centre of Excellence (No. 822746).

© 2018 The Author(s)
Eurographics Proceedings © 2018 The Eurographics Association.

References

[AMST11] AIGNER W., MIKSCH S., SCHUMANN H., TOMINSKI C.: *Visualization of time-oriented data*. Springer Science & Business Media, 2011. 1

[CCM*14] CARMINATI M., CARON R., MAGGI F., EPIFANI I., ZANERO S.: Banksealer: an online banking fraud analysis and decision support system. In *ICT Systems Security and Privacy Protection*. Springer, 2014, pp. 380–394. 2

[CGK*07] CHANG R., GHONIEM M., KOSARA R., RIBARSKY W., YANG J., SUMA E., ZIEMKIEWICZ C., KERN D., SUDJIANTO A.: Wirevis: Visualization of categorical, time-varying data from financial transactions. In *Visual Analytics Science and Technology. VAST. IEEE Symposium on* (2007), IEEE, pp. 155–162. 1

[HLN09] HUANG M. L., LIANG J., NGUYEN Q. V.: A visualization approach for frauds detection in financial market. In *Information Visualisation, 13th International Conference* (2009), IEEE, pp. 197–202. 1

[KSH*99] KIRKLAND J. D., SENATOR T. E., HAYDEN J. J., DYBALA T., GOLDBERG H. G., SHYR P.: The nasd regulation advanced-detection system (ads). *AI Magazine 20*, 1 (1999), 55. 1

[LGM*18] LEITE R. A., GSCHWANDTNER T., MIKSCH S., KRIGLSTEIN S., POHL M., GSTREIN E., KUNTNER J.: Eva: Visual analytics to identify fraudulent events. *IEEE transactions on visualization and computer graphics 24*, 1 (2018), 330–339. 1, 2

[MA14] MIKSCH S., AIGNER W.: A matter of time: Applying a data–users–tasks design triangle to visual analytics of time-oriented data. *Computers & Graphics 38* (2014), 286–290. 2

[RR96] RAMALINGAM G., REPS T.: On the computational complexity of dynamic graph problems. *Theoretical Computer Science 158*, 1-2 (1996), 233–277. 2

© 2018 The Author(s)
Eurographics Proceedings © 2018 The Eurographics Association.