A Study of Bitcoin De-Anonymization: Graph and Multidimensional Data Analysis

Xingyu Lv
School of Computer Science
Guangzhou University
Guangzhou, China
2111806039@e.gzhu.edu.cn

Ye Zhong
School of Computer Science
Guangzhou University
Guangzhou, China
2111906109@ e.gzhu.edu.cn

Qingfeng Tan
School of Computer Science
Guangzhou University
Guangzhou, China
tqf528@gzhu.edu.cn

***Abstract*—Bitcoin was designed to be a decentralized global electronic payment system that does not require verification by a third-party intermediary platform and can be used by anyone originally. Due to its anonymity and globalization, bitcoin has achieved great success and attracted the attention of various illegal traders. In recent years, the number of illegal transactions of bitcoin has been increasing. Although bitcoin can support a certain amount of privacy, the bitcoin users and entity information can be linked by tracking the on-chain information of bitcoin users and combining the public off-chain information. Through bitcoin users de-anonymization, we can obtain some valuable intelligence information, which plays an important role in combating bitcoin-related crimes. In this paper, we build a visual analysis system for bitcoin transactions based on a graph database and use real-world multi-dimensional data sources to analyze the entity information of bitcoin transactions on the chain to achieve the effect of de-anonymization. Besides, we adopt a supervised learning method in our system to predict the legitimacy of unknown bitcoin transactions. Experiments and analyses show that our system can achieve good correlation analysis and de-anonymization. Finally, we put forward the future research direction of the bitcoin de-anonymization field.**

Keywords—bitcoin, multidimensional data, de-anonymization, graph, machine learning

I. INTRODUCTION

Bitcoin is a cryptocurrency proposed by S. Nakamoto in 2008[1]. It mainly realizes capital flow through pseudonyms, that is, anonymous bitcoin address, without revealing any personal information. A bitcoin address consists of a string of numbers and letters and it can be shared with anyone who wants to transfer you bitcoin. The anonymity of bitcoin protects the privacy of users to a certain extent so that it has attracted the attention of many criminals and has been widely used in illegal transactions such as dark web, ransomware, and so on, which makes it difficult to combat crime. Foley found that about 25% of bitcoin users and 44% of bitcoin transactions are linked to illegal economic activity[2]. According to a new survey by the PeckShield security team[3], the total number of bitcoin flowing into the dark web was approximately 330000 in 2018 and raised to about 540000 in 2019.

However, bitcoin is semi-anonymous and researchers can still link to bitcoin account entity information through various methods. After you participated in a transaction the bitcoin network, you left two traces which can be classified as on-chain information and off-chain information[4]. The on-chain and off-chain information can not only reveal the relationship between yours and others' transactions but also associate your entity identity information with transaction information to achieve the de-anonymization of bitcoin transactions. Therefore, many studies have begun to analyze bitcoin transactions and achieve bitcoin users de-anonymization with the help of necessary data sources such as internet sites, dark web, ransomware, etc. By this means, some valuable information has been obtained, which helps crackdown on some unlawful acts related to bitcoin such as criminal hacking, the use of digital currencies for money laundering, and other illegal and criminal acts, thereby guiding the healthy development of the blockchain. However, some deficiencies exist in most of these studies, because the data source they obtained of the bitcoin on-chain data and the off-chain information is not complete and the overall thinking about bitcoin de-anonymization is lacking, etc.

A. Contribution

In this paper, we mainly discuss bitcoin transaction analysis and the de-anonymization of bitcoin users. Some previous surveys of bitcoin anonymity and privacy have also been conducted. In [5], the authors provided a discussion of various cryptocurrencies and a preliminary overview of the advantages and disadvantages of using bitcoin. [6] presented a comprehensive technical survey on decentralized digital currencies, mainly focusing on bitcoin and its related technologies. Reid analyzed bitcoin's anonymity and explored the considerations of bitcoin anonymity[7]. The authors provided a survey of the security and privacy of bitcoin in [8]. Overall, their investigations focus on the elaboration of the principles of bitcoin privacy and anonymity, which have some enlightening effects on subsequent researches. However, a summary of technical work related to bitcoin transaction analysis is lacking. Therefore, we firmly believe that a comprehensive study of the bitcoin de-anonymization technic-al work is vital for those who plan to research this direction. Particularly, the main contributions of our work are as follows.

- We discussed the characteristics of bitcoin transaction and its anonymity, and relatively comprehensively summarized the technical work related to bitcoin transaction analysis and de-anonymization, including graph data analysis, address clustering, and related machine learning methods.
- We build a bitcoin transaction analysis system based on the neo4j graph database to discuss the key role of graph analysis and multidimensional data sources in obtaining the intelligence behind bitcoin transactions and also make an experimental analysis.

978-1-7281-9558-2/20/$31.00 ©2020 IEEE
DOI 10.1109/DSC50466.2020.00059

- We train machine learning models based on the characteristics of the graph data to predict the legitimacy of unknown bitcoin transactions. This method can reduce the workload of manually identifying abnormal transactions in bitcoin, which is a feasible research direction in the future.

B. Organization

The rest of this paper is organized as follows. In Section II, we briefly introduced the related work of bitcoin transaction analysis and de-anonymization. In Section III, we discuss the background concepts including transaction and anonymity of the bitcoin system. In Section IV, we discuss the key technologies for de-anonymizing Bitcoin. We provide a case of de-anonymization through graph analysis combined with multidimensional data and we obtained some valuable intelligence information. At the same time, we also train a dataset of bitcoin addresses with specific labels in machine learning models to explore the models' effect on bitcoin de-anonymization and propose the future research direction. Finally, we conclude the paper in Section V.

II. RELATED WORK

Bitcoin transaction analysis and de-anonymity are conducive to obtaining and offering high-value intelligence information related to illegal transactions for related departments, therefore there are some studies at home and abroad. The related work in this paper can be summarized into two parts: graph analysis and machine learning related methods including address clustering.

A. Graph analysis

Bitcoin transactions are interrelated, so researchers transformed the analysis of transaction data into graph analysis and formed the bitcoin's transactions graph, users graph, and network graph. After analyzing previously mentioned graphs, we can further analyze and de-anonymize bitcoin transactions and calculate the shortest path between bitcoin addresses, etc.

Reid established a "transactions graph" and a "users graph" through the topological structure to de-anonymize the bitcoin flow presented in the graph and combined these structures with external information and network measurement technologies such as traffic analysis, aiming at exploring the privacy issues brought about by the pseudo-anonymity of the bitcoin system[9]. His work verified the feasibility of the network graph analysis method. Ron analyzed all bitcoin transaction data as of May 13, 2012 and introduced the "Entity" concept. He got the "Entity Network" through using the transaction graph model based on a parallel check algorithm[10]. Based on predecessors, Fleder proposed a graphical analysis framework, which combines with publicly available information such as bitcoin forum users to de-anonymize the user's identity information[11]. Haslhofer proposed an open-source framework named GraphSense [12] which is a solution that applies a graph-centric view to digital currency transactions and it integrates some analysis tools such as BlockSci[13]. It supports path and map mode search that allows users to explore transactions, track the flow of funds, and facilitate analysis by enriching transaction graphs semantically. GraphSense proved to be effective and was used in subsequent researches. For instance, Masarah implemented a bitcoin transaction analysis on GraphSense [14].

B. machine learning methods including address clustering

Address clustering is mainly an attempt to build a one-to-many mapping from entities to addresses in the bitcoin system through grouping address cluster, its main methods include heuristic rules such as multi-in-transaction, change address, etc.

[15] proposed a tool named "Bitiodine" that realized two heuristic clustering algorithms("N-1" transactions and change transactions) for account addresses and combined web crawlers to achieve traceability of transactions and de-anonymization of account addresses. In [16], Meiklejohn used a stain analysis technology and heuristic clustering algorithm on the base of Bitiodine and summarized the rules of address clustering. BTCScope was proposed by Zhen Zhang, which uses network graph analysis to perform topology clustering and community detection to analyze bitcoin transaction data[17]. Remy used the community discovery method in a complex network and he combined the heuristic clustering method with the Louvain algorithm to analyze the bitcoin transaction situation, which was proved to be effective [18].

In the field of machine learning methods, Jason combined the K-means algorithm with the unsupervised learning algorithm "RolX" to analyze the dataset of bitcoin transactions. This work not only summarized some inter-esting abnormal behaviors but also explored the application of machine learning algorithms in bitcoin transaction anal-ysis[19]. There is also a study with novel methods such as supervised machine learning to reduce the anonymity of the bitcoin blockchain to predict unidentified entity types[20]. At the same time, other machine learning methods such as unsupervised clustering are also applied to the deanonym-ization of bitcoin users[21].

These works above have accumulated a lot of useful features for bitcoin users clustering and correlation analysis, therefore we can combine them with graph and multidimen-sional data analysis. The de-anonymization effect of these methods depends to a large extent on the choice of machine learning algorithm models by researchers, however, it does not affect the machine learning work to become a popular direction in this field. We will discuss this in Part IV.

III. BACKGROUND

A. Bitcoin Transaction

Bitcoin transactions are like lines in a double-entry book. In simple terms, each transaction contains one or more "inputs", which represent the transfer of bitcoin from the bitcoin account. A transaction also has one or more "outputs" that are credited to the Bitcoin account as credits. A transaction also contains the proof of ownership of each transferred bitcoin which exists in the form of the owner's digital signature.

The transaction is the flow of bitcoin from the inputs to the outputs. The inputs refer to the source of the bitcoin which usually comes from the outputs of a previous transaction, and the outputs of the transaction mean that the new owner receives the bitcoin by associating keys. A signature is required to redeem bitcoins in future transactions to prevent double-spending attacks on bitcoin.

The output of a transaction can be used as the input of another new transaction, thus forming a chain of ownership as bitcoin is moved from one address to another[24].

B. Bitcoin address and Anonymity

Bitcoin address is a string of numbers and letters that can be shared with anyone who wants to transfer you bitcoin. For example, address"13tPtyd8itHf6cNGzauttfcrpYQoXsDXyy" is a bitcoin address for ransomware. Only such addresses are displayed in the public on-chain transaction data and the specific information of the entities behind is unknown to others. However, many recent studies prove that bitcoin anonymity is limited because they have confirmed that it is possible to effectively de-anonymize bitcoin users by analyzing a wealth of information from the continuous transaction. Bitcoin transaction information is fully public and everyone including attackers can obtain the address information of bitcoin senders and receivers by deploying a client or public blockchain browsers[25]. Attackers can de-anonymize bitcoin users rely on the address information and necessary data sources such as Internet websites or Dark Web[26]. The destruction of the famous Dark Web Silk Road website is a successful example[27]. As a result, bitcoin is not fully anonymous and transparent completely. Challenges of bitcoin anonymity and privacy are in a gray area: the exposure of users' financial activities ultimately depends on the capabilities of the investigators and the complexity of the tools selected by the investigators.

C. Graph database

The graph database is non-relational and it uses graph theory to store the relationship information between entities. The graph database has two main components, the nodes set and the relationship among nodes. The most common example in bitcoin is the relationship between users in the bitcoin transactions network. Common graph databases include Neo4j,FlockDB, GraphDB, etc. In many intelligence and user traceability work, it is important to look for a pattern of events. A single node may seem harmless, but the relationship between it and other nodes can reveal a lot of valuable information.

D. Multilayer perceptron and Random Forset

Multi-layer perceptron (Multilayer Perceptron, MLP) is also called Artificial Neural Network (ANN)., there are multiple hidden layers in between the input and output layers in MLP. The simplest MLP needs a hidden layer, it is called a simple neural network. The most basic problem to be solved by neural networks is the classification problem. We pass the eigenvalues into the hidden layer and use the data with the results to train the parameters of the neural network (W, weight; b, bias) to make the output values consistent with the results we give, which can be used to predict new input value. The common activation functions of MLP include ReLU in equation 1 and sigmoid functions in equation 2.

$$\mathrm{ReLU}(x) = \max(x,0) \tag{1}$$

$$\mathrm{Sigmoid}(x) = 1/(1+\exp(-x)) \tag{2}$$

Random forest is a more advanced algorithm based on the decision tree. Like decision trees, random forests can be used for both regression and classification. As you can see from the name, random forest is a forest constructed in a random way, and this forest is composed of many unrelated decision trees. Random forest in real time belongs to a very important branch of machine learning called ensemble learning. Ensemble learning solves a single prediction problem by establishing a combination of several models. It works by generating multiple classifiers or models to make predictions independently. These predictions are finally combined into a single prediction, so it is better than any single classification to make predictions.

E. Accuracy,Recall and F1-score

In the bitcoin transactions classification tasks, that is, the problem of predicting discrete values, the most commonly used evaluation indicators is the accuracy. The accuracy is the ratio of the number of correctly classified samples to the total number of samples.For m samples,the formula for accuracy is as equation 3:

$$\mathrm{acc}(f; D) = \frac{1}{m}\sum_{i=1}^{m} I(f(x_i) = y_i) \tag{3}$$

The accuracy rate is based on our prediction results. It indicates how many of the positive prediction samples are true positive samples. Then there are two possibilities for predicting positive, one is to predict the positive class as the positive class (TP), and the other is to predict the negative class as the positive class (FP),which is:

$$P=TP/(TP+FP) \tag{4}$$

The recall rate refers to our original sample, which indicates how many positive examples in the sample were predicted correctly. There are also two possibilities, one is to predict the original positive class as a positive class (TP), and the other is to predict the original positive class as a negative class (FN),which is described as equation 5:

$$R=TP/(TP+FN) \tag{5}$$

Accuracy and recall are a pair of contradictory measures. If you want to make the recognition result as close as possible to the bitcoin users transaction behavior, you can only recognize some criminal behavior with high grasp, so that it may be easy to miss some valuable information. For novel crimes such as using digital currency to launder money, the recall rate is low at this time; if we want all crimes to be recognized, the recognition system will be particularly sensitive at this time, and will be recognized as criminal behaviors for some unusual behaviors. The accuracy rate is very low. Although it is possible to identify rare and novel criminal acts, it greatly increases the follow-up workload.

The P in equation 4 and the R in equation 5 indicators sometimes conflict, so we need to consider them comprehensively, the most common method is F-Measure, also known as F-Score. F-Measure is the weighted harmonic average of P and R, namely:

$$F_\beta = \frac{(1+\beta^2)\times P\times R}{(\beta^2\times P)+R} \tag{6}$$

When β=1, which is a common F1 metric,it is the harmonic average of P and R, when F1 is higher, the better the performance of the model. Among them, M is the total number of samples.

$$F1 = \frac{2\times P\times R}{P+R} = \frac{2\times TP}{M+TP-TN} \tag{7}$$

IV. EXPERIMENTS AND ANALYSIS

In this section, we firstly use an example to verify the effectiveness of the bitcoin transaction graph visualization system we constructed, which highlights the key role of graph analysis and multidimensional data analysis in obtaining bitcoin transaction intelligence. At the same time, we use some machine learning models in our system to train a famous dataset and we explore future research directions in the end.

A. Graph analysis & multi-scale data analysis

Firstly, we deployed a full bitcoin node on a Linux machine(RAM>256G, Storage capacity>30T) and we took advantage of Greg's work[28] to synchronize bitcoin transactions to the neo4j database. Through the graph visualization, we can easily calculate the path between two addresses, as shown in *Fig. 1*.

Then, we use the graph-based system to analyze a cold-storage wallet address(12rU….dj7) we obtained from an illegal trading forum. According to the statistics of the top-n transactions we did, we learned that this address has been active in the recent period and there is a possibility of involving a Ponzi scheme. We obtain the transactions and related addresses associated with the target address through the method of graph analysis as shown in Table 1. For instance, we can get the address, transferred amount. Additionally, we also combine multidimensional data such as wallet fingerprint, exchange information[29], forum Information[30]to carry out the de-anonymization of the bitcoin address.

After sorting out funds at two addresses (19xHiA…and 1PFtrR…), the bitcoins were decentralized and transferred to multiple exchanges. The flow of the first few exchanges

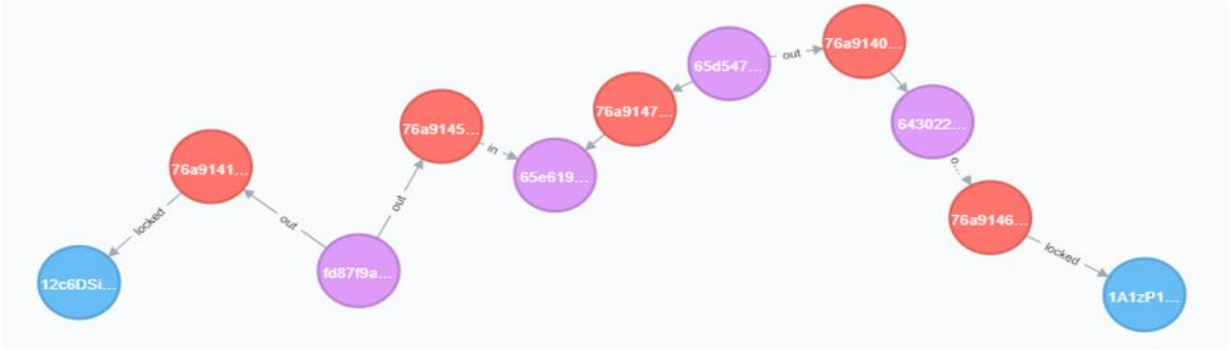

Fig. 1. The path between addresses

TABLE.1 TYPE STYLES TARGET ADDRESS CAPITAL FLOW INFORMATION OBTAINED THROUGH GRAPH ANALYSIS AND MULTIDIMENSIONAL DATA SOURCES

Address_ In	*Address _Out*	*Value (BTC)*	*Wallet fingerprint*	*Active time*
12rU….dj7	1PFtrR…..r6Q8Q	11107	877e…4a8	2018/05/22~2020/02/13
12rU….dj7	19xHiA…..Uembv	8009	0009..1dc	2018/10/12~2020/01/29
12rU….dj7	363sZd…..XRQuv	55	1486...173	2020/02/13~2020/02/29

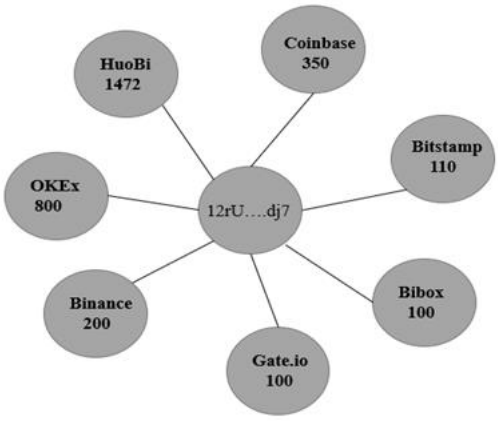

Fig. 2. Exchange flow of target address

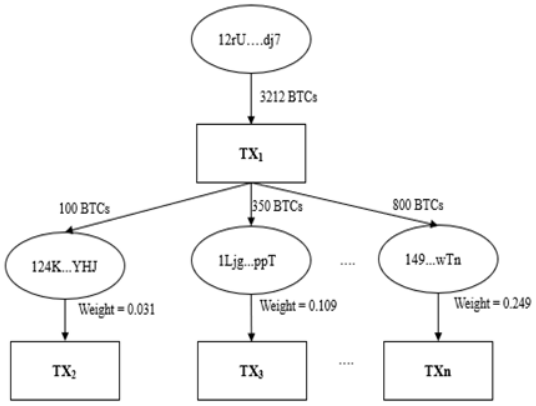

Fig. 3. Weight analysis of target address with equation 8

Fig. 4. Crawl the forum information leaked by the target address. For security reasons, the picture has been blurred.

in *Fig. 2*, it should be noted that each bitcoin amount was transferred to the exchange is huge. In the end, these bitcoins make profits through offline transactions. In addition, in order to more clearly quantify the flow of funds transferred to this bitcoin address and mine the address transaction mode associated with it. We introduce a bitcoin address weight analysis, which we define as the percentage of bitcoin transferred from the input bitcoin address to each target bitcoin address in equation 8.

$$\text{Weight\_addr}_{m,t} = \sum_{j} \Pi_{pt \in N_j^{t,m}} \frac{outtx_{pt,next}}{\sum_i outtx_{pt,i}} \quad (8)$$

(**Note:**N_j: the *j*th transaction in transactions to address m, pt is a transaction in $N_j^{t,m}$, $outtx_{pt,i}$ is the bitcoin amount of the *i*th transaction in the output transactions, $outtx_{pt,next}$ is the amount of the output transaction in the next transaction)

In the equation 8, for the unspent transaction outputs at address a, we can calculate the corresponding weight if we know the transferred amount between it and the address m, so that we can focus on the analysis of some transactions associated with the address, to obtain some valuable information. Because there are many transactions for an address, it is more meaningful for us to only obtain transactions for a certain period for analysis, and the address is not fixed, there are alternative addresses, so this weight analysis method brings us to the de-anonymization work great convenience.

In Figure 3, we conducted a weighted analysis of the flow of funds to the target address and judged the address of focus by the proportion of the next transfer amount in the total expenditure amount. Our de-anonymization work mainly focuses on the associated addresses with relatively large weights, and through further data mining to connect the entity information such as social media information, a better de-anonymization effect can be achieved.

In the meantime, we get the relevant information of the target address in the forum through data mining as shown in Fig. 4. We can learn that the owner of the target address operated a centralized bitcoin fundraising project and was suspected of defrauding investors' bitcoins. It is conceivable that if timely warnings and prompts can be made, the relevant investors will withdraw bitcoin funds from the project, thereby reducing losses.

We briefly summarize the processing steps of this experiment. (1) **Transaction graph visualization:** That is graph visualization of Bitcoin's on-chain transactions. Here we choose the graph database. Compared with traditional relational databases, graph databases have some path analysis and other benefits. (2) **Bitcoin addresses Mining**: The source of the target address for the de-anonymization experiment we did is mainly through third-party data sources, page extraction, and news reports.(3)**User address association:** The main goal is to associate multiple cryptocurrency addresses of the same entity. (4) **Bitcoin users de-anonymization:** By associating entity information, we tracked the capital flow of the exchange, which achieved our desired goal, because in actual supervision of many exchanges will carry out corresponding KYC certification. As long as we can obtain the exchanges' flow of bitcoin users' funds, the de-anonymization of Bitcoin has achieved great success. As can be seen from the above case, graph analysis is a simple operation to make visualizations of bitcoin transactions. Through graph analysis, it is possible to more directly obtain the association between bitcoin transactions, users, and entities. Surely, graph analysis alone is not enough, we can also use external information such as website forums, the dark web, and other multidimensional data for analysis and apply some natural language processing techniques for semantic and association analysis. Therefore, we hope to explore the de-anonymization of Bitcoin users by combining machine learning methods, which is why we will do the next part of the experiment.

B. Training based on transaction graph data

In this experiment, we used the dataset from ELLIPTIC[31] which includes a time series graph containing more than 200000 bitcoin transactions (nodes), 234000 directed payment streams (edges), and 166 node features including non-public based characteristics of the data to train the machine learning model. Suppose that the bitcoin transaction graph from the dataset is described as equation 9, Among them, N is the node transactions set, and E is the edges set representing the BTC flow.:

$$G=(N,E) \quad (9)$$

We regard the structure of the entire transaction as a network graph, which is conducive to graph analysis. We trained two supervised classifiers based on Pytorch including multilayer perceptron and random forest models for training features based on single transaction and features based on transaction graph. Our results are 10x cross-validated based on mean squared error and we use layered random sampling, split the training set and validation set. The training result is shown in Table 2 and is represented by a line chart as shown in Fig. 5.

As can be seen from the above results, after constructing features based on the transaction graph, the effect of accuracy in equation 3 and F1-score in equation 7 in classifiers improve obviously. Consequently, we can get the point that the features based on transaction graph analysis are more effective in the training of machine learning models. From the above experiment, we learn that we can train a classifier that automatically recognizes the behavior of bitcoin users if we construct a transaction graph of bitcoin users addresses and obtain the source of multidimensional data. But the accuracy of the classifier still depends on how effective the multidimensional data labels we can get and the machine learning model we choose. Moreover, we believe that machine learning has at least several applications in the work

of de-anonymizing bitcoin users' addresses: (1) **Application of knowledge maps:** Constructing a knowledge map of digital currency user addresses by transforming raw data into structured graph storage is helpful for further association analysis. When we search for an address, it is very easy to obtain the entity information of the user address, which greatly improves the efficiency of de-anonymization work. (2) **Improve illegal address identification:** For example, based on the regular extraction of cryptocurrencies users addresses, we can determine whether the digital currency address belongs to an illegal service provider which base on local semantic features and global non-semantic features. (3) **Users addresses correlation analysis field:** By using the graph convolutional neural networks for deep learning of cryptocurrency transaction graphs, we can automatically learn the individual and structural characteristics of transaction graphs (4) **Malicious traffic analysis:**The machine learning method is used to analyze the monitored blockchain network behavior and traffic, which is helpful to identify abnormal transactions in the bitcoin network behavior and so on. Predictively, with the continuous development of cryptocurrency technology and the need for regulatory supervision, combining machine learning models with bitcoin de-anonymization will be a research hotspot shortly in the future. This method aims to build an automated intelligence system to analyze bitcoin transactions.

TABLE.2 MODEL TRAINING RESULTS

Features	Model	Precision	Recall	F1-score
Features based on single transaction	Multilayer Perceptron	0.638	0.661	0.649
	Random Forest	0.803	0.611	0.694
Features based on transaction graph	Multilayer Perceptron	0.840	0.57	0.679
	Random Forest	0.983	0.651	0.782

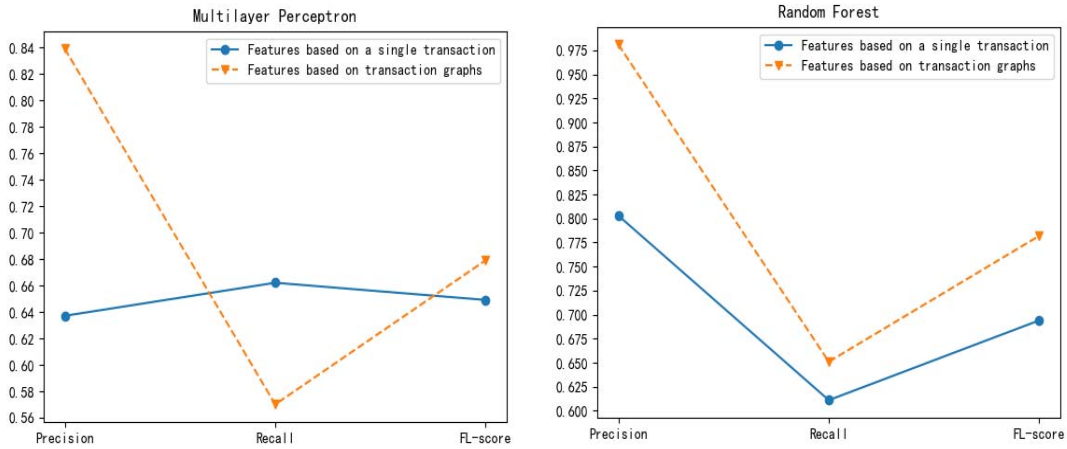

Fig. 5. Comparison of training effect between single transaction feature and transaction graph feature

V. CONCLUSION

In this paper, we focus on the study of bitcoin transaction analysis and the de-anonymization of bitcoin users. Firstly, we outline the basic concepts of the bitcoin transaction and bitcoin anonymity. Secondly, we study a case of suspected scam bitcoin address through the graph analysis system we build with external multidimensional data sources. At the same time, we also use the weight analysis method to obtain the associated users who need to focus on, to reduce the complexity of the de-anonymization work. Moreover, we explore the feasibility of machine learning algorithms for de-anonymizing bitcoin users and corroborate them with experiments and the method achieves good prediction results. Finally, we propose future research directions.

We highlight the key role of graph analysis and multi-dimensional data sources in obtaining the intelligence behind digital currency transactions. And we believe that in the future, the work of cryptocurrencies de-anonymization especially bitcoin will be more automated and effective.

REFERENCES

[1] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System" Online Available: http://bitcoin.org/bitcoin.pdf, 2008.

[2] Foley, Sean,Karlsen, Jonathan,Putnins, Talis, "Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed through Cryptocurrencies?" Review of Financial Studies,2019

[3] PeckShield Security Team, "2019 global digital asset anti-Money lau ndering research report", 2019.

[4] F Reid, M Harrigan, "analysis of anonymity in the bitcoin system Security and privacy in social networks" Springer, 2013.

[5] J. Bonneau, “SoK: Research perspectives and challenges for Bitcoin and cryptocurrencies” Proc. IEEE Symp. Security Privacy, San Jose, CA,USA, pp. 104–121, May 2015.

[6] F. Tschorsch and B. Scheuermann, “Bitcoin and beyond: A technical survey on decentralized digital currencies,” IEEE Commun. Surveys Tuts., vol. 18, no. 3, pp. 2084–2123, 3rd Quart, 2016.

[7] Reid F, Harrigan M. “An Analysis of Anonymity in the Bitcoin System” IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing, pp. 1318-1326,2011.

[8] Mauro Conti , E. Sandeep Kumar,Chhagan Lal , “A Survey on Security and Privacy Issues of Bitcoin” IEEE Communications Surveys & Tutorials, Volume: 20, Issue: 4, Fourthquarter ,2018.

[9] Reid F, Harrigan M, “An Analysis of Anonymity in the Bitcoin System” IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing. IEEE,pp. 1318-1326, 2011

[10] Ron D, Shamir, “A. Quantitative Analysis of the Full Bitcoin Transaction Graph” International Conference on Financial Cryptography and Data Security,Springer:Berlin, Heidelberg, 2013

[11] Michael Fleder, Michael S. Kester,Sudeep Pillai,“Bitcoin Transaction Graph Analysis” CoRR abs/1502.01657,2015.

[12] Graphsense,Online Available:https://graphsense.info/

[13] Harry A. Kalodner, Steven Goldfeder, Alishah Chator, Malte Möser, Arvind Narayanan, “BlockSci: Design and applications of a blockchain analysis platform” CoRR abs/1709.02489,2017.

[14] Paquet-Clouston M,Haslhofer B,Dupont B, “Ransomware payments in the bitcoin ecosystem.” Journal of Cybersecurity.

[15] Spagnuolo M, Maggi F, Zanero S,“Bitiodine: Extracting intelligence from the bitcoinnetwork” International Conference on Financial Cryptography and Data Security, Springer:Berlin,Heidelberg,pp.457-468, 2014.

[16] Meiklejohn S, Pomarole M, Jordan G, “A Fistful of Bitcoins: Charact erizing Payments among Men with No Names” Proceedings of the 20 13 conference on Internet measurement conference ACM,pp.127-140, 2013.

[17] Zhen Zhang, Tianyi Zhou, “BITSCOPE: Scaling Bitcoin Address De-anonymization using Multi-Resolution Clustering”, Proceedings of the 51st Hawaii International Conference on System Sciences,2018.

[18] Remy C, Rym B, Matthieu L, “Tracking bitcoin users activity using community detection on a network of weak signals” International conference on complex networks and their applications,Springer: Cham,pp. 166-177, 2017.

[19] Hirshman,Jason, “Unsupervised Approaches to Detecting Anomalous Behavior in the Bitcoin Transaction Network.”, 2013.

[20] Mikkel Alexander Harlev, Haohua Sun Yin, “Breaking Bad: De-Anonymising Entity Types on the Bitcoin Blockchain Using Supervised Machine Learning”, HICSS,pp.1-10,2018.

[21] Thai Pham, Steven Lee, “Anomaly Detection in Bitcoin Network Using Unsupervised Learning Methods” CoRR abs/1611.03941, 2016.

[22] Rahouti, Mohamed Xiong, Kaiqi Ghani, Nasir, “Bitcoin Concepts, Threats, and Machine-Learning Security Solutions” IEEE Access,PP. 1-1. 10.1109,2018.

[23] Tanwar, Sudeep , “Machine Learning Adoption in Blockchain-Based Smart Applications: The Challenges, and a Way Forward.” IEEE Access 8,pp. 474-488,2020.

[24] Andreas M. Antonopoulos, “Mastering bitcoin Nanjing Southeast University Press”, 2018

[25] blockchain browser, online avalible: https://www.blockchain.com/

[26] Lee Seunghyeon et al, “Cybercriminal Minds: An investigative study of cryptocurrency abuses in the Dark Web” NDSS ,2019.

[27] Silk_Road of darknet,online avalible:https://en.wikipedia.org/wiki/Silk Road_(marketplace)

[28] bitcoin-to-neo4j, online avalible: https://learnmeabitcoin.com/

[29] wallet fingerprint, online avalible:https://www.walletexplorer.com

[30] address tag, online avalible: https://bitcoinwhoswho.com/

[31] https://www.kaggle.com/ellipticco/elliptic-data-set