Identification of money laundering accounts based on weighted capital flow network

Jianying Xiong, Haifeng Zhong
Security Management Department of Jiangxi Police College, Nanchang, China

Corresponding author and e-mail: Jianying Xiong, special8212@sohu.com

Abstract. Money laundering is an activity of transferring money through fraud and concealment. And its behaviour is different from normal transaction. The key to identifying money laundering crime is to find illegal account from massive capital flow data. The paper studies a method to identify the money laundering nodes by constructing the structural features of the capital flow network. According to the characteristics, we define the calculation methods of node outgoing, incoming and connectivity in the capital flow network, and identify the key nodes. The experimental results show that this method can find abnormal account from large-scale capital transaction data. Although the accuracy needs to be improved, it can help extract evidence of money laundering by ranking the weight and analyzing the capital flow from high-weight account.

1. Introduction
Money laundering refers to the act of camouflage and cleansing the source of black money through commercial banks, investment banks, insurance companies and other financial institutions. In this way, illegal income can be legalized [1]. A typical and complete money laundering process can be divided into three main stages: placement stage, separation stage and merger stage. In the actual operation process of money laundering, the three stages are sometimes obvious. But in most cases, they are cross used and difficult to be separated completely [2]. In general, the placement phase has an obvious feature that there is a huge amount of money on the account. Then entering the layering stage, the total amount of money no longer changes dramatically, and the frequency of capital deposit and withdrawal becomes more and more serious. The money transfer in and out from an account is frequently, and the flow direction is extremely complex [3].

Capital transaction information refers to a large number of customer data and transaction behavior data. Data statistics is the most common analysis method for capital transaction data [4]. We can directly search out the accounts with money transactions from the collected data, and make accurate statistical calculation for the total amount, frequency, average amount of single transaction and other statistics between them. The suspicious accounts based on expert’s experience after obtaining this statistical information. This method ignores the structural information of the capital transaction network, and requires manual judgment based on experience, so it is powerless in the face of large-scale data. The network data includes the liquidity of money transaction and the volatility of transaction times, how to
identify suspicious transactions from these massive data has become an effective means to combat money laundering crimes.

2. Analysis of money laundering

2.1. Money laundering process
The money laundering process is divided into three stages. The first stage is the placement stage of the money laundering process, which is to put the criminal income into the cleaning system. Then, in the layering stage, criminals transfer money frequently and repeatedly through complex transactions, so as to blur the illegal characteristics of criminal income and cover up the source and destination of criminal income. In the integration stage, the money is transferred to the accounts of legal organizations or individuals that have no obvious connection with criminal organizations or individuals, and the scattered money are reunited and become the legitimate taxable income.

2.2. Using improved RFM to evaluate account transaction value
There are abnormal features in money laundering. The financial behavior between nodes is similar to the analysis of user transaction behavior in transaction data of financial network. RFM model is an important method to measure customer value. RFM used to describe the value of a customer through three indicators: the recent purchase behavior, the overall frequency of purchase and the total consumption. In capital trading, the proximity, frequency and amount of transactions are also important indicators to measure money transfer. If there is a significant difference between the RFM value of different account, there may be abnormal transaction activities and a certain probability of money laundering. Consumer domain used five grades RFM score to identify different value users. Due to the difference between the objectives of transaction behavior evaluation and user management, the five grades cannot be marked with fine-grained, so the calculation rules are redefined [5].

R (Regency): the interval between the last transactions, it indicates that there are also ongoing transactions in the near future. The larger the R value is, the longer the transaction takes place; otherwise, the closer the transaction takes place. It is inversely proportional to the value of money transfer.

F (frequency): the number of transactions of nodes in the latest period. The larger the F value is, the more frequent the node money transfer is; otherwise, the less active the node transaction is. It is proportional to the money transfer value.

M (monetary): the amount of a node's transaction in the latest period of time. The larger the m value is, the higher the transfer money are; otherwise, the lower the transfer amount is. it is in direct proportion to the transfer value.

We defines $V_{ac}$ is account transaction value, and calculation rules are as follows:

$$V_{ac} = F \times M / R$$

3. Analysis of money laundering

3.1. Money laundering process
The process of identifying Suspicious Money Laundering nodes based on capital transfer graph mining is as Figure 1. We first need to select the appropriate transaction record attributes. Then design the corresponding graph data storage structure, in order to transform the transaction data flow of the item structure into the graph data structure. At last, we use the set correlation parameters to complete the calculation of the association characteristics and the determination of the suspicious degree of the association characteristics in the graph data structure.
The key attributes of transaction records are the premise of algorithm implementation. According to the analysis of suspicious association characteristics, the key attributes screened from the financial transaction data flow of commercial banks are shown in Table 1.

Figure 1. Suspicious node identification of Money Laundering.

Table 1. Selection of key attributes in financial transaction data flow.

| Property field      | field type | Representation symbol |
|---------------------|------------|-----------------------|
| Transaction time    | Datetime   | Dt                    |
| Transaction amount  | text       | Oc                    |
| counterparty Account| text       | Ic                    |
| Transaction amount  | numerical  | Mc                    |
| Transaction type    | text       | Pc                    |

3.2. Storage model of money transfer graph

In financial network, money is recorded in bank accounts in real name and transferred from one account to another [6]. This process of money transfer is accompanied by legal economic activities, which may also be the money launderers' operation. The capital transfer path between accounts in the banking system is directional, and these accounts constitute an open financial network. In the financial network topology, nodes represent accounts in the financial network. Adjacency table structure is chosen as the basic schema structure of financial transaction data flow. The specific structure design is as follows.

**Nodes in the graph.** If V is the account node, then \( V = \{ V_i | i=1,2,\ldots N \} \), \( n \) represents the number of nodes in the graph set.

**Edge in graph.** Capital transaction graph is a kind of directed structure. The edge of graph can be recorded as \( E(G_f) = \{ (V_k, V_m) | k,m=1,2,\ldots N \} \). \( G_f \) indicates that there has been a transaction between two node accounts.

**Weight of edge.** The weight of edge is the value index of money transfer, which is a three-dimensional vector \( (R_{km}, F_{km}, M_{km}) \) corresponding to the three measures of proximity, frequency and quota in the improved RFM model. That is, \( W(\langle u,v \rangle) = \text{Vac} \).

**Adjacency table design.** \( G = \langle V, E \rangle \) is a weight graph of capital flow network. Each edge \( u,v \) \( \in E \), \( w(\langle u,v \rangle) \) represents the transfer weights of nodes from \( u \) to \( V \), and the weight value is also stored in the adjacency table.
4. Characteristic index calculation
In the money laundering financial network, node $V_1$ is the starting point, which is the illegal asset to be disposed in the money placement stage. The node $V_n$ is consolidation node, which is the money are collected after the legal transaction turnover. There are many scattered nodes $V_x$ in the middle, which are the turnover nodes of money transfer. So the out degree of starting nodes are very high, and in degree of the merging nodes are very high. And the intermediate turnover nodes are characterized by high-frequency complexity and large amount of money transfer. If the larger the money transfer weight is, the more important the bridge node is. Tracing the fund source for this node is conducive to finding out the money laundering path.

The following indicators are used to classify different node features:

Out-degree $O_d$: it is the number of times that a vertex of the digraph serves as the starting point. This node belongs to the source point of capital flow and plays the role of capital dispersion. $O_d^k$ is the outgoing value of the k-th node, and $w(V_i, V_k)$ is the weight of nodes $i$ to $K$. The sum of the weights of all $N$ nodes is calculated.

$$O_d^k = \sum_{i=1}^{N} w(V_i, V_k)$$ (2)

In-degree $I_d$: It is the number of times that a vertex of a digraph is the end point. This node belongs to the end point of capital flow and plays the role of capital merging. $I_d^k$ is the output value of the k-th node, $w(V_k, V_i)$ is the weight of nodes $K$ to $I$, and the sum of the weights of all $N$ nodes is calculated.

$$I_d^k = \sum_{i=1}^{N} w(V_k, V_i)$$ (3)

Connectivity indicator $C_d$: The centrality of an individual reflects the importance of the point in the network. The more a node occupies such a position in the network, the more representative it is of high intermediate centrality. It can be considered that this node plays an important role in the capital network flow. $C_d^k$ is the connection value of the k-th node, but if it is only determined by one of the values, it will degenerate into a single indicator reflecting output or entry degree. It should be defined as an end node rather than an intermediate transfer node. Because both $O_d^k$ and $I_d^k$ have a positive impact on connectivity, this value is defined as the balance between the sum of the two indicators and the deviation.

$$C_d^k = O_d^k + I_d^k - |O_d^k - I_d^k|$$ (4)

5. Experiment
5.1. Data preparation
The experimental data in this paper comes from the illegal money analysis case. We extract the relevant fields, 180559 records were consolidated into a transaction flow information table, and these data is import to SQL server 2012. Some non transaction data were filtered out, duplication, blank records, small amount transactions are removed through pre-processing, 168507 data records and 2660 account numbers were sorted out. The transaction form is shown in the table 2:
Table 2. Transaction record form.

| No. | Transaction account | Transaction amount | counterparty Account | transaction date   |
|-----|---------------------|--------------------|----------------------|-------------------|
| 1   | 6214                | 4800000            | 6225                 | 2017/4/26 10:52  |
| 2   | 6236                | 3020000            | 4201                 | 2017/9/13 11:15  |
| 3   | 6226                | 2930000            | 6214                 | 2017/4/14 14:53  |
| 4   | 2710                | 2120000            | 6236                 | 2017/11/24 16:07 |

5.2. Identification results

In the experiment, the data need to be further filtered and screened due to the large number of transaction data and account nodes. By calculating the weights between nodes, set the weight filtering threshold T. in the experiment, filter the weights below 0.001max, and Max is the maximum value of FM / R in the network. The whole calculation process is completed in about ten seconds. 149 key capital outflow accounts, 164 capital flow inflow accounts and 234 non repetitive accounts were obtained.

Classify and identify these accounts, calculate the out-degree, in-degree and connectivity of different nodes through the money transfer graph index, and identify the key accounts in the money laundering network. The scatter diagram of outgoing and incoming is shown in the figure 2.

Figure 2. Money laundering financial network.

The calculation results of some typical nodes are as Table 2.

Table 3. Transaction record form.

| No. | transaction account | Out-degree | In-degree | connectivity |
|-----|---------------------|------------|-----------|--------------|
| 1   | 621691170406        | 621691170406 | 370566939006 | 741133878012 |
| 2   | 6236692080000       | 36849155261 | 144834000432 | 73698310522  |
| 3   | 713014180000        | 0          | 54566683695  | 0            |
| 4   | 6227003711110       | 136742811317| 171061863154 | 273485622634 |
| 5   | 4420150590000       | 19684284312 | 0          | 0            |
| 6   | 208822368177        | 0          | 2675081326  | 0            |
| 7   | 621691170422        | 6406933218  | 44587289892  | 89174579784  |
| 8   | 330016167830        | 41315180120 | 0          | 0            |

Through weight index sorting, different types of accounts identified through calculation are as follows:
### Table 4. Transaction record form.

|                      | merging nodes | starting nodes | intermediate turnover nodes |
|----------------------|---------------|----------------|-----------------------------|
| Number of ecognition | 31            | 13             | 210                         |
| Manual check         | 25            | 10             | 170                         |
| Accuracy rate        | 80.6%         | 76.9%          | 81.1%                       |

Some of the nodes filtered by the model are miscalculated after manual verification, and the accuracy rate is not ideal from a scientific point of view. Because the capital data flow is very complex, it is very difficult to find clues from the massive capital network. However, it is very time-consuming to find clues from the traditional methods, which requires manpower to conduct target checks one by one. Therefore, the algorithm can better filter information and assist investigation and evidence collection. Of course, there may be suspicious nodes filtered out by rules in the experiment, and the recall rate is difficult to determine. However, due to the interoperability of the money transaction network, the key information in the screening can be further associated with the account.

### 6. Conclusions

The research on anti-money laundering technology is important to financial system in modern society. In the money transaction network, which has a large amount or frequent money flows, are usually abnormal accounts. Through the analysis of the characteristics of money laundering money deposit and withdrawal, the attributes of the method should include the description of the amount of money deposit and withdrawal, the frequency of money deposit and withdrawal, and the transaction situation. In this paper, we construct account transaction network through money transaction relationship, compute the weighted graph with the transaction characteristics in the money transaction network, and introduce the value index of money transfer between accounts as the weight. It realized the identification of money laundering nodes by calculation of weighted network node characteristics.

### Acknowledgement

This research was financially supported by Jiangxi province education department, the science and technology plan project, NO.GJJ191032.

### References

[1] Shen Ze-jun, Yang Wen-yuan. Application of BP neural network under granular computing thinking in financial trend prediction. Journal of Chinese Computer Systems, 2019,40 (3):527-532.

[2] Liu Ya, Li Hua, Zheng Bing, et al. Risk identification method of guarantee circle combined with feature selection and logistic regression. Journal of Chinese Computer Systems, 2019, 40(8): 1781-1787.

[3] Simon D. Norton. Suspicion of money laundering reporting obligations: Auditor compliance, or sceptical failure to engage? Critical Perspectives on Accounting .2018, 50:56–66.
[4] Sabau AS. Survey of clustering based financial fraud detection research. Informatica Economica Journal, 2012,16(1):110-122.

[5] Foley S, Karlsen J R, Putnins TJ. Sex, drugs, and bitcoin: how much illegal activity is financed through cryptocurrencies. The Review of Financial Studies, 2019, 32(5):1798-1853.

[6] Han J G, et al. Distributed deep learning based language technologies to augment anti-money laundering investigation, Proceedings of ACL 2018, System Demonstrations, 2018:37-42.