OntoFiC : an ontology for financial fraud detection and customer behavior modeling.

1st Lylia Abrouk
LIB- University of Burgundy
Email: Lylia.Abrouk@u-bourgogne.fr

2nd Hamza Chergui
LIB - SKAIZen Group
Email: hchergui@skaizengroup.fr

3rd Hamid Ahaggach
LIB- University of Burgundy
Email: hamid.ahaggach@u-bourgogne.fr

***Abstract*—Fraud detection is a complex issue for financial institutions. They must have tools for the prevention and detection of fraud. In this article, we present our approach to detect fraudulent transactions in SWIFT network based on the domain ontology. Firstly, we present the OntoFiC ontology constructed for the modeling of SWIFT transactions and actors. This ontology is populated with a real dataset. We developed our rules-based approach with rules associated to fraud scenarios to label our transactions as legitimate or fraudulent. Finally, we made SPARQL requests to visualize these transactions through graphs. Our work is part of a collaboration project with a financial company, SKAIZen Group.**

***Index Terms*—Fraud detection, SWIFT network, domain ontology, ontology population, rules-based approach, SPARQL**

I. Introduction

Fraud Detection is a complex problem and represents a real challenge for financial institutions. Fraudulent transactions are not frequent, but the consequences for these institutions can range from reimbursing the amount to the customer to facing significant fines, for instance, in cases of money laundering. Banks are at the heart of financial exchanges and must be vigilant regarding operations that may involve money laundering or terrorism financing. To address this, financial institutions need tools for fraud prevention and detection.

SKAIZen Group is developing a research and innovation project aimed at modeling customers and financial institutions by populating a knowledge base from various data sources. This base will optimize fraud detection engines in financial transactions. The work presented in this article is part of our collaboration with SKAIZen Group in the "France Relance" project [1], which aims to build a knowledge base fueled by transactional data, taking into account financial data by type (customer, account, financial institution) and by relationship (beneficiary, debtor, account/customer, etc.).

In this work, we propose a specialized ontology for financial information, modeling the information from banking transactions in the SWIFT network and information about actors: customers and financial institutions. This work is based on the SWIFT model, which has become an ISO 20022 standard [1], and our KYC customer knowledge base [2], [3]. This ontology is populated with banking transactions from heterogeneous sources. Creating a specialized ontology for financial information allows querying the knowledge base regarding relationships between different entities (individuals or organizations) and detecting fraud based on rules.

The rest of the article is organized as follows: In Section II, we provide an overview of techniques for detecting financial fraud. We focus on semantic techniques based on ontologies and conclude this section with a summary. We present our approach in Section III. First, we introduce the domain ontology built in the context of our collaboration to model the financial domain, especially financial transactions in the SWIFT network. Secondly, we present our fraud detection approach based on rules. To validate our approach, we conducted experiments using real data, which we present in Section IV. We also showcase ontology querying and visualization. We conclude our work in Section V and provide some perspectives.

II. State of the Art

In recent years, several studies have focused on financial fraud detection. We present below the techniques based on machine learning, followed by semantic techniques using ontologies, and finally the techniques that combine ontology usage with machine learning methods.

A. Machine Learning-based Techniques

In recent years, several studies have compared machine learning models based on data and volume [4]–[6]. Supervised learning approaches are used in fraud detection problems to predict suspicious or fraudulent transactions. [7] use four algorithms for fraud detection in credit cards and the SMOTE (Synthetic Minority Oversampling Technique) algorithm to address data imbalance due to the low number of fraudulent transactions. [8] utilize multiple algorithms for credit card fraud detection, including Logistic Regression, Random Forest (RF), K-Nearest Neighbors (KNN), and neural networks. Evaluation on real-world data showed that the KNN algorithm produced the best results. [9] employ the Random Forest algorithm to detect fraudulent transactions and classify alerts by severity. Unsupervised learning approaches are used for detecting fraudulent transactions. [10] present an approach to detect rare activities related to money laundering. They propose a clustering algorithm and calculate an AICAF (Anomaly Index Calculation for Anti-Money Laundering) index based on principal component analysis and the K-means algorithm. The objective is to model user behavior and detect suspicious behavior. [11] use the K-means algorithm to create fraudulent

[1]https://www.iso20022.org/iso-20022-message-definitions

and legitimate classes. Semi-supervised learning approaches leverage unsupervised learning techniques to label data with classes generated by the model. Then, these labeled data feed a supervised learning algorithm [12]–[14].

B. Semantic-based Techniques

For several years, ontologies have been used to represent knowledge in various domains, including commerce, health, biology, and finance. In recent years, several works have employed ontologies in the financial domain for knowledge representation and fraud detection. Ontologies facilitate domain representation and rule definition, enabling the analysis of transactions to detect fraud or suspicious behaviors. In [15], an ontology is proposed to model banking frauds. The TF-IDF method is used to identify sentences with fraud-related experience, and the Latent Dirichlet Allocation (LDA) model is utilized to extract key domain terms for ontology creation, along with WordNet thesaurus for defining relations. The ontology serves as a reference for detecting new fraudulent transactions. [16] present an approach to prevent and detect fraudulent transactions in electronic payment systems using an ontology containing activities, anti-fraud rules, risks, and transactional activities. The authors outline a semi-automated fraud detection process but do not provide details on ontology construction or population. [17] propose an alert generation algorithm based on severity-ranked rules. They create the *Financial Fraud Detection* (FFD) ontology and develop anti-money laundering fraud rules and an alert generation algorithm. [18] develop an ontology to define financial frauds for identifying patterns for prevention and detection. Their approach involves defining fraud types, modeling user behavior based on existing ontologies, and validating the approach on multiple datasets.

C. Hybrid Techniques

Other approaches combine the usage of ontologies with machine learning approaches. [19] construct an ontology and use the decision tree algorithm to generate fraud patterns, which are then transformed into SWRL rules used in the knowledge base. This ontology models companies and financial concepts related to companies, such as profit, revenue, and cash.

Fraud detection should not degrade the relationship with the customer by blocking legitimate clients' transactions. Therefore, the fraud detection process must also consider the customer by incorporating their profile and usage patterns. To our knowledge, there have been no studies on transactions within the SWIFT network and the modeling of different actors (clients or agents). Developing an approach based on rule-based systems with a domain ontology that considers customer profiles and the specificities of the SWIFT network could assist financial institutions in enhancing their fraud prevention systems.

D. Synthesis

While recent works have studied financial fraud detection, there are evident gaps in the existing research. Machine learning methods, although prevalent, often lack of interpretability or knowledge describing. Semantic techniques, like ontologies, offer a way to model domain knowledge. However, to date, no study has addressed transactions within the SWIFT network while accounting for varied actors, such as clients or agents. This paper seeks to introduce a rule-based system coupled with a domain-specific ontology, aiming to enhance fraud detection and prevention within financial institutions.

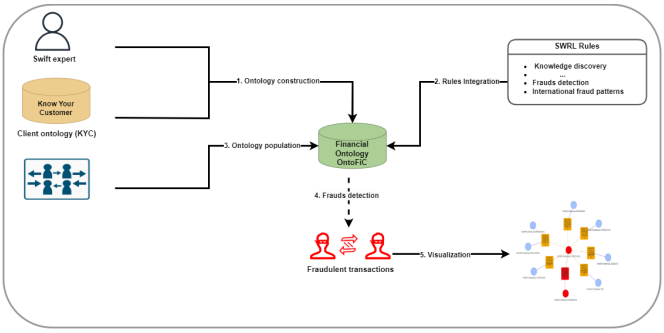

Fig. 1. General Approach.

III. Approach

In this section, we present the ontology of the domain developed within the scope of our work. This ontology is utilized for fraudulent transaction detection. Figure 1 illustrates the steps of our work. Our approach is composed of five steps, firstly, we construct the domain ontology (concepts and properties). Then, with financial experts assistance specialized in fraud prevention and detection, we define fraud rules based on transactions and customers. Afterward, we populate our ontology with SWIFT transactions from a private dataset. Next, we execute queries to identify fraudulent transactions, and finally we present them to the experts in a structured form using a graph-based visualization tool.

A. The OntoFiC Financial Ontology

The first step of our work is the construction of an ontology modeling the financial domain with concepts related to transactions and actors. This ontology builds upon the work of SKAIZen Group on extracting knowledge from financial articles for populating a KYC (Know Your Customer) ontology for banks. SWIFT transactions are messages that comply with a standard called ISO, which describes the formatting standards for these messages. In our work, we focus on international banking transactions represented by the ISO 20022 standard ("pacs.008" for MX messages). Figure 2 represents the transaction in the SWIFT network.

With the help of domain experts, we have selected the most relevant information for fraud detection. Information regarding customers has been selected from the KYC knowledge base. The following sections describe the concepts and properties of the ontology.

1) Concepts: In the proposed ontology, the concepts presented in Table I are divided into three main concepts: the **customer**, the financial institution which is an **agent** acting as a debtor, creditor, or intermediary, and the **transaction**.

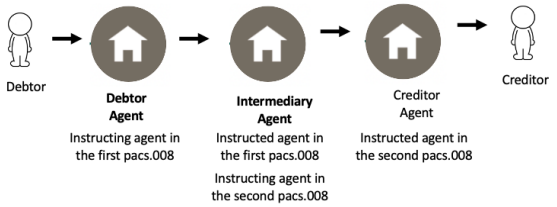

Fig. 2. Transaction in the SWIFT network

TABLE I
CONCEPTS

Name	Description
Activity	The domain of activity
Agent	Financial institution (customer's bank, intermediary bank)
Agent:CreditorAgent	Financial institution associated with the creditor customer
Agent:DebtorAgent	Financial institution associated with the debtor customer
Agent:IntermediaryAgent	Financial institution acting as an intermediary between two financial institutions
Customer	Creditor and debtor customer (not a financial institution)
Customer:Organization	Entity engaged in selling goods and services
Customer:Organization:Association	Organizations, associations, and NGOs. Not used due to legal ambiguity.
Customer:Organization:Company	Public or private company
Customer:Person	Individual
Transaction	Money transfer between two customers

2) Data Properties: Table II describes part of the data properties, and Table III describes the object properties.

TABLE II
DATA PROPERTIES

Name	Description	Concept
Amount	Amount exchanged between two customers	Transaction
BICFI	Code representing an Agent	Agent
ClrSysMmbId	Identifier of the system members	Agent
CreDtTm	Creation date and time	Message
CtryOfRes	Country of residence	Customer
Currency	Currency	Transaction
EndToEnd	End-to-end identifier	Transaction
hasPersonalData	Personal data	Customer
Id	Customer identifier	Customer
InstrId	Instruction identifier, a point-to-point reference	Transaction
LEI	Legal Entity Identifier	Agent
Name	Name of the agent or customer	Agent, Customer
PmtId	Payment identifier	Transaction

B. Rule with SWRL for Fraud Detection

In collaboration with our experts, we create some inference rules presented in Table IV using the SWRL language. The rules are designed to target fraud patterns associated to the international dimension of SWIFT transactions.

TABLE III
OBJECT PROPERTIES

Name	Description	Domain	Range
hasActivity	Indicates the activity associated with an organization	Organization	Activity
hasCreditor	Indicates the creditor of a transaction	Transaction	Customer
hasDebtor	Indicates the debtor of a transaction	Transaction	Customer
hasCreditorAgent	Indicates the agent associated with the creditor in a transaction	Transaction	Agent
hasDebtorAgent	Indicates the agent associated with the debtor in a transaction	Transaction	Agent
hasInstructedAgent	Indicates the intermediary agent in a transaction	Transaction	Agent

TABLE IV
RULES FOR FRAUD DETECTION

Name	Description
Rule 1: ForeignCountry	The transaction is conducted in a country different from the agent's country.
Rule 2: EU-USZone	The client is located in Europe or the United States and conducts a transaction with a currency other than the euro or dollar.
Rule 3: DailyAmount	Examines the total amount of transactions made in a day.
Rule 4: CurrencyCount	Transactions involving more than 5 currencies.
Rule 5: Expiration	Verifies the date and amount.
Rule 6: Triangular	Checks for money transfer.

As an illustration, let's consider these rules representing below:

Rule 1:	$\textbf{Transaction}(?t) \wedge \textbf{Currency}(?t, ?c) \wedge \textbf{TransactionCountry}(?t, ?tc) \wedge (\textbf{notEqual}(?c, \text{"euro"}) \vee \textbf{notEqual}(?c, \text{"dollar"})) \wedge (\textbf{equal}(?tc, \text{"France"}) \vee \textbf{equal}(?tc, \text{"USA"})) \rightarrow \textbf{IsFraud}(?t, \text{true})$
Rule 2:	$\textbf{Transaction}(?t) \wedge \textbf{Currency}(?t, ?c) \wedge \textbf{TransactionCountry}(?t, ?tc) \wedge (\textbf{notEqual}(?c, \text{"euro"}) \vee \textbf{notEqual}(?c, \text{"dollar"})) \wedge (\textbf{equal}(?tc, \text{"France"}) \vee \textbf{equal}(?tc, \text{"USA"})) \rightarrow \textbf{IsFraud}(?t, \text{true})$
Rule 3:	$\textbf{Transaction}(?t) \wedge \textbf{amount}(?t, ?amt) \wedge \textbf{Agent}(?a) \wedge \textbf{DailyAmount}(?a, ?dam) \wedge \textbf{lessThan}(?amt, 100) \wedge \textbf{greaterThan}(?dam, 10000) \rightarrow \textbf{IsFraud}(?t, true)$
Rule 4:	$\textbf{Transaction}(?t) \wedge \textbf{Agent}(?a) \wedge \textbf{AgentCurrencyCount}(?a, ?count) \wedge \textbf{greaterThan}(?count, 5) \rightarrow \textbf{IsFraud}(?t, true)$
Rule 5:	$\textbf{Transaction}(?t) \wedge \textbf{Agent}(?a) \wedge \textbf{LastTransactionDate}(?ltd, ?a) \wedge \textbf{greaterThan}(monthsBetween(now(), ?ltd), 6) \wedge \textbf{amount}(?t, ?amt) \wedge \textbf{greaterThan}(?amt, 1000) \rightarrow \textbf{IsFraud}(?t, true)$

Rule 6: **Transaction**(?t1) ∧ **TransactionDate**(?t1, ?td1) ∧ **DebitorAgent**(?da) ∧ **hasTransWithSameCredAndDate**(?da, ?td1, ?t2List) ∧ListSize(?t2List, ?size) ∧ **GreaterThan**(?size, 0) ∧ **Transaction**(?t2) ∧ **CreditorAgent**(?t2, ?ca) ∧ **TransactionDate**(?t2, ?td2) ∧**CreditorAgent**(?ca) ∧ Equals(?ca, ?da) ∧ **Equals**(?td2, ?td1) ∧ **Amount**(?t1, ?amount1) ∧ ListContains(?t2List, ?t2) ∧ **Amount**(?t2, ?amount2) ∧ **Equals**(?amount1, ?amount2) → **IsFraud**(?t1, true)

IV. Experimentation

The experiments were conducted using a private dataset of $1,000,000$ transactions from the SWIFT network. These transactions were SWIFT messages from which we extracted relevant fields related to the transaction and the client. Due to the dataset's private nature, we cannot disclose information regarding its distribution.

A. Ontology Evaluation

Before using our ontology, it was crucial to ensure its validity and coherence. To achieve this, we subjected our ontology to evaluation by financial experts, who assessed its vocabulary, concepts, data hierarchy, and semantics. The ontology was created with Protégé 5.5.0.

Furthermore, we employed the reasoners Fact++ and HermiT to verify the consistency and coherence of our financial ontology. The reasoners were used to confirm that the data had been appropriately linked and that the logical inferences were consistent with the data within the ontology.

B. Ontology Population and Visualization

Populating the financial ontology was a crucial step for analyzing financial data and identifying suspicious and fraudulent transactions. This process involved creating instances, concepts, object properties, and data properties within the ontology. Using *Owlready2* [2], we populated our ontology with $1,000,000$ financial transactions. After populating the ontology, we utilized *SPARQL* queries to identify suspicious or fraudulent transactions (figure 3). Finally, we employed the *Pyvis* library [3] to display the results of our *SPARQL* queries in the form of a graph, facilitating the visualization of relationships between transactions and enabling more straightforward data analysis.

C. Discussion and perspectives

Through our experiments, we visualized connections between actors in fraudulent transactions, offering new insights and directions for future research. We aim to apply graph theory for deeper insights in the future. While ontology is beneficial for financial systems, it faces scalability challenges with large datasets. Prioritizing potentially fraudulent transactions, like those with high amounts, can improve robustness. Our focus is on SWIFT transactions, distinguished by their international and interbank characteristics. Given the challenges associated with the handling of sensitive data in our study, future works will focus on exploring methodologies within ontologies to uphold strict ethical and privacy standards, ensuring the protection of sensitives information.

[2]http://owlready2.readthedocs.io

[3]https://pyvis.readthedocs.io

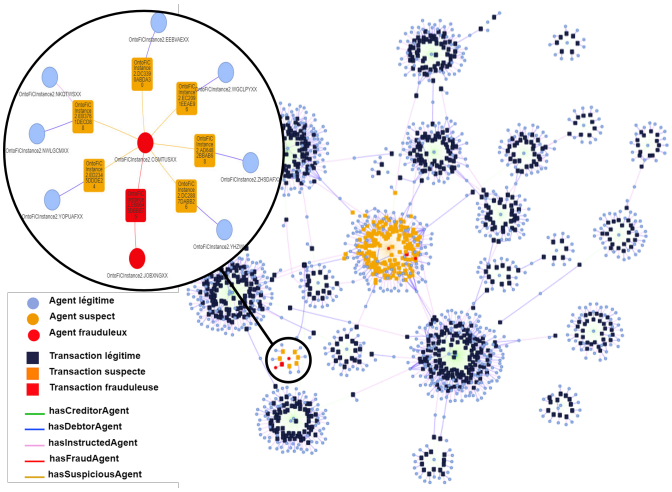

Fig. 3. Visualization of Fraudulent Transactions

V. Conclusion

In this paper, we have proposed a domain ontology for interbank transactions in the SWIFT network, incorporating client information from KYC records. Our work is part of a collaborative project with SKAIZen Group. The construction of the OntoFiC ontology enables the modeling of SWIFT transactions and clients, populated with real-world transaction data. Our fraud detection approach relies on rule-based reasoning, and we also offer visualization of SPARQL queries using the *pyvis* library. The validation of our approach was performed through experiments, yielding promising results. This approach can complement fraud prevention and detection tools for financial institutions.

For our future work, we plan to expand the ontology by dynamically generating new rules for detecting fraudulent transactions using machine learning techniques. We also aim to analyze client information to facilitate the work of experts responsible for monitoring blocked transactions.

Acknowledgments: This work is supported by SKAIZen Group, ANRT, and ANR (France Relance).

References

[1] B. Auger, H. Chergui, Y. Chehade, J. E. Kadri, L. Abrouk, N. Cabioch, Construction d'une ontologie dans le domaine financier pour la détection de fraudes, in: INFORSID, 2022, pp. 157–162.

[2] A. Jabbari, O. Sauvage, N. Cabioch, Towards a knowledge base of financial relations: Overview and project description, in: 2019 IEEE Second International Conference on Artificial Intelligence and Knowledge Engineering (AIKE), IEEE, 2019, pp. 313–316.

[3] A. Jabbari, O. Sauvage, H. Zeine, H. Chergui, A french corpus and annotation schema for named entity recognition and relation extraction of financial news, in: Proceedings of the 12th Language Resources and Evaluation Conference, 2020, pp. 2293–2299.

[4] Y. Zhang, P. Trubey, Machine learning and sampling scheme: An empirical study of money laundering detection, Computational Economics 54 (3) (2019) 1043–1063.

[5] J. Lorenz, M. I. Silva, D. Aparício, J. T. Ascensão, P. Bizarro, Machine learning methods to detect money laundering in the bitcoin blockchain in the presence of label scarcity, in: Proceedings of the First ACM International Conference on AI in Finance, 2020, pp. 1–8.

[6] B. Bestami Yuksel, S. Bahtiyar, A. Yilmazer, Credit card fraud detection with nca dimensionality reduction, in: 13th International Conference on Security of Information and Networks, 2020, pp. 1–7.

[7] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic, A. Anderla, Credit card fraud detection - machine learning methods, in: 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH), 2019, pp. 1–5. doi:10.1109/INFOTEH.2019.8717766.

[8] A. Mehbodniya, I. Alam, S. Pande, R. Neware, K. P. Rane, M. Shabaz, M. V. Madhavan, Financial fraud detection in healthcare using machine learning and deep learning techniques, Security and Communication Networks 2021 (2021).

[9] R. More, C. Awati, S. Shirgave, R. Deshmukh, S. Patil, Credit card fraud detection using supervised learning approach, International Journal of Scientific Technology Research 9 (2021) 216–219.

[10] A. S. Larik, S. Haider, Clustering based anomalous transaction reporting, Procedia Computer Science 3 (2011) 606–610.

[11] U. Porwal, S. Mukund, Credit card fraud detection in e-commerce: An outlier detection approach, arXiv preprint arXiv:1811.02196 (2018).

[12] N. A. Le Khac, M.-T. Kechadi, Application of data mining for anti-money laundering detection: A case study, in: 2010 IEEE International Conference on Data Mining Workshops, IEEE, 2010, pp. 577–584.

[13] S. Raza, S. Haider, Suspicious activity reporting using dynamic bayesian networks, Procedia Computer Science 3 (2011) 987–991.

[14] F. Carcillo, Y.-A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, G. Bontempi, Combining unsupervised and supervised learning in credit card fraud detection, Information sciences 557 (2021) 317–331.

[15] G. Attigeri, M. M M, R. Pai, R. Kulkarni, Knowledge base ontology building for fraud detection using topic modeling, Procedia Computer Science 135 (2018) 369–376. doi:10.1016/j.procs.2018.08.186.

[16] A. El Orche, M. Bahaj, S. Ain Alhayat, Ontology based on electronic payment fraud prevention, Faculty of Sciences Technologies HASSAN 1 (2018).

[17] M. Ahmed, K. Ansar, C. B. Muckley, A. Khan, A. Anjum, M. Talha, A semantic rule based digital fraud detection, PeerJ Computer Science (2021).

[18] A. Hussaini, Z. Guessoum, E. Laurent, Elaboration of financial fraud ontology, 2022, pp. 277–285. doi:10.15439/2022F35.

[19] X.-B. Tang, G.-C. Liu, J. Yang, W. Wei, Knowledge based financial statement fraud detection system: Based on an ontology and a decision tree, National Natural Sciences Foundation of China "Research on Intellience Consulting Services Based on the Semantic Analysis of Trxt and Web" (2018).