1-14-2021

Designing a relational model to identify relationships between suspicious customers in anti-money laundering (AML) using social network analysis (SNA)

Abdul Khalique Shaikh

Malik Al-Shamli

Amril Nazir

Journal of Big Data

RESEARCH

Open Access

Designing a relational model to identify relationships between suspicious customers in anti-money laundering (AML) using social network analysis (SNA)

Abdul Khalique Shaikh[1*], Malik Al-Shamli[2] and Amril Nazir[3]

*Correspondence:
shaikh@squ.edu.om
[1] Department of Information Systems, College of Economics & Political Science, Sultan Qaboos University, Muscat, Oman
Full list of author information is available at the end of the article

Abstract

The stability of the economy and political system of any country highly depends on the policy of anti-money laundering (AML). If government policies are incapable of handling money laundering activities in an appropriate way, the control of the economy can be transferred to criminals. The current literature provides various technical solutions, such as clustering-based anomaly detection techniques, rule-based systems, and a decision tree algorithm, to control such activities that can aid in identifying suspicious customers or transactions. However, the literature provides no effective and appropriate solutions that could aid in identifying relationships between suspicious customers or transactions. The current challenge in the field is to identify associated links between suspicious customers who are involved in money laundering. To consider this challenge, this paper discusses the challenges associated with identifying relationships such as business and family relationships and proposes a model to identify links between suspicious customers using social network analysis (SNA). The proposed model aims to identify various mafias and groups involved in money laundering activities, thereby aiding in preventing money laundering activities and potential terrorist financing. The proposed model is based on relational data of customer profiles and social networking functions metrics to identify suspicious customers and transactions. A series of experiments are conducted with financial data, and the results of these experiments show promising results for financial institutions who can gain real benefits from the proposed model.

Keywords: Social network analysis, Anti-money laundering, Relational analysis, Relationships, Customer profile

Introduction

Money laundering activities within financial institutions not only affect the growth of national economy but also weaken the political stability of a country by transferring economic power from government to criminals. To stabilize national economy and a political system of the country, it is essential for the government to prevent these types of crimes in their jurisdiction.

Many existing anti-money laundry (AML) solutions are available in preventing money laundering, such as clustering-based anomaly detection techniques [1], rule-based decision-tree algorithm [2], supervised learning technique [3], and various other statistical methods. One of the review papers [4] studied models such as rule-based approach, clustering-based approach, classification-based approach, and model-based approach and claimed clustering techniques can be best for detecting money laundering activities. The research paper [5] states money laundering has now become a universal concern and technological measures can be very helpful in preventing it. A recent article [6] stated the current AML system solutions do not fulfil the requirements of the latest economic conditions. However, the AML transaction monitoring process can improve detection and resolve issues related to the time factor to identify suspicious transactions. The AML data analysis paper [7] states that if a transaction pattern is found to be unusual, then such unusual transaction activity can be reported to a law force agency such as Financial Agency Unit for further investigation. Most of the current literature reviews on money laundering focus on identifying individual suspicious customers and transactions. However, the researchers have found a big gap in the utilization of relations data and identified the impact of links of suspicious customers. This paper aims to identify the relationships and associations of suspicious customers by using social networking functions, namely identifying mafia groups involved in money laundering activities, and to prevent money laundering activities and potential terrorist financing. This paper is an extension of our previous work [8] where we identified suspicious customers based on the dynamic trends of transactions, history of transactions, and profiles of customers. However, the proposed model seeks to identify the relationships and associations linked with suspicious customers. The model make use the data on suspicious customers from our previous model [8] and identifies the various kinds of relationships that are significant in controlling money laundering. This proposed model utilizes social networks analysis functions due to their increasing popularity and their ability in capturing the relationships of suspicious customers to identify the mafia groups involved the money laundering. These relationships include those of common owner, business, spouse, parent/child, family, and likewise. We have utilized relational data and social networking functions such as degree of centrality, ego group, and profile-based criteria. Each customer has been considered as a node, and each node has a single or multiple accounts. The nodes are connected to each other with the transactions carried out between these accounts. The transactions between the customers are recorded as relations within a network. The overall performance and effectiveness of our proposed model depend entirely on the customer's profile information, which is provided at the time of opening a bank account. We have developed various kind of rules that can be executed under customer profiles which can be used to identify existing relationships of suspicious customers.

The main contributions of this paper are as follows: (i) to develop and implement a model to identify the links to suspicious customers in a money-laundering network; (ii) to utilize social networking functions to investigate the effectiveness of the proposed model; and (iii) to capture the hidden relationships between suspicious customers and their transactions from the given customer profile data.

The remaining sections of the paper are organized as follows:

"Literature review" section briefly discusses the literature, "Research methodology" section explains the methodology of the proposed approach, and experimental results are demonstrated in "Results and discussion" section. Finally, "Conclusion and future work" section concludes the paper with possible future work.

Literature review

This section reviews the related work and summarizes its similarities and the differences compared to our work.

The research study [9] initially compared the different methods used to build social networks and discussed the techniques employed by other researchers for building social networks and ontologies from unstructured, dirty, and conflicting datasets. Further, [10] also introduced a method that could build ontology-based social networks by determining the relationships/links of users and resources. The method is supported by the research study [11] that present a semantic-based centralized resource discovery model for improvement of matching resources. From these relationships, inferences could then be extracted. However, less emphasis was made on extracting social relationships. Instead, the authors greatly emphasized on user access management.

The research article [12] applied SNA to identify the neighbors between nodes to form network communities, whereby individuals could be grouped based on their social links. They discovered that groups within the network with similar behaviors and characteristics to be critical in detecting money laundering activities. Further, analyzing the variation in different types of relationships of people from their social network links can lead to discovering suspicious activities. These social links normally encompass members of a family, relatives, work colleagues, friends, connected friends, and other 2nd and 3rd-level connections. However, the paper more focuses on social learning content and utilizes users' cookies data that can be a challenge to the ethical use of data.

Another research study [7] proposed an integrated system of customer relationship management (CRM) and AML to detect suspicious data reporting in commercial banks with a specific goal of identifying suspicious transactions more effectively and reducing false alarms. The system would perform a customer background check and conduct customer analysis, customer identification, customer business analysis, and customer visit. However, such systems did not have capabilities to detect whether an illegal transaction belonged to any of the launderer groups or not.

The research article [3] made use of supervised learning using social networks to detect illegal transactions. The social networks were built combining explicit transaction and implicit relationships. The business knowledge was extracted from networks of small and meaningful communities. Multiple relationships from the networks were identified based on the node edges, shared accounts, connection similarities, and locations. Both random forest and support vector machines (SVMs) were used for supervised learning to detect suspicious transactions. This is perhaps of the earliest methods that could make use of both SNA and supervised learning to detect money-laundering activities.

Soltani et al. [13] proposed a novel approach based on a clustering method to group abnormal users as potential money launderers. First, they made use of feature reduction to reduce the input data set. Next, the reduced data was used to group a set of outlier users with similar attributes and characteristics. The preliminary experimental results were promising. However, the approach was unable to analyze the social network's links from these transactions.

A new approach was proposed by [14] to sort and map relational data and present predictive models based on SNA. The researchers used the real financial dataset to identify the risks associated with customer profiles using social network metrics and showed the importance of using a network-based approach. However, this paper could not identify the hidden relationships from the network.

A research study [15] is presented to describe public health practitioners' experiences with a social network where 13 public health practitioners interviews are recorded and transcribed. The paper utilized SNA to identify the relational dynamics and then applied the knowledge gained from the reflection to improve their practice. The paper measures relational dynamics in the service of health systems as authors in this study suggested that SNA can be a potentially useful reflective tool to self-assess the overall composition of their networks for strengthening their network's composition and other relations.

The research paper by [16] proposed a conceptual model based on the multi-agent for AML. The model could control detection and prevention processes and support major phases of decision making to identify the specific activities involved in each decision-making phase for AML. The approach defined in this research focuses more on business values in terms of the enhancement of operation and reduction of costs. However, the profile of money launders has not being considered in the detection and decision making processes.

Recently, [17] introduced SNA to prevent money laundering. They mapped the relational data with their proposed model and presented a predictive model that is based on network metrics, to identify the risks associated with the economic market. They discovered that social networking functions are critical in predicting the risk profile of the clients, and their experimental results showed that social network metrics are important to consider when assessing risk profiles. The work provides an early analysis of the importance of social network metrics in effectively identifying suspicious customers/transactions using social networking functions and the results of a qualitative study [18] demonstrate that the impact of social media and relationships on socio-political dimension. The research paper [19] had also conducted an extensive study on the behavior of Tancent QQ social network accounts to understand money laundering activities. They considered the additional features of the account viability, transaction sequences, and spatial correlation among accounts and fed these features into a statistical classifier to improve the model's performance. The model indicated some social networking functions that could be effectively used to identify suspicious customers or transactions. In addition, The paper [20] uses SNA as a methodological approach for identifying collaborative networks and examining the social dimension of co-author relationships using social networking functions that shows a positive impact of SNA to analyses structure, relationships, and interactions of any network.

By considering the literature review and to overcome the above-mentioned limitations and challenges, this paper proposes a relational model to identify the relationships and associations with suspicious customers in AML using SNA, which can build a social network of customer profiles by utilizing the extracted financial transactions. Our research methodology is described in the next section.

Research methodology

This section presents the research methodology of the proposed model, which includes the system architecture, flow chart, and details of how the components are connected to each other. The proposed model seeks to come up with a preeminent technique that can detect maximum relationships and measure the profile data of customers within a network. The social network is built using the customer's profile data and their transaction information. Social network functions such as degree centrality, ego network, clustering functions are then applied to the network. These functions are used to identify explicit relationships of each suspicious customer within a network and establish the relations and key nodes within a network. The built network represents a customer as a node and transaction as an edge. The research work by [21] focused on the importance of application of SNA to intelligence and the study of criminal organizations.

To evaluate the effectiveness of our proposed model, we conducted our experiments under a different number of customers and transactions. The model can detect all possible relations between suspicious and non-suspicious customers under different network parameter scenarios. In order to simulate this scenario, we run multiple sets of experiments with different sizes of nodes, such as 100, 500, 750, and 1000, under a fixed transaction per node and, in another set, different sizes of transactions with a fixed size of nodes. The details of the algorithms are described as follows.

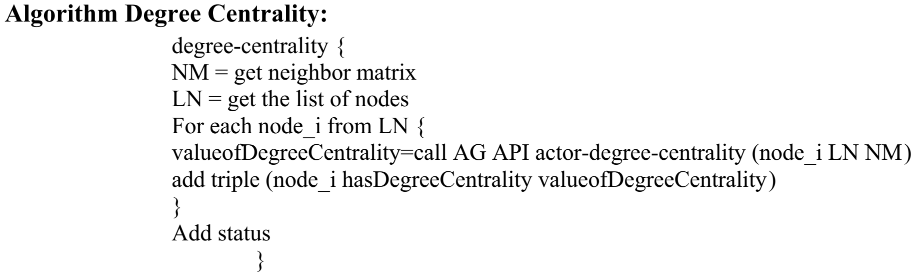
Degree centrality

To find out the key and influential customers within a network, we implement a degree centrality algorithm that uses a number of transactions associated with nodes as a proxy of importance. The mathematical representation of the calculation of degree centrality is as follows:

$$\text{Degree centrality(node\_i)} = \text{Number of edges(node\_i)} / \text{N} - 1$$

where N is the total number of nodes in the network.

In our proposed model, N is the total number of customers and the number of edges is counted based on the transactions of specific customers with other customers in the network. The pseudo-code of the degree centrality algorithm is as follows:



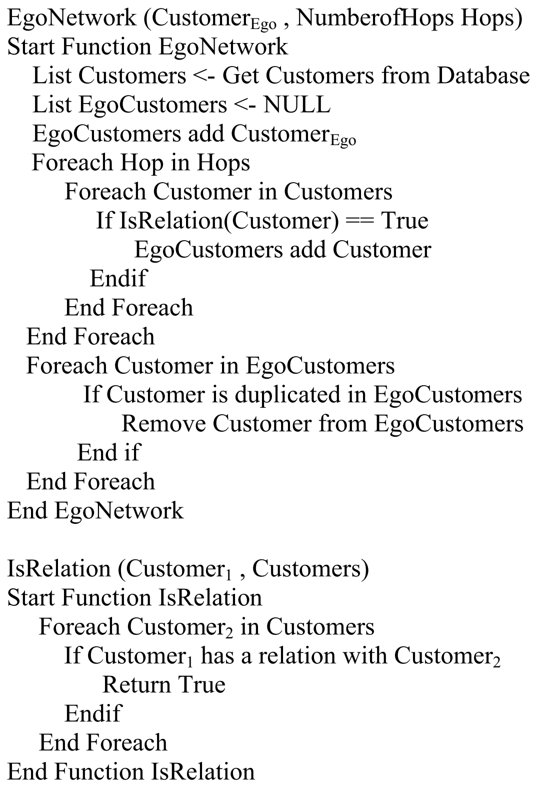
```
Algorithm Degree Centrality:
    degree-centrality {
        NM = get neighbor matrix
        LN = get the list of nodes
        For each node_i from LN {
            valueofDegreeCentrality=call AG API actor-degree-centrality (node_i LN NM)
            add triple (node_i hasDegreeCentrality valueofDegreeCentrality)
        }
        Add status
    }
```

We can identify influential suspicious customers with the help of degree centrality values. A high degree of centrality value indicates the node has a high influence on the network, and the associated customer is highly risk-oriented. The details of the implementation of the degree centrality algorithm are explained in the experiment section.

Ego network

This SNA function focuses on the individual network of a specific node within an entire network. The function can be used not only to establish the link of the targeted person under direct connection but also it can provide the information on other nodes that are connected to the direct connection node. For example, we can find friends of friends in the network. Since we aim to identify of relationships of individual suspicious customers, this ego-centric approach can be used to achieve the target as it is focused on individuals, groups, and communities. The approach was employed in a prominent study on the adoption of innovation of a particular drug and used to understand the differences in the personal and social characteristics of the medical staff in the United Kingdom [22]. According to this approach, the suspicious customers in the relationship are referred to as the "ego" and the links or connections to the "ego" as "alters". With the help of an ego network, we can identify the direct and indirect connections of a suspicious customer. The details of the implementation of the ego network are explained in the experiment section.

The pseudo-code of the ego network algorithm is listed below:



```
EgoNetwork (Customer$_{Ego}$ , NumberofHops Hops)
Start Function EgoNetwork
    List Customers <- Get Customers from Database
    List EgoCustomers <- NULL
    EgoCustomers add Customer$_{Ego}$
    Foreach Hop in Hops
        Foreach Customer in Customers
            If IsRelation(Customer) == True
                EgoCustomers add Customer
            Endif
        End Foreach
    End Foreach
    Foreach Customer in EgoCustomers
        If Customer is duplicated in EgoCustomers
            Remove Customer from EgoCustomers
        End if
    End Foreach
End EgoNetwork

IsRelation (Customer$_1$ , Customers)
Start Function IsRelation
    Foreach Customer$_2$ in Customers
        If Customer$_1$ has a relation with Customer$_2$
            Return True
        Endif
    End Foreach
End Function IsRelation
```

Table 1 Rules and associated conditions

Rules	Conditions	Outcomes
1	∃t(A1, A2) and C1 ∈ A1 & C2 ∈ A2 && (GC1 = GC2) &&(C1of A1 ∩ C2 of A2) ‖ (C1 of A1 = C2 of A2) → Rc(O(C1), O(C2)	Common owner relationship
2	∃t(A1, A2) and C1 ∈ A1 & C2 ∈ A2 && if (A1& A2) ∈ Type "B") → Rb(O(C1), O(C2))	Business relationship
3	∃t(A1, A2) and C1 ∈ A1 & C2 ∈ A2 C1 ∈ A1 & C2 ∈ A2 && if (Address, Lastname & Father Name of C1 == Address, Lastname & Father name of C2) && AgeDiff of C1 and C2 ≤ 10 && Father Name of C1 ! = to Firstname of C2 && Firstname of C1 ! = to Father name of C1 → Rs(O(C1), O(C2))	Sibling relationship
4	∃t(A1, A2) and C1 ∈ A1 & C2 ∈ A2 C1 ∈ A1 & C2 ∈ A2 && if (Add & Lastname of C1 == Add & Lastname of C2) && AgeDiff ≤ 20 && (Gender of C1 ≠ Gender of C2 && First name and FatherName of C1 ≠ Firstname and Father name of C2) → Rs(O(C1), O(C2))	Spouse relationship
5	∃t(A1, A2) && if (Address & LastName of C1 == Addess & LastName C2) && AgeDiff ≥ 20 && (FirstName of C1 ≠ FirstName of C2) && First name of C1 is not equal to Father Name of C2 && father name of C1 is not equal to First name of C1 && agediff is > 20 (if C1 is elder than C2 && gender is female then C1 is mother otherwise C2 is mother) → Rs(O(C1), O(C2))	Mother/Child relationship
6	∃t(A1, A2) and C1 ∈ A1 & C2 ∈ A2 && if (lastName, address DOB & Gender of C1 == LastName, address DOB & Gender of C2) → Rsa(O(C1), O(C2))	SameAs relationship
7	∃t(A1, A2) and C1 ∈ A1 & C2 ∈ A2 && if (lastname, address & race of C1 == lastname, address & race of C2) → Rf(O(C1), O(C2))	Family relationship

(*A* account, *C* customer, *G* gender, *R* relationship)

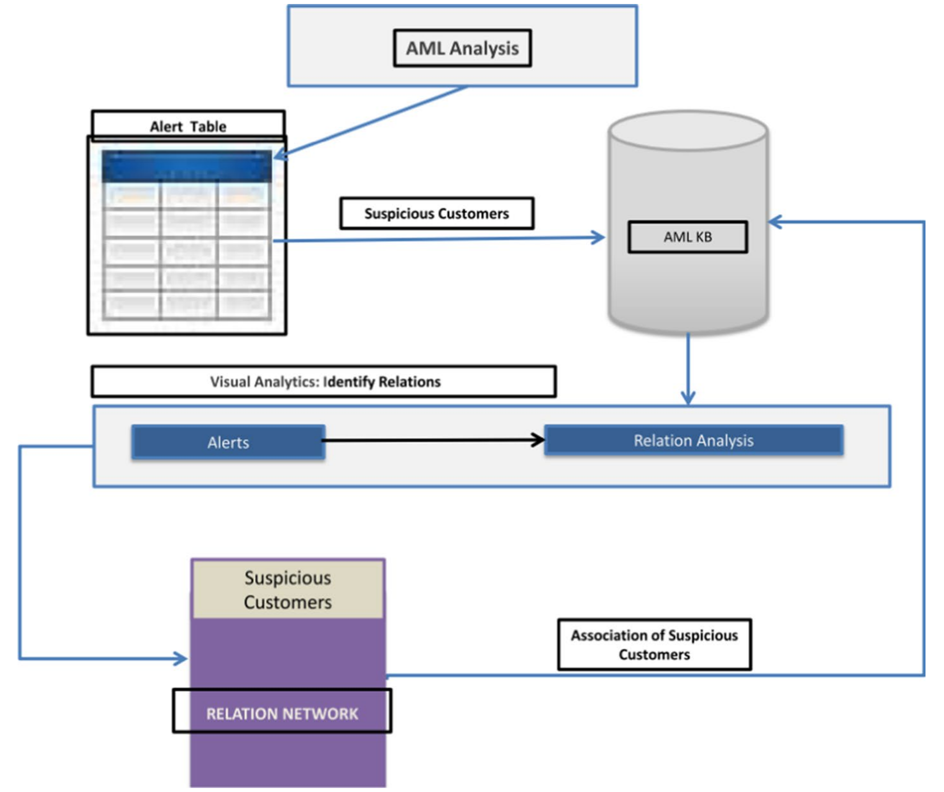


Fig. 1 Architecture of the proposed model

Clustering

Clustering is used to identify and separate similar characteristics of customers within a network. In the proposed model, customers are separated based on their profile data, such as gender, city, age, and country, etc. In addition to this, suspicious and non-suspicious customers can be separated to measure the ratio between them. Our goal is to explore and identify the hidden relationships of a suspicious customer within a network. The paper [23] and [24] described the methods available for clustering and categorized these as partitioning methods, hierarchical methods, SOM, and density-based methods. However, since K-means and k-medoids are popular partitioning algorithms, we have utilized a K-means algorithm [25] for clustering the customers based on their profile and status. Moreover, the research study [26] suggest that the K-means algorithm is well appropriate in finding similarities between entities based on distance measures.

Rule-based approach

Our proposed model uses the conditions mentioned in Table 1 to identify different kinds of relationships. The conditions are implemented as a rule-based approach in.NET framework, where the rules are triggered if the proper condition is met. The details of the conditions, criteria, and outcomes are given below:

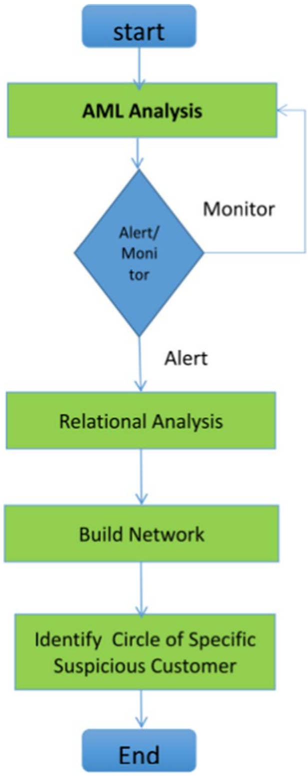


Fig. 2 Flow chart of the proposed model

System architecture of the proposed model

The architecture of the proposed model is shown in Fig. 1, where the AML analysis component feeds the input of suspicious customers and transactions as data to the relational analysis module [8]. The customers were identified as suspects or suspicious customers in our previous work [8] where we identified suspicious customers based on the dynamic trends of transactions, history of transactions, and profiles of customers. However, the proposed model seeks to identify the relationships and associations linked with suspicious customers. All suspicious customers and unsuspicious customers' account and transaction data are stored in the AML knowledgebase (KB). The AML analysis generates suspicious customer data and displays alert and monitor lists. The difference between alert and monitor customers list is a risk factor. The alert list is already recognized as suspect-based, whereas a customer monitor list is for customers under observation, which confirmed as suspicious customer that is taken from source i.e., trusted financial institutions. We use a dynamic approach to identifying suspicious customers in money transactions and the alert table is generated and updated periodically to alert suspicious customers. We execute the relational analysis module based on customer profiles and the rule-based approach and we then identify links of suspicious customers within a network. These relations include those of common owner, business, sibling, spouse, mother–child, family, etc., and it can be identified by the relational analysis module. With the help of these relations, we can identify mafia groups involved in money laundering activities. The data associated with the customer and their accounts

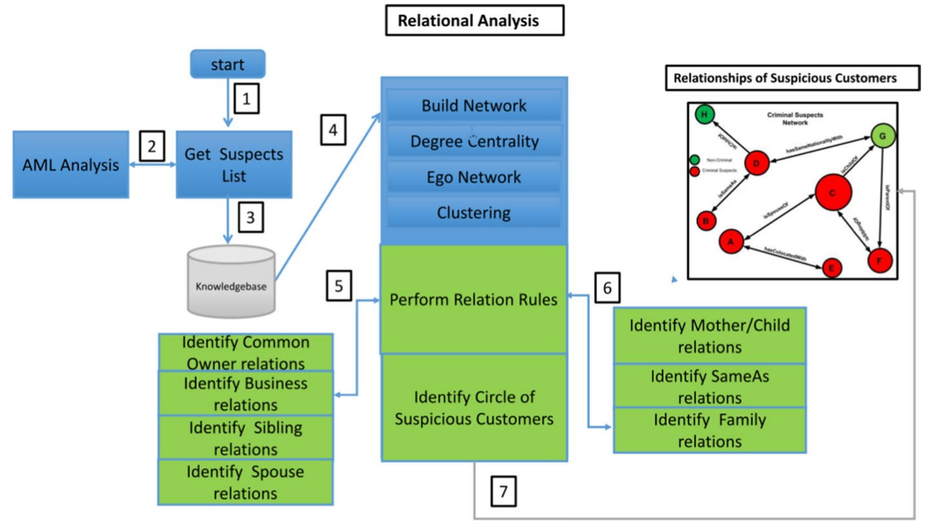

Relational Analysis
start
1
2
AML Analysis
Get Suspects List
3
Knowledgebase
4
Build Network
Degree Centrality
Ego Network
Clustering
5
Perform Relation Rules
Identify Common Owner relations
Identify Business relations
Identify Sibling relations
Identify Spouse relations
Identify Circle of Suspicious Customers
6
Identify Mother/Child relations
Identify SameAs relations
Identify Family relations
7
Relationships of Suspicious Customers
Criminal Suspects Network
Non-Criminal
Criminal Suspects
H
G
D
C
B
A
E
F
isChildOf
hasSameNationalityWith
isChildOf
isSameAs
isSpouseOf
isParentOf
isSiblingOf
hasColocatedWith

Fig. 3 Flow chart of relational analysis

and transactions can be found in AML KB and present the required data of suspicious customer on any display. The flow chart below shows a step-by-step process to identify the links of suspicious customers.

The flow chart of the proposed model is shown in Fig. 2. The AML analysis generates suspicious customer data and displays alert and monitor lists. The difference between alert and monitor is a risk factor. The alert list is already recognized as suspect-based, whereas a customer monitor list is for customers under observation.

The relational analysis module takes the data of suspicious customers and links it with that of normal customers and builds the network. The social network shows the circles of each suspicious customer within a network. The relation analysis model is explained in detail as follows.

Relation analysis

The relational analysis module extracts the relations and associations of suspicious customers using social networking functions and rules. The relationships are identified based on profile information, such as the suspect's last name, address, gender age, race, etc. The flow chart of relational analysis is presented below:

Figure 3 shows a step-by-step procedure of how relations are identified through relation identification strategies such as network analysis functions and the rule-based approach. The module produces various types of relations such as that of the common owner, family, business, etc. The conditions of the rules are mentioned in Table 1. The above flow chart first assesses the input of suspicious customer profiles from the AML analysis module and stores the data in the AML KB where all the other customer data is stored as well. With the help of customer profile data, such as customers' profile, account, and transaction information, the network is built and the degree centrality function executed to find the most influential customer who has had many transactions with other customers. Subsequently, the SNA function "ego network" is executed to find out the direct and indirect relations of the specific

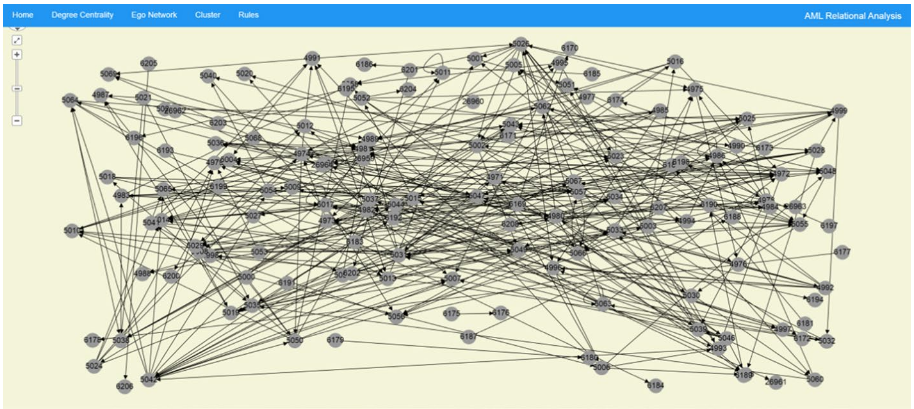
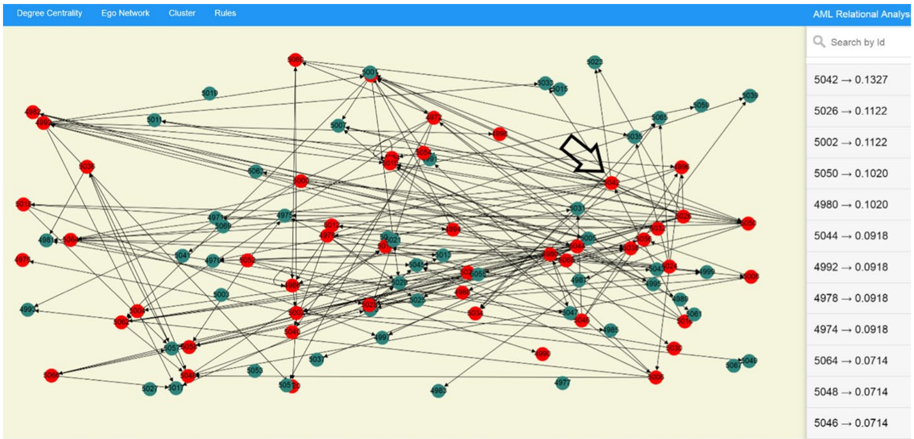


Fig. 4 Network of 100 suspicious and non-suspicious customers



5042 → 0.1327
5026 → 0.1122
5002 → 0.1122
5050 → 0.1020
4980 → 0.1020
5044 → 0.0918
4992 → 0.0918
4978 → 0.0918
4974 → 0.0918
5064 → 0.0714
5048 → 0.0714
5046 → 0.0714

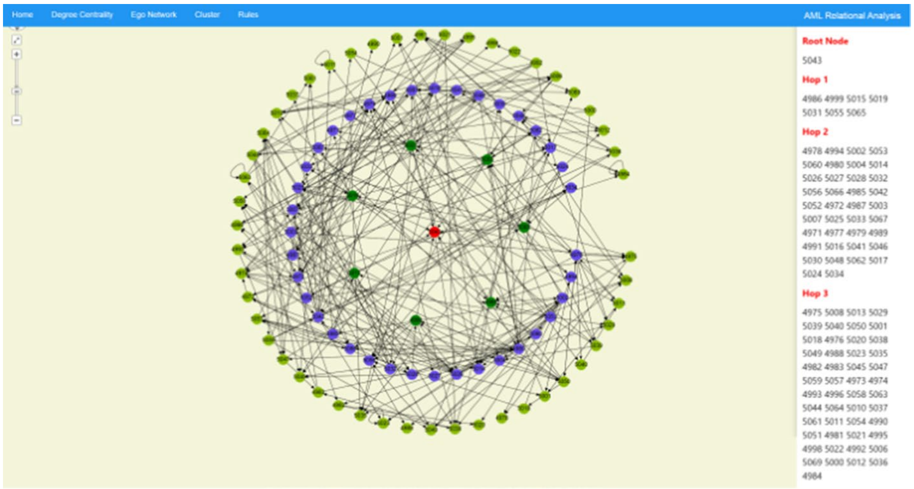
Fig. 5 Degree centrality of 100 Suspicious and unsuspicious customers

suspicious customer. The clustering function is utilized to separate the group based on customer profile such as city, country, gender, and risk status. Subsequently, the model executes the key part of the system, i.e., the relational rules module, where seven rules are triggered based on the customer profile. These rules capture seven relationships as mentioned in Table 1 and display the graph as mentioned in Fig. 3. The following section discusses the experiment and the results of the proposed model.

Results and discussion

To implement the algorithms and rules mentioned in "Research methodology" section, the.NET environment is set up to utilize Cytoscape, an open-source software platform [27] for visualizing the required relations. This section explains the details of the experiment steps along with results and discussion.

The data related to customers' accounts and transactions was imported to a database known as the SQL server. We generated all the relevant data via a software tool using C# language. The reason for using a system-generated data is financial



Root Node

5043

Hop 1

4986 4999 5015 5019
5031 5055 5065

Hop 2

4978 4994 5002 5053
5060 4980 5004 5014
5026 5027 5028 5032
5056 5066 4985 5042
5052 4972 4987 5003
5007 5025 5033 5067
4971 4977 4979 4989
4991 5016 5041 5046
5030 5048 5062 5017
5024 5034

Hop 3

4975 5008 5013 5029
5039 5040 5050 5001
5018 4976 5020 5038
5049 4988 5023 5035
4982 4983 5045 5047
5059 5057 4973 4974
4993 4996 5058 5063
5044 5064 5010 5037
5061 5011 5054 4990
5051 4981 5021 4995
4998 5022 4992 5006
5069 5000 5012 5036
4984

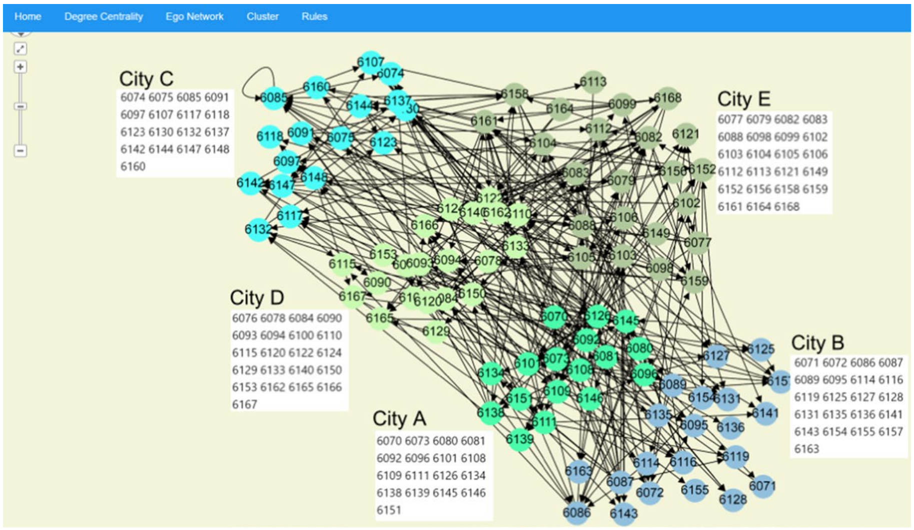
Fig. 6 Ego network

institutions are very reluctant to share the real customer's data for some security reasons. However, we use the same realistic pattern of customer bank data using the software tool. The data used for this research is from January 2010 to December 2018. The total raw data was created for 1000 customers with 138,114 transactions. However, for the experiment purpose, we selected 10% of customers sample data from the total customers of raw data. Finally, the sample data of this research was 100 customers with 14,253 transactions.

This data is of both normal customers and suspicious customers. However, suspicious customers are identified from AML analysis module using a dynamic approach, which is explained in our previous work [8]. By using both suspicious and normal customer profiles, we developed a social network of both types of customers and linked their profile data with their accounts and transactions. The aim of establishing a social network was to identify the relations of suspicious customers who could be connected with other suspicious and unsuspicious customers. The network of 100 nodes for both suspicious and unsuspicious customers is presented below:

Figure 4 shows the network of 100 suspicious and unsuspicious customers. Each node has a unique ID displayed on top of the node. The links in the network show the transactions between the customers; so we can identify which customer has made a credit or a debit transaction from the customer account. Under this network, we implemented and applied the degree centrality algorithm to measure the degree centrality of each suspicious customer. The degree of centrality of each customer is mentioned below:

Figure 5 indicates the network of 100 customers with their transactions, and all suspicious customers within a network are highlighted in red. The degree centrality of each suspicious customer is measured and displayed on the right-hand side of the frame. The frame shows the top of the node IDs with a maximum value of degree centrality of the network. With the help of degree centrality, we can identify the influential suspicious customers who have made maximum transactions with other customers. Subsequently, we can take specific nodes with a high degree of centrality value and determine the relation of those specific nodes. As Node ID 5042 has made the maximum transactions, so



City C
6074 6075 6085 6091
6097 6107 6117 6118
6123 6130 6132 6137
6142 6144 6147 6148
6160

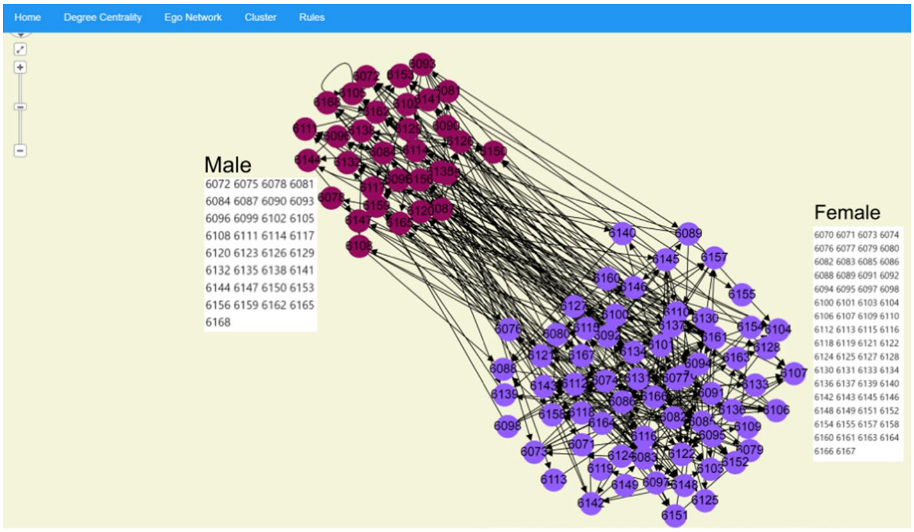
City E
6077 6079 6082 6083
6088 6098 6099 6102
6103 6104 6105 6106
6112 6113 6121 6149
6152 6156 6158 6159
6161 6164 6168

City D
6076 6078 6084 6090
6093 6094 6100 6110
6115 6120 6122 6124
6129 6133 6140 6150
6153 6162 6165 6166
6167

City B
6071 6072 6086 6087
6089 6095 6114 6116
6119 6125 6127 6128
6131 6135 6136 6141
6143 6154 6155 6157
6163

City A
6070 6073 6080 6081
6092 6096 6101 6108
6109 6111 6126 6134
6138 6139 6145 6146
6151

Fig. 7 Clustering of suspicious customers based on their city



Male
6072 6075 6078 6081
6084 6087 6090 6093
6096 6099 6102 6105
6108 6111 6114 6117
6120 6123 6126 6129
6132 6135 6138 6141
6144 6147 6150 6153
6156 6159 6162 6165
6168

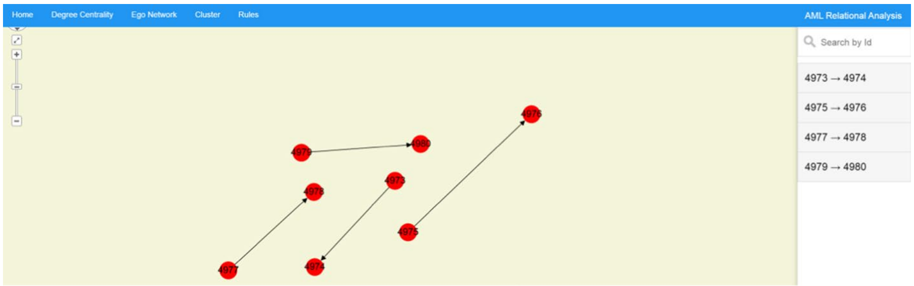
Female
6070 6071 6073 6074
6076 6077 6079 6080
6082 6083 6085 6086
6088 6089 6091 6092
6094 6095 6097 6098
6100 6101 6103 6104
6106 6107 6109 6110
6112 6113 6115 6116
6118 6119 6121 6122
6124 6125 6127 6128
6130 6131 6133 6134
6136 6137 6139 6140
6142 6143 6145 6146
6148 6149 6151 6152
6154 6155 6157 6158
6160 6161 6163 6164
6166 6167

Fig. 8 Clustering of suspicious customers based on their gender

the value of its degree centrality is the highest, i.e., 0.1327 and is highlighted with an arrow within the network.

In the 2nd step, we have implemented the ego network algorithm under the 100-node network to identify direct connections and indirect connections of suspicious customers. The following figure shows the ego network of Node ID 5043.

Figure 6 shows the ego network of Node ID 5043, which is directly connected to 7 other nodes i.e., node IDs 4986, 4999, 5015, 5019, 5031, 5055, and 5065. This direct connection means the root node is connected with under one hop. However, Node ID 5043 also has an indirect connection with other nodes highlighted in blue and light green. The node IDs are displayed in the right frame. The blue nodes are indirect



4973 → 4974
4975 → 4976
4977 → 4978
4979 → 4980

Fig. 9 Rule 1: Common ownership

connections that are under two hops with the root node and the light green nodes are under three hops away from the root node. In the same way, we can find out indirect connections up to N number hops within a network. However, we have identified indirect connections in Fig. 6 of up to 3 hops. These connections can be used to further investigate any member of a suspicious customer network and can help in identifying each suspicious customer within a gang, and each ego network will be treated as a separate case.

In the 3rd step, we have implemented a clustering algorithm to classify the sub-groups of suspicious customers based on their profiles. The following figure shows that various sub-groups of suspicious customers based on different cities.

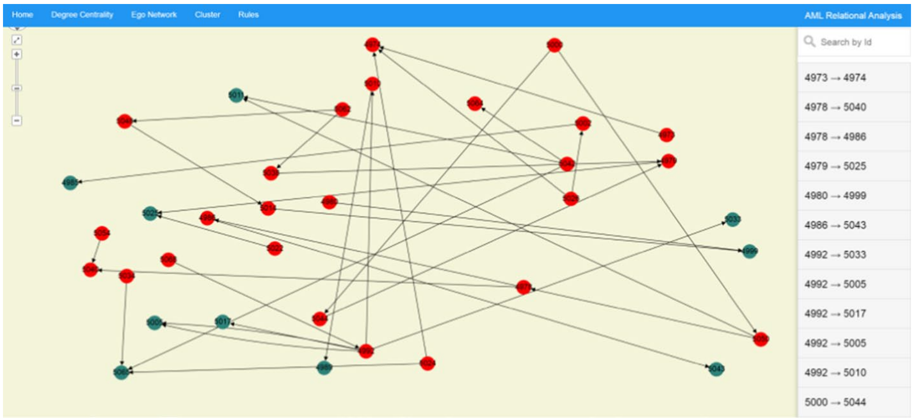
Figures 7 and 8 show the network of suspicious customers are categorized based on city and gender respectively. However, the model can also make a cluster-based on any other profile parameter, such as country, age, etc. This will help identify the characteristics and features of each sub-group, and the information can be used to predict the behavior of the group in the future. In Fig. 7, all suspicious customers are categorized into five different cities: City A, City B, City C, City D, and City E. The male and female groups are categorized in Fig. 8. The experiment utilizes a K-means algorithm for clustering the suspicious customers based on their profile and status.

In the 4th step, we have implemented the rules based on the criteria listed in Table 1. The purpose for the implementation is to identify the relation between suspicious customers and unsuspicious customers based on the customer profile criteria. The details of the implementation are as follows:

Rule 1: Common owner relationship

Under this this rule, we identify different accounts that run under the common or same ownership, meaning the same person running multiple accounts with different profiles. For instance, if two different accounts such as A1 and A2 belong to two different customers C1 and C2 in existing record and at least one of them has been identified as a suspicious customer, then the model considers a common owner relationship if there is at least one transaction between C1 and C2 and if either one of the following conditions is true:

1. if ((C1.Phone == C2.Phone && C1.FirstName + C1.LastName == C2.FirstName + C2.LastName) && (Gender of C1 = = Gender of C2) || ((C1.Address == C2.



4973 → 4974
4978 → 5040
4978 → 4986
4979 → 5025
4980 → 4999
4986 → 5043
4992 → 5033
4992 → 5005
4992 → 5017
4992 → 5005
4992 → 5010
5000 → 5044

Fig. 10 Rule 2: Business relationship

Address && C1.FirstName + C1.LastName == C2.FirstName + C2.LastName) && (Gender of C1 == Gender of C2))

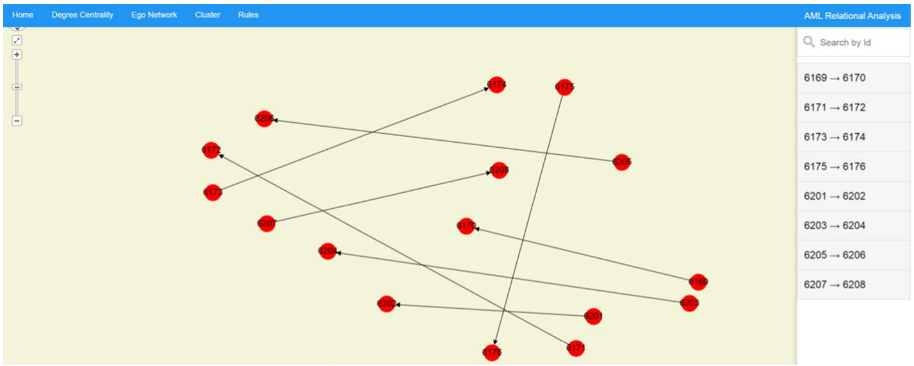
2. if ((C1.Phone == C2.Phone && C1.FirstName + C1.LastName == C2.LastName + C2.FirstName) && (Gender of C1 = = Gender of C2)) || ((C1.Address == C2.Address && C1.FirstName + C1.LastName == C2.LastName + C2.FirstName) && (Gender of C1 == Gender of C2)).

The following figure shows the results after triggering the rule:

Figure 9 shows a common owner relationship between customers 4973 and 4974, 4975 and 4976, 4977 and 4978, and customers 4979 and 4980. This rule is triggered as there is a common factor in its profile under Rule 1, either 1 or 2. As mentioned in Table 1, the proposed model identifies four pairs of customers who have utilized two different accounts, but each pair has one ownership. For instance, Customer ID 4973 is a suspicious customer and has made a transaction with Customer ID 4974. However, the first name, last name, gender, and phone number of both the customers are same; based on these criteria, the model identified both the accounts under the same ownership. In the same way, customers 4975 and 4976 have utilized two different accounts but both have a relationship of common ownership because it matches criteria 2 of Rule 1, as mentioned in Table 1, such as address with first name, last name, and gender. Under this rule, we identified the common ownership by swapping the first name with the last name as people can swap their first name with their last name and vice versa to hide their identity (ID). However, the proposed model monitored these kinds of changes and identified common ownership.

Rule 2: Business relationship

To identify the business relationship between the customers in the network, both customers should have account type B (Business), the transactions should be carried out between these two customers, and either one should be a suspicious customer. For instance, two different accounts A1 and A2 belong to two different customers C1 and C2, A1 for C1 and A2 for C2, and suppose C1 is a suspicious customer, then the model identifies the business relationship between these customers if there is at least one transaction between C1 and



6169 → 6170
6171 → 6172
6173 → 6174
6175 → 6176
6201 → 6202
6203 → 6204
6205 → 6206
6207 → 6208

Fig. 11 Rule 3: Sibling relationship

C2 and that both the customers have account type B (Business). The criteria can be represented as follows:

$$\text{if (Account of C1.Type} == \text{"B" \&\& Account of C2.Type} == \text{"B"), there is a transaction between}$$
$$\text{C1 and C2, and either one is a suspicious customer.}$$

In the above condition, both accounts A1 and A2 have an account of business type "B".

The diagram below shows the business relations between the suspicious customer and suspicious/unsuspicious customers.

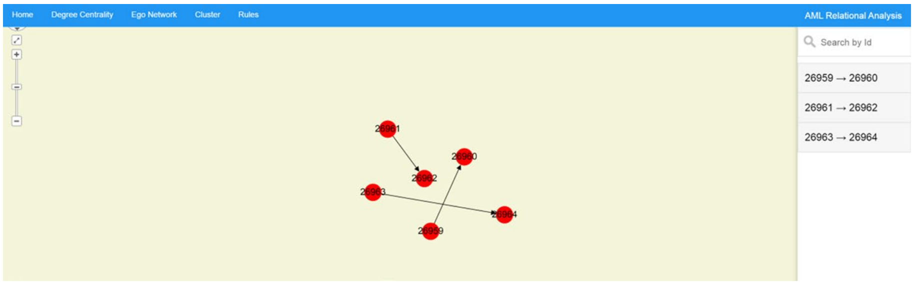
Figure 10 shows the network of suspicious and unsuspicious customers with business relations and who have performed some transactions. All the transactions between these customers are shown in Fig. 10. The arrow shows the transaction direction from the sender to the receiver. The customers are represented with their IDs. For example, Customer ID 4973 has sent money to Customer ID 4974 and both the customers have an account type B; so, based on this criteria, the model can find the business relations of suspicious customers. The model can find specific search with specific customer ID input. The results can be filtered based on the user search. With the help of this rule, we can identify the customers with a business relationship. The list of business relationships between nodes is mentioned in the right frame of Fig. 10.

Rule 3: Sibling relationship

The third rule is to identify the sibling relationship. Suppose we have two different customers C1 and C2 and at least one of them is a suspicious customer, we can say there is a sibling relationship between these customers if there is at least one transaction between C1 and C2 and the following condition is true:

$$\text{if (Address \& Lastname \& Father Name of C1} == \text{Address \& Lastname \& Father name of C2)}$$
$$\text{\&\& AgeDiff of customer 1 and customer 2} \leq 10 \text{ \&\& Father Name of C1 is not equal to First}$$
$$\text{name of C2 \&\& First name of C1 is not equal to Father name of C1} \rightarrow \text{Rs(O(C1), O(C2)).}$$

In the above condition, we set the age difference between the customers as 10 years since most of the time, the difference between the siblings is not more than 10 years. The following figure shows the results after triggering the rule:

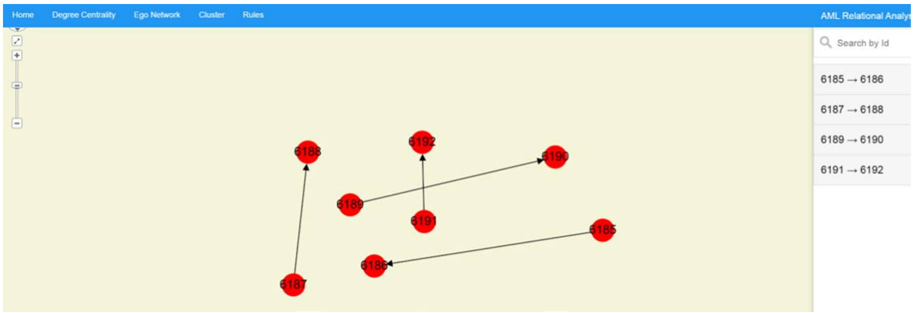


26959 → 26960

26961 → 26962

26963 → 26964

Fig. 12 Rule 4: Spouse relationship



6185 → 6186

6187 → 6188

6189 → 6190

6191 → 6192

Fig. 13 Rule 5: Mother–child relationship

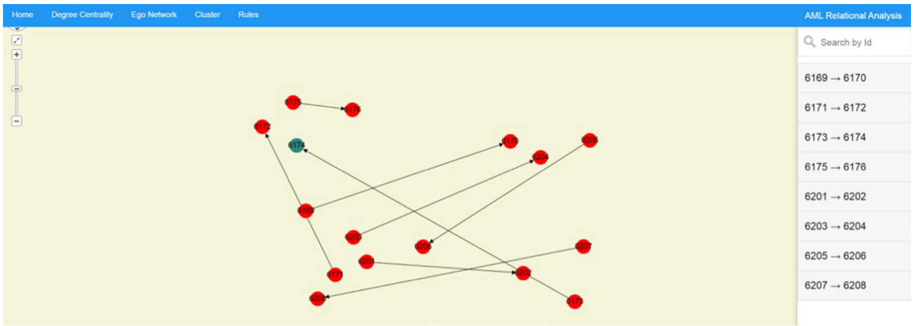
Figure 11 shows the sibling relationship between eight suspicious customers. For example, the suspicious Customer Node ID 6169 has a transaction with Node 6170 and both the customers have been identified as siblings as Node 6169 has shared the same address, last name and father name as Node 6170. Based on the criteria, Node 6171 has a sibling relationship with Node 6172 as both the nodes share the same last name, father name, address, and their age difference is under 10. Coincidently, both the customers who are siblings are suspicious in this scenario, but it is not always true that one of the sibling members could be an unsuspicious customer.

Rule 4: Spouse relationship

In this rule, the model identifies a spouse relationship of suspicious customers, For instance, we have two different customers C1 and C2 and C1 is a suspicious customer, then there is a spouse relationship if there is at least one transaction between C1 and C2 and the following condition is true:

$$\exists t(A1,\ A2)\ \&\&\ \text{if (Add \& Lastname of C1} \ == \ \text{Add \& Lastname of C2)}\ \&\&\ \text{AgeDiff} \le 20 \\ \&\&\ \text{(Gender of C1} \neq \text{ Gender of C}\ \&\&\ \text{First name and FatherName of C1} \neq \text{ Firstname and} \\ \text{Father name of C2} \rightarrow \text{Rs(O(C1),\ O(C2)).}$$

In the above condition, the age difference between C1 and C2 must be equal or less than 20 as we cover only those suspicious spouse relationships with an age difference of not more than 20 because according to the U.S. Survey [28], in heterosexual married couples, 98.7% of the population, the age difference of spouse is lesser than 20. The following figure shows the results after triggering the rule:



```
6169 → 6170
6171 → 6172
6173 → 6174
6175 → 6176
6201 → 6202
6203 → 6204
6205 → 6206
6207 → 6208
```

Fig. 14 Rule 6: SameAs relationship

Figure 12 shows a spouse relationship between Node 26,959 and Node 26,960 as both the customers have the same last name, address, age difference of 5 years, and opposite gender with different father name as the father name distinguish them from the sibling relationship. Under the same criteria, Node 26,961 and Node 26,962 and Node 26,963 and Node 26,964 exhibited spouse relationships.

Rule 5: Mother–Child relationship

This rule identifies the mother–child relationship of a suspicious customer. For instance, we have two different customers C1 and C2, and one of them is the suspicious customer, then there is a mother/child relationship if there is at least one transaction between C1 and C2 and the following condition is true:

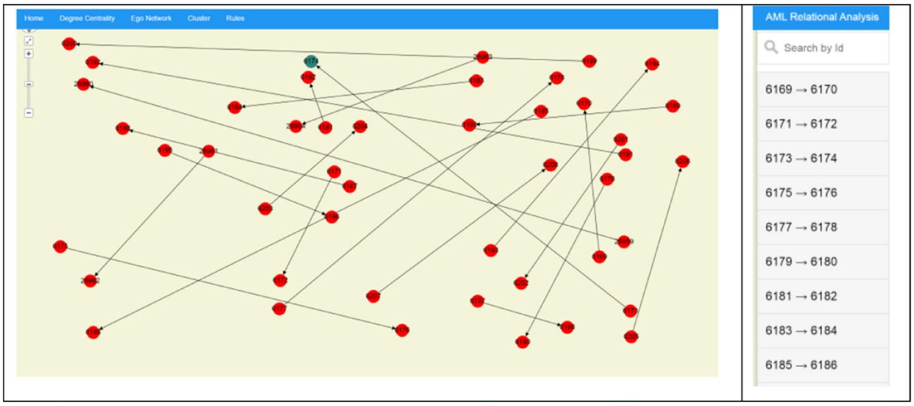
$$\begin{aligned}&\text{if (Address \& LastName of C1} \; == \; \text{Address \& LastName C2) \&\& AgeDiff} \; \geq 20\\&\text{\&\& (FirstName of C1} \neq \text{ FirstName of C2) \&\& First name of C1} \neq \text{to Father Name of C2}\\&\text{\&\& father name of C1} \neq \text{to First name of C1 \&\& agediff is} > 20\\&\left(\text{if C1 is elder than C2 \&\& age is female then C1 is mother otherwise C2 is mother}\right).\end{aligned}$$

To trigger the rule based on the above condition, the address and the last name of both the customers should be the same, the age difference between both not less than 20, and the first name of both should be different. Besides, the first name of C1 should not be equal to the father name of C2, and the father name of C1 should not be equal to the first name of C2 and the age difference between both customers should be more than 20.

The following figure shows the results after triggering Rule 5:

Figure 13 shows the result of Rule 5, where the rule is identified four mother–child relationships of suspicious customers. For example, suspicious customer Node ID 6186 was identified as the mother of customer Node ID 6185 because the former was found to be older than the latter and have a gender of a female; further customer node IDs 6186 and 6185 were seen to have the same address and last name and met the conditions defined in Table 1 for Rule 5. On the same criteria, other mother–child relations were established with the following customers:

Customer 6188 is the mother of Customer 6187, Customer 6190 is the mother of customer 6189, and Customer 6191 is the mother of Customer 6192.



AML Relational Analysis

6169 → 6170
6171 → 6172
6173 → 6174
6175 → 6176
6177 → 6178
6179 → 6180
6181 → 6182
6183 → 6184
6185 → 6186

Fig. 15 Rule 7: Family relationship

Rule 6: SameAs relationship

The SameAs relationship is identified by the model based on part of the suspicious customer profile. For instance, we have two different customers C1 and C2 and C1 is a suspicious customer, then there is a SameAs relationship between both the customers if there is at least one transaction between C1 and C2 and the following condition is true:

$$\exists t(A1,\ A2) \text{ and } C1 \in A1\ \&\ C2 \in A2\ \&\&\ \text{if}$$
$$\left(\text{LastName, address DOB \& Gender of C1} == \text{LastName, address DOB \& Gender of C2}\right)$$
$$\rightarrow \text{Rsa}(O(C1),\ O(C2))$$

The following figure shows the results after triggering Rule 6:

Figure 14 shows the SameAs relationships. The relationships are identified by triggering the rule. For instance, Customer 6169 is the same as Customer 6170 as both share the same last name, address, DOB, and gender. The model considered these customers as same but as using different profiles to open their account. In the same way, the following customers are the same:

Customer 6169 is the same as Customer 6170, and Customer 6171 is the same as Customer 6172.

Rule 7: Family relationship

The Family relationship is identified based on the customers' profile. For instance, if we have two different customers C1 and C2 and at least one of them is a suspicious customer, then we can say there is a family relationship if there is at least one transaction between C1 and C2 and the following condition is true:

$$\exists t(A1,\ A2) \text{ and } C1 \in A1\ \&\ C2 \in A2\ \&\&\ \text{if}$$
$$\left(\text{LastName, address \& race of C1} == \text{LastName, address \& race of C2}\right)$$
$$\rightarrow \text{Rf}(O(C1),\ O(C2)).$$

After triggering the rule, the following family relationships were identified:

Figure 15 shows many family relationships of suspicious customers within a network. However, we have listed only nine family relationships on the right frame such as the relationships between Customer Node ID 6169 and Customer Node ID 6170, etc., as suspicious customers' profile properties such as family name, address, and race matched with those of other customers. Similarly, the other listed nodes in the diagram were seen to have the same matched properties.

Discussion on policy implications

The Financial Action Task Force (FATF) is an official policy-making body to control money laundering globally. The efforts made by FTFS bring legislative and regulatory reforms. Due to the importance of having effective AML systems, banks have begun to venture into artificial intelligence (AI) and machine learning-based solutions for their AML solutions [29]. However, the strength of artificial intelligence and machine learning-based solutions is essentially influenced by the nature of financial data as well as the algorithms used. It is therefore crucial to apprehend the strengths and consequences of the algorithm when applied to money laundering data. This requires a separate independent review by policymakers to evaluate the risk and impact of potential failure that can occur in employing artificial intelligence, machine learning including our Social Network Analysis (SNA) based solutions. As our study utilises social network functions to identify groups involved in the money laundering, the FTFA should consider this factor when they make policy for financial institutions. The financial institutions must obtain consent from account holders regarding the usage of their data for analytics purpose so that customer profiles and their transactions can run via social network functions to identify groups who involved in Money laundering.

Conclusion and future work

Due to the current challenge in money laundering to identify the mafia of suspicious customers, this paper proposed a relational model using social networking functions. The model can identify the relationships and links of suspicious customers who are involved in the money laundering with groups. The model implemented the social network functions such as degree centrality, clustering, and ego network under the.NET environment that utilized Cytoscape, an open-source software platform, using financial data such as customer profile and their transactions to identify the groups involved in money laundering. A series of experiments was conducted to identify the relationships of suspicious customers, such as common owner relationships, business relationships, family relationships, and sibling, spouse, mother–child relationships. The results of the experiment are promising for financial institutions that can benefit from the proposed model. However, the proposed relational model has some limitations such as lack of real-world financial data and dependency of the proposed model on AML analysis module that feed the list of suspicious customers. To address these limitations, we aim to implement the proposed relational model using real-world financial institutional data and compare the results of the experiment in real-time under different network densities. It is also possible to apply the proposed model under different domain such criminal detection system and compare the results with AML relations.

Abbreviations

AML: Anti-money laundering; CRM: Customer relationship management; FATF: Financial Action Task Force; ID: Identity; KB: Knowledgebase; SNA: Social network analysis.

Acknowledgements
We would like to thank Sultan Qaboos University for granting this research and for their support in hosting our experiment on the Cloud computing environment.

Author contributions
The authors hereby confirm that we all have made a substantial contribution towards designing and implementation of this research. Primary author has contributed to writing the paper, designing and analysis sections. However, co-authors have contributed to the implementation part of this research experiment and results. We read and approved the final manuscript.

Funding
The grant number IG/EPS/INFS/18/01 is received by the primary author from an internal grant of Sultan Qaboos University Muscat to promote the academic research and to achieve the research and educational objectives of the University. The ideas and views contained in this article are from the authors and should not be interpreted as official from Sultan Qaboos University.

Availability of data and materials
The datasets used and analysed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Department of Information Systems, College of Economics & Political Science, Sultan Qaboos University, Muscat, Oman. [2] Department of Information Studies, College of Arts and Social Sciences, Sultan Qaboos University, Muscat, Oman. [3] College of Technological Innovation, Zayed University, Abu Dhabi, United Arab Emirates.

Received: 5 November 2020 Accepted: 4 January 2021
Published online: 14 January 2021

References

1. Ahmed M, Mahmood AN, Islam MR. A survey of anomaly detection techniques in financial domain. Fut Gen Comput Syst. 2016;55:278–88.
2. Rui L. et al. Research on anti-money laundering based on core decision tree algorithm. In: Control and Decision Conference (CCDC), 2011 Chinese. 2011.
3. Savage D. et al. Detection of money laundering groups using supervised learning in networks. arXiv preprint. 2016.
4. Rohit KD, Patel DB. Review on detection of suspicious transaction in anti-money laundering using data mining framework. Int J Innov Res Sci Technol. 2015;1(8):129–33.
5. Vaithilingam S, Nair M. Mapping global money laundering trends: lessons from the pace setters. Res Int Bus Finan. 2009;23(1):18–30.
6. IBM. Big data and analytical hub. 2019. https://www.ibmbigdatahub.com/tag/5474. Accessed 20 Nov 2019.
7. Tang J, Ai L. The system integration of anti-money laundering data reporting and customer relationship management in commercial banks. J Money Laundering Control. 2013;16(3):231–7.
8. Shaikh AK, Nazir A. A novel dynamic approach to identifying suspicious customers in money transactions. Int J Bus Intell Data Mining. 2020;17(2):143–58.
9. Jamali, M. and H. Abolhassani. Different aspects of social network analysis. In: IEEE/WIC/ACM international conference on web intelligence. 2006. New York: IEEE.
10. Carminati B. et al. A semantic web based framework for social network access control. In: Proceedings of the 14th ACM symposium on Access control models and technologies. New York: ACM; 2009.
11. Shaikh AK, Alhashmi SM, Parthiban R. A semantic-based centralized resource discovery model for grid computing. in: International conference on grid and distributed computing. Berlin: Springer; 2011.
12. Shum SB, Ferguson R. Social learning analytics. J Educ Technol Soc. 2012;15(3):3.
13. Soltani R. et al. A new algorithm for money laundering detection based on structural similarity. In: IEEE annual ubiquitous computing, electronics & mobile communication conference (UEMCON). New York: IEEE; 2016.
14. Fronzetti Colladon A, Remondi E. Using social network analysis to prevent money laundering. Expert Syst Appl. 2017;67:49–58.
15. Kothari A, et al. Exploring community collaborations: social network analysis as a reflective tool for public health. Syst Pract Action Res. 2014;27(2):123–37.
16. Gao S, Xu D. Conceptual modeling and development of an intelligent agent-assisted decision support system for anti-money laundering. Expert Syst Appl. 2009;36(2):1493–504.

17. Colladon AF, Remondi E. Using social network analysis to prevent money laundering. Expert Syst Appl. 2017;67:49–58.
18. Baron LF, Gomez R. Relationships and connectedness: Weak ties that help social inclusion through public access computing. Inf Technol Dev. 2013;19(4):271–95.
19. Zhou Y, et al. Proguard: Detecting malicious accounts in social-network-based online promotions. IEEE Access. 2017;5:1990–9.
20. Mukerji M, Chauhan U. A social network analysis of ICTD conferences (2006–2017). Information Technology for Development, 2019: p. 1–23.
21. Ferrara E, et al. Detecting criminal organizations in mobile phone networks. Expert Syst Appl. 2014;41(13):5733–50.
22. Chung KK, Hossain L Davis J. Exploring sociocentric and egocentric approaches for social network analysis. In: Proceedings of the 2nd international conference on knowledge management in Asia Pacific. 2005.
23. Witten DM, Tibshirani R. A framework for feature selection in clustering. J Am Stat Assoc. 2010;105(490):713–26.
24. González PC, Velásquez JD. Characterization and detection of taxpayers with false invoices using data mining techniques. Expert Syst Appl. 2013;40(5):1427–36.
25. Krishna K, Murty MN. Genetic K-means algorithm. IEEE Trans Syst Man Cybern Part B. 1999;29(3):433–9.
26. Sreedhar C, Kasiviswanath N, Reddy PC. Clustering large datasets using K-means modified inter and intra clustering (KM-I2C) in Hadoop. J Big Data. 2017;4(1):27.
27. CytoscapeConsortium. Cytoscape. 2018. https://cytoscape.org/. Accessed Nov 2018
28. Bureau USC. Current population survey, 2013 annual social and economic supplement. MD: US Census Bureau Suitland; 2013.
29. Singh C, Lin W. Can artificial intelligence, RegTech and CharityTech provide effective solutions for anti-money laundering and counter-terror financing initiatives in charitable fundraising. J Money Laundering Control. 2020.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.