National Blockchain Project

Towards Malicious address identification in Bitcoin Blockchain

IEEE Blockchain-2021

Workshop on Blockchain security, Application and Performance in 4th IEEE International conference on Blockchain Melbourne, Australia | 06-08 December 2021

Authors :
Deepesh Chaudhari,
Dr. Rachit Agarwal,
Prof. Sandeep Kumar Shukla

Motivation

- Bitcoin is a crypto-currency that is prone to cyber-attacks, scams and ransom payments.
- Bitcoin transactions constitute ever-increasing socio-temporal interaction graph. Such temporal aspects of graph help us to understand the behavior (which could be illicit) of accounts.
- There exist several approaches[1] using temporal features to detect illicit activities on different Blockchains such as Ethereum, but in case of Bitcoin, temporal aspect are not studied for malicious account detection.
  - Agarwal et al.[1] state that their approach is valid for all permissionless blockchain. But they did not validate on Bitcoin.
- We are motivated to validate the applicability of already existing temporal features on the Bitcoin. Which is currently known for other blockchains.

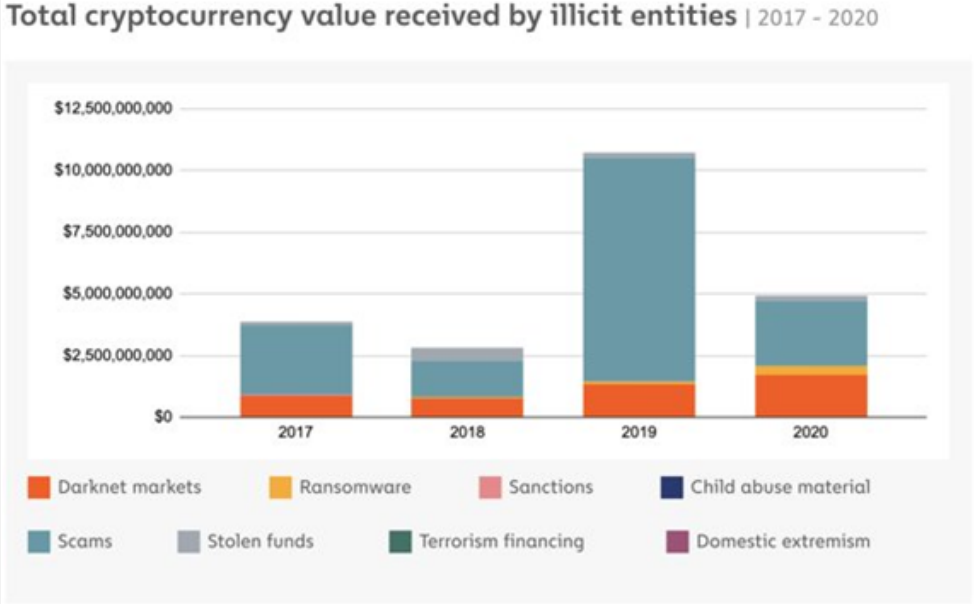


Image source: https://go.chainalysis.com/2021-Crypto-Crime-Report.html

[1] Agarwal, R., Barve, S. & Shukla, S.K. Detecting malicious accounts in permissionless blockchains using temporal graph properties. *Appl Netw Sci* **6,** 9 (2021).

Motivation

Hence, the research questions (RQ) that we ask are:

- ***(RQ1)*** Are the features identified in state-of-the art approach[1] targeting permissionless blockchains such as Ethereum applicable in Bitcoin or not?
- ***(RQ2)*** Can we detect malicious accounts in the Bitcoin using ML techniques while considering the temporal features of the Blockchain?
- ***(RQ3)*** What changes occur in the result after the change address clustering?
- ***(RQ4)*** Does behavior change exists in Bitcoin accounts?

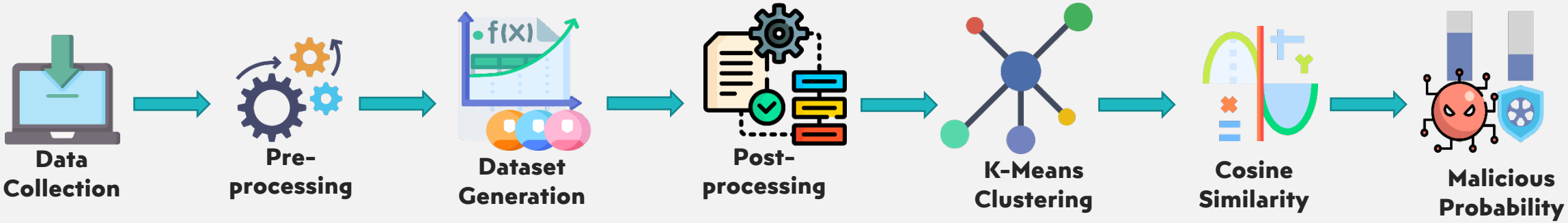
[1] Agarwal, R., Barve, S. & Shukla, S.K. Detecting malicious accounts in permissionless blockchains using temporal graph properties. *Appl Netw Sci* **6,** 9 (2021).

Related work

		Used features based on											
	B/C	AS	iD	oD	Bal	TF	BB	A	CC	IET	ML Algo Used	Dataset	Hyperparameters	Performance
[1]	PL(ETH)	✓	✓	✓	✓	✓	✓	✓	✓	✓	AutoML (ETC the best)	700k	estimators = 200	0.88^{ac}, $>0.29^{p}$, $>0.78^{r}$
											K-means (best)		k ∈ [3,24]	k_{opt}= 9,10
[2]	BTC	✓	✓	✓	✓	-	-	-	✓	✓	K-Means	100k[a]	k ∈ [1, 14]	k_{opt}=7, 8
											Mahalanobis Distance		x	0.0256^{MDE}
											SVM		V = 0.005	0.1441^{MDE}
[3]	BTC	✓	✓	✓	✓	-	-	-	✓	✓	Local Outlier Factor	6.3M[a]	K=8	0.55^{MDE}
[4]	BTC	-	✓	✓	✓	-	-	-	✓	-	K-Means	1M[a]	k ∈ [1, 14]	K_{opt} = 8
											Trimmed K-Means		k ∈ [1, 15], α = 0.01	k_{opt} = 8
[5]	BTC	-	✓	✓	✓	✓	-	-	-	-	Adaboost	1000M[a]	estimators = 50, rate = 1	$>0.2^{r}$
											Random Forest		estimators = 10	$>0.85^{r}$
											Gradient boosting		estimators = 100, rate = 0.1 depth = 3	$>0.93^{r}$

Indegree (iD), Outdegree (oD), Balance (Bal), Transaction Flow (TF), Bursty Behavior (BB), Attractiveness (A), Clustering Coefficient (CC), Inter-event-time (IET), Permissionless Blockchain (PL), Extra tree classifier(ETC), ^{ac} accuracy, ^{p} precision, ^{r} recall, ^{MDE}Dual Evaluation Metric, Bitcoin(BTC), Ethereum(ETH)

Methodology : Process pipeline



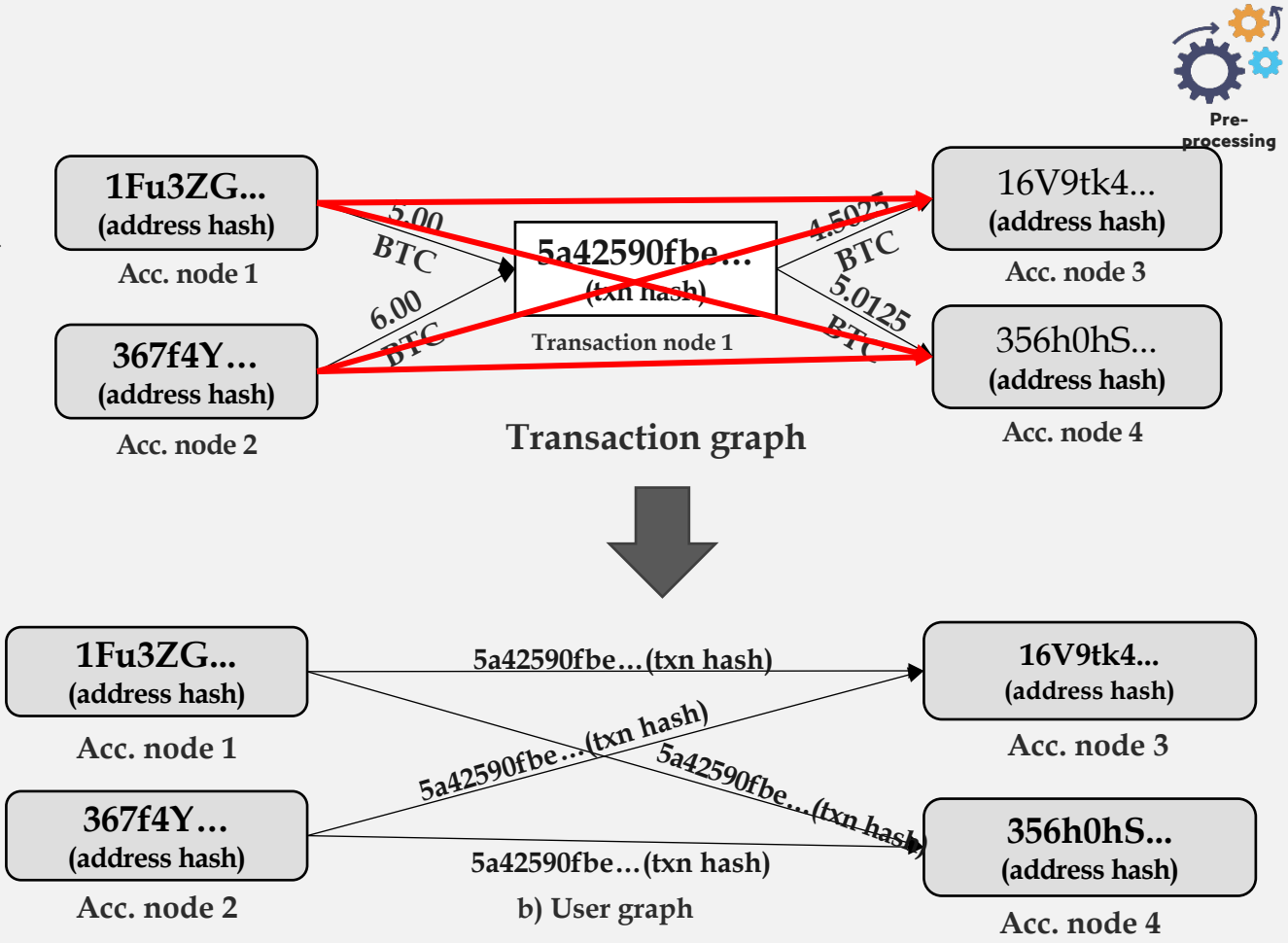
(Process Pipeline)

Methodology : Pre-processing (Graph Generation)


```
{
    "hash": "5a42590fbe0a90ee8e8747244d6c84f0db1a3a24e8f1...",
    "ver": 1,
    "vin_sz": 2,
    "vout_sz": 1,
    "lock_time": 0,
    "size": 404,
    "in": [ {
            "prev_out": {
                "hash": "3be4ac9728a0823cf5e2deb2e86fc0bd...",
                "n": 0
            },
            "scriptSig": "30440..."
        }, {
            "prev_out": {
                "hash": "7508e6ab259b4df0fd5147bab0c949d81473d...",
                "n": 0
            },
            "scriptSig": "3f3a4ce81...."
        } ],
    "out": [ {
            "value":"4.5025",
            "scriptPubKey":"OP_DUP OP_HASH160 69e02e18b5705a05dd6b28ed517716
             OP_EQUALVERIFY OP_CHECKSIG"
        }, {
            "value":"5.0125",
            "scriptPubKey":"OP_DUP OP_HASH160 5dd6b28ed51771669e02e18b5705a0
             OP_EQUALVERIFY OP_CHECKSIG"
        } ]
}
```

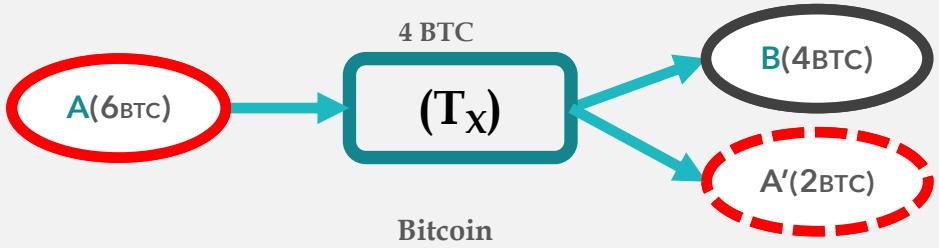

Hash & output index of previous transaction

Bitcoin transaction structure



Methodology : Post-Processing (Concept of Change Addresses)

- In Bitcoin, if a user makes a transaction, then he needs to transfer the total amount of BTCs to the output's accounts.
- If the user wants to store the remaining amount (change), then he has to create a new account address and send the remaining amount to that address.
  - This new address is known as **Change Address** of the user.



- We need to identify such change addresses of the user in order to extract temporal aspects.

Methodology : Post-Processing (How to identify Change Addresses)

- There are 4 heuristics methods to identify change address:
  - *multi-input address heuristics*[7, 8]
  - *change address heuristics*[7, 8]
  - *change address heuristics with exception* that are based on *value* and *growth*[2]
  - *modified change address heuristics* based on address reuse[8]
- This heuristics is basically used to identify the accounts used by same users.

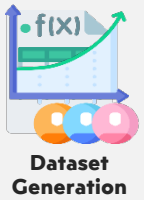
Neudecker et al.[2] compared the different heuristics methods[2, 7, 8] and found that ***multi-input*** together with ***change address heuristics*** provides a better results.

[2] T. Neudecker and H. Hartenstein, "Could network information facilitate address clustering in bitcoin?" In *Financial Cryptography and Data Security*, Cham: Springer International Publishing, 2017, pp. 155–169,

[7] S. Meiklejohn, M. Pomarole, G. Jordan, K. Levchenko, D. McCoy, G. Voelker, and S. Savage, "A fistful of bitcoins: Characterizing payments among men with no names," in Conference on Internet Measurement Conference, (Barcelona, Spain), pp. 127–140, ACM, 10 2013.

[8] Y. Zhang, J. Wang, and J. Luo, "Heuristic-based address clustering in bitcoin," IEEE Access, vol. 8, pp. 210582–210591, 11 2020.

Methodology : Feature Extraction



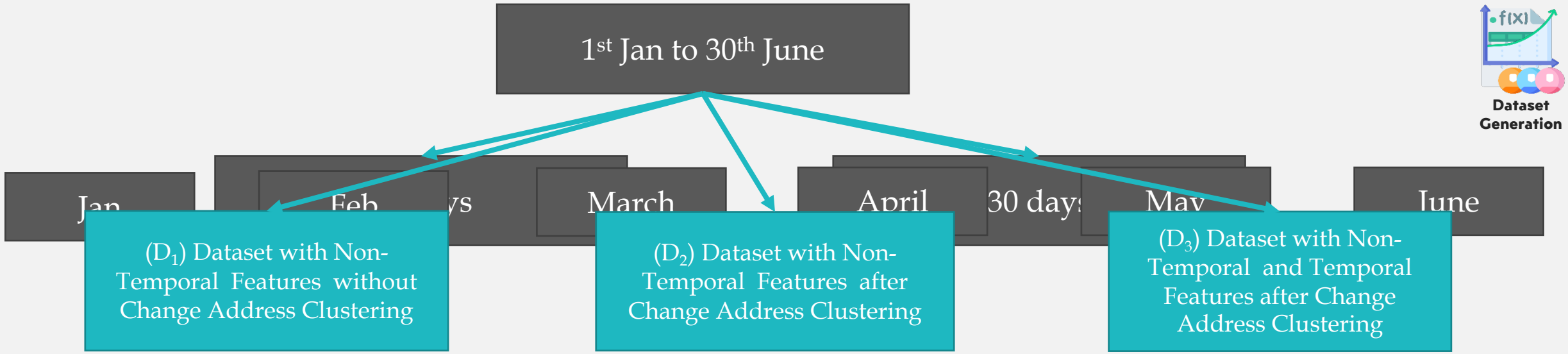
Non-temporal Features

- Non-temporal features are features that are independent of time and are extracted using transaction graph and user graph.
  - *indegree, unique indegree, outdegree, unique outdegree, clustering coefficient*

Temporal Features

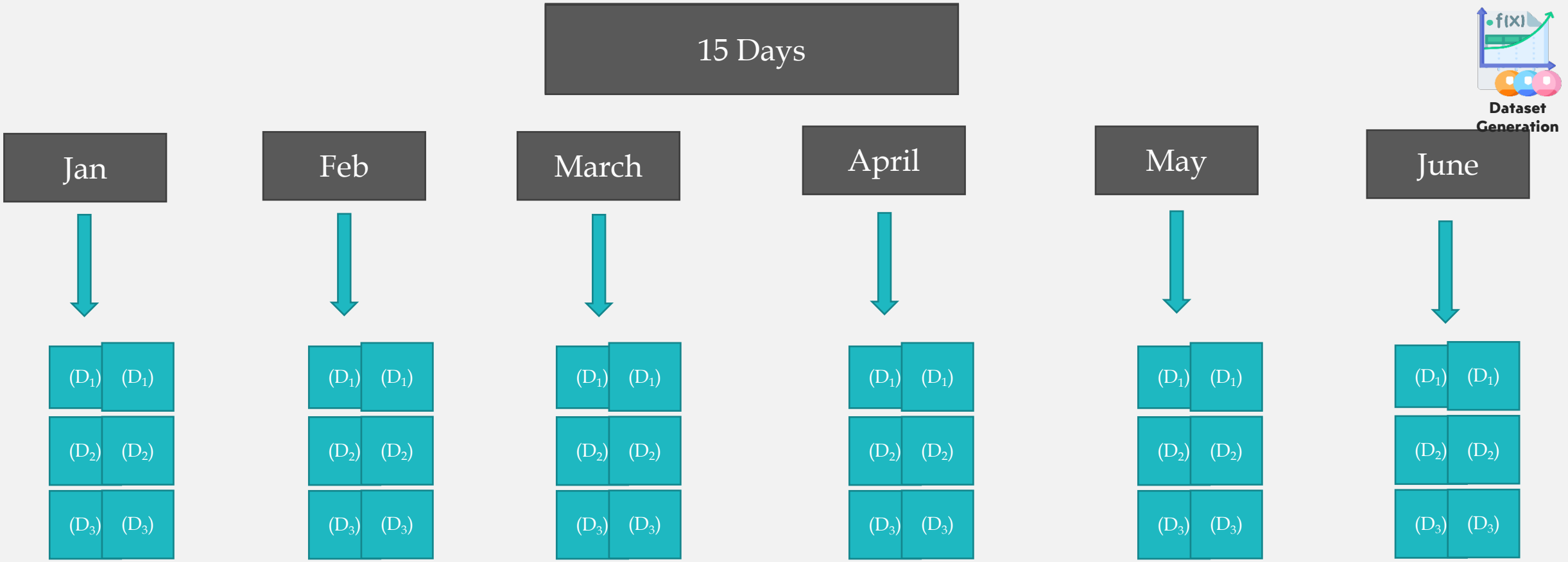
- Temporal features are time dependent features, we extract these features after the address clustering of Bitcoin.
  - ***inter-event time***[1]***, burstiness*** behavior of ***indegree, outdegree, balance***[1] ***and attractiveness***[1]

Methodology : Data Configuration



Dataset Generation

Methodology : Data Configuration



Methodology : Machine Learning

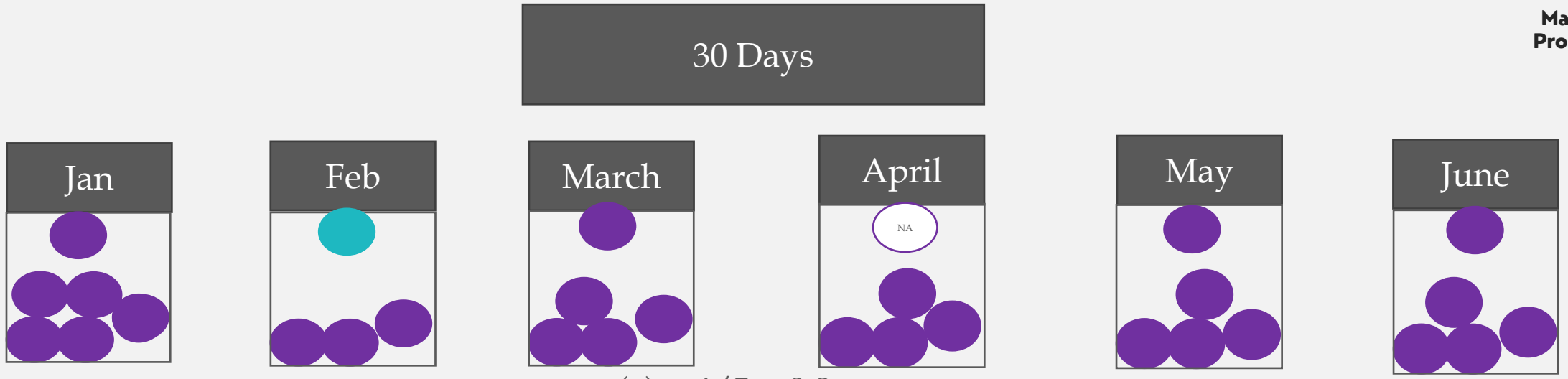
K-means Clustering

- After generating all the different datasets, we apply *K-Means* with K = 10 to validate the state-of-the-art approach[1].
- After that, we take accounts of the most maliciously tagged cluster and then find cosine similarity between tagged and unlabeled accounts to identify whether their behavior is similar (within $\varepsilon \rightarrow 0$).
- After calculating cosine similarity for all the different datasets for different time granularities we calculate probability of an account to be malicious.

[1] Agarwal, R., Barve, S. & Shukla, S.K. Detecting malicious accounts in permissionless blockchains using temporal graph properties. *Appl Netw Sci* **6,** 9 (2021).

Methodology : Machine Learning

Probability of account x to be Malicious[1], p(x)



$p(x) = 4/5 = 0.8$

Malicious Probability

- we calculate the probability p^k of a particular account k to be malicious in a given T_g. $p^k = \frac{\sum_{d \in SD^j(T_g)} M_d^k}{n^k}$

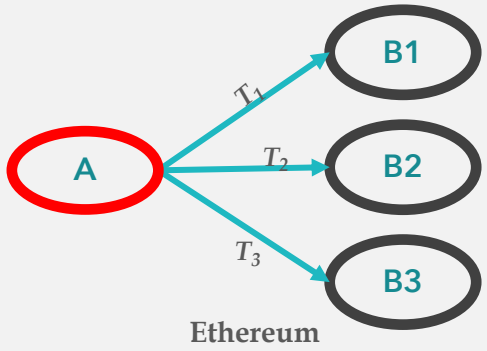
  Here, $M^k{}_d$ depicts whether the address k was identified as suspect in the d^{th} sub-dataset $(SD^j(T_g))$ using j features at time granularity T_g, and n_i is the total number of SDs in which address k transacted.

Un-tagged(Ben.) Account Suspicious Account

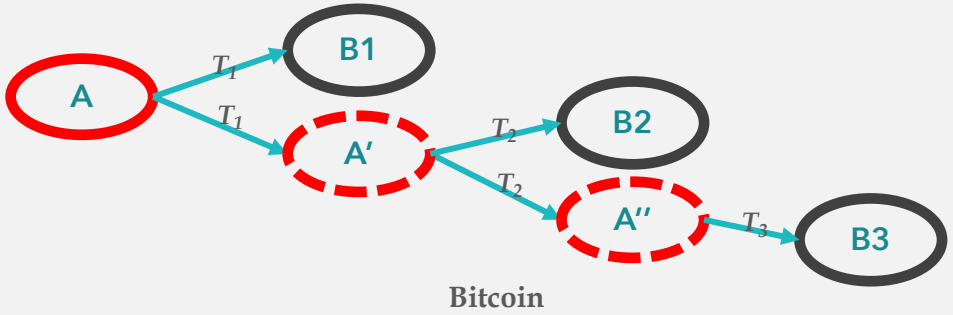
Evaluation : Result Analysis

(RQ1) Are the features identified in state-of-the art approaches targeting permissionless blockchains such as Ethereum applicable in Bitcoin blockchain or not?

- Bitcoin blockchain has a different architecture than Ethereum blockchain.
- After address clustering, the total number of users reduces to **[0.96 to 1.75]**% of the total number of address.



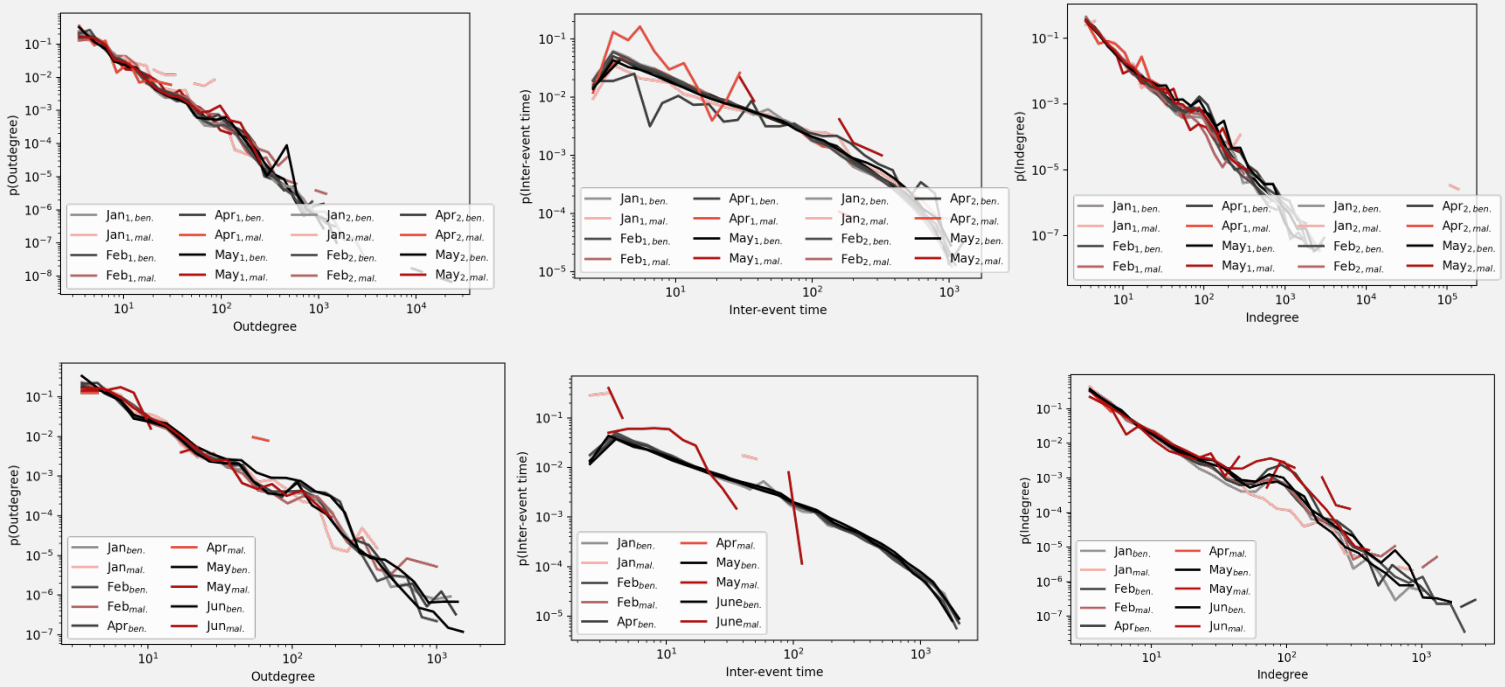
Ethereum



Bitcoin

Evaluation : Result Analysis

(RQ1) Are the features identified in state-of-the art approaches targeting permissionless blockchains such as Ethereum applicable in Bitcoin blockchain or not?



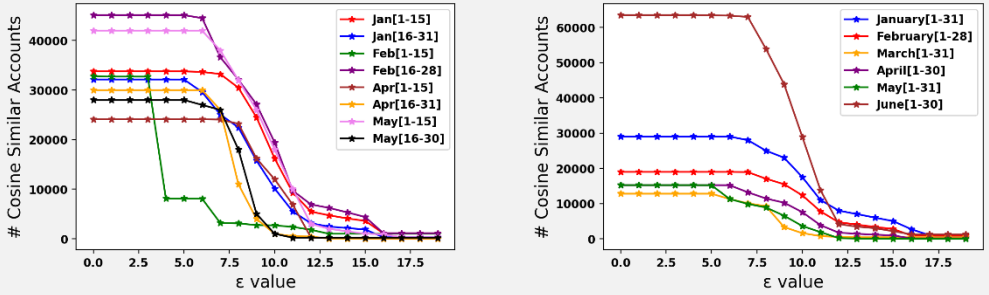
KL-Divergence between Ethereum & Bitcoin		
	15 days	30 days
Indegree	0.051	0.174
Outdegree	0.124	0.059

Hence, the method and features identified in state-of-the-art targeting Ethereum are **also applicable** in Bitcoin blockchain but after change address clustering.

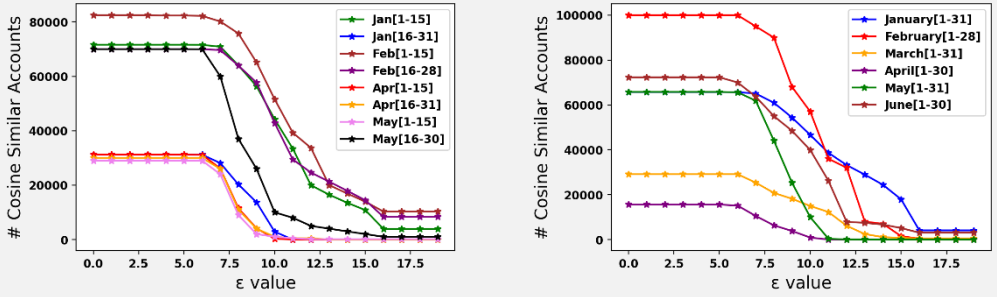
Evaluation : Result Analysis

(RQ2) Can we detect malicious accounts in the Bitcoin blockchain using ML techniques and consider the temporal evolution of the graph?

	15 days		30 days	
	Mal. account	ε	Mal. account	ε
Non-temporal features	***43,366***	[7,16]	***6,575***	[12,16]
Temporal features	***87,907***	[10,16]	***68,215***	[11,16]



Number of benign accounts having high Cosine Similarity with malicious accounts for different values of with non-temporal features.

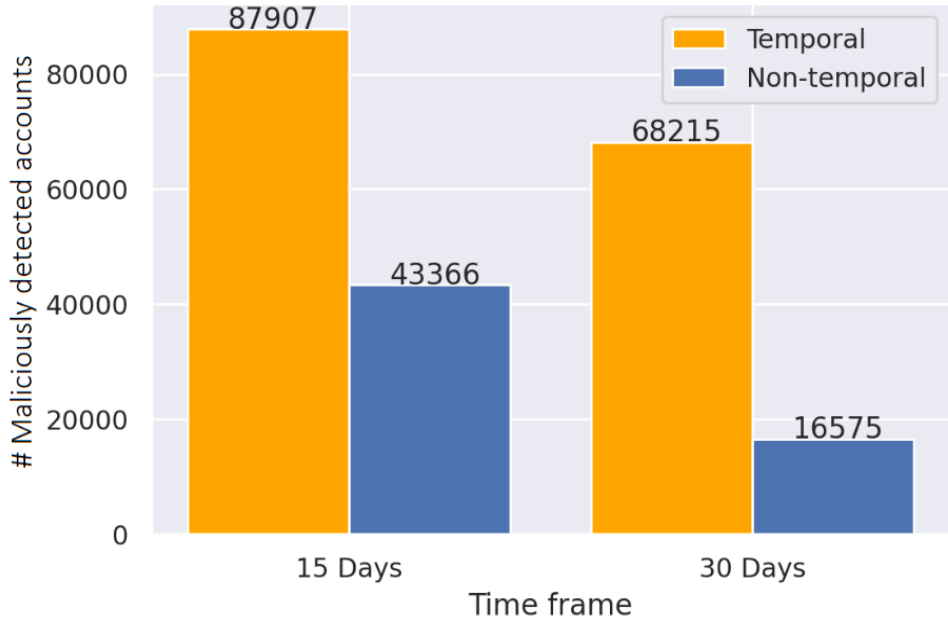


Number of benign accounts having high Cosine Similarity with malicious accounts for different values of with both temporal and non-temporal features.

Evaluation : Result Analysis

(RQ3) What changes occur in the result after the change address clustering?

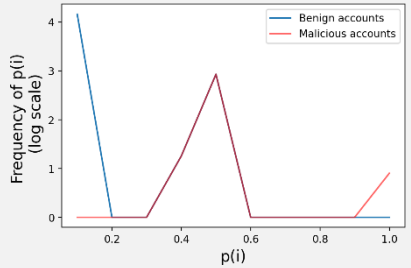
- After including temporal feature, we successfully detected ***44,541*** more malicious accounts in 15 days time frame and ***51,640*** more malicious accounts in 1-month temporal granularity datasets when using non-temporal features.



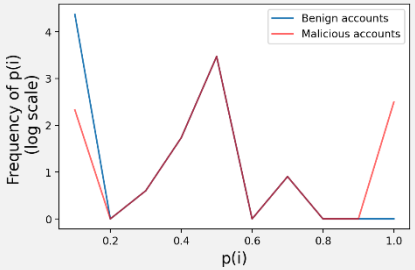
Maliciously detected accounts with temporal and non-temporal features

Evaluation : Result Analysis

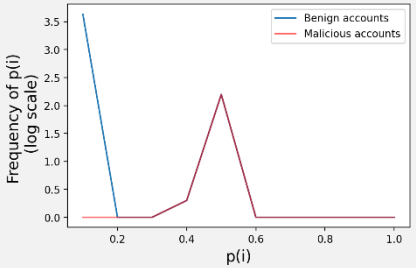
- *(RQ4)* Does behavior change exists in Bitcoin accounts?



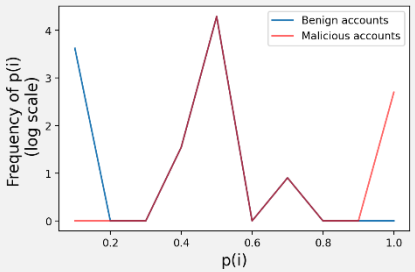
15 days



15 days



30 days



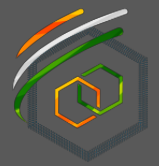
30 days

Probability distribution of accounts to be malicious with D2

Probability distribution of accounts to be malicious with D3

- For non-temp features we find that 868 and 159 accounts change their behavior between mal. to ben for 15 days and 30 days time granularities, respectively.
- For temp features we find that 3273 and 19712 accounts change their behavior between mal. to ben for 15 days to 30 days time granularities, respectively.
- For non temporal we find very less accounts with $p^k=1$ and in contrast in temporal we detects 313 accounts in 15 days and 501 accounts for 30 days with $p^k=1$.
- find 3 suspect accounts that were detected with high probability across different time granularities.

Conclusion and Future work



NATIONAL BLOCKCHAIN PROJECT

Conclusion

- Address clustering of bitcoin addresses is essential in the detection of malicious accounts.
- We find that in Bitcoin, behavior of addresses is similar to those in Ethereum with respect to features such as in-degree, outdegree and inter-event time.
- We detect behavior change in accounts using temporal features in different time granularities 313 (15 days) and 501 (30 days)
- Find 3 suspect accounts that were detected with high probability across different time granularities.

Future Work

- In the future, we would like to test the state-of-the-art method with more data to provide more robust results.
- Validation for other crypto-currency such as Algorand, Binance Coin (BNB), Cardano (ADA), Chainlink (LINK), Dogecoin (DOGE), Litecoin (LTC), Polkadot (DOT), and Ripple (XRP) is still needed.

National Blockchain Project

Thank You

Any Question?