Temporal Networks based Industry Identification for Bitcoin Users

Weili Han[1], Dingjie Chen[1], Jun Pang[2], Kai Wang[1], Chen Chen[1], Dapeng Huang[1], and Zhijie Fan[1]

[1] School of Software, Fudan University, China
[2] Department of Computer Science, University of Luxembourg, Luxembourg

Abstract. With the development of Bitcoin, many thriving activities have developed into stable industries, such as Miner. Identifying and analyzing the transaction behaviors of users within these industries helps to understand the Bitcoin ecosystem from a macro perspective. Currently, industry identification mainly faces two issues. First, the anonymity of Bitcoin makes it difficult to identify the industry identifiers of users who participate in activities through different addresses. Second, since users usually engage in multiple industries at different periods, both the identification of their dynamically changing industry identifiers and the detection of their mostly engaged industry are challenging research tasks. In this paper, we propose an industry identification approach for Bitcoin users. First, we develop a fine-grained address clustering method to mine the relationship between addresses and their owners. Compared with existing methods, this method improves 0.18 in accuracy and 0.60 in recall. Based on temporal networks, we then train a multi-label classification model to identify the dynamically changing industry identifiers of users with an average accuracy of 0.92. With respect to multi-industry users, we further propose a major industry identifier detection method to identify the industry where users are mostly engaged. Applying this approach, we reproduce the major activity trajectories of users across the industries, which provides us with an opportunity to analyze the transaction behaviors of users within the industries.

Keywords: Bitcoin activity · address clustering · industry identifier identification · temporal network.

1 Introduction

The Bitcoin system (Bitcoin for short) has attracted a large number of users to participate in various activities, generating over 600 million transactions in total. Up to January 2021, the digital currency bitcoin has become the fifth-largest world currency [1]. As the number and value of transactions have grown rapidly, many thriving activities have developed into relatively stable modular structures. By constructing a large-scale transaction network, we observe that Bitcoin users with similar activity purposes exhibit tighter connectivity, resulting in stable modular structures for these activities. Referring to the classification of

activities in macroeconomics [2], we define these modular structures as Bitcoin industries, i.e., the Bitcoin industry is a group of activities with similar purposes. Here, we introduce five Bitcoin industries, including Darknet, Exchange, Gambling, Investment and Miner. For example, the Exchange industry provides digital currency exchange activities for Bitcoin users. Identifying and analyzing the transaction behaviors of users in such modular structures can deepen the understanding of Bitcoin from a macro perspective [3].

Currently, there are mainly two issues with industry identification. First, the anonymity of Bitcoin allows users to participate in various activities through different addresses, which makes it challenging to accurately identify their industry identifiers. Before identifying the industry identifiers of users, we shall first master the many-to-one relationship between addresses and their owners. However, existing address clustering methods indiscriminately apply coarse-grained heuristic rules to different transaction patterns, which mistakenly associates the addresses of multiple Bitcoin users into a single cluster, i.e., causing the problem of over-merging.

Second, since users can engage in various industries at different periods, it is difficult to identify their dynamically changing industry identifiers and detect the industry where they are mostly engaged (i.e., major industry). Driven by personal interests, the activity participation of Bitcoin users presents similar overlaps and migrations to that of social network users [4]. More specifically, they may change their current activities or perform activities in various industries, leaving their industry identifiers uncertain. Therefore, it is unpractical to classify the changing activity patterns into a single fixed industry identifier. However, in a short period (such as a week), a Bitcoin user tends to focus on certain industries and exhibits a relatively stable activity pattern, which provides us with an opportunity to accurately identify their dynamic industry identifiers and major industry identifier.

To solve these issues, we propose an industry identification approach based on temporal networks. We cluster addresses into users, classify the dynamically changing industry identifiers of users, and detect the industry where users are mostly engaged. Based on this approach, we can reproduce the major activity trajectories of users across the industries.

The main contributions of this paper are summarized below:

- We develop a fine-grained address clustering method to mine the relationship between addresses and Bitcoin users. Compared with existing address clustering methods, our method has improved precision by 0.18 and recall by 0.60, mitigating the problem of over-merging.
- Based on temporal networks, we train a multi-label classification model to identify the dynamic industry identifiers of users with an average accuracy of 0.92. Among them, about a quarter (23.35%) of users have multiple industry identifiers within a short period, called *multi-industry users*.
- For multi-industry users, we propose a major industry identifier detection method. By comparing the active scores of users in different industries, this method identifies the industry where users are mostly engaged.

The rest of the paper is organized as follows. First, we present the background knowledge in Section 2. Next, we discuss the collection and preparation of datasets in Section 3 and introduce industry identification in Section 4. The related work is discussed in Section 5. Finally, we conclude the paper in Section 6.

2 Background

2.1 Transaction patterns in Bitcoin

Bitcoin supports users to complete transactions in an open computing environment. In a typical transaction pattern, the sender sends bitcoins from his addresses to the recipients and pays some bitcoins to the miners as miner fees. When the number of bitcoins sent exceeds the sum of the recipients' expectation and the miner fee, the extra bitcoins will be sent back to the address predefined by the sender. The extra bitcoins are called *changes* and the predefined address is called *change address*. In addition to the typical transaction pattern, the following four special transaction patterns are also considered in our work.

• *Coinbase transaction*: Bitcoin uses the transaction to reward miners who submit new blocks, thus all recipients of this transaction can be regarded as miners.

• *Mixing transaction*: This transaction packages multiple remittance transactions into one single transaction to obfuscate the address association among different Bitcoin users. Some Bitcoin mixers offer this type of service, such as Bitblender.

• *Peeling chain transaction*: The transaction consists of one input address and two output addresses. The sender peels off a small number of bitcoins to one recipient and sends the remaining bitcoins to the other recipient. The latter recipient then follows this pattern and conducts the next peeling chain transaction.

• *Locktime transaction*: The transaction uses the optional field *Locktime* to preset its effective time, i.e., to take effect at a specific block height or at a specific timestamp. Generally, Bitcoin users have their own setting preferences.

2.2 Address association in Bitcoin

In practice, many Bitcoin users often reuse their addresses in multiple transactions for convenience, which may expose the potential address association. Since only the sender can use the private key to unlock the balance in the addresses, all input addresses of the transaction should belong to the same sender. Once the sender reuses these input addresses in other transactions, the reused addresses will serve as bridges to associate other addresses together. In addition, the study [5] states that the transaction preferences of Bitcoin users can reflect the relationship between addresses and their owners. In other words, personal behaviors in transactions, particularly the usage of change addresses, may become an important entry point for address association detection.

Based on the above observations, we consider the effect of special transaction patterns when performing address clustering in Section 4.1. In particular, we aim to improve the association of addresses involved in two transaction patterns: peeling chain and locktime, which are often ignored in previous studies.

2.3 Industries in Bitcoin

Referring to the classification of activities in macroeconomics [2], we define the concept of the Bitcoin industry as follows: a Bitcoin industry consists of activities that provide goods or services for similar purposes. Specifically, we divide activities into five industries: (1) Darknet, where trading smuggling or illegal service through bitcoins, e.g., SilkRoad. (2) Exchange, where conducting exchanges between bitcoin and other currencies, e.g., Mt.Gox. (3) Gambling, where gambling with bitcoins, e.g., SatoshiDice. (4) Investment, where offering the services of bitcoin returns and management, including bitcoin lending (e.g., Nexo), bitcoin faucet (e.g., Cointiply) and wallet management (e.g., Trezor). (5) Miner, where generating new blocks and distributing rewards to miners, e.g., F2Pool.

Based on the activity purposes and patterns of Bitcoin users in the industries, we describe industry members in two roles: *organizer* and *participant*. As organizers, Bitcoin users provide goods or services to participants in their activities, such as drug traffickers. This paper describes industry organizers as darknet vendors, exchange sites, gambling bankers, investment merchants, and miner pools, respectively. Their corresponding participants are darknet customers, exchange buyers, gamblers, individual investors, and individual miners. These industry roles are treated as class labels for industry identification in Section 4.2.

3 Datasets

We collect Bitcoin transaction data and entity labels of addresses as datasets for our work. The dataset *Transactions* records all historical transaction data of Bitcoin users. The dataset *Entity Identities* stores the addresses of well-known entities, mapping anonymous addresses to their real-world identities. Below, we detail the collection and preparation of each dataset.

(1) *Transactions*: We download raw Bitcoin transaction data from the genesis block to 12/31/2020, parse the data into address-based transactions. In total, we obtain 601,452,574 transactions and 759,091,687 addresses.

(2) *Entity Identities*: We collect entity labels of addresses from website *WalletExplorer* [6] and *Ethonym* [7], where the former has been used as ground truth in the study [8]. In the preparation step, we perform data cleansing of addresses and classify them into industry roles. We first exclude addresses that are duplicated or failed in validation checks. Based on the service rules of different activities, we then classify these addresses into industry organizers and participants. We treat the addresses of wallet management as participants and classify the remaining addresses as organizers. Moreover, we identify other participants from organizer-related transactions and coinbase transactions to enrich the dataset.

Consequently, this dataset covers 382 entities, including 21,057,772 organizers and 130,145,529 participants, accounting for 2.77% and 17.14% of the total addresses. Table 1 details the number of addresses in different industries. In particular, the relationship between entities and their containing addresses helps to evaluate the address clustering method in Section 4.1; the labels of organizers and participants are used to train an industry identifier classifier in Section 4.2.

Table 1: The number of addresses in five Bitcoin industries.

Industry	# of Organizers	# of Participants
Darknet	2,332,854	5,657,783
Exchange	9,967,932	87,932,289
Gambling	3,098,500	14,451,596
Investment	5,619,822	21,420,157
Miner	38,664	683,704

4 Industry Identification

In this section, we introduce an industry identification approach for Bitcoin users. This approach consists of three steps: address clustering, multi-industry identifier classification, and major industry identifier detection.

To break the protection of anonymous payment mechanism, Section 4.1 proposes an address clustering method to capture the hidden associations among Bitcoin addresses and cluster them as users. After that, Section 4.2 takes these users as the basic units and designs a multi-label classification model to master dynamic industry identifiers of users within certain periods. To understand the major activity purposes of multi-identifier users, Section 4.3 devises a quantitative method to determine their major industry identifier.

4.1 Bitcoin address clustering

Protected by anonymous transactions, it is hard to figure out the whole activity intent of Bitcoin users if we just analyze their transactions based on individual addresses. Therefore, we develop an address clustering method to mine address association before capturing multiple industry identifiers of Bitcoin users.

Discussion of existing methods. Several heuristic rules are widely used in existing address clustering methods. (1) *MI* (Multiple Input) [9]: all input addresses of a transaction belong to one user; (2) *MX* [5, 10]: excluding mixing transactions before applying method *MI*, where *X* denotes the mixing transactions; (3) *NA* (New Address) [9, 11] : an address which first appears as an output address belongs to the change address of the sender; (4) *DP* (Decimal Points) [10]: in a two-output transaction, an output address which has three more decimal points than the other output value belongs to the change address of the sender; (5) *SP* (SPecial) [5]: the addresses in two consecutive transactions with the same transaction pattern belong to one user.

However, indiscriminately applying these coarse-grained heuristic rules to different transaction patterns may lead to the over-merging of clusters.

Transaction pattern observation. To mitigate the problem of over-merging, we observe the features of two special transaction patterns: peeling chain and locktime. (1) Peeling chain pattern is a typical pattern, with 43.11% of transactions matching this pattern. Moreover, 83.82% of the output addresses are used

for one-time bitcoin transfers. Combined with the peeling off behaviors of bitcoin transfers, we argue that the features of the new addresses and the number of received bitcoins can help mine address association in this pattern. (2) For locktime transactions, Bitcoin users usually generate them under the same type of effective condition, i.e., at a specific block height or timestamp. Also, 89.19% of these transactions have spent all output bitcoins for subsequent payments. These preferences can help mine address association in locktime transactions.

Based on these observations, we design a series of experiments to develop a fine-grained address clustering method with high precision and recall.

Our method. The address clustering method consists of three heuristic rules. We apply *MX* as a basic rule[3] to eliminate the interference of mixing transaction JoinMarket [5] and CoinJoinMess [12]. The other two heuristic rules as follows.

• *Heuristic rule 1:* The output address of a peeling chain transaction is the change address of the sender if it meets three features: (1) the address is a new address, (2) the address receives a larger number of bitcoins, and (3) the number of bitcoins received in this address has three more decimal points than that in the other output address.

• *Heuristic rule 2:* The input addresses of two consecutive locktime transactions belong to the same user if each transaction meets two features: (1) all the outputs of the transaction have been spent, and (2) the transaction specifies the effective time in the same way, i.e., a specific block number or a specific timestamp.

Evaluation. We use three existing address clustering methods [5, 9, 10] as baselines to evaluate the quality of our method. We take the address association of entities in *Entity Identities* as the evaluation dataset and measure the clusters in two aspects. First, we evaluate the number of identified entities, including the number of entities successfully identified (indicator N) and the number of entities incorrectly identified into one cluster (indicator E). Second, we use four indicators to evaluate the quality of the clusters, including *Precise* (P), *Recall* (R), *Weighted Precise* (WP) and *Weighted Recall* (WR). The first two indicators have been used in the study [13]. Considering that clusters with a larger number of addresses usually contain more information, we further take the number of addresses per cluster as weight and propose the latter two indicators.

$$WP = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij}\frac{|o_{ij}|}{|S_i|} \tag{1}$$

$$WR = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n} w_{ij}\frac{|o_{ij}|}{|E_i|} \tag{2}$$

Equation 1 and Equation 2 introduce the indicators WP and WR, where E_i is the ith entity and S_i is the group of identified clusters mapping to E_i. The number of entities and the number of clusters in S_i are denoted by m and n. In

[3] Addresses excluded in mixing transactions can be associated with addresses through other normal transactions or recorded as isolated users.

Table 2: Evaluation of several address clustering methods.

Method	N	E	P	R	WP	WR
$MI + NA$	336	154	0.15	0.02	0.07	0.03
$MX + NA + DP$	339	96	0.43	0.09	0.18	0.13
$MX + SP$	355	37	0.80	0.60	0.28	0.20
Our method	**366**	**17**	**0.94**	**0.96**	**0.31**	**0.31**

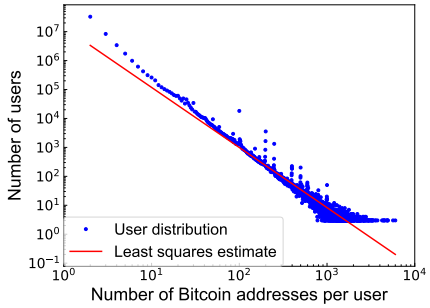

Fig. 1: User distribution follows Zipf's law.

addition, s_{ij} is the jth cluster of S_i, o_{ij} is the overlap between E_i and s_{ij}, and w_{ij} is the proportion of the number of addresses in s_{ij} to that in S_i.

Table 2 presents the evaluation results. We observe that our method can cluster more (95.81% of the total) entities and reduce the over-merging of entities by 20.59% on average. Moreover, the evaluation results of indicators P and R both exceed 0.90 in our method. Compared to the best values of the baselines, indicators P, R, WP and WR have increased by 0.18, 0.60, 0.11 and 0.55, respectively. These improvements show that our fine-grained method can mitigate the problem of over-merging and provide clusters with high precision and recall.

Results analysis. Since the clusters can well reflect the transaction behaviors of Bitcoin users, we call them *users*. As a result, we generate a total of 337,158,548 users, of which 81.67% have one address (called *isolated users*), 17.66% have 2-10 addresses, and 0.04% have more than 100 addresses. Figure 1 describes the distribution between users and addresses[4] and further performs a linear regression on this distribution. The coefficient of determination R^2 is 0.95, which indicates this distribution largely follows Zipf's law [14]. In the next step, we use these users as the basic unit to classify industry identifiers.

4.2 Multi-industry identifier classification

Similar to social network users [4], Bitcoin users can change their current activities and engage in activities of other industries (i.e., activity migrations), or

[4] We group users by the number of addresses they hold and filter out the group with less than three users.

Table 3: Comparison of graph embedding algorithms in multi-label classification.

Algorithm	Macro-F1			Micro-F1		
	10%	20%	30%	10%	20%	30%
DeepWalk	0.62	0.72	0.76	0.90	0.91	0.92
GraphSAGE	**0.70**	**0.75**	**0.77**	**0.90**	**0.91**	**0.93**
LINE	0.33	0.35	0.35	0.81	0.81	0.81
Matrix Factorization	0.30	0.33	0.42	0.80	0.80	0.82
Node2Vec	0.47	0.52	0.62	0.84	0.87	0.88
SDNE	0.46	0.54	0.54	0.79	0.81	0.81

perform various activities across multiple industries (i.e., activity overlaps). In a short period, a Bitcoin user tends to concentrate on specific industries and exhibits a stable activity pattern, which allows us to identify his dynamically changing industry identifiers within a certain period (e.g., one week).

Based on temporal networks, we design a multi-label classification model to identify the dynamic industry identifiers of users. We construct a transaction graph to describe user interactions, extract temporal activity patterns of users, and train an industry identifier classifier for industry identification.

Graph construction. We construct a directed graph *User-Transaction*, where each node represents a user and each edge represents the transaction interaction from a sender to a recipient. In addition, we record the timestamp and the number of bitcoins received by the recipients as edge annotations.

Feature extraction. Since some special Bitcoin events may lead to imbalances among different industries in training data, we extract features from the temporal networks of several Bitcoin events to improve the robustness of the model. Based on the search popularity in Google Trends [15], we select five events related to the industries, including SatoshiDice game released in Gambling, Liberty Reserve unsealed in Inverstment, SilkRoad shut down in Darkent, Mt.Gox disappeared in Exchange and BTC Guild announced to shut down in Miner. We then extract sub-graphs of *User-Transaction* before and after the events as temporal networks.

In each temporal network, the proportion of known industry identifier labels is rather limited, accounting for 10%-30% of the total users. To ensure the quality of the features extracted at such proportions of recognized labels, we test the performance of six graph embedding algorithms listed in the study [16], i.e., *DeepWalk*, *GraphSAGE*, *LINE*, *Matrix Factorization*, *Node2Vec* and *SDNE*. We apply one-vs-rest logistic regression to evaluate the performance of these graph embedding algorithms. Specifically, we randomly sample 10%, 20% and 30% of labeled users as training data and the rest of labeled users as the testing data. To eliminate the contingency, we repeat this process ten times and record the average results in Table 3. The results indicate that *GraphSAGE* [17] performs better accuracy and stability, so we choose it to extract the features of users.

Model training and evaluation. We apply MLP model [18] to build the classification model. We first filter out labeled users involved in less than three transactions to ensure the quality of training data. After that, we define the size

Table 4: Evaluations for temporal networks in five events.

Industry	Network Time Span	Accuracy	Macro-F1	Micro-F1
Gambling	04/01/2012 – 04/07/2012	0.92	0.88	0.94
	05/22/2012 – 05/28/2012	0.96	0.89	0.97
Investment	04/04/2013 – 04/10/2013	0.92	0.90	0.94
	05/28/2013 – 06/03/2013	0.91	0.88	0.94
Darknet	09/18/2013 – 09/24/2013	0.89	0.87	0.93
	10/04/2013 – 10/10/2013	0.90	0.89	0.93
Exchange	01/02/2014 – 01/08/2014	0.92	0.85	0.94
	02/12/2014 – 02/18/2014	0.94	0.87	0.95
Miner	03/02/2015 – 03/08/2015	0.93	0.87	0.94
	03/24/2015 – 03/30/2015	0.91	0.88	0.94

of the training data as 0.67 and adopt 3-fold cross-validation to train the model. Based on the selected event of each industry, we evaluate the performance of the model in different temporal networks (see Table 4). We observe that our model presents high accuracy with an average of 0.92, which can well identify multiple industry identifiers of users in a certain period.

4.3 Major industry identifier detection

We observe that 23.35% of users engage in multiple industries during a week and call them multi-industry users. Such a non-negligible proportion further motivates us to detect the industry where they are mostly engaged, i.e., to detect their major industry identifier. Being an industry member, the user is active inside the industry and rarely participates in other activities outside the industry. If a user devotes more participation frequency and bitcoin traffic to a specific industry, we determine this industry identifier as his major industry identifier.

Based on these observations, we define the indicators of participation frequency and bitcoin traffic, determine their weights, and calculate active scores of users in different industries to detect the major industry identifier.

Indicator extraction. For multi-industry users, we treat the transactions they perform within a single industry as internal transactions and extract indicator values from these transactions. In each internal transaction, the participation frequency (f) is the reciprocal of the time difference between the current transaction and the previous internal transaction conducted in the same industry. The bitcoin traffic (v) is the number of bitcoins transferred to the industry. When both parties of a transaction are multi-industry users, if the senders and recipients have at least one same industry identifier, we treat their interaction in the transaction branch as an internal transaction of this industry. Based on the above extraction rules, we obtain the indicator sequences F and V.

Weight calculation. In general, sequences with higher entropy contain richer information and should be given more weight to determine the major industry identifier. Therefore, we apply entropy weight method (EWM) [19] to calculate

the weights of these indicators. Specifically, we normalize the sequences, compute their entropy values (e_F and e_V) and obtain the weights through Equation 3.

$$w_i = \frac{1 - e_i}{\sum_{j \in \{F,V\}}(1 - e_j)},\ i \in \{F, V\} \tag{3}$$

Active score calculation. We use Equation 4 to calculate the active scores of users in different industries. For industry i, we calculate its active score S_i based on the behaviors of internal transactions (A_i) and the prediction probability of the industry identifier (P_i). Among them, A_i is calculated from the average time of participation frequency (t_{iF}) and the total sum of bitcoin traffic (t_{iV}).

$$S_i = P_i * A_i = P_i * (t_{iF} * w_F + t_{iV} * w_V) \tag{4}$$

Finally, we rank the active score of each industry identifier and determine the industry identifier with the highest score as the major industry identifier.

4.4 Summary

In short, our approach clusters addresses into users with high accuracy, identifies dynamic industry identifiers of users, and detects major industry identifier to reproduce the major activity trajectories of users across the industries.

5 Related Work

In this section, we introduce studies that are closely related to address clustering and activity identifier classification in Bitcoin.

Address clustering. As we discussed in Section 4.1, many heuristic rules are proposed to mine address association in Bitcoin. Some studies focus on the association of input addresses. Interfered by mixing transactions, the original method *MI* generates users with relatively low recall [13]. On this basis, an improved method *MX* is proposed to filter out mixing transactions before applying the method *MI*, which has been widely used for address clustering. However, the exclusion of addresses involved in mixing transactions somewhat reduces the recall of clustering results. Other studies mine the association of output addresses through several patterns, such as method *NA* and method *DP*. Currently, many of these methods have been extended to detect associated addresses in other cryptocurrencies, such as Zcash [20]. In practice, some transactions may mismatch the patterns, resulting in incorrect address association. For example, in the ransomware activity Locky, criminals use the new output address of the ransom payment transaction to receive the ransoms [21]. However, the method *NA* would treat this new output address as the change address of the victim.

To mitigate these problems, we develop several fine-grained heuristic rules and further mine the association of addresses involved in mixing transactions.

Activity identifier classification. Classifying the activity identifiers of Bitcoin users is essential to explore their behavior purposes across various activities.

At present, many studies apply supervised machine learning techniques to detect the addresses of the activities. Usually, researchers extract features from the transaction behaviors of the addresses, such as the number of bitcoins transferred in a transaction. For instance, Toyoda et al. [22] analyze the transfer features of addresses in high yield investment programs (HYIP) and design a classifier with an accuracy of 0.94. Moreover, a few studies introduce features of different dimensions to improve the quality of identification. Li et al. [23] extract features from three dimensions to identify addresses involved in illegal activities, including transaction, topology and time. Besides, other studies exploit graph techniques to detect anomalous addresses. For example, Chen et al. [24] build the transaction graphs of the exchange site Mt.Gox and calculate singular value decomposition to identify abnormal accounts related to market manipulation.

Most studies assume that Bitcoin users are only active in a single activity and ignore to classify users with multiple activity identifiers. In this paper, we train a multi-label classification model from an industry perspective, which can accurately describe the whole behaviors of Bitcoin users across various activities.

6 Conclusion and Future Work

In this paper, we have proposed a practical approach for identifying dynamic industry identifiers of Bitcoin users based on temporal networks. First, we developed a fine-grained address clustering method to mitigate the problem of over-merging, which improved over existing methods 0.18 in precision and 0.60 in recall. We then trained an industry identifier classification model to identify dynamic industry identifiers of users with an average accuracy of 0.92. For multi-industry users, we further calculated the active scores in different industries to detect their major industry identifier. Based on this approach, we captured the major activity trajectories of users across the industries. In the future, we will study more transaction patterns in address clustering and apply our approach for analyzing the interactions and migrations of users across the industries.

Acknowledgement. This paper has been supported by the National Key R&D Program of China (2018YFC0830900), Natural Science Foundation of China (U1836207), and China Postdoctoral Science Foundation (2020M670998).

References

1. Phillips, D.: Bitcoin is now the 5th largest world currency. `https://decrypt.co/39425/bitcoin-is-now-the-5th-largest-world-currency` (2021)
2. Syverson, C.: Macroeconomics and market power: Context, implications, and open questions. Journal of Economic Perspectives **33**(3), 23–43 (2019)
3. Quiles, M.G., Macau, E.E., Rubido, N.: Dynamical detection of network communities. Scientific Reports **6**, 25570 (2016)
4. Newell, E., Jurgens, D., Saleem, H.M., Vala, H., Sassine, J., Armstrong, C., Ruths, D.: User migration in online social networks: A case study on reddit during a period of community unrest. In: Proceedings of the 10th International Conference on Web and Social Media. pp. 279–288 (2016)

5. Kalodner, H.A., Möser, M., Lee, K., Goldfeder, S., Plattner, M., Chator, A., Narayanan, A.: Blocksci: Design and applications of a blockchain analysis platform. In: Proceedings of the 29th USENIX Conference on Security Symposium. pp. 2721–2738 (2020)
6. Walletexplorer. `https://www.walletexplorer.com` (2021)
7. Ethonym. `https://ethonym.com` (2021)
8. Foley, S., Karlsen, J.R., Putniņš, T.J.: Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? The Review of Financial Studies **32**(5), 1798–1853 (2019)
9. Spagnuolo, M., Maggi, F., Zanero, S.: Bitiodine: Extracting intelligence from the bitcoin network. In: Proceedings of the 18th International Conference on Financial Cryptography and Data Security. pp. 457–468 (2014)
10. Athey, S., Parashkevov, I., Sarukkai, V., Xia, J.: Bitcoin pricing, adoption, and usage: Theory and evidence. Stanford institute for Economic Policy Research **13**(4), 675–746 (2016)
11. Meiklejohn, S., Pomarole, M., Jordan, G., Levchenko, K., Savage, S.: A fistful of bitcoins: characterizing payments among men with no names. In: Proceedings of the 13th Conference on Internet Measurement. pp. 127–140 (2013)
12. Coinjoinmess. `https://www.walletexplorer.com/wallet/CoinJoinMess` (2021)
13. Cazabet, R., Baccour, R., Latapy, M.: Tracking bitcoin users activity using community detection on a network of weak signals. In: Proceedings of the 6th Conference on Complex Networks and Their Applications. pp. 166–177 (2017)
14. Newman, M.E.: Power laws, pareto distributions and zipf's law. Contemporary physics **46**(5), 323–351 (2005)
15. Google trends. `https://trends.google.com` (2021)
16. Cai, H., Zheng, V.W., Chang, K.C.C.: A comprehensive survey of graph embedding: Problems, techniques, and applications. IEEE Transactions on Knowledge and Data Engineering **30**(9), 1616–1637 (2018)
17. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: Proceedings of the 31st Conference on Neural Information Processing Systems. pp. 1024–1034 (2017)
18. Bishop, C.M., et al.: Neural networks for pattern recognition. Oxford university press (1995)
19. Cheng, Q.: Structure entropy weight method to confirm the weight of evaluating index. Systems Engineering Theory & Practice **30**(7), 1225–1228 (2010)
20. Kappos, G., Yousaf, H., Maller, M., Meiklejohn, S.: An empirical analysis of anonymity in zcash. In: Proceedings of the 27th USENIX Conference on Security Symposium. pp. 463–477 (2018)
21. Huang, D.Y., Aliapoulios, M.M., Li, V.G., Invernizzi, L., Bursztein, E., McRoberts, K., Levin, J., Levchenko, K., Snoeren, A.C., McCoy, D.: Tracking ransomware end-to-end. In: Proceedings of the 39th IEEE Symposium on Security and Privacy. pp. 618–631 (2018)
22. Toyoda, K., Mathiopoulos, P.T., Ohtsuki, T.: A novel methodology for HYIP operators' bitcoin addresses identification. IEEE Access **7**, 74835–74848 (2019)
23. Li, Y., Cai, Y., Tian, H., Xue, G., Zheng, Z.: Identifying illicit addresses in bitcoin network. In: Proceedings of the 2nd Conference on Blockchain and Trustworthy Systems. pp. 99–111 (2020)
24. Chen, W., Wu, J., Zheng, Z., Chen, C., Zhou, Y.: Market manipulation of bitcoin: Evidence from mining the Mt. Gox transaction network. In: Proceedings of the 38th IEEE Conference on Computer Communications (2019)