Graph Mining for Detection of a Large Class of Financial Crimes

Czeslaw Jedrzejek[1], Maciej Falkowski[1], Jaroslaw Bak[1]

Poznan University of Technology,
Institute of Control and Information Engineering,
Curie Sq. 5, 60-965 Poznan, Poland
{Czeslaw.Jedrzejek, Maciej.Falkowski, Jaroslaw.Bak}@put.poznan.pl

Abstract. Financial crime perpetrators use many different and sophisticated types of schemes, techniques and transactions to accomplish their goals. However, for a large class of financial crimes, such as doing harm to a company, they cannot escape a powerful principle: illegal proceeds have to return to or be under control of managers to achieve a personal gain. This circular flow of transaction attributes is characteristic of another type of financial crimes: such as a VAT carousel or a Polish fuel mafia scheme. In this work we propose a minimal model of descriptions of a doing harm to a company crime, combined with money laundering. Such a model uses sufficient ontology to build evidence and assign legal qualifications to criminal activities and nothing more. The scheme can be described by using 8 layers of concepts and relations that follow in logical order of uncovering a crime. For example, on the first level that describes money transfers there are only 6 parameters necessary assuming that certain operations can be grouped. Using conceptual graphs with subsumption and negation operations, one can reason on people involvement in a crime and choose between strategies of building a case. The model captures over 90% of relevant information for a typical use case of issuing fictitious invoices, the so called Hydra case and is able effectively reason over relevant facts, which means that legal qualification for this case is basically correct. To what extent the model can be generalized to more complex schemes will be a subject of a further study.

1 Introduction

1.1 Financial Crimes

Money laundering (ML) as a criminal activity accompanying most serious crimes is seriously undermining economic and social order. There are many schemes of complex nature that are recognized by FATF [5] and new schemes appear exploiting ever increasing richness of money forms and economic activity. Altogether currently up to possibly 50 basic schemes could be identified. Trade-based money laundering is defined as "the process of disguising the proceeds of crime and moving value through the use of trade transactions in an attempt to legitimize their illicit origins". For money laundering [11] FATF developed over the

years indicators [5-7], that are frequently observed signs of suspicious activity. Representation of facts is important for a financial crime description in terms of uncovering mechanisms (modus operandi) and data collection for building evidence. In this work we propose focusing analysis of a large class of financial crimes on looking for chain graphs representing flows of money, invoices and goods/services. We demonstrate on a few examples that such a representation largely facilitates asking relevant questions on connections between financial entities and people associated with them, which is conducive to evidence building and a crime qualification. To study these relations we propose a data model (called a minimal model), based on conceptual graphs [13]. This means that an ontology is crafted to a task rather that attempting to describe whole conceivable space of concepts and relations (top ontologies). The methodology consist of several steps:

1. Design of a hierarchical data representation with minimal ontology, constructed in sequence of uncovering of a crime scheme. In the first stage, goods/services transfer data is analyzed with relation to 3 basic flows: money, invoices, and documents (i.e. confirming that the service or goods have been delivered - particularly important for a fuel mafia type of crimes). In addition, responsible or relevant people within companies are associated with particular illegal activities..
2. Construction of a multigraph of the data flow and looking for cycles or potential cycles when a chain of transactions strongly indicates a "closure path".
3. Provision of a framework in which the graph building process and queries are executed
4. Relating answers to queries with crime qualifications.

This approach is preliminary and limited, but provides an essential model for evidence building of a very important class of financial crimes: among them money laundering and acting to a harm to a company. Our approach is an extension of an approach used by Badia and Kantardzic [1] to analysis of the Enron [4] dataset.

1.2 Data and Red Flags Used in Financial Crime Investigations

The analysis begins with information from banks on suspicious operations - symptoms of financial crimes (red flags) - that are transferred to countries' Financial Intelligence Units (FIUs). The common red flags detected by financial institutions and designated non-financial businesses and professions using data mining of transaction records are:

- Unconventionally large currency transactions, particularly in exchange for negotiable instruments or for the direct purchase of fund transfer services;
- Apparent structuring of transactions to avoid identification requirements or regulatory record-keeping and reporting thresholds;

– Introduction of a client by an overseas associate or financial institution based in a country or jurisdiction known for drug trafficking and production, other financial crimes and "bank secrecy".

FIUs employ data mining techniques of suspicious activity patterns (over 750 in case of the Polish FIU - GIIF). These employ rule-based systems, customer profiling, and statistical techniques [9]. Other FATF indicators require financial audits, analysis of tax information, testimonies and more difficult to obtain or correlate information, such as:

– Transactions that appear inconsistent with a client's known legitimate (business or personal) activities or means; unusual deviations from normal account and transaction patterns;
– Any situation where personal identity is difficult to determine;
– Unauthorized or improperly recorded transactions; inadequate audit trails;
– Transactions passed through intermediaries for no apparent business reason.

2 General Crime Model for Selected Crimes

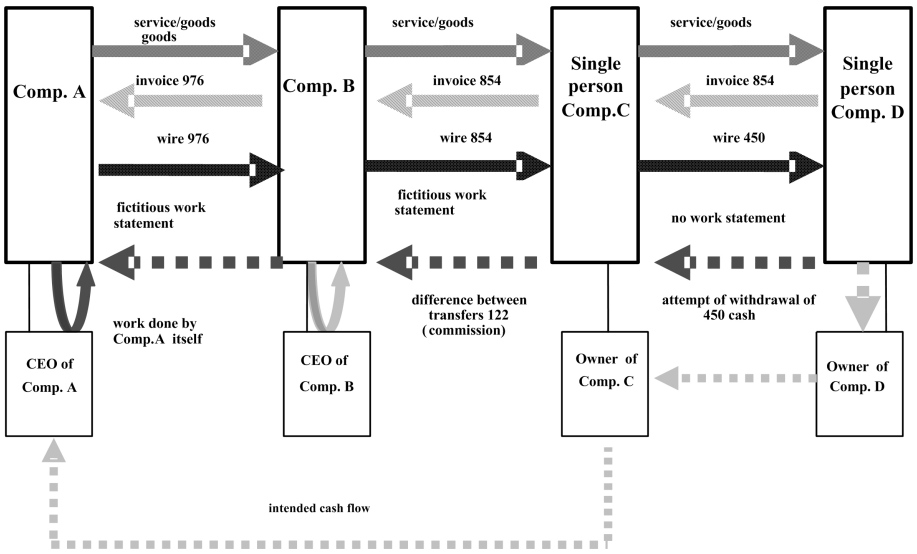

Fig. 1. A basic schema of transaction flows for a real case of acting to harm a company

We define concepts only to the level necessary to ask questions and reason on the main mechanisms of financial crimes. We assume from the start that we capture 80-95% of relevant knowledge - but this type of knowledge will be universal. In some cases deciding whether a work has been done or not is left to

a human. Although less ambitious, such a procedure opens a provision of using an opinion of an expert (i.e. expert accountant). The class of crimes we consider is presented in Fig. 1 (modeled after a real case, the so called Hydra case with real data).The fraudulent scheme is based on invoicing work not done.

The CEO of company A subcontracts construction work which the company A does itself. The work is then consecutively subcontracted though a chain of phony companies. Each of them is getting a commission for money laundering and commits that contracted work has been done with fictitious statements. At the end of a chain an owner of a single person company D attempts to withdraw cash, and there is a suspicion that this cash reaches "under the table" the management of company A (possibly through a trusted intermediary, here associated with company C). People associated with company D may not know full details of a case. In this scenario money leaves companies and goes to physical persons.

The knowledge base (KB) is modeled at several (here 8) levels to describe a mechanism of crime, that are presented in Section 3.

The concepts come in a hierarchy according to a prevailing mode of gathering data.

3 Hierarchies of Submodels

The facts for a case we analyze (the Hydra case) can be presented in 7 pages of a natural language text. This case is one of the easiest for which we have data (only 4 companies and 10 people involved) but contains all elements of the financial crime mechanisms. The full analysis is beyond the scope of the paper, however, in the minimal model the first 3 layers are described by the following top level concepts:

General concepts related to layers 1-3:

- General Flow (From Entity, To Entity, Date, Flow),
- Flow (Money Transfer, Invoice, Work/Service).

Specifically layer 1 concepts are modeled as follows:

- MoneyTransfer (From Entity, To Entity, Method of Transfer, Date, Value, Title of transfer, i.e. for Work, registered/unregistered)
- MethodOfTransfer (Electronic transfer, Cash)
- ElectronicTransfer(Account1, Account2)
- Account (Time of setting, History)

The unregistered transfer could be cash or a transfer to tax heavens banks. We model the remaining layers in a similar way (to be presented in future publications). At each level we assign concerns, and type of analysis. Answers to questions infer assignment of relevant penalty as sanctioned by articles of the Polish Penal Code ("Kodeks Karny" in Polish).

Table 1: Hierarchy of models representing data involved in a criminal activity (sanctions are presented in bold print)

Level of Analysis	Concern type	Type of Analysis/ Tools/ Visualization	What can be deduced?	Potential legal qualifications
1	Levels Levels 1-3 - companies, institutions, single person companies. Levels 1-2 ML - type of a scheme	Money transfer between companies. Tool for company transaction relations (selecting all relevant attributes for case of Company A).	Abnormal transaction pattern for a chain of companies.	Suspicion of a crime sanctioned by **art. 299** of the Polish penal code (abbreviation kk), that is money laundering and unknown base crime (they always come together).
2		Invoice flow between companies. Tax statements. Tool for invoice analysis and tax statement analysis.	Abnormal company operation: a) establishing correspondence of invoices and money transfers b) tax statements values much smaller than real turnover c) single-transaction companies d) detection if invoices are for the same goods/services	Reinforcement of suspicion of **art. 299** crimes and unknown base crime.
3		Work/Services flow Activities: Inspection of work (in case of Company A inspection of site). Question: X; Value of work as invoiced Y: Market value of work Event calculus tool.	Establishing whether work has been done: a) find a chain of companies involved in operations for the same goods/services. Conditions: b) investigate whether goods were sold/work performed; if no c) find the last company in a chain that did not do a work. If yes, all companies in the chain implicated. d) alternate procedure: find the first company in a chain that ordered work.	All companies involved in chain transactions that is a part of a scheme. Initial legal qualification: fraud, **art. 286 1 kk** [10] in connection with money laundering Later after testimonies of accomplices **art. 296 1** and 2 and 3 kk and **art. 284 2 kk** - the legal qualification changes to do-harm-to-a- company crime (at significant or large degree) and acting for a personal gain.

Level of Analysis	Concern type	Type of Analysis/ Tools/ Visualization	What can be deduced?	Potential legal qualifications
4	Assign roles to people in potentially implicated companies (in some cases scheme organizers can act and not be directly associated with transacting companies): a) logical decision for ordering work; if authorized person did not order nor accept; nor authorized payment the work find who forged the document b) who physically did invoices, money transfers c) who accepted work in chain of command	Tool for associating people with companies. Investigate documents stating that the work was done. Roles: management, proxies, legal advisors. Administration: executing activities ordered by superiors.	Who are the persons that were responsible for false documents?	
5	Mapping potential roles coming from positions in companies to a contribution to a financial crime. Correlate potentially implicated people, and testimonies of people, or their role as indicated by documents.	In the Hydra case, suspect's testimonies accused of crime related to art. 286 1 kk implicated CEO of the company A as a mastermind of a scheme. Change of qualification of CEOs of companies A and B from **art. 286 1 kk** to **art. 296 1 and 2, and 3 kk (using art. 212 kk)**	Did accused people worked conscientiously? Did they work together? (constituted a crime group)? Can this be only a matter of negligence?	Making false statements: Sanctioned by **art. 271 1 and 3 kk**.
6	People not related to companies but being a part of crime. Other relations of people.			Sanctioned by **art. 18 1 kk in relation to art. 299 1 and 5 kk using art. 12 kk**.
7	Information about people; whether sentenced in the last 10 years, criminal connection; school business etc.	Discovering invoice of company A waiting to repeating execution of a scheme (after completing the first cycle).		Detailing legal qualification. Owner of company D got an increase of sentenced using **art. 64 1 kk**.
8	Additional factors	An invoice waiting for payment. That is why initially Company C transferred only half of money to Company D testing whether a withdrawal will be blocked.		**Art. 12.** Repeated crime (counts as one). **Art. 13. 1** - intent to commit crime.

4 Money Transfer Flow

In this Section we show details of modeling and data querying for layers 1-3. The source of data is twofold: structured documents, eg. Excel or Calc spreadsheets containing flows data and additional, manually entered data about persons and companies. That data is stored in relational database and is further converted to Conceptual Graphs relations.[13, 14]. We use elements of graph theory to effectively process and query details of financial data. Our goal is to find suspicious behavior, described by patterns (discussed later). There are three basic parts in our system: a knowledge base, a set of rules and a set of patterns. The knowledge base contains facts about companies: their cash flow, invoice flow, work or goods flow and related physical persons. Facts are represented by entity nodes connected with relation nodes. Fig. 2 contains visualization of a fragment of knowledge base. Formally, we can describe Fig. 2 with a formula:

KB = MoneyTransfer(CompanyA, CompanyB, 01.02.2007, "Invoice no 18/07", 500000) Nodes can be typed such as Company, Transfer, Person and other. This

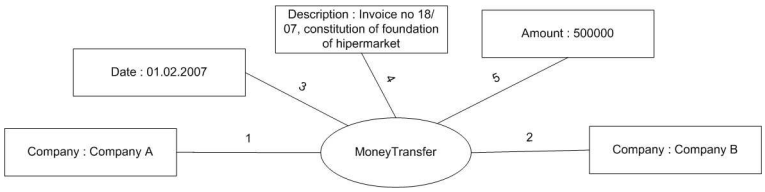

Fig. 2. A fragment of KB

kind of data description is an extension of conceptual graphs [13] that preserves its soundness and decidability. We use simple query and ontological constructs (subsumption and negation) that are needed for our data analysis and computing crime qualifications. Knowledge base can contain thousands of facts about transactions and for our analysis the crucial thing is the transaction flow. Such flows can be described by rules of form as in Fig. 3. The main part of a rule is a pattern, which can contain variables. Pattern can be complemented with additional variable constraints (date d1 is earlier than d2). Rule in Fig. 3 states that if a company transfers money to another company, and that company transfers later the same amount of money to a third company, then the middle company acts as a broker and in fact the first company transfers money to the third company (minus a provision). This pattern can be easily extended to include more intermediary companies, as shown in Fig. 4. Application of such rules to the KB produces a web of transfers and shows the real starting and ending points of a money flow. Presented transitive rules can be set to include only transfers above minimum value or within a period of time. For the sake of clarity we used simple less and equal operators to compare dates and values, but in fact they can be far more complex. We developed sophisticated algorithms that can aggregate small

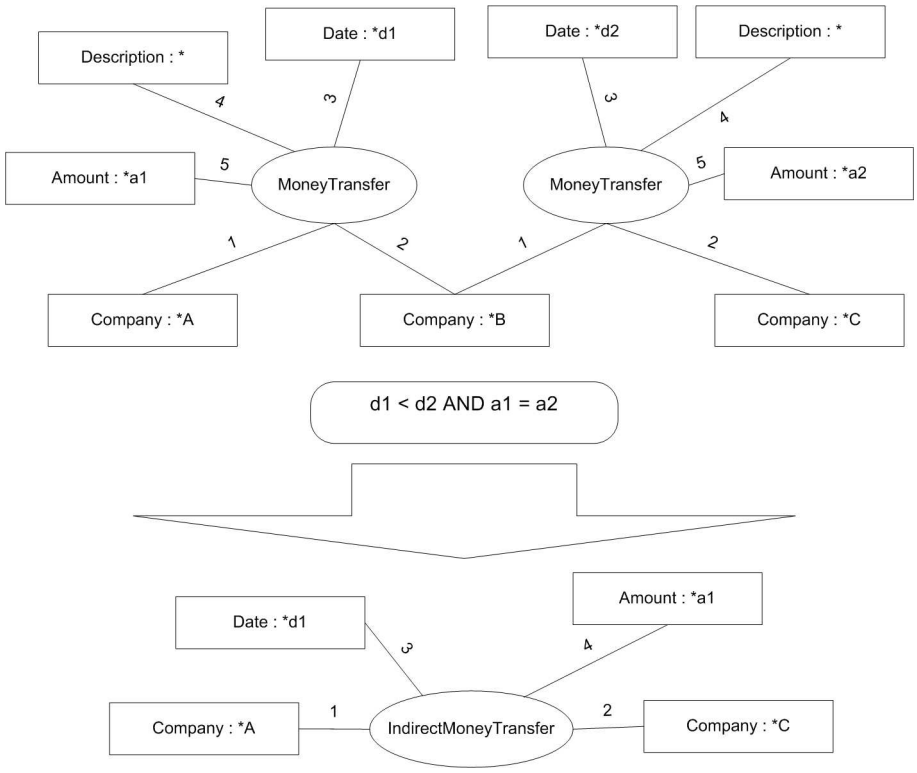

Fig. 3. Transfer chain rule

transfers (technique often used to disguise real money flow), not discussed here. One of patterns that exploits such a transitive use of broker companies is money laundering where additional physical persons and unregistered money transfer are involved. Such a pattern is shown in Fig. 5: Transfer from one person to another can also be more complex and involve intermediary persons.

After applying specified rules to the KB, the crime detection phase takes place. In this phase we try to find homomorphism between data collected in KB and defined crime patterns. Our computational machinery searches for crime symptoms described by a pattern (like VAT fraud or money laundering). Data sets that are closest to match the pattern are presented to a user. At this level the user can judge probability of crime and provide additional data, such as bank transfers and invoices from other companies that look suspicious, or data from expert analysis (e.g. that work or service was fictitious). System works in cycles: data input, analysis, propositions of directions of further investigation. After gathering enough information additional functionality is triggered, that is a crime qualification for humans involved. The qualifications are proposed based on roles persons play in a discovered crime schema.

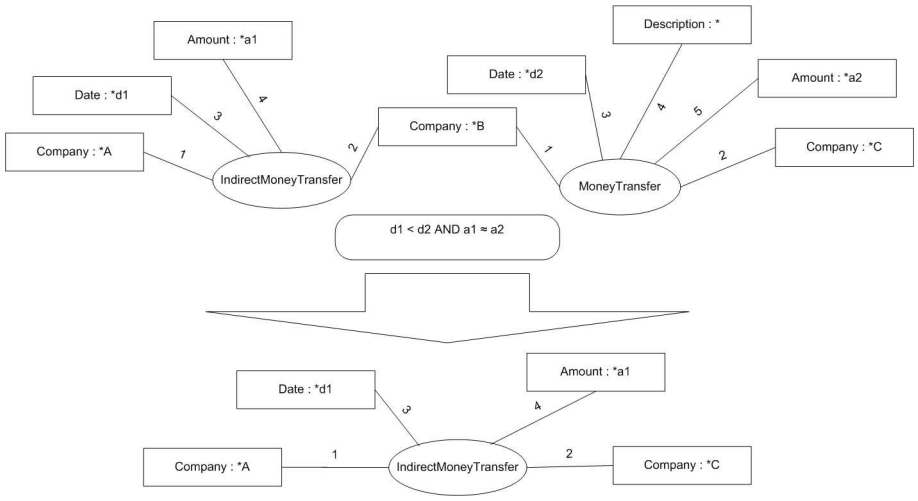

Fig. 4. Transitive transfer chain rule

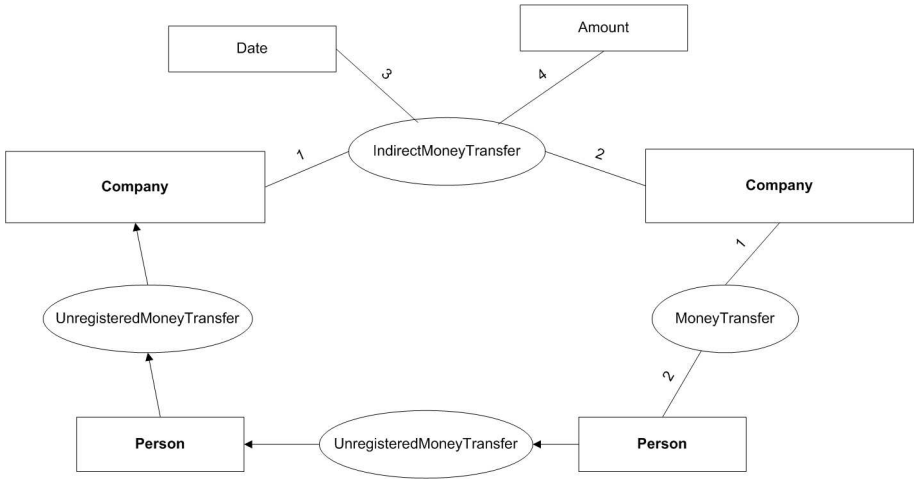

Fig. 5. Crime flow pattern

5 Conclusions

In this paper we have presented a computational background and an use case of our tool that is still being developed. Although it is not fully completed yet, we have obtained first promising practical results. The success is due to several reasons:

1. Guidance of legal experts as to legal procedures and crime schemes [12].
2. Existence of an invariant in the scheme (the cycle of money transfers).

3. The minimal model description.

The first feature bootstraps relevant questions. The second one introduces clarity to a description, at a price of some decrease of power of expression. The problem with complete ontologies is that a formulation relevant questions becomes very difficult [2, 3]. Take an invoice, for example. Company A issues an invoice, and company B receives an invoice. These operations are described as 2 RDF triples. We assume that both activities are coupled, so "invoices" becomes a predicate, A invoices B. This is the only concept related to invoice. If invoice is lost or destroyed and it is important for understanding a scheme our model cannot use such information. We are working on implementation of the rest of analysis levels and a few more financial crime schemas. One of the most desired schema that we are developing and that our tool aims to help with, is the VAT fraud schema. Our practice shows that the current implementation can handle up to with 50 thousand of VAT invoices on a standard PC machine. Currently we are working on optimization of querying algorithms, which we applied.

Acknowledgments. Our research is supported by the Polish Ministry of Science and Higher Education, Polish Technological Security Platform grant 0014/R/2/T00/06/02. We thank drs. J. Cybulka and Jacek Martinek; and Mr. Z. Wieckowski and Mr T. Dela for guiding information and discussions.

References

1. Badia A., M. Kantardzic M. M.: Link Analysis Tools for Intelligence and Counterterrorism. ISI 2005, pp. 49-59
2. Cybulka J., Martinek J.: The core ontology of criminal processes and investigation procedures. Submitted for publication, 2008.
3. Cybulka J., Jdrzejek C., Martinek J., Police Investigation Management System Based on the Workflow Technology, Chapter in: Legal Knowledge and Information Systems, Frontiers Artificial Intelligence and Applications, vol. 189, IOS Press, Amsterdam, Berlin, Oxford, Tokyo, Washington DC, 2008, pp. 150-159
4. Enron data set. available at http://www-2.cs.cmu.edu/ Enron
5. FATF (The Financial Action Task Force) Report: Trade Based Money Laundering, 23 July 2006; http://www.fatf-gafi.org/
6. FATF (The Financial Action Task Force) Report: Complex Money Laundering Techniques: A Regional View, 23 February 2007; http://www.fatf-gafi.org/
7. FATF (The Financial Action Task Force) Report: Laundering the Proceeds of VAT Carousel Fraud, 23 February 2007; http://www.fatf-gafi.org/
8. Jedrzejek C., Martinek J.: On the modelling of money laundering techniques as courses of events. Proceedings of 3rd Language & Technology Conference, October 5-7, 2007, Pozna, Poland, pp. 544-548.
9. Mena, J. (2003) Investigative Data Mining for Security and Criminal Detection, Butterworth Heinemann, 2003.
10. The Penal Code (in Polish). Ustawa z dnia 6 czerwca 1997 r. Kodeks karny.
11. Unger B., Busuioc, E. M.: The Scale and Impacts of Money Laundering, Publisher:Edward Elgar, May 2007

12. Wieckowski J.: Regional Prosecutors Bureau in Katowice, Poland, private communication, February-May 2008.
13. Chein M., Mugnier M., Graph-based Knowledge Representation Computational Foundations of Conceptual Graphs, Springer London 2008
14. Allemang D., Hendler J.: Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL, Morgan Kaufmann Publishers 2008