Egonet based anomaly detection in E-bank transaction networks

Yuan Wang¹,², Liming Wang¹,* and Jing Yang¹

¹Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
²School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China

*Email: wangliming@iie.ac.cn

Abstract. While most of the risks to online banking come from outside, some abnormal accounts can also cause losses to banks or their customers. In this work, we used a data set of 3,411,486 customers collected from a top Chinese bank to mine the abnormal accounts of online banking customers. In order to better express the transaction activities between user accounts, the concept of egonet is used to represent the account relationship in the transaction network. Then, we found that the characteristics of the egonet-based account model conform to the power law distribution, and based on this, anomaly detection is performed. When this method is applied to actual transaction data, we find abnormal account relationships from the data.

1. Introduction

Today, online banking (E-banking) has become one of the financial services that most people visit frequently. At the same time, various security issues are increasing. Although people have done a lot of work on identity authentication and communication security in e-banking [1], there are still some security issues that these solutions cannot solve.

Authentication is at least in the following two situations: an account that is stolen by phishing software or created by a malicious user. Most solutions detect these anomalous behaviours by detecting deviations from previously established baselines [2], [3], [4], [5], [6]. These jobs really help solve the first problem, that is, some users have lost certificates, or their accounts have been stolen [7]. However, it does not recognize abnormal accounts for special purposes. Therefore, based on the assumption that abnormal behaviour is less than normal behaviours, the anomaly detection work in this paper analyzes whether the behaviour of the account is like that of other accounts.

Due to cooperation with a large Chinese bank, we could analyze online banking customer behaviour based on anonymous ground truth data sets. In the analysis, it is found that these transactions between accounts naturally constitute a network; one time of transaction is two nodes and an edge between them, because the transaction is only meaningful to both parties. In this paper, we propose a method to detect abnormal transaction behaviour by checking the characteristics of the egonets of user accounts in the transaction network, and then use the Mahalanobis distance to measure the degree of deviation of the user account from normal behaviour.

Main contributions of this work are:

1) By constructing a transaction network between transaction accounts, the behaviour of an account
is represented as an egonet model.

2) It is found that the characteristics of the user’s egonet model obey the power-law distribution, which gives us the confidence to declare outliers that are abnormally deviated.

3) We apply the method to real transaction data and show that it indeed spots the abnormal account relationships.

The reminder of this paper is organized as follows: In Section 2, we describe the data set and the reference methods. Section 3 examines the account behaviour and summarizes the findings we derived and implications. Section 4 presents the anomaly detection results. Section 5 discusses the related work, while Section 6 concludes this paper.

2. Data description

Online banking customers can generally access online banking through a browser to perform various financial services for individual customers or corporate customers, like various inquiries, money transfers, fee payments, investments. During the login process, each request from the customer is stored as a record in the data.

Our data comes from a very large online banking system in China that provides online services to millions of customers every day. The data contains various information about the customer’s behaviour, such as time stamp, account ID, payer ID, payee ID, operation (e.g. login, money transfer), amount login IP, login area, operation status (success or failure). Our data includes 3,411,486 customers and 23,852,308 sessions in total. A session is a period during which a client logs in to log out, using a session ID assigned by the system. We call a session as a transaction session if there exists at least one transaction (money-moving operation) in a session; otherwise call it as a non-transaction session.

3. Analysis of transaction accounts

In this section, we analyze the user accounts in the trading network and extract the egonet attributes of the user accounts from the transaction network.

3.1. Transaction network

In this section, we use the concept of interaction graph in social networks to model transaction relationships as transaction graphs. The transaction relationship is usually used to describe the transaction activity between the payer and the payee. Payer and payee are envisioned as nodes and each node can act as either a payer or a payee in one transaction activity. For two nodes x and y, the activity of node x paying money to node y can be denoted as a directed edge e = (x, y). Then, we represent the transaction network as a directed graph G = (V, E), where V is the set of all nodes and E is the set of all edges. The outdegree and indegree of node x are the number of edges going out of x and the number of edges coming into x in G, respectively. Consider that our data only includes the transfer-out record without the transfer-in record for interbank transactions, we only build the transaction graph for inner-bank transactions.

The distribution of outdegree and indegree is shown in figure 1. In this transaction graph, the outdegree follows a power-law distribution fitted with parameter a = 1.646, and the indegree with parameters a = 1.724, meaning also a heavy-tailed distribution. This power-law distribution of indegree and outdegree indicate that the transaction graph is a scale-free network in complex network theory [8]. We find that 64.83% accounts have higher outdegree than indegree. The maximum outdegree is 1,525, which means that an account has transferred to thousands of payees in 12 days. The maximum indegree is 440, which indicates that many accounts transfer funds to an account.
3.2. Feature extraction based on egonet

What statistical information/characteristics extracted from the transaction network have a significant impact on the detection results. Intuitively, features which help spot the type of anomalies we are interested in should be selected. As mentioned above, a transaction combines two accounts in the transaction network, which is applicable to the egonet model, namely the 1-step neighbourhood of a node. In this paper, we want to find user accounts with unusual operations in transaction activities. To accommodate the detection of all these anomalies, we extract 4 features based on egonet from account-level information as follows.

1) \( N_i \): number of neighbours (degree) of egonet \( i \).
2) \( E_i \): number of edges in egonet \( i \).
3) \( W_i \): total weight of egonet \( i \).
4) \( \lambda_{w,i} \): principal eigenvalue of the weighted adjacency matrix of egonet \( i \).

3.3. Laws and observations

Due to the lack of abnormal samples, the second important issue in anomaly analysis is the accurate description of normal behaviour. Therefore, it is important to find patterns for egonet-based transaction neighbours and then use them to calculate deviations.

As shown in figure 2, all features we extract follow a power law or have a power law tail at least. The power-law-type distribution is called heavy-tailed or fat-tailed if \( a < 2 \). In this case, the variance of the random variable is infinite. Furthermore, when \( a < 1 \), the mean of the random variable is also infinite [9].

In our experiments, the \( a \) in the number of nodes of every egonet is about 1.92 (Figure 2(a)); the \( a \) in the number of edges of every egonet is about 1.76(Figure 2(b)); the \( a \) in total weight of every egonet is about 1.53(Figure 2(c)); the \( a \) in principal eigenvalue of the weighted adjacency matrix of every egonet is about 1.12(Figure 2(d)). Therefore, it can be concluded that all of these features have theoretical averages, and the existence of these mean values is important because we can use them as criteria for judging user behavior anomalies.
Figure 2. Description of Transaction Network: (a) No. of Nodes of every egonet; (b) No. of Edges of every egonet; (c) Total weight of every egonet; (d) Principal eigenvalue of the weighted adjacency matrix of every egonet.

4. Abnormal transaction account detection

4.1. Anomaly detection

Usually, it is a good idea to use LOF-based detection algorithm to detect outliers far away from communities. However, as mentioned in Section 3, we find that the characteristics of accounts behaviours follow power law distributions and it also tells that we can get mean values of these characteristic variables in theory. Intuitively, the Mahalanobis distance-based algorithm has more power to detect and rank all these anomalies.

After extracting the accounts, Mahalanobis-based abnormal scores are used to estimate the abnormal of online banking accounts. Intuitively speaking, it is believed that the farther away an account’s characteristics values are from the mean, the more abnormal it will be.

4.2. Result analysis

To evaluate the detection result, we manually analyze the top 1000 results according to Mahalanobis-based abnormal scores and use the LOF-based abnormal scores in comparison.

The AUC of the ROC plot of Mahalanobis-based detection in figure 3 is 0.998, while the AUC for LOF-based is 0.821. It is shown that Mahalanobis-based detection method has a better performance than LOF-based. After analysis, it tells us an unexpected result: some accounts are tightly connected to form a community where users have abnormal money moving. These behaviours are abnormal for common customers but will not be detected by LOF-based algorithms.

Figure 3. ROC plot of Mahalanobis-based and LOF-based.

5. Related work

Because of privacy, confidentiality and business interests, there is little research on customer behaviour analysis of online banking.
Most research work is based on customer behaviour analysis to detect online banking fraud [10], [11], [12], [13], [6]. Their research usually models the behaviour of each customer and monitors whether it deviates from normal behaviour [14]. However, only a small percentage of these efforts systematically analyze customer behaviour based on actual data and perform anomaly detection. Wei et al. [6] introduced a systematic online banking fraud detection method using transaction data from a large Australian bank. His detection method is mainly through the analysis of web logs of users accessing online banking. Carminati et al. [11] developed a semi-supervised and unsupervised fraud and anomaly detection method based on a real-world dataset of a large Italian national bank.

Compared to these studies, our research relies on the transaction dataset that includes more details about the transaction. Our method focuses on building transaction relationships in transaction networks, but their approaches are based primarily on the user’s own historical behaviour. In addition, most of these works use fraudulent data generated by simulations due to the lack of publicly available real-world fraud samples. Our work based on ground truth data reveals some of the anomalous behaviour that is occurring in the online banking system.

6. Conclusion
In this work, we use the data sets collected from a top Chinese bank to mine the abnormal accounts of online banking customers. The online banking accounts were analyzed under transaction network, which focuses on account transaction activities. Then, we propose a method that based on egonet model to extract the egonet characteristics of each account. It shows that egonets of accounts obey some patterns, which gives us the confidence to declare as outliers the ones that deviate. We apply the method to real transaction data and show that it indeed spots the abnormal behaviours like abnormal community connected by transactions.

References
[1] S. Kiljan, K. Simoens, D. D. Cock, M. V. Eekelen, and H. Vranken. A survey of authentication and communications security in online banking. ACM Computing Surveys (CSUR), 49(4):61, 2017.
[2] A. Abdallah, M. A. Maarof, and A. Zainal. Fraud detection system: A survey. Journal of Network and Computer Applications, 68:90–113, 2016.
[3] N. A. Azeem Ush Shan Khan and M. N. Qureshi. Real-time creditcard fraud detection using artificial neural network tuned by simulated annealing algorithm. Institute of Doctors Engineers and Scientists, ACEEE (A Computer division of IDES), 2014.
[4] S. Pulakkazhy and R. Balan. Data mining in banking and its applications-a review. Journal of Computer Science, 9(10):1252–59, 2013.
[5] J. T. S. Quah and M. Sriganesh. Real-time credit card fraud detection using computational intelligence. Expert Syst. Appl., 35(4):1721–32, 2008.
[6] W. Wei, J. Li, L. Cao, Y. Ou, and J. Chen. Effective detection of sophisticated online banking fraud on extremely imbalanced data. World Wide Web, 16(4):449–75, 2013.
[7] Li S, Romdhani I, Buchanan W. Password pattern and vulnerability analysis for web and mobile applications[J]. ZTE Communications, 2016, 14(S0): 32-36.
[8] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. 2007.
[9] J. Gao, Y. Cao, W.-w. Tung, and J. Hu. Multiscale analysis of complex time series: integration of chaos and random fractal theory, and beyond. John Wiley & Sons, 2007.
[10] G. Cabanes, Y. Bennani, and N. Grozavu. Unsupervised learning for analyzing the dynamic behavior of online banking fraud. In 2013 IEEE 13th International Conference on Data Mining Workshops (ICDMW), pages 513–20. IEEE, 2013.
[11] M. Carminati, R. Caron, F. Maggi, I. Epifani, and S. Zanero. Banksealer: A decision support system for online banking fraud analysis and investigation. Computers & Security, 53:175–86, 2015.
[12] K. N. Karlsen and T. Killingberg. Profile based intrusion detection for internet banking systems. 2008.

[13] S. Kovach and W. V. Ruggiero. Online banking fraud detection based on local and global behavior. In Proc. of the Fifth International Conference on Digital Society, Guadeloupe, France, pages 166–71, 2011.

[14] V. Jyothsna, V. R. Prasad, and K. M. Prasad. A review of anomaly based intrusion detection systems. International Journal of Computer Applications, 28(7):26–35, 2011.