

区块链交易网络挖掘和 行为识别

吴嘉婧

副教授

中山大学 计算机学院

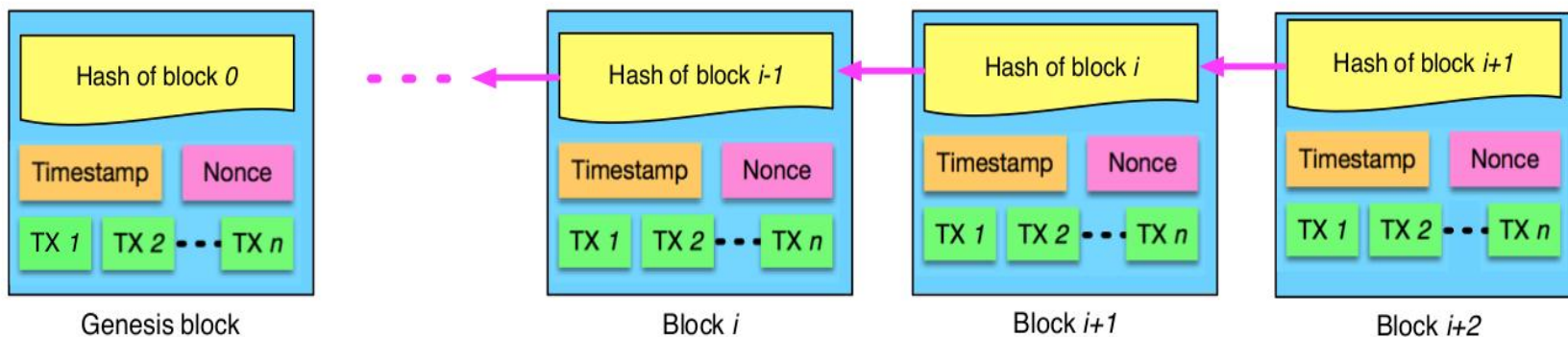


中山大學
SUN YAT-SEN UNIVERSITY



- 1** **区块链数据**
- 2 交易网络构建
- 3 网络分析与挖掘
- 4 交易行为识别
- 5 其他工作

区块链



- 区块链是一个**分布式的**账本数据库
- 区块里面存储的是转账记录
- 按照时间顺序组织区块
- 难以篡改，可追溯等特征

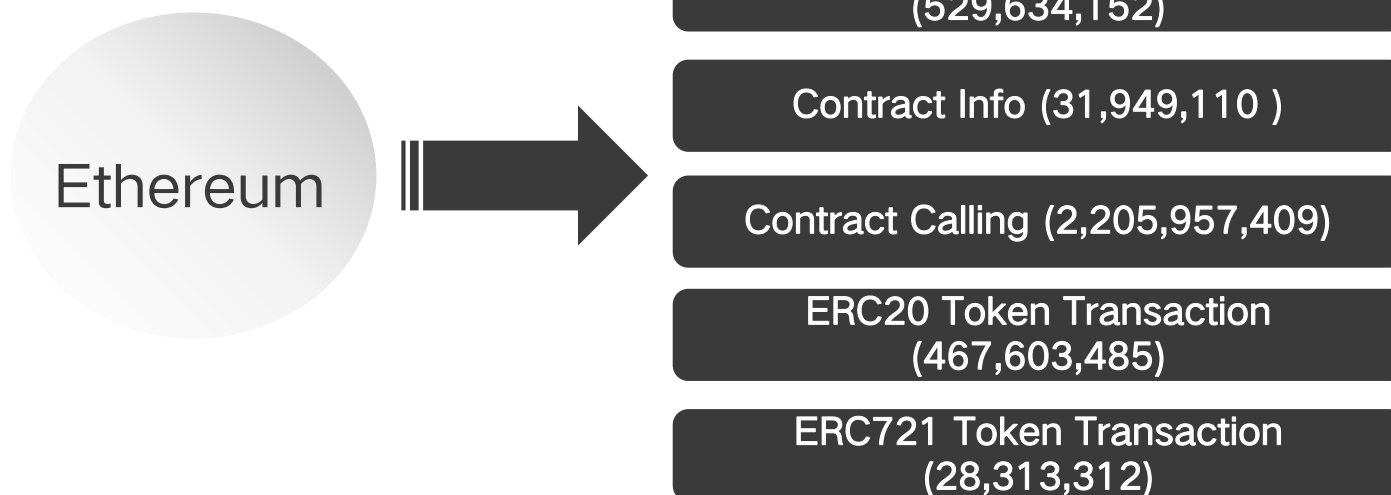


丰富的链上数据



- 2019年第三季度，比特币数据将近242GB
- 截至2020.04，EOSIO上的区块数量高达8983万
- 截至2020.10，以太坊上的区块数量高达1100万

如：截至2020.10的以太坊数据





比特币/以太坊的链上数据
交易分析, 诈骗识别

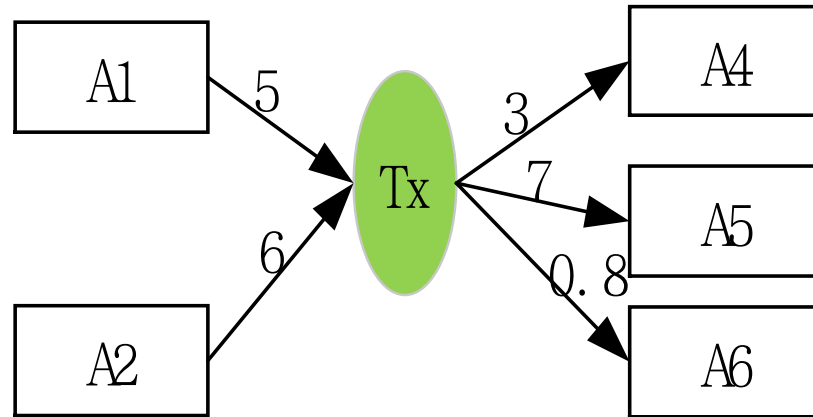
实现合约相关的代码数据
合约被触发的交易数据

交易所数据、白皮书
交易评论、论坛

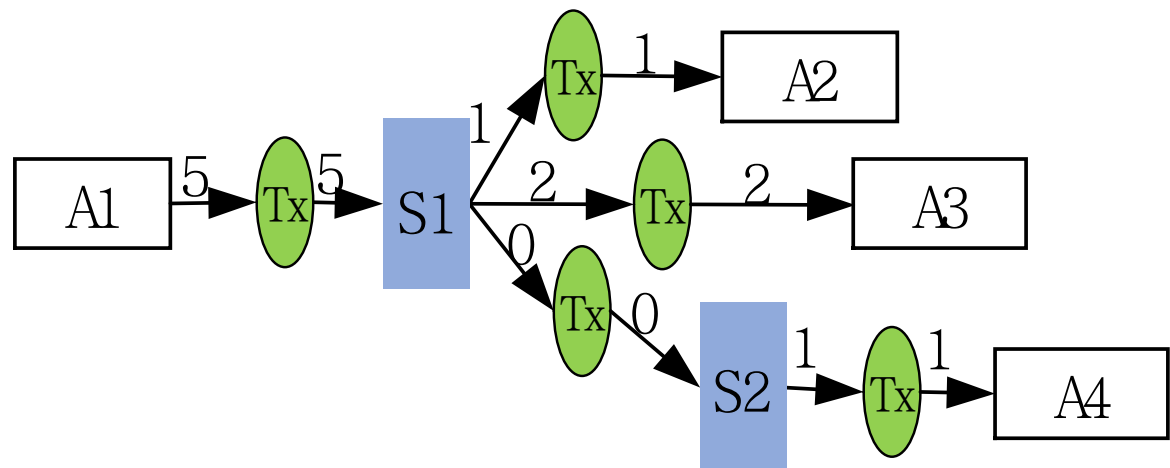
交易数据



典型的比特币交易



典型的智能合约平台 (以太坊) 交易



合约数据



合约交易数据

区块记录

Transactions	Internal Transactions	Token Transfers	Code
Latest 25 txns from a total Of 362 transactions			
TxHash	Block	Age	From
0xb49c0addffb36ed...	5521073	26 days 15 hrs ago	0x0e
0xef4750a3fa00179...	5521043	26 days 15 hrs ago	0x0e
0xfbef5310257cb1f1...	5509833	28 days 14 hrs ago	0x70
0x83a1e9108e809d...	5509822	28 days 14 hrs ago	0x70
0x4d7d08dd0010b6...	5497316	30 days 17 hrs ago	0x29
0x5d32453faca7546...	5497306	30 days 17 hrs ago	0x29
0x13103c80ff9fe654...	5497280	30 days 18 hrs ago	0x29

实现合约的数据:

源代码: Solidity (<1%)

```
contract Rubixi {  
  
    //Declare variables for storage critical  
    uint private balance = 0;  
    uint private collectedFees = 0;  
    uint private feePercent = 10;  
    uint private pyramidMultiplier = 300;  
    uint private payoutOrder = 0;  
  
    address private creator;  
  
    //Sets creator  
    function DynamicPyramid() {  
        creator = msg.sender;  
    }  
}
```

字节码: 只对虚拟机有意义
可反编译为操作码

Contract Creation Code

```
606060405260008080556001819055600a60025561012c60  
d578063253459e31461011c5780634229616d1461013d578  
61022e5780639dbc4f9b14610260578063a26dbf26146102  
192146103ab575b6103d66103d86000670de0b6b3a764000  
82526103da94670de0b6b3a7640000900493926107d29083  
00090600160a060020a03908116339091161415610595576  
815481101561000257925260029190910260008051602061  
a039081163390911614156103d857600154600014156104e  
013990509091565b6103da60408051602081810183526000  
102e857600680548490811015610002575080548183526000  
0b6b3a764000010005000101561000257508054849081101
```

交易所数据



白皮书

BLOGS EVENTS PUBLICATIONS

Blockchain: practical technology for financial

Share this   

11 APR 2016
Webinar

DOWNLOADS

交易评论

Make sure to use the "downvote" button for any spammy posts

5 Comments [Etherscan](#)

 Recommend  Share



Join the discussion...

LOG IN WITH



OR SIGN UP WITH DISQUS



[criticalCommentator](#) · 3 months ago

SCAM ADDRESS?!

1 ^ | v · Reply · Share >



[Sudheer Tummala](#) · 3 months ago

i Lost 1 ETH to this address, this address impe

^ | v · Reply · Share >



[Sudheer Tummala](#) · 3 months ago


scammer address,
dont release it to this address

^ | v · Reply · Share >



[Thatguy](#) · 3 months ago

论坛



交易数据

➤ 安装客户端同步交易数据

- 比特币 BitcoinCore
- 以太坊 Geth、OpenEthereum
-

➤ 通过区块链浏览器爬取交易数据

Blockchain.com、Etherscan.com、.....

标签数据

➤ 结合论坛、标签网站等信息进行身份识别

交易数据

➤ 安装客户端同步交易数据

- 比特币 BitcoinCore
- 以太坊 Geth、OpenEthereum
-

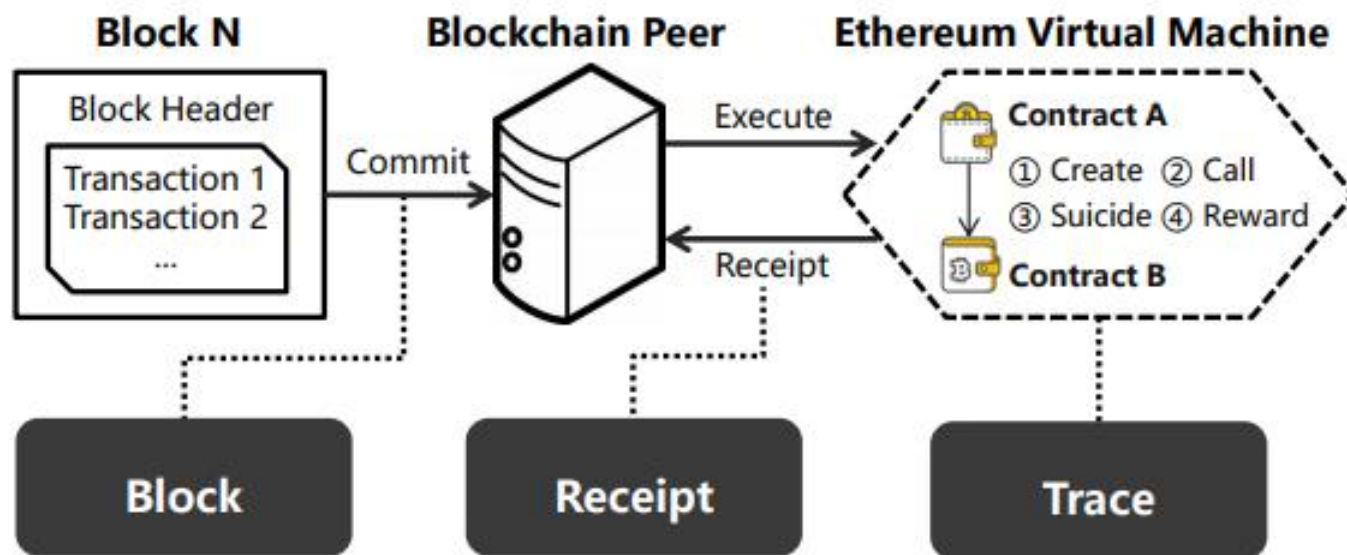
➤ 通过区块链浏览器爬取交易数据

Blockchain.com、Etherscan.com、.....

标签数据

➤ 结合论坛、标签网站等信息进行身份识别

以太坊交易数据获取：通过客户端同步



- 由智能合约触发的交易需要通过以太坊虚拟机交易重演的trace数据获取
- 使用OpenEthereum的trace模块获取

以太坊交易数据获取：通过区块链浏览器



- Etherscan 是以太坊官方支持的区块链浏览器
- Etherscan国内站点界面（cn.etherscan.com）
 - **交易信息查询**——只要你知道钱包地址，轻松查询其所有交易信息，尤其你可以轻松查到大户地址，**跟踪大户动作**
 - **ERC代币查询**——目前大部分的token，都基于以太发行，所以只要是ERC代币，将你想查找的Token名称输入到地址栏，会看到持有该币的地址数量，作为你对**该币市场热度**的一个重要参考点
 - **合约代码查询**——有的项目会持续发行Token，具体进展可以从合约代码中查到
- 提供Ethereum Developer APIs



以太坊交易数据获取



◆ Etherscan首页显示当前以太坊网络的概览



Home Blockchain Tokens Resources More Sign In

The Ethereum Blockchain Explorer

All Filters Search by Address / Txn Hash / Block / Token / Ens

Sponsored: AAX - AAX Futures trading fees as low as 0.02%. Visit [AAX.com](https://aax.com) now!

Ad
AAX / Year 1 Club
Trade crypto futures
Enjoy fees as low as 0.02%

ETHER PRICE
\$465.85 @ 0.02783 BTC (+2.34%)

TRANSACTIONS
904.84 M (13.5 TPS)

MED GAS PRICE
48 Gwei (\$0.47)

ETHEREUM TRANSACTION HISTORY IN 14 DAYS



MARKET CAP
\$52,855,440,269

DIFFICULTY
3,346.86 TH

HASH RATE
266,499.55 GH/s

Latest Blocks

Bk	11274041 30 secs ago	Miner xnpool 205 txns in 19 secs	2.77023 Eth
Bk	11274040 49 secs ago	Miner BTC.com Pool 177 txns in 15 secs	2.72817 Eth
Bk	11274039 1 min ago	Miner Ethermine 236 txns in 18 secs	2.67442 Eth

Latest Transactions

Tx	0x7f415f9d8ae9... 30 secs ago	From 0xd9ad1de7905dc39ac... To 0xf68151affd0935fae6b...	0.64 Eth
Tx	0x231badc386... 30 secs ago	From 0xc6f9e2dbb990662c6... To 0x00000000441378008...	0 Eth
Tx	0x232626c8206... 30 secs ago	From 0x66e9d84290e21a8eb... To 0x90aaf33a10df1365cd...	0.02 Eth

以太坊交易数据获取



◆ Etherscan 提供直观的可视化数据统计模块

Ethereum Charts & Statistics

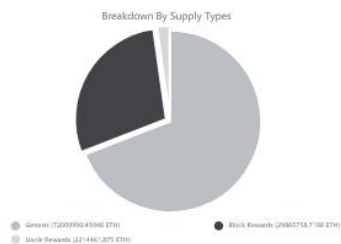
Charts & Stats

Market Data

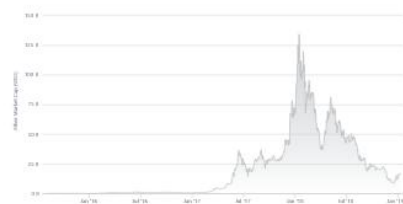
Ether Daily Price (USD) Chart



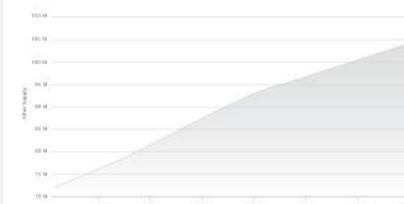
Total Supply & Market Cap Chart



Ether Market Capitalization Chart

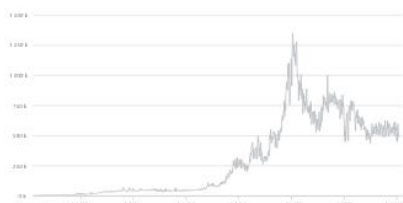


Ether Supply Growth Chart

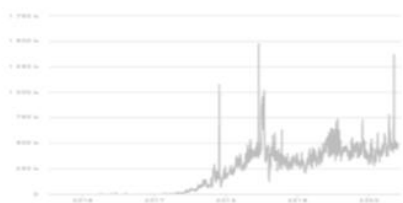


Blockchain Data

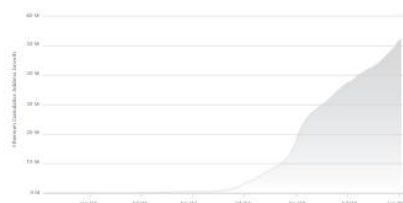
Daily Transactions Chart



ERC20 Daily Token Transfer Chart



Unique Addresses Chart



Average Block Size Chart



Average Block Time Chart

Average Gas Price Chart

Average Gas Limit Chart

Daily Gas Used Chart

以太坊交易数据获取



◆ Etherscan交易详情部分截图

The screenshot shows the 'Overview' tab of an Etherscan transaction. The transaction is successful and occurred 40 seconds ago on November 14, 2020. It involved a transfer of 0.05 Ether from a sender to a contract (Forsage.io) with a value of \$22.94. The transaction fee was 0.0026153215 Ether (\$1.20) and the gas price was 0.0000000145 Ether (14.5 Gwei).

Field	Value
Transaction Hash	0x30c20cd4f79d54a22b8cb767d39adf7e896afcd5b124d2ea6945d7d3237871bf
Status	Success
Block	11255356 (2 Block Confirmations)
Timestamp	40 secs ago (Nov-14-2020 10:43:29 AM +UTC) Confirmed within 5 mins:47 secs
From	0x567d87b8f2387d51664b49a8c369daae790f4c43
To	Contract 0x5acc84a3e955bdd76467d3348077d003f00ffb97 (Forsage.io) L TRANSFER 0.05 Ether From Forsage.io To → 0x76c5ff8a3352d5ea131fd5d...
Value	0.05 Ether (\$22.94)
Transaction Fee	0.0026153215 Ether (\$1.20)
Gas Price	0.0000000145 Ether (14.5 Gwei)

以太坊标签数据



◆ Etherscan提供部分地址标签

➤ 如下图标签为 Phish/Hack

Accounts Phish / Hack Label Cloud / Phish / Hack

Sponsored: [DeFi Yield Protocol: DeFi gem with anti-manipulation features and Staking with ETH Rewards. Join Now!](#)

Related labels: Transactions (1) Tokens (25)

A list of addresses related to phishing and hacks

A total of 4,657 accounts found First < Page 1 of 187 > Last

Address	Name Tag	Balance	Txn Count
0x9f26ae5cd245bfeeb5926d61497550f79d9c6c1c	Akropolis Hacker 1	0 Ether	29
0xbceaa0040764009fdcff407e82ad1f06465fd2c4	Bancor Hacker	0 Ether	3
0x03b70dc31abf9cf6c1cf80bfeeb322e8d3dbb4ca	Browser Extension Hack	0 Ether	64
0x4639cd8cd52ec1cf2e496a606ce28d8afb1c792f	CBDAO: BREE Token	0 Ether	6,389
0xf6884686a999f5ae6c1af03db92bab9c6d7dc8de	Coinrail Hacker	0 Ether	31

I Journal of Network and Computer Applications

Journal of Network and Computer Applications 190 (2021) 103139



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Network and Computer Applications

journal homepage: www.elsevier.com/locate/jnca



Review

Analysis of cryptocurrency transactions from a network perspective: An overview

Jiajing Wu^a, Jieli Liu^{a,b}, Yijing Zhao^c, Zibin Zheng^{a,b,*}

^a School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510006, China

^b School of Software Engineering, Sun Yat-sen University, Zhuhai 519082, China

^c School of Electronics and Communication Engineering, Sun Yat-sen University, Guangzhou 510006, China



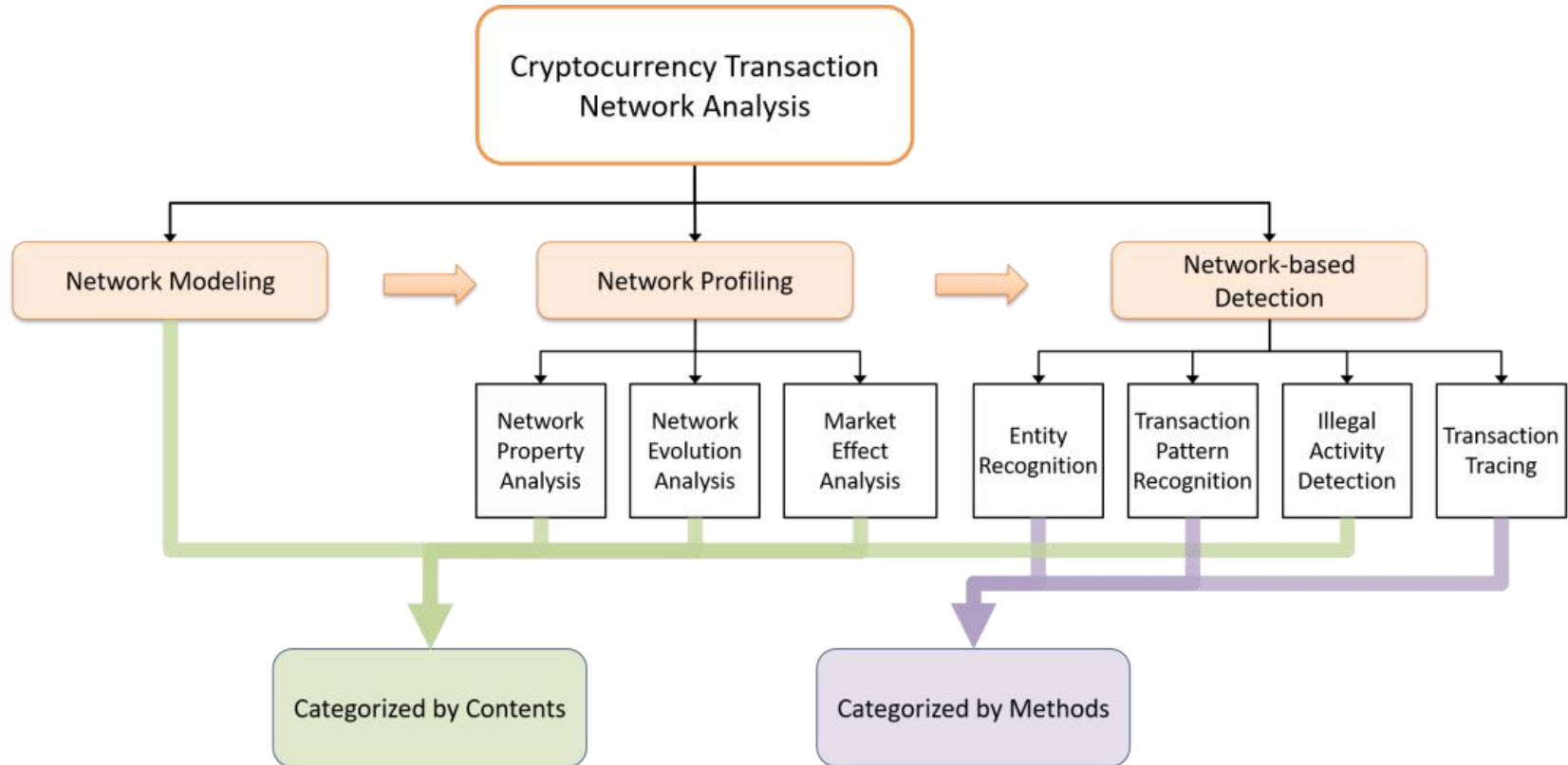
ARTICLE INFO

Keywords:

Cryptocurrency
Blockchain
Transaction records
Complex networks
Data mining

ABSTRACT

As one of the most important and famous applications of blockchain technology, cryptocurrency has attracted extensive attention recently. Empowered by blockchain technology, all the transaction records of cryptocurrencies are irreversible and recorded in blocks. These transaction records containing rich information and complete traces of financial activities are publicly accessible, thus providing researchers with unprecedented opportunities for data mining and knowledge discovery in this area. Networks are a general language for describing interacting systems in the real world, and a considerable part of existing work on cryptocurrency



- 1 区块链数据
- 2 交易网络构建**
- 3 网络分析与挖掘
- 4 交易行为识别
- 5 其他工作

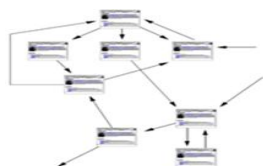
工作1：以太坊账户交易关系建模



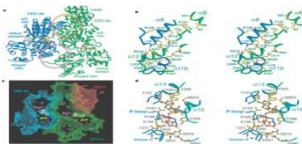
- ◆ 网络常用于描述物体之间的关系，可以通过金融网络描述账户之间的转账关系



社交网络



网页网络

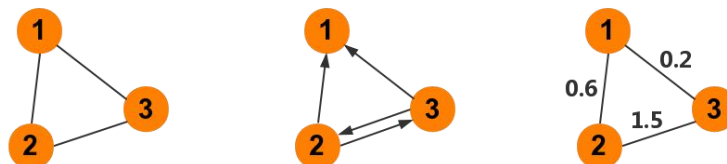


生物网络



金融网络

- ◆ 传统的分析通常将账户之间的交易关系建模为一个静态的简单图



- ◆ 构建保留交易时间和金额信息的网络模型

交易数据： (发送方, 接收方, 交易金额, 交易时间)

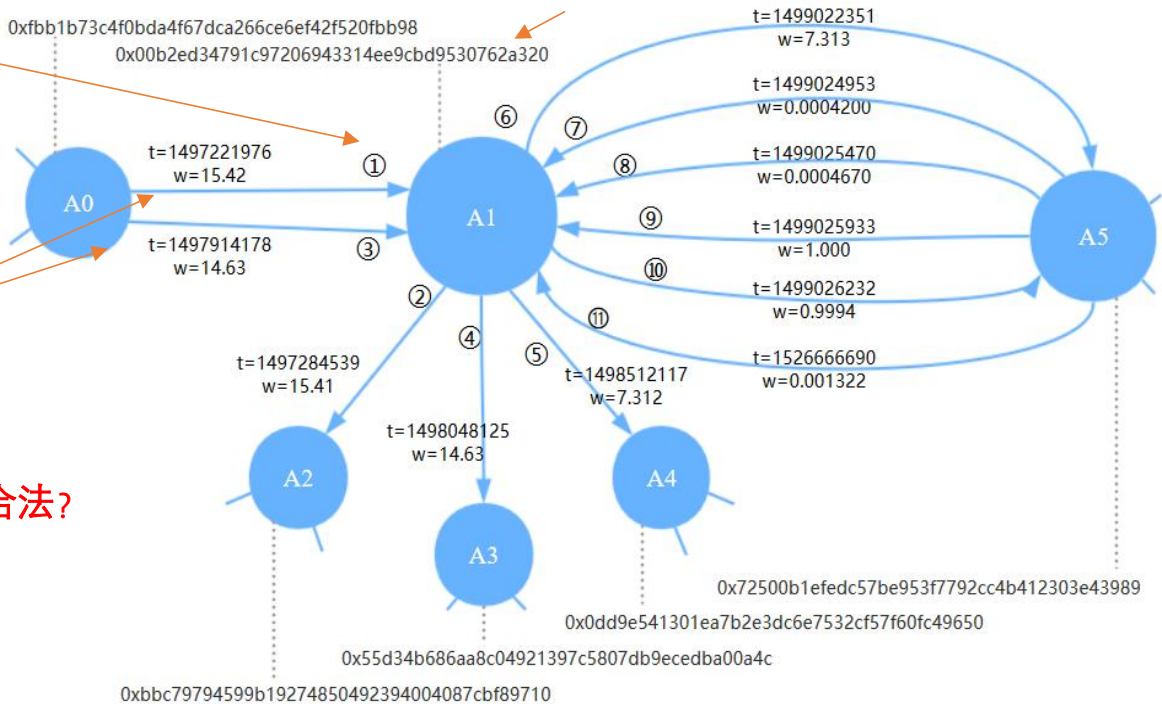
设计目标： 尽可能保留交易网络信息，区分节点间的多次交易

工作1：以太坊网络嵌入



以太坊账户

按时间戳从小到大标号



时序网络的重要性

从A0到A1存在2条路径
传统游走表示: {A0, A1, A2}

多重时序网络下, A2节点是否合法?

- $\{e_1, e_2\}, \{A0, A1, A2\}$ [✓]
- $\{e_3, e_2\}, \{A0, A1, A2\}$ [✗]

- 数学定义: 网络 $G=(V,E)$, 边 $e=(u,v,w,t)$
- 主要特点: 允许存在平行边, 每个节点表示一个以太坊账户, 每条边表示一次交易, 包含交易时间 t 和交易金额 w , 并按照 t 从小到大对边进行标号

工作1：以太坊账户交易关系建模



网络建模：时序加权多重有向网络

↓ 保留交易的时间、金额信息

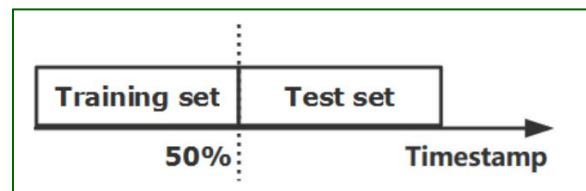
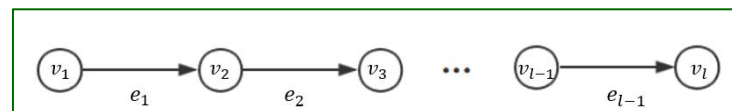
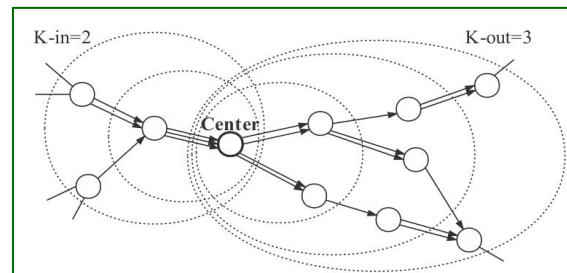
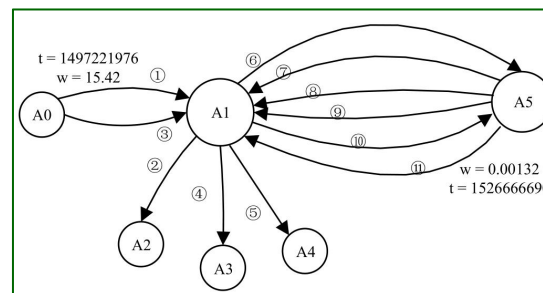
信息聚焦： K 阶有向子网络采样

↓ 空间密集型交易数据

网络刻画：带偏好的游走方法

↓ 学习网络嵌入向量

实验验证：时序链路预测



工作1：以太坊账户交易关系建模



■ IEEE Transactions on Circuits and Systems

IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: EXPRESS BRIEFS, VOL. 67, NO. 11, NOVEMBER 2020

2737

Modeling and Understanding Ethereum Transaction Records via a Complex Network Approach

Dan Lin, Jiajing Wu^{ID}, *Senior Member, IEEE*, Qi Yuan, and Zibin Zheng^{ID}, *Senior Member, IEEE*

Abstract—As the largest public blockchain-based platform supporting smart contracts, Ethereum has accumulated a large number of user transaction records since its debut in 2014. Analysis of Ethereum transaction records, however, is still relatively unexplored till now. Modeling the transaction records as a static simple graph, existing methods are unable to accurately characterize the temporal and multiplex features of the edges. In this brief, we first model the Ethereum transaction records as a complex network by incorporating time and amount features of the transactions, and then design several flexible temporal walk strategies for random-walk based graph representation of this large-scale network. Experiments of temporal link prediction on real Ethereum data demonstrate that temporal information and multiplicity characteristic of edges are indispensable for accurate modeling and understanding of Ethereum transaction networks.

Index Terms—Ethereum, blockchain, complex networks, graph representation, cryptocurrency, transaction network.

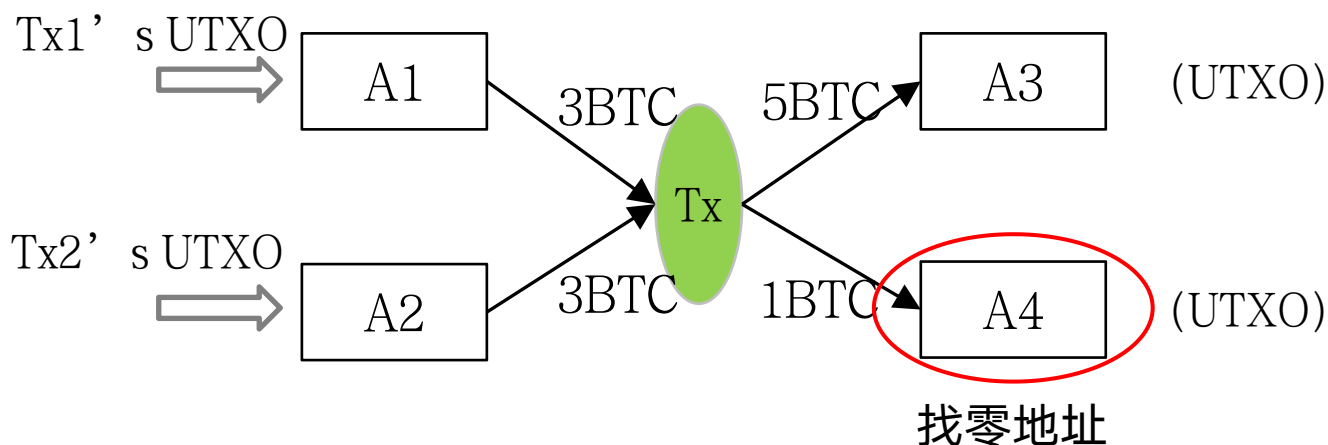
existing studies, transaction records are modeled as a simple graph, where multiple transactions between a pair of addresses are merged as a one-time transaction in the graph construction procedure.

Different from other large-scale complex networks, each edge in the Ethereum transaction network represents a particular Ether transaction, and thus contains some unique information such as the direction, amount value and timestamp of a particular transaction. It is essential to incorporate the aforementioned information for accurate modeling, characterization, and understanding of transaction network data. In addition, multiple transactions between two users are expected and it is more comprehensive to model a transaction network as a *multidigraph* rather than a simple graph. In graph theory, a multigraph (in contrast to a simple graph) is a graph which is permitted to have self-loops and multiple edges (also called parallel edges). A multidigraph is a directed multigraph.

工作2：比特币地址关系建模



➤ 比特币交易基于UTXO（未被使用的交易输出）模型



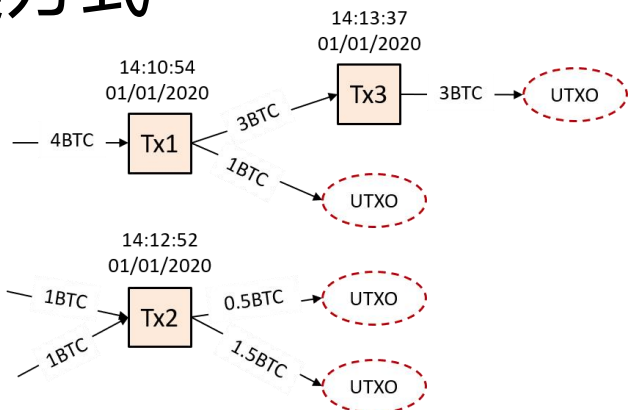
➤ 比特币交易特点

- 基于UTXO（未被使用的交易输出）模型
- 可能涉及到多个输入和输出
- 自动生成找零地址

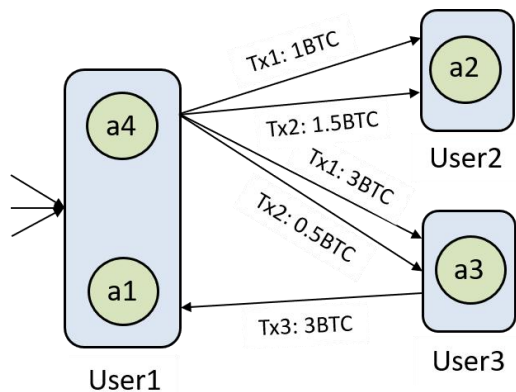
工作2：比特币地址关系建模



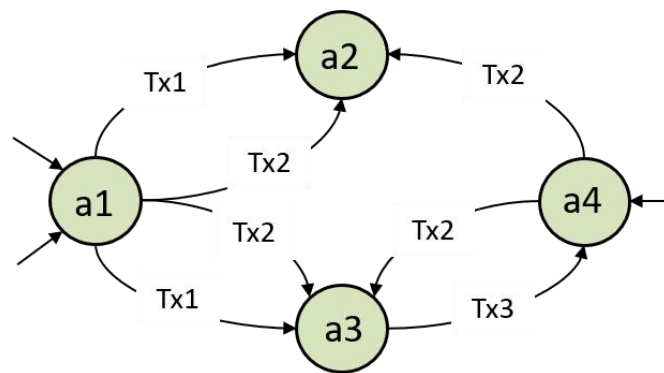
建模方式



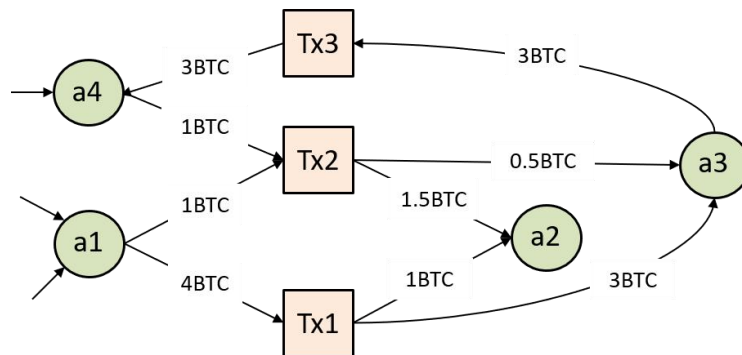
交易作为节点
较为直观的体现资金流的方式



用户使用多个地址
先将用户的地址聚合



地址作为节点
难以体现交易由多个账户参与的关系



交易、地址作为节点
更好地体现地址之间的交易关系

工作6：比特币混币服务检测



- IEEE Transactions on Systems, Man and Cybernetics: Systems

Detecting Mixing Services via Mining Bitcoin Transaction Network with Hybrid Motifs

Jiajing Wu, *Member, IEEE*, Jieli Liu, Weili Chen, Huawei Huang, *Member, IEEE*, Zibin Zheng, *Senior Member, IEEE*, and Yan Zhang, *Fellow, IEEE*

Abstract—As the first decentralized peer-to-peer (P2P) cryptocurrency system allowing people to trade with pseudonymous addresses, Bitcoin has become increasingly popular in recent years. However, the P2P and pseudonymous nature of Bitcoin make transactions on this platform very difficult to track, thus triggering the emergence of various illegal activities in the Bitcoin ecosystem. Particularly, *mixing services* in Bitcoin, originally designed to enhance transaction anonymity, have been widely employed for money laundry to complicate trailing illicit fund. In this paper, we focus on the detection of the addresses belonging to mixing services, which is an important task for anti-money laundering in Bitcoin. Specifically, we provide a feature-based network analysis framework to identify statistical properties of mixing services from three levels, namely, network level, account level and transaction level. To better characterize the transaction patterns of different types of addresses, we propose the concept of Attributed Temporal Heterogeneous motifs (ATH motifs). Moreover, to deal with the issue of imperfect labeling, we tackle the mixing detection task as a Positive and Unlabeled learning (PU learning) problem and build a detection model by leveraging the considered features. Experiments on real Bitcoin datasets demonstrate the effectiveness of our detection model and the importance of hybrid motifs including ATH motifs in mixing detection.

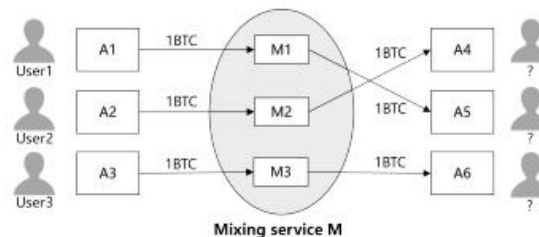


Fig. 1. An example of mixing services, which can conceal the identity of users and complicate fund tracing by participating in a transaction with multiple users.

be laundered into “clean” Bitcoins by some techniques before they are cashed out. It has been demonstrated that, mixing services such as BitLaundry, Helix Light, Bitcoin Fog, etc., have involved in this process of *money laundry* [5] and can be regarded as significant tools for concealing illicit profits in Bitcoin.

Bitcoin mixing services are originally designed to enhance the anonymity of transactions and make the sources of funds

3v1 [cs.SI] 15 Jan 2020

➤ arXiv:2011.09318v1 ,

Analysis of Cryptocurrency Transactions from a Network Perspective: An Overview

Jiajing Wu, *Senior Member, IEEE*, Jieli Liu, Yijing Zhao, and Zibin Zheng, *Senior Member, IEEE*

Abstract—As one of the most important and famous applications of blockchain technology, cryptocurrency has attracted extensive attention recently. Empowered by blockchain technology, all the transaction records of cryptocurrencies are irreversible and recorded in the blocks. These transaction records containing rich information and complete traces of financial activities are publicly accessible, thus providing researchers with unprecedented opportunities for data mining and knowledge discovery in this area. Networks are a general language for describing interacting systems in the real world, and a considerable part of existing work on cryptocurrency transactions is studied from a network perspective. This survey aims to analyze and summarize the existing literature on analyzing and understanding cryptocurrency transactions from a network perspective. Aiming to provide a systematic guideline for researchers and engineers, we present the background information of cryptocurrency transaction network analysis and review existing research in terms of three aspects, i.e., network modeling, network profiling, and network-based detection. For each aspect, we introduce the research issues, summarize the methods, and discuss the results and findings given in the literature. Furthermore, we present the main challenges and several future directions in this area.

Index Terms—Cryptocurrency, blockchain, transaction records, complex networks, data mining

all the transaction records of cryptocurrencies are irreversible and recorded in the blocks, which are linked in chronological order.

Due to the open and transparent nature of blockchain, these transaction records containing rich information and complete traces of financial activities are publicly accessible, thus providing researchers with unprecedented opportunities for data mining in this area. The main value of analyzing and mining the transaction data of cryptocurrencies is twofold: 1) Transaction records in traditional financial scenarios are relatively unexplored in existing studies as transaction records are usually not publicly accessible for the sake of security and interest. Through analysis and mining of transaction information of cryptocurrencies, we can extensively explore trading behaviors, wealth distribution, and generative mechanism of a transaction system, as well as infer reasons for fluctuations in the financial market of cryptocurrencies. This study can also provide a reference for knowledge discovery in other financial systems. 2) Due to the anonymity of blockchain systems and the lack of authority, various types of cybercrimes have arisen on the blockchain ecosystem in recent years.

8v1 [cs.SI] 18 Nov 2020

- 1 区块链数据
- 2 交易网络构建
- 3 网络分析与挖掘**
- 4 交易行为识别
- 5 其他工作

工作3：以太坊网络嵌入



数据收集

通过一个区块链浏览器Etherscan爬取以太坊交易数据

网络构建

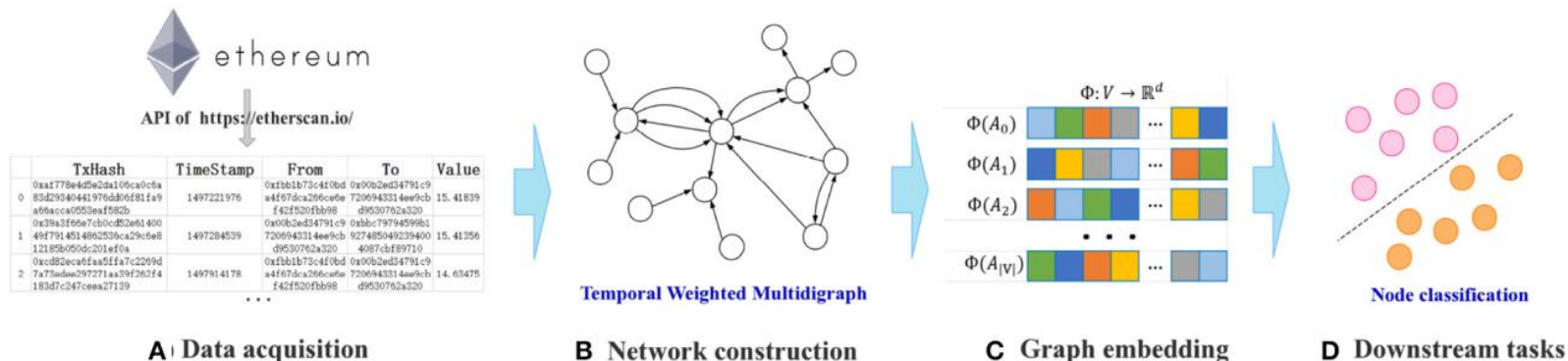
将获取的交易数据构造为一个复杂网络结构

网络嵌入

将构建的大规模网络用提出的T-EDGE嵌入到低维空间

下游任务

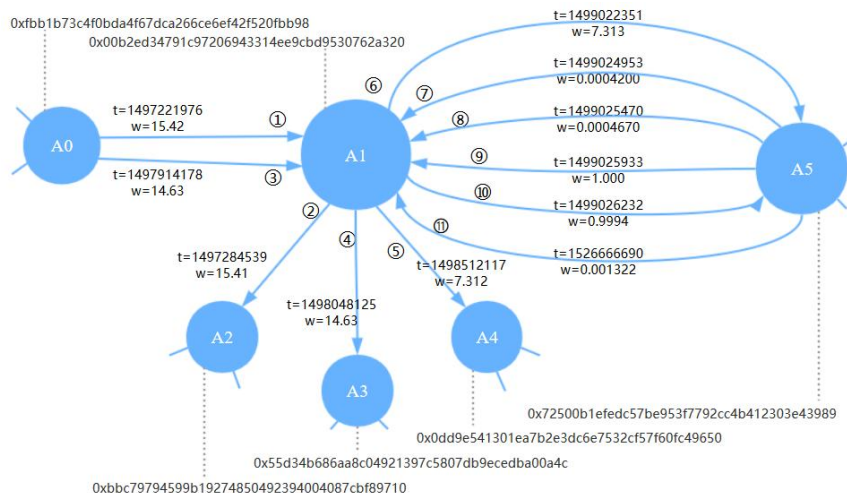
将节点嵌入向量应用到以太坊上的具体任务——钓鱼节点分类



工作3：以太坊网络嵌入



➤ 传统网络嵌入算法中随机游走的局限性

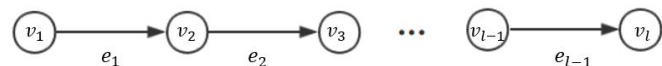


*注：按照t从小到大对边进行排序与标号

- **DeepWalk, node2vec:**
游走序列{A5, A1, A2}不符合交易网络的时序逻辑
- **CTDNE:**
A0到A1的游走被表示为节点集合 {A0, A1}, 不能确定节点A2是否为合法的下一个节点

➤ T-EDGE中时序游走的表示

- 目的：游走序列为有效的资金转移路径
- 定义：节点集合 + 时序递增的边集合
- 示意图：



➤ T-EDGE的时序游走策略

算法表示	时间域		权重域	
	无偏好	有偏好	无偏好	有偏好
T-EDGE	✓		✓	
T-EDGE (TBS)		✓	✓	
T-EDGE (WBS)	✓			✓
T-EDGE (TBS+WBS)		✓		✓

- 时间域的偏好：考虑交易的频率
- 权重域的偏好：考虑交易的金额大小

工作3：以太坊网络嵌入



■ Frontier in Physics



ORIGINAL RESEARCH
published: 30 June 2020
doi: 10.3389/fphy.2020.00204



T-EDGE: Temporal WEighted MultiDiGraph Embedding for Ethereum Transaction Network Analysis

Dan Lin ^{1,2}, Jiajing Wu ^{1,2*}, Qi Yuan ^{1,2} and Zibin Zheng ^{1,2}

¹ School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China, ² National Engineering Research Center of Digital Life, Sun Yat-sen University, Guangzhou, China

Recently, graph embedding techniques have been widely used in the analysis of various networks, but most of the existing embedding methods omit the network dynamics and the multiplicity of edges, so it is difficult to accurately describe the detailed characteristics of the transaction networks. Ethereum is a blockchain-based platform supporting smart contracts. The open nature of blockchain makes the transaction data on Ethereum completely public and also brings unprecedented opportunities for transaction network analysis. By taking the realistic rules and features of transaction networks into consideration, we first model the Ethereum transaction network as a Temporal Weighted Multidigraph (TWMDG) where each node is a unique Ethereum account and each edge represents a transaction weighted by amount and assigned a timestamp. We then define the problem of Temporal Weighted Multidigraph Embedding (T-EDGE) by incorporating both temporal and weighted information of the edges, the purpose being to capture more comprehensive properties of dynamic transaction networks. To evaluate the effectiveness of the proposed embedding method, we conduct experiments of node classification on real-world transaction data collected from Ethereum. Experimental results demonstrate that T-EDGE outperforms baseline embedding methods, indicating that time-dependent walks and the multiplicity characteristic of edges are informative and essential for time-sensitive transaction networks.

Keywords: network embedding, ethereum, machine learning, temporal network, transaction network

OPEN ACCESS

Edited by:

Jianguo Liu,
Shanghai University of Finance and
Economics, China

Reviewed by:

Shiyuan Wang,
Southwest University, China
Yongxiang Xia,
Hangzhou Dianzi University, China
Zhihai Rong,
University of Electronic Science and
Technology of China, China

*Correspondence:

Jiajing Wu
wujiajing@mail.sysu.edu.cn

Specialty section:

This article was submitted to

工作4：通过复杂网络分析EOSIO生态



➤ EOSIO是2018年上线的一款为商用分布式应用设计的区块链系统，具有秒级出块的特点

➤ EOSIO成立了仅三个月，其DApp交易量就超过了以太坊

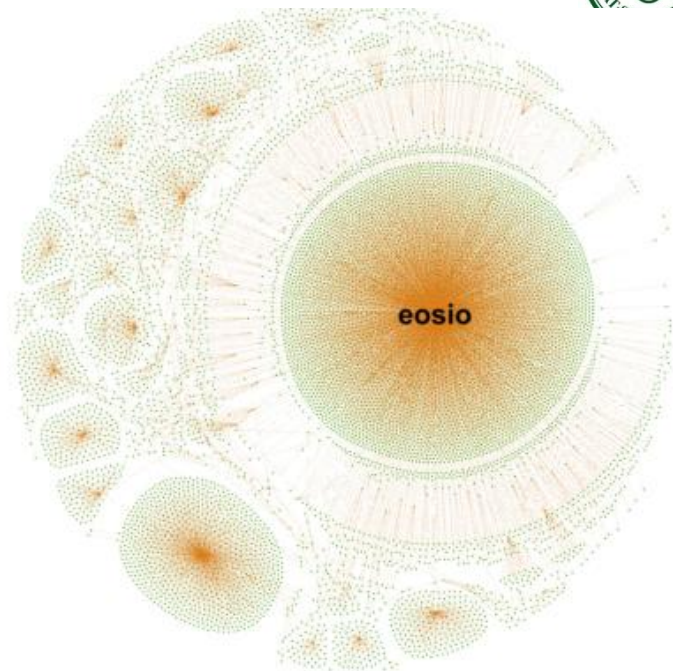
➤ EOSIO现有的分析较少，本工作基于一些复杂网络指标对EOSIO的生态进行分析



工作4：通过复杂网络分析EOSIO生态



- | 分析账户的四种行为：账户创建、投票、转账、授权
- | 对应构建四种网络
- | 通过网络可视化、度分布、聚类系数、连通分量等进行测量分析，帮助用户了解这个系统



EOSIO账户创建网络

- | 一些现象：账户转账网络是小世界网络；投票网络中存在一定规模的强连通分量，可能存在投票团伙

工作4：通过复杂网络分析EOSIO生态



■ International Conference on Blockchain and Trustworthy system

Exploring EOSIO via Graph Characterization

Yijing Zhao^{1,2,3}, Jieli Liu^{1,2}, Qing Han^{1,2}, Weilin Zheng^{1,2}, and Jiajing Wu^{1,2}

¹ School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China

² National Engineering Research Center of Digital Life, Sun Yat-sen University, Guangzhou, China

³ School of Electronics and Communication Engineering, Sun Yat-sen University, Guangzhou 510006, China

{zhaoyj53, liujli7, hanq25, zhengwlin}@mail2.sysu.edu.cn

{wujiajing}@mail.sysu.edu.cn

Abstract. Designed for commercial decentralized applications (DApps), EOSIO is a Delegated Proof-of-Stake (DPoS) based blockchain system. It has overcome some shortages of the traditional blockchain systems like Bitcoin and Ethereum with its outstanding features (e.g., free for usage, high throughput and eco-friendly), and thus becomes one of the mainstream blockchain systems. Though there exist billions of transactions in EOSIO, the ecosystem of EOSIO is still relatively unexplored. To fill this gap, we conduct a systematic graph analysis on the early EOSIO by investigating its four major activities, namely account creation, account state, money transfer and contract authorization. We obtain some

- 1 区块链数据
- 2 交易网络构建
- 3 网络分析与挖掘
- 4 交易行为识别**
- 5 其他工作

工作5：以太坊钓鱼诈骗检测



区块链的另一面

The Rise of Cryptocurrency Ponzi Schemes

Scammers are making big money off people who want in on a digital gold rush but don't understand how the tech works.

DAVID Z. MORRIS | MAY 31, 2017 | TECHNOLOGY

Cyber Criminals Have Stolen More Than \$1 Billion Worth of Ethereum Through This Year

'Market manipulation 101': 'Worms' and 'pump and dump' scams plague crypto



Oscar Williams-Grut
Nov. 14, 2017, 7:00 AM 26,789



Sylvain Ribes
Trader and investor. Follow me on Twitter @ArtPlaie
Mar 10 · 11 min read

Chasing fake volume: a crypto-plague

Why I believe more than \$3 billion of all cryptoassets' volume is fake, and how OKex, #1 exchange rated by volume, is the biggest culprit.

Ponzi Scheme Operator Hit \$1 Billion Judgment

花式骗局全记录 | 钛媒体深度

得得研报 钛媒体

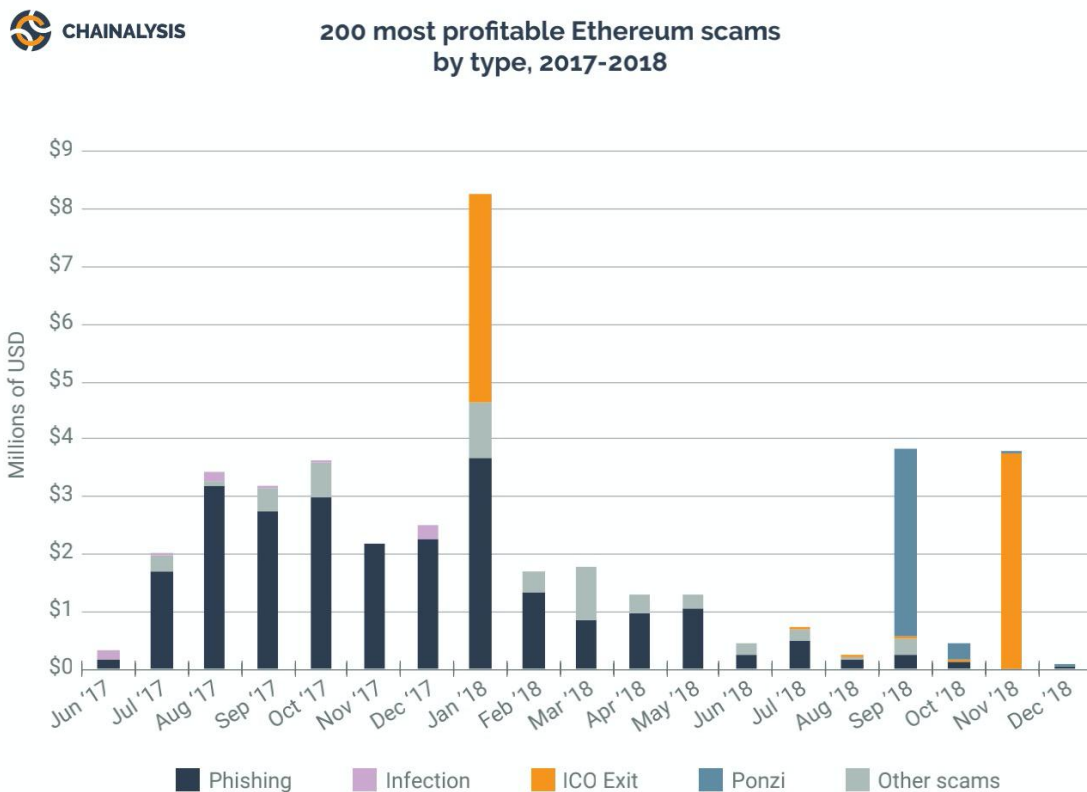


ICO遭遇网络钓鱼诈骗 25小时用户损失近100万美元

工作5：以太坊钓鱼诈骗检测



各欺诈造成的非法收益对比：从一份来自CHAINALYSIS的报告可知，以太坊上钓鱼诈骗造成的非法收益(深灰色柱)占了较大的比重，仅在一个月就可超过3百万美元



工作5：以太坊钓鱼诈骗检测



一个典型的例子—Bee Token:

- 国际象棋组织
- 2018年，在ICO发布前，向潜在的投资者发送钓鱼邮件，承诺在未来6小时内给所有投资者100%的红利，以及两个月内Bee Token的价值将翻倍
- 该钓鱼诈骗最终在**25小时内**累计骗取资金接近**100万美金**



ICO CROWDSALE IS NOW OPEN!

After much waiting, The Bee Token is proud to announce that our crowdsale is **NOW OPEN!** You used this address to register to our newsletter so we thought we'd give you some instructions on how to participate.

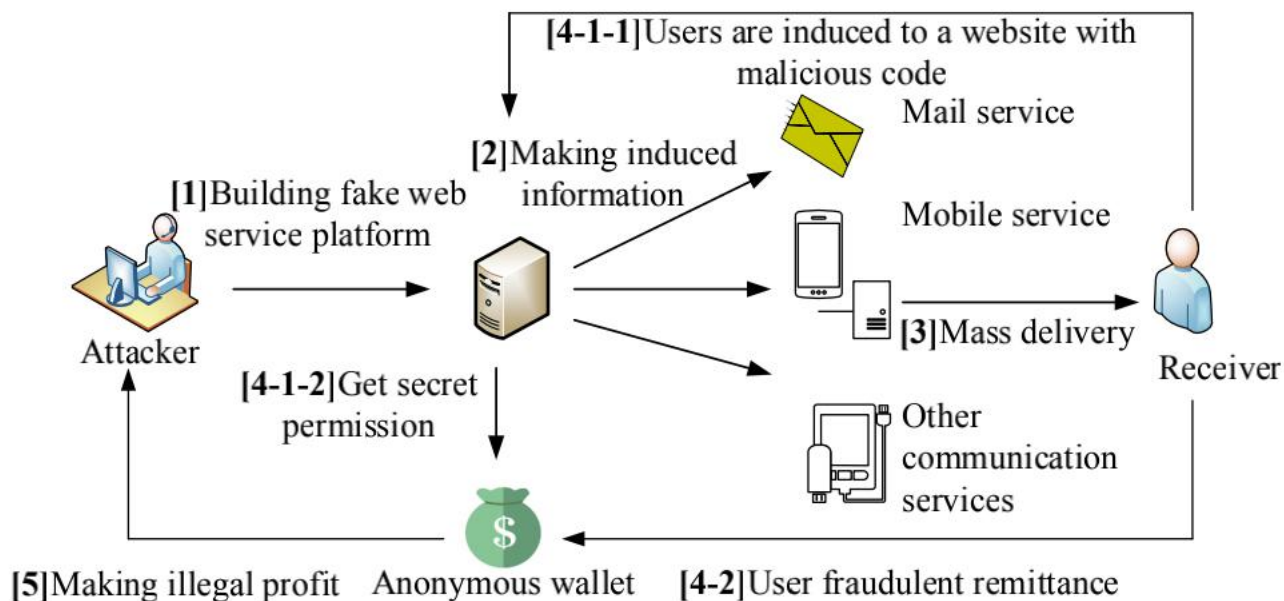
Firstly, we have modified your contribution maximum limit to be **104.43 ETH** so please don't send more than this or you won't be compensated. If you are receiving this email then you are permitted to join the ICO but your contribution limit is only guaranteed for 24 hours so don't miss out!

Secondly, to celebrate our **NEW** partnership with Microsoft thought we'd give you a **100% BONUS** on all tokens sent within the next 6 hours. We guarantee that The Bee Token will double in value within 2 months or we'll give you your Ethereum back!

◆ Our Ethereum ICO address is:

0x2A6D8021861f27aB992572D8689017b7A83C989D

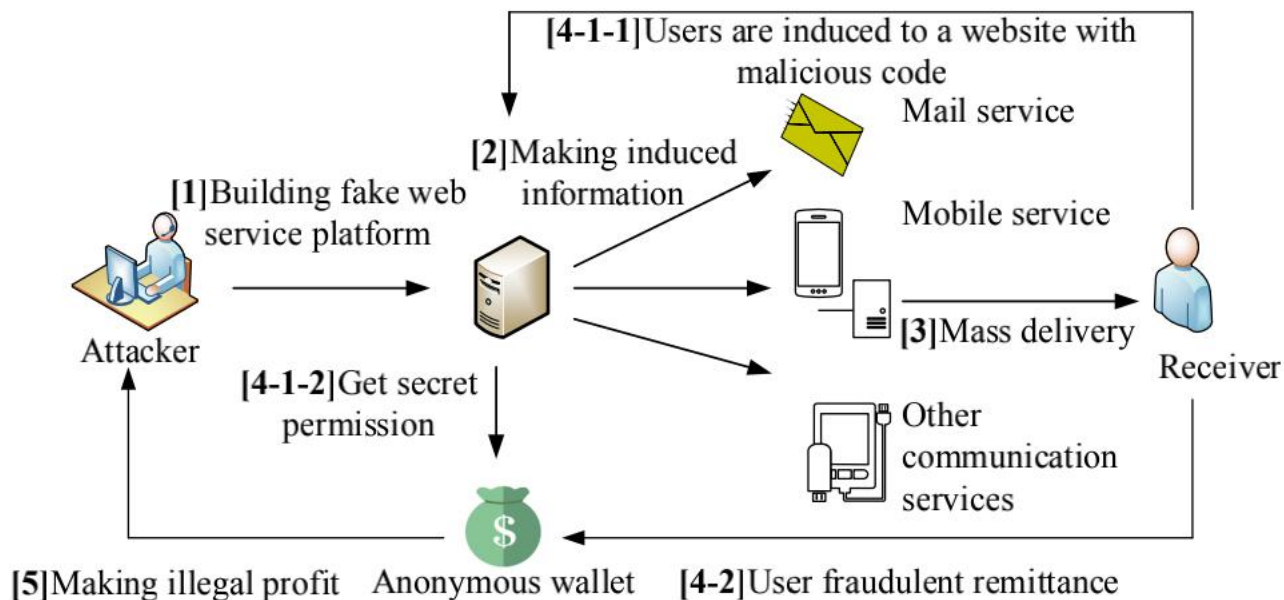
工作5：以太坊钓鱼诈骗检测



典型钓鱼诈骗过程：

1. 建立虚假的服务平台（高仿网站、软件等）
2. 设计诱导的信息内容
3. 诱导信息发放，比如邮件、短信等
4. 获利阶段

工作5：以太坊钓鱼诈骗检测

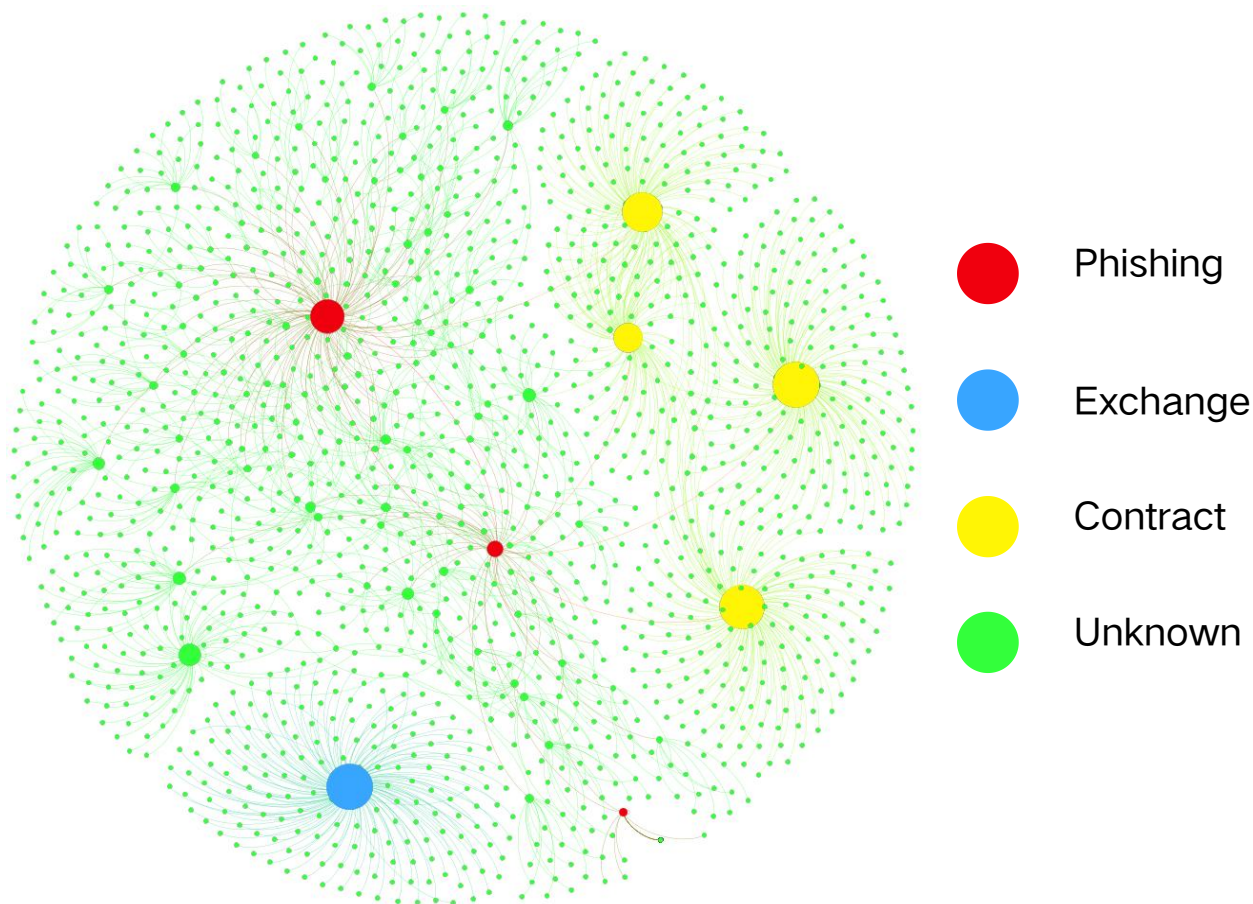


- 普通地址、尤其是真实ICO地址与钓鱼地址特征上有什么区别？（特征工程）
- 是否还有一些钓鱼地址，没有被披露？（模型识别）
- 如何解决后钓鱼阶段问题？（reputation）

工作5：以太坊钓鱼诈骗检测



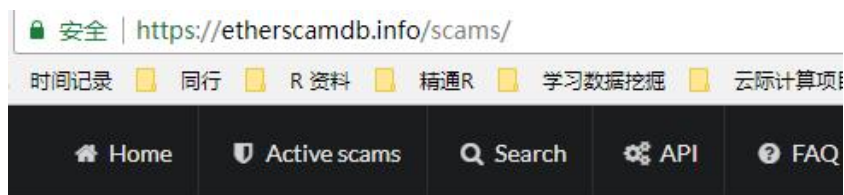
交易数据



工作5：以太坊钓鱼诈骗检测



Github 数据



3,859
TOTAL SCAMS

554
ACTIVE SCAMS



Scams

Category	Subcategory
Scamming	

用户评论

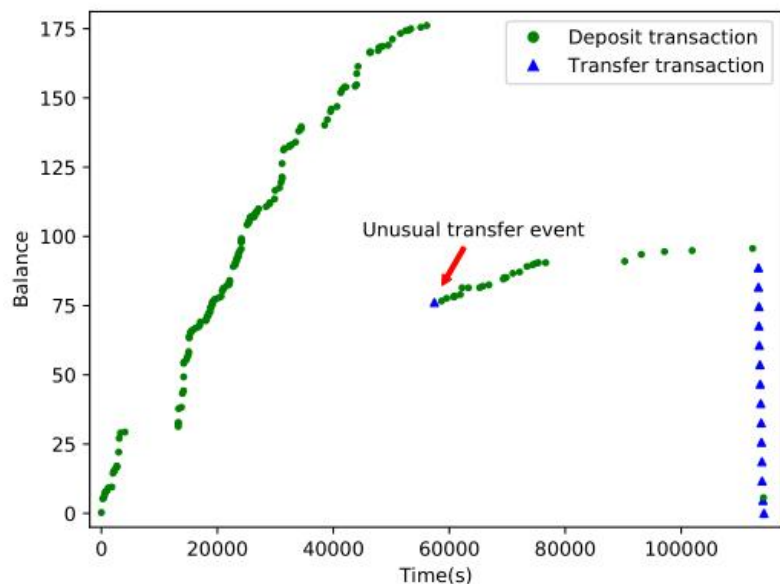
 **Slothlover**
February 1, 2018
[Reply](#)

I can't believe they scammed me out of \$200 🙄 the email looked really official. And they're still scamming people right now!! We need to share this before they steal even more money!!

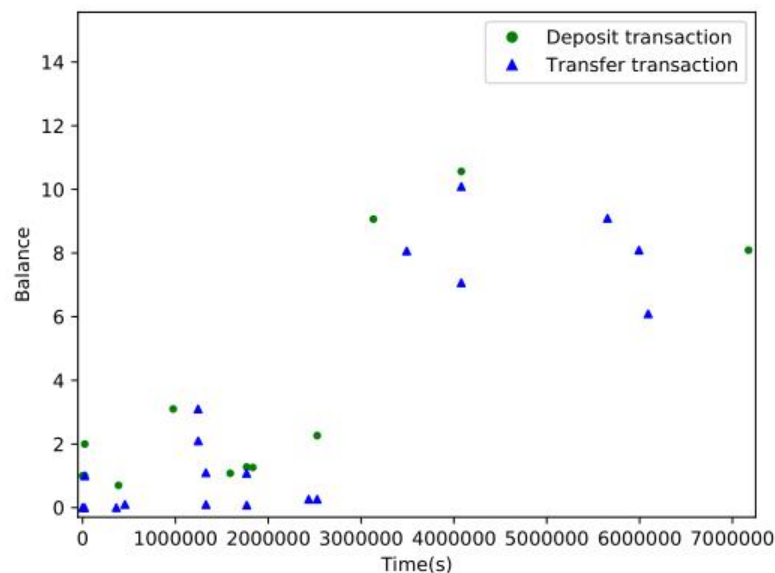
工作5：以太坊钓鱼诈骗检测



存取模式对比



一个典型钓鱼账户



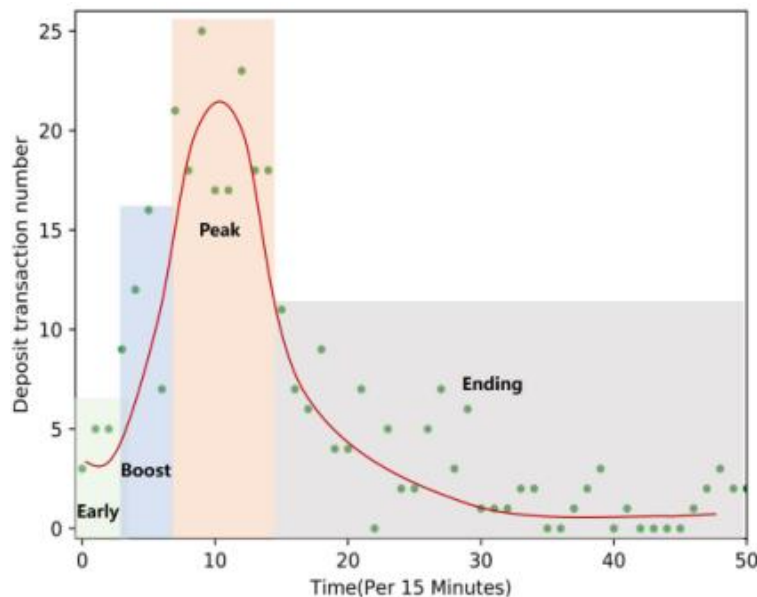
一个典型非钓鱼账户

- 钓鱼账户在开始时基本只会接受转入，从某个时刻开始，会将资金大额转出（可能开始洗钱、套现动作）
- 非钓鱼账户，存取动作表现更均匀一些

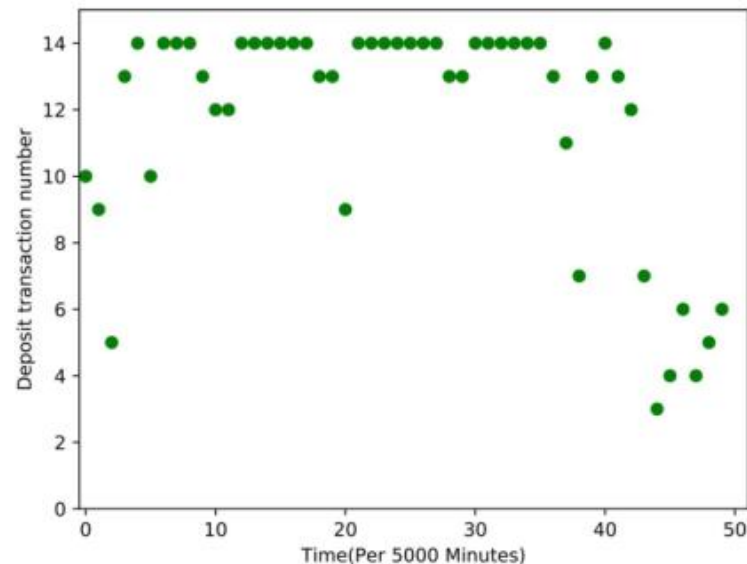
工作5：以太坊钓鱼诈骗检测



资金流入频率对比



Bee Token 钓鱼账户



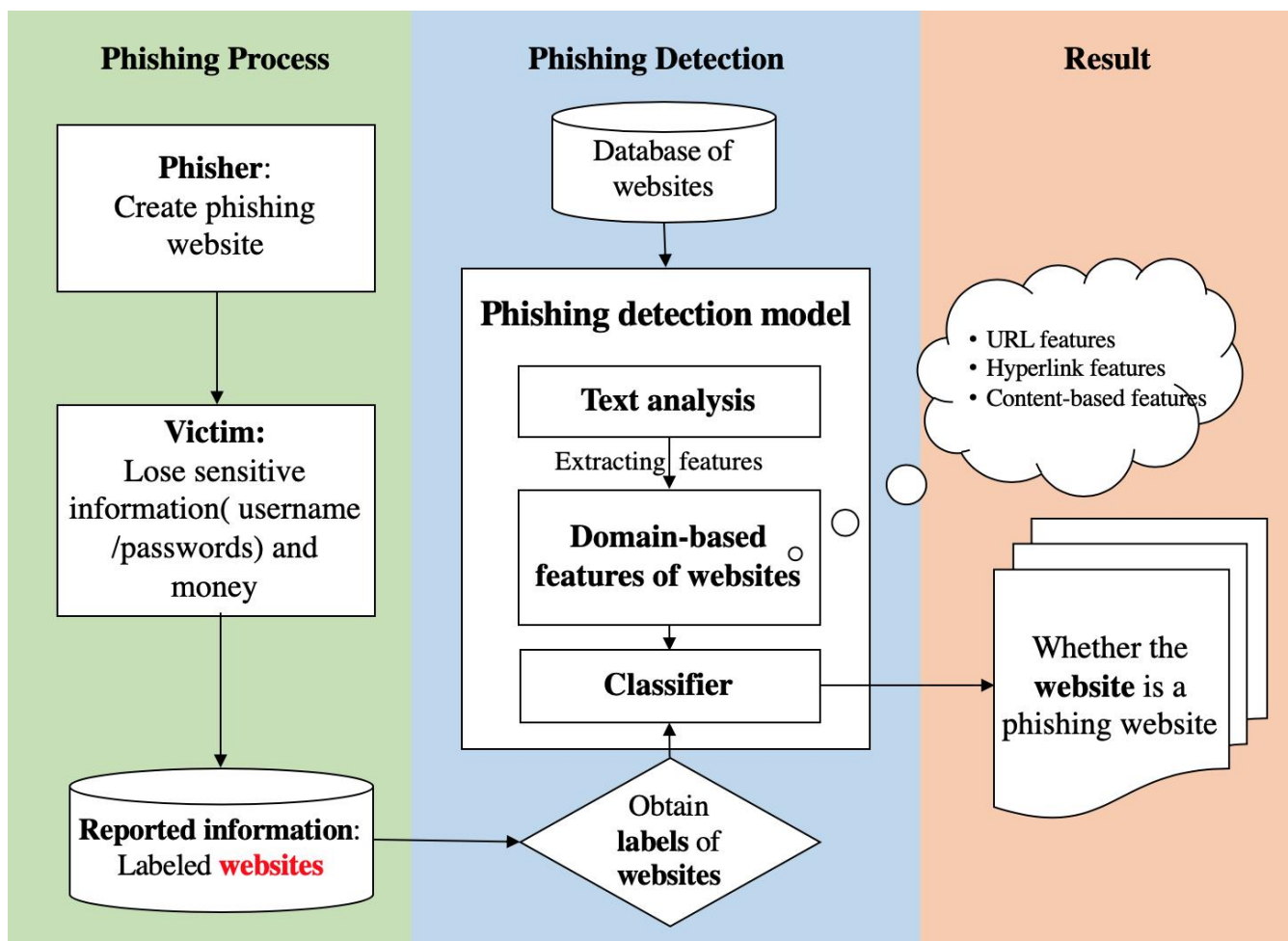
一个典型非钓鱼账户

- 典型钓鱼账户在开始阶段较少，然后会快速爆发，最后慢慢趋于平静
- 非钓鱼账户，资金流入相对稳定

工作5：以太坊钓鱼诈骗检测



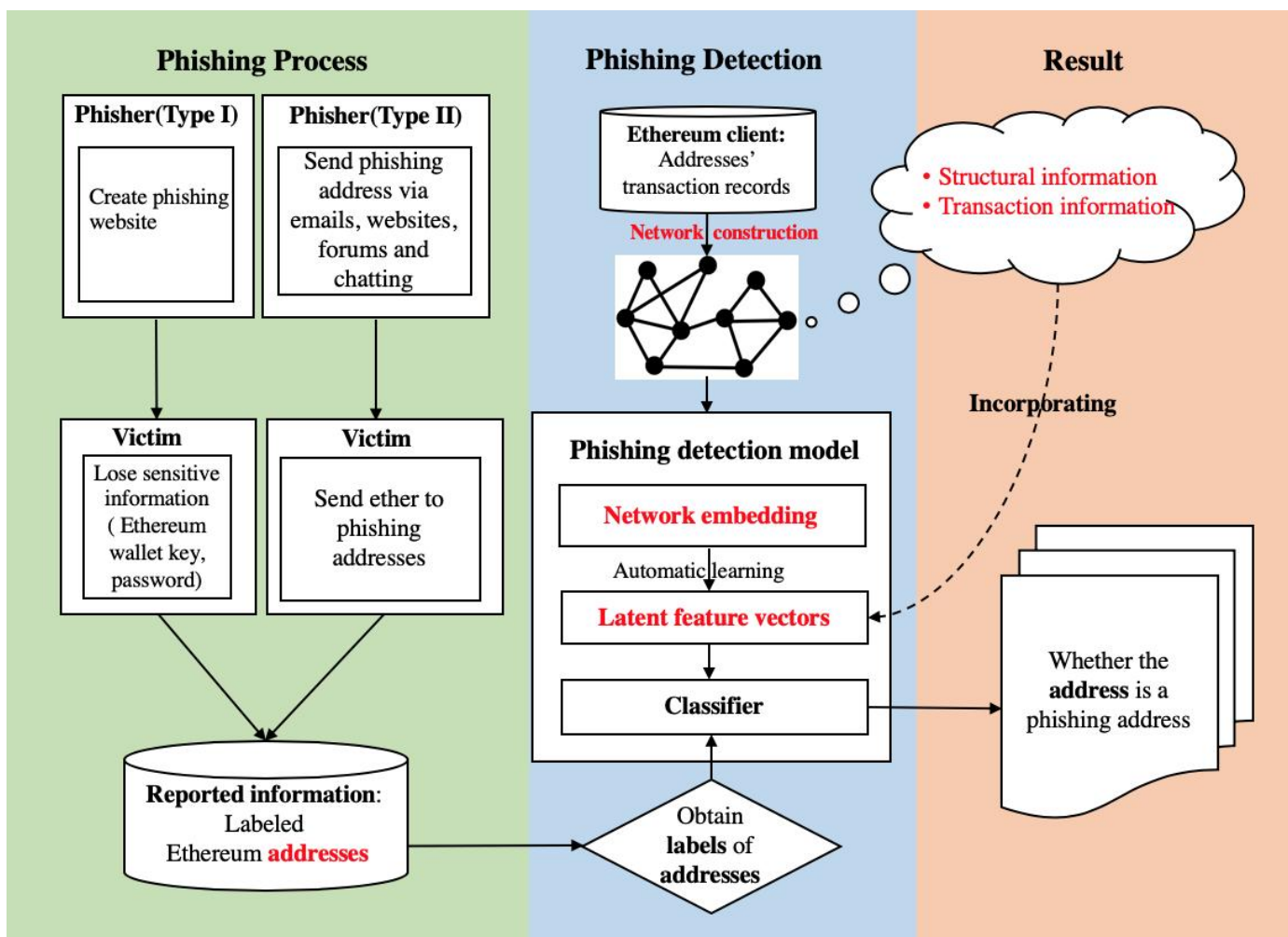
传统钓鱼诈骗识别过程



工作5：以太坊钓鱼诈骗检测



基于网络表示学习方法的以太坊钓鱼识别过程

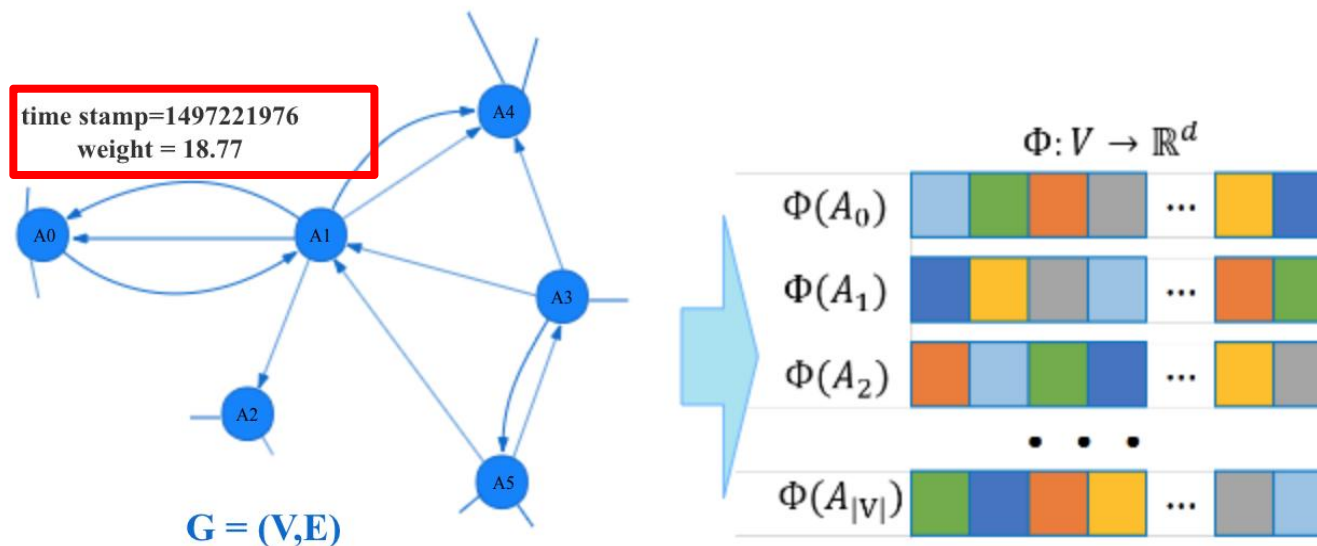


工作5：以太坊钓鱼诈骗检测



以太坊交易网络：含有时间戳和交易金额信息

trans2vec：融合了时间戳和交易金额信息的网络表示学习方法



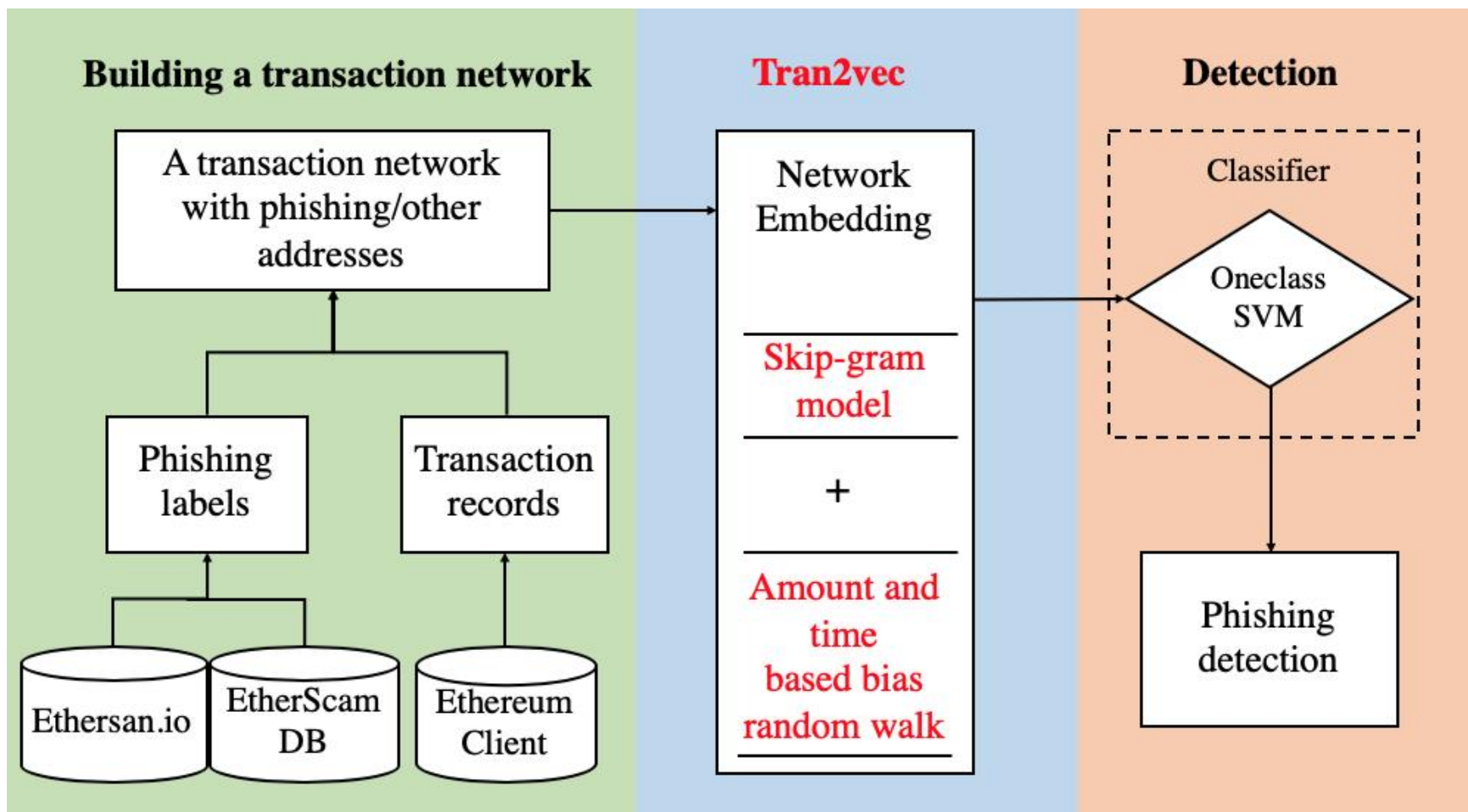
Ethereum network

Network embedding

工作5：以太坊钓鱼诈骗检测



基于trans2vec的识别模型



工作5：以太坊钓鱼诈骗检测



评价指标

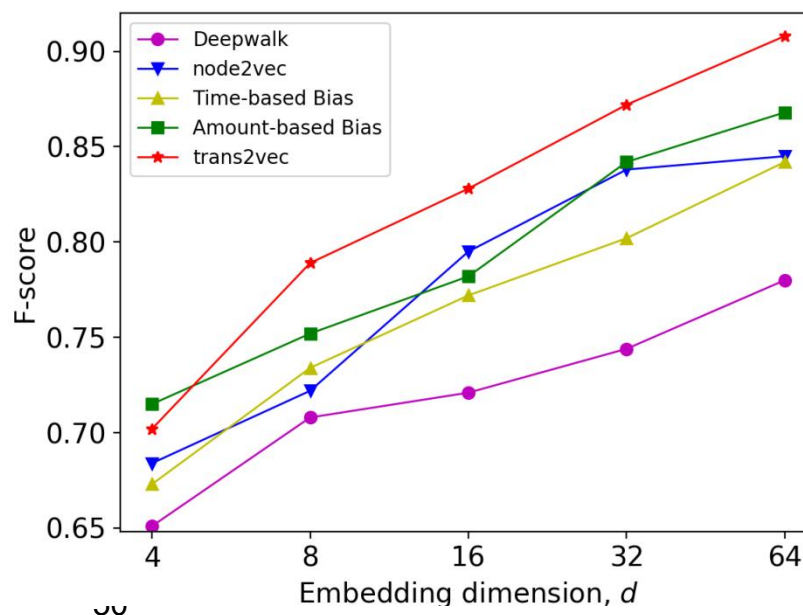
$$\text{Precision} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}}$$

$$\text{Recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}}$$

$$\text{F-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

结果

Method	Precision	Recall	F-score
Deepwalk	0.799	0.762	0.780
Node2vec	0.870	0.822	0.845
Time-based Bias	0.864	0.822	0.842
Amount-based Bias	0.883	0.855	0.868
trans2vec	0.927	0.893	0.908



工作5：以太坊钓鱼诈骗检测



➤ IEEE Transactions on Systems, Man and Cybernetics: Systems

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS

1

Who Are the Phishers? Phishing Scam Detection on Ethereum via Network Embedding

Jiajing Wu¹, Senior Member, IEEE, Qi Yuan, Dan Lin², Wei You, Weili Chen³,
Chuan Chen⁴, Member, IEEE, and Zibin Zheng⁵, Senior Member, IEEE

Abstract—Recently, blockchain technology has become a topic in the spotlight but also a hotbed of various cybercrimes. Among them, phishing scams on blockchain have been found to make a notable amount of money, thus emerging as a serious threat to the trading security of the blockchain ecosystem. In order to create a favorable environment for investment, an effective method for detecting phishing scams is urgently needed in the blockchain ecosystem. To this end, this article proposes an approach to detect phishing scams on Ethereum by mining its transaction records. Specifically, we first crawl the labeled phishing addresses from two authorized websites and reconstruct the transaction network according to the collected transaction records. Then, by taking the transaction amount and timestamp into consideration, we propose a novel network embedding algorithm called *trans2vec* to extract the features of the addresses for subsequent phishing identification. Finally, we adopt the one-class support vector machine (SVM) to classify the nodes into normal and phishing ones. Experimental results demonstrate that the phishing detection method works effectively on Ethereum, and indicate the efficacy of *trans2vec* over existing state-of-the-art algorithms

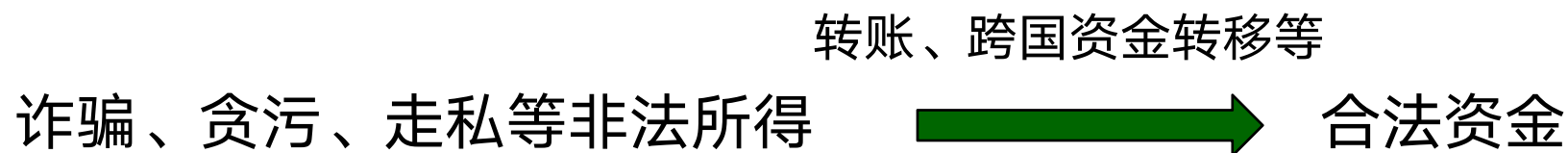
application of blockchain and the first practical implementation of cryptocurrency.

Ethereum is currently the largest blockchain platform that supports smart contracts and the corresponding cryptocurrency *ether* is the second-largest cryptocurrency [3]. However, along with its high-speed development, Ethereum has also become a hotbed of various cybercrimes [4]. Initial coin offering (ICO) is a financing method for the blockchain industry, which refers to financing through the issuance of tokens. However, till now, more than 10% of ICOs released on Ethereum have been reported to be suffer from a variety of scams, including phishing, Ponzi schemes, etc. [5]. According to a report of *Chainalysis*, a provider of investigation and risk management software for virtual currencies, there were 30 287 victims losing \$225 million in the first half of 2017 [6], indicating that financial security has become a critical issue in the blockchain ecosystem.

工作6：比特币混币服务检测



➤ 洗钱：将非法所得合法化的过程



➤ 三个阶段：

- 安置：犯罪分子向金融系统注入黑钱
- 分层：洗钱者进行复杂的多层的金融交易使黑钱最大程度地分散开，并让黑钱和合法的钱融为一体
- 整合：清洁后的资金在看似合法的状态下重新进入金融系统

工作6：比特币混币服务检测



➤ 区块链交易特点：

- 注册简单，无需身份认证
- 使用假名
- 低手续费
- 跨国交易方便
- 匿名币交易的不可追踪

➤ 区块链上的非法所得：从链下注入的非法所得、诈骗所得、资金盘跑路资金、窃币所得等

工作6：比特币混币服务检测

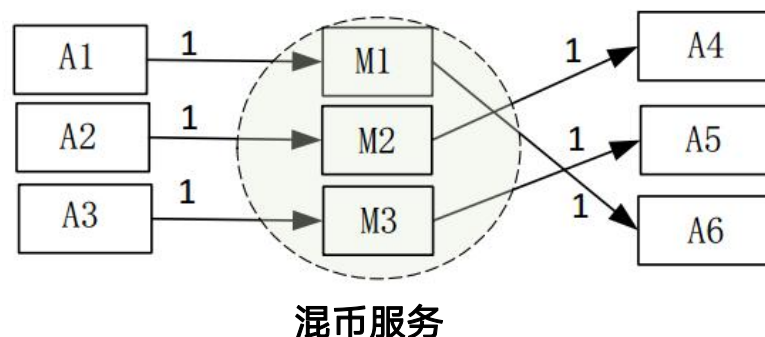


➤ 区块链上的洗钱途径：

赌博、混币服务、廉价的区块链转账、交易所币币兑换等
其中，交易所和混币服务是两个关键的区块链洗钱组件

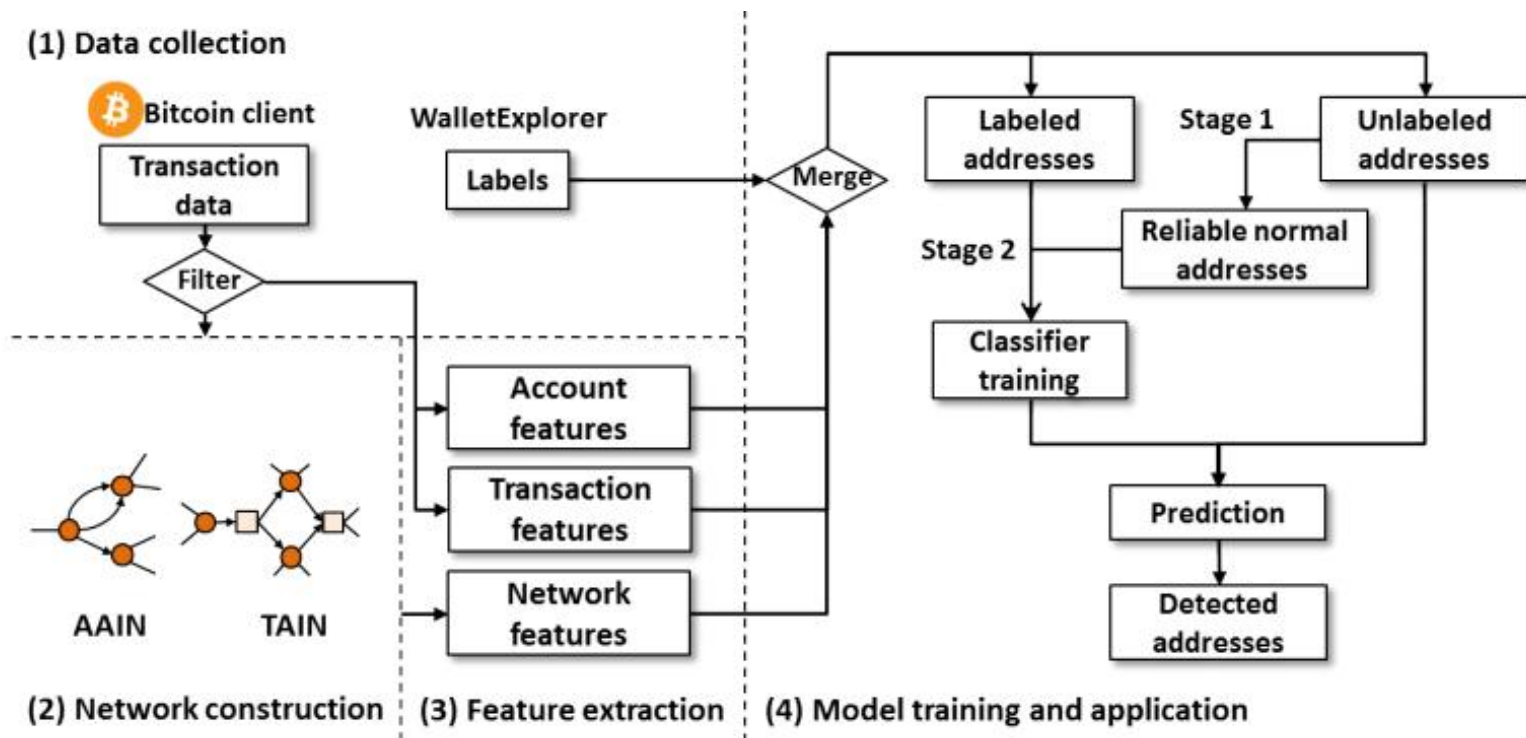
➤ 混币服务

- 目的：加强交易匿名性
- 方法：多个用户间的资金快速高效混合
账户之间创建随机的映射关系
- 成为比特币**洗钱**的一种手段



研究目的：检测参与混币的地址，进一步可以分析相关用户，判断它们是否参与犯罪行为，规范加密货币市场

工作6：比特币混币服务检测



方法：

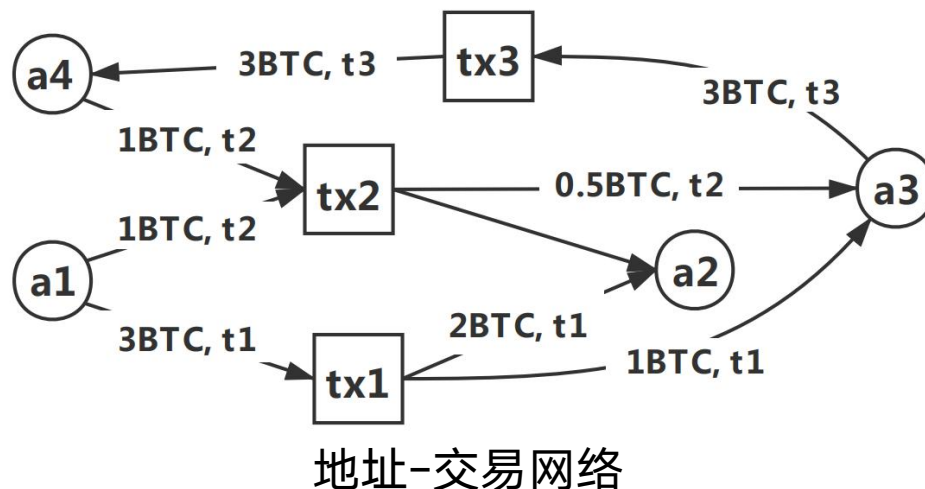
- 数据收集
- 交易网络构建
- 特征提取（时序网络模型+异构网络模型作为主要特征）
- 模型训练

工作6：比特币混币服务检测



网络构建

- 地址-地址网络
- 地址-交易网络



网络模体

- 网络中反复出现的子图或相互作用的模式，指定了节点间交互的特定模式，揭示了复杂网络的基本构建块
- 利用的模体：
 - 时序网络模体 (Temporal motif)
 - 带属性的时序网络模体 (Attributed temporal motif)

工作6：比特币混币服务检测



TABLE II
AVERAGE FRACTION OF δ -TEMPORAL MOTIFS ($\delta = 3$ HOURS).

Temporal motif						
Labeled address	0.2552	0.0051	0.5902	0.1465	0.0000	0.0030
Unlabeled address	0.2320	0.0576	0.4016	0.2395	0.0003	0.0690

TABLE III
AVERAGE FRACTION OF δ -ATH MOTIFS ($\delta = 3$ HOURS).

ATH motif				
Labeled address	0.4957	0.4916	0.0057	0.0069
Unlabeled address	0.6557	0.3148	0.0069	0.0225

- ◆ 发现一： a_1 的比重远远大于 a_2 ， b_4 （输入金额大于等于输出金额，输入时间小于等于输出时间）的比重远远大于同作为既有输入又有输出的 b_5 和 b_6 模式，说明混币服务交易更遵从与先输入后输出、结余不小于0的模式
- ◆ 发现二： a_5 和 a_6 的占比较小，说明混币服务地址相对普通地址来说更少进行地址重用
- ◆ 发现三： a_3 的占比很多可能是因为找零地址模式

工作6：比特币混币服务检测



TABLE VI
PERFORMANCE COMPARISON OF DIFFERENT FEATURES (WITH STANDARD DEVIATION).

Dataset	Metric	Basic features	Temporal motifs	ATH motifs	Hybrid motifs*	Basic features & Temporal motifs	Basic features & ATH motifs	Basic features & Hybrid motifs*
2014	TPR	0.8744±0.0145	0.8728±0.0070	0.7059±0.0111	0.8912±0.0064	0.9032±0.0064	0.8797±0.0089	0.9165±0.0060
	FPR	0.1779±0.0128	0.0455±0.0009	0.1508±0.0013	0.0318±0.0007	0.0362±0.0014	0.1350±0.0091	0.0334±0.0010
	G-Mean	0.8479±0.0120	0.9127±0.0036	0.7742±0.0059	0.9289±0.0032	0.9330±0.0033	0.8723±0.0075	0.9412±0.0029
2015	TPR	0.8146±0.0115	0.8453±0.0098	0.8426±0.0100	0.8823±0.0088	0.8864±0.0092	0.8543±0.0095	0.9149±0.0081
	FPR	0.1388±0.0038	0.1423±0.0024	0.0716±0.0009	0.0667±0.0020	0.0878±0.0079	0.0852±0.0018	0.0379±0.0016
	G-Mean	0.8376±0.0064	0.8515±0.0043	0.8845±0.0051	0.9074±0.0041	0.8992±0.0065	0.8840±0.0048	0.9382±0.0038
2016	TPR	0.6442±0.0317	0.9271±0.0073	0.6639±0.0145	0.9123±0.0071	0.9335±0.0067	0.8150±0.0112	0.9318±0.0066
	FPR	0.3812±0.0077	0.0584±0.0012	0.3154±0.0043	0.0356±0.0011	0.0508±0.0011	0.1995±0.0047	0.0356±0.0010
	G-Mean	0.6311±0.0129	0.9343±0.0035	0.6741±0.0061	0.9380±0.0034	0.9413±0.0031	0.8077±0.0047	0.9479±0.0031

* Hybrid motifs are a combination of Temporal and ATH motifs.

数据：在2014，2015，2016年各取连续的150万条交易

标签来源：Walletexplorer.com

结论：网络模体使模型在原有的检测效果上有所提升

工作6：比特币混币服务检测



- IEEE Transactions on Systems, Man and Cybernetics: Systems

Detecting Mixing Services via Mining Bitcoin Transaction Network with Hybrid Motifs

Jiajing Wu, *Member, IEEE*, Jieli Liu, Weili Chen, Huawei Huang, *Member, IEEE*, Zibin Zheng, *Senior Member, IEEE*, and Yan Zhang, *Fellow, IEEE*

Abstract—As the first decentralized peer-to-peer (P2P) cryptocurrency system allowing people to trade with pseudonymous addresses, Bitcoin has become increasingly popular in recent years. However, the P2P and pseudonymous nature of Bitcoin make transactions on this platform very difficult to track, thus triggering the emergence of various illegal activities in the Bitcoin ecosystem. Particularly, *mixing services* in Bitcoin, originally designed to enhance transaction anonymity, have been widely employed for money laundry to complicate trailing illicit fund. In this paper, we focus on the detection of the addresses belonging to mixing services, which is an important task for anti-money laundering in Bitcoin. Specifically, we provide a feature-based network analysis framework to identify statistical properties of mixing services from three levels, namely, network level, account level and transaction level. To better characterize the transaction patterns of different types of addresses, we propose the concept of Attributed Temporal Heterogeneous motifs (ATH motifs). Moreover, to deal with the issue of imperfect labeling, we tackle the mixing detection task as a Positive and Unlabeled learning (PU learning) problem and build a detection model by leveraging the considered features. Experiments on real Bitcoin datasets demonstrate the effectiveness of our detection model and the importance of hybrid motifs including ATH motifs in mixing detection.

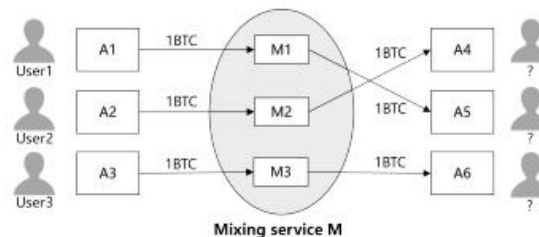


Fig. 1. An example of mixing services, which can conceal the identity of users and complicate fund tracing by participating in a transaction with multiple users.

be laundered into “clean” Bitcoins by some techniques before they are cashed out. It has been demonstrated that, mixing services such as BitLaundry, Helix Light, Bitcoin Fog, etc., have involved in this process of *money laundry* [5] and can be regarded as significant tools for concealing illicit profits in Bitcoin.

Bitcoin mixing services are originally designed to enhance the anonymity of transactions and make the sources of funds

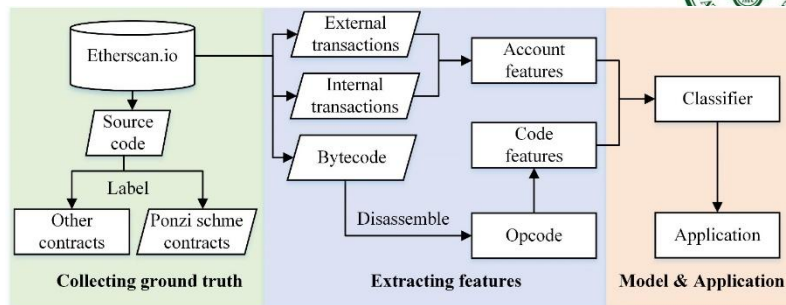
3v1 [cs.SI] 15 Jan 2020

- 1 区块链数据
- 2 交易网络构建
- 3 网络分析与挖掘
- 4 交易行为识别
- 5 其他工作

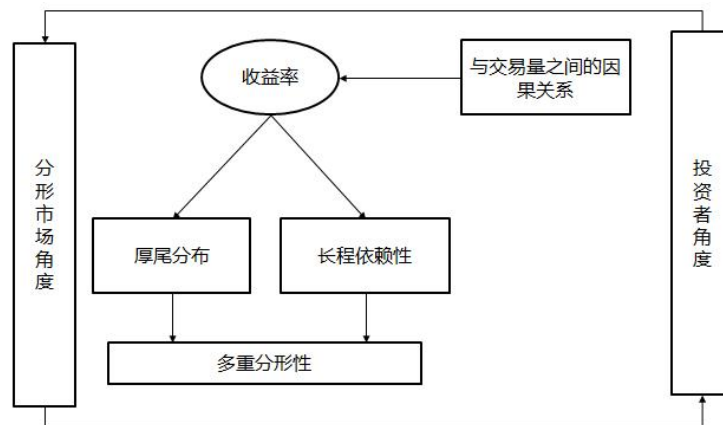
其他工作



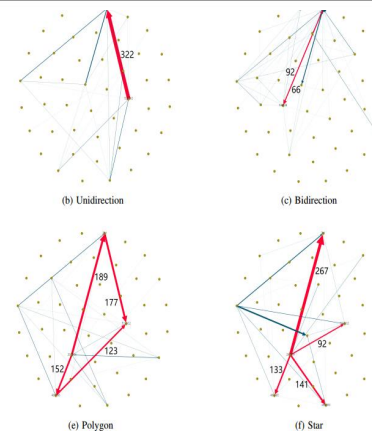
- **庞氏骗局检测**。基于字节码和账户行为对庞氏骗局进行识别



- **以太坊市场数据分析**。从分形市场的角度出发，发现收益率的厚尾分布、长程依赖性和不同时间尺度上的多重分形性



- **比特币市场操纵检测**。基于复杂网络分析和矩阵分解对比特币交易所 Mt.Gox 的交易数据进行分析

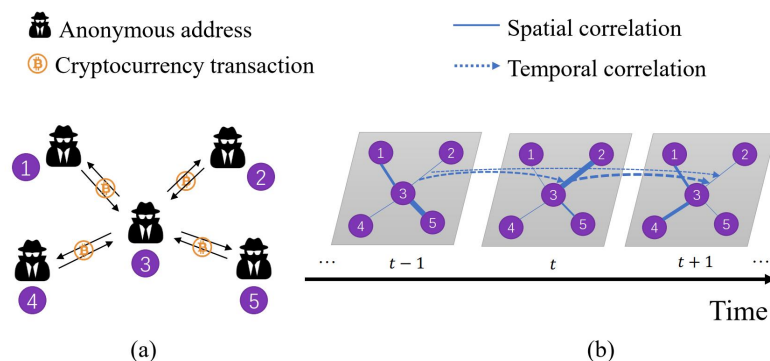


操纵模式

其他工作



- **基于混合时空表示学习的以太坊交易预测。**将交易预测建模为链路权重预测问题，即预测交易网络中节点对的权重变化



- **基于网络传播方程的KYC/KKT。**根据资金是否受到可疑路径的污染，或是否与已知的异常账户相关，计算地址和交易的风险等级

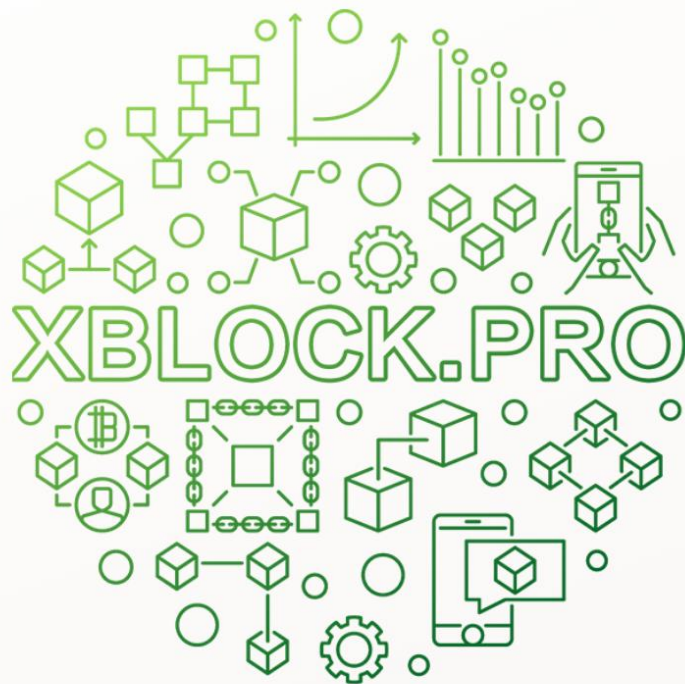


图片来源 CipherTrace

| <http://xblock.pro/>



[Home](#) [Transaction-Dataset](#) [Contract-Dataset](#) [Market-Dataset](#) [Related Papers](#) [About](#) [Help](#)



eXplore **Blockchain** Reliability

XBLOCK.PRO

XBlock collects the current mainstream blockchain data and is one of the blockchain data platforms in the academic community.

All blockchain datasets have been cleaned and classified in a standardized way, which can be easily downloaded into a standard and consistent format.



区块链数据网站



I <http://xblock.pro/>

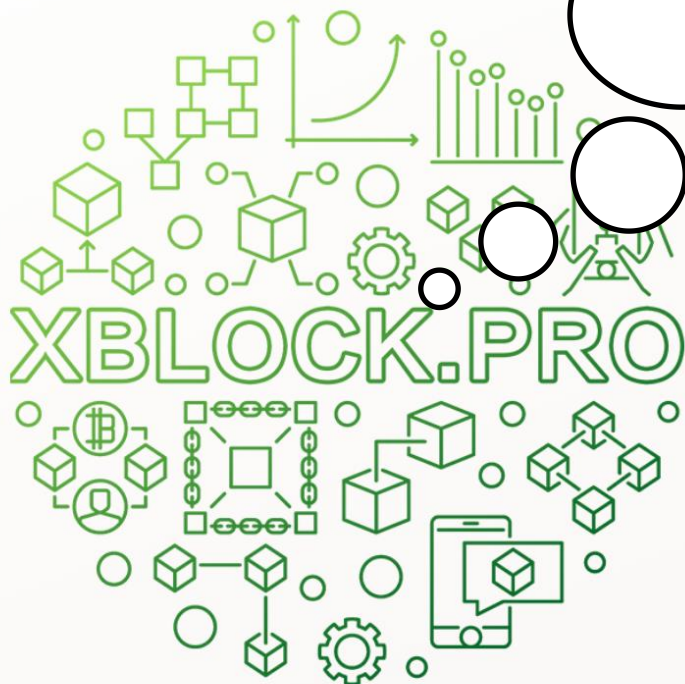
 **XBlock**

Ethereum: 128G压缩包
810万区块
4.9亿交易
5.4千万账户
1.7千万智能合约
11亿次合约调用

XBLOCK.PRO

XBlock collects the current mainstream blockchain data and is one of the blockchain data platforms in the academic community.

All blockchain datasets have been cleaned and classified in a standardized way, which can be easily downloaded into a standard and consistent format.



Transaction Dataset

钓鱼诈骗数据集

部分交易快照，含混币服务地址标签

Labeled Dataset



First-order Transaction Network of Phishing Nodes



Second-order Transaction Network of Phishing Nodes



Ethereum Phishing Transaction Network



Bitcoin Partial Transaction Dataset

Unlabeled Dataset



Ethereum On-chain Data



EOSIO On-chain Data



Ethereum Partial Transaction Dataset

Transaction Dataset

Labeled Dataset



First-order Transaction Network of Phishing Nodes



Second-order Transaction Network of Phishing Nodes



Ethereum Phishing Transaction Network



Bitcoin Partial Transaction Dataset

Unlabeled Dataset



Ethereum On-chain Data



EOSIO On-chain Data



Ethereum Partial Transaction Dataset

通过以太坊全节点
获得链上数据。

EOSIO全节点获
得的链上数据。

便于分析的
较小数据集。

➤ Contract Dataset

庞氏骗局是一种欺骗性的投资操作，经营者通过新投资者支付的收入为老投资者创造回报。

开源合约数据集包含约14000个合约

Labeled Dataset

Smart Ponzi Scheme Labels

Unlabeled Dataset

Smart Contract Attribute Dataset

➤ Market Dataset

- 价格及数量数据集
 - ◆ Ether Price and Volume Dataset
 - ◆ Bitcoin Price and Volume Dataset
- mt.gox交易数据
 - ◆ MtGox Leaked Transaction
- DApp数据
 - ◆ Activity Information of DApps

Unlabeled Dataset



Ether Price and Volume Dataset



Bitcoin Price and Volume Dataset



Mt.Gox Leaked Transaction



Activity Information of DApps

➤ Related Papers



Survey



Anomaly
Detection



Network Portrait



Smart
Contract



Entity
Recognition



Transaction
Pattern
Recognition



Market
Analysis



Transaction
Tracking



Other
Cryptocurrency



- | [Book] Zibin Zheng, Hongning Dai, **Jiajing Wu**, Blockchain Intelligence, Springer Singapore, 2021.
- | [JNCA] **Jiajing Wu**, Jieli Liu, Yijing Zhao and Zibin Zheng, "Analysis of cryptocurrency transactions from a network perspective: An overview", Journal of Network and Computer Applications
- | [TSMC] **Jiajing Wu**, Jieli Liu, Weili Chen, Huawei Huang, Zibin Zheng and Yan Zhang, "Detecting mixing services via mining Bitcoin transaction network with hybrid motifs", IEEE Transactions on Systems, Man, and Cybernetics: Systems
- | [TSMC] **Jiajing Wu***, Qi Yuan, Dan Lin, Wei You, Weili Chen, Chuan Chen and Zibin Zheng, "Who are the phishers? Phishing scam detection on ethereum via network embedding", IEEE Transactions on Systems, Man, and Cybernetics: Systems
- | [IEEE TCSS] Dan Lin, Jialan Chen, **Jiajing Wu**, Zibin Zheng Evolution of Ethereum Transaction Relationships: Toward Understanding Global Driving Factors From Microscopic Patterns, IEEE Transactions on Computational Social Systems
- | [IEEE TCSII] Dan Lin, **Jiajing Wu***, Qi Yuan, Zibin Zheng, "Modeling and understanding ethereum transaction records via a complex network approach", IEEE Transactions on Circuits and Systems II: Express Briefs
- | [IEEE TCSII] Dan Lin, **Jiajing Wu***, Qi Yuan, Zibin Zheng, "Modeling and understanding ethereum transaction records via a complex network approach", IEEE Transactions on Circuits and Systems II: Express Briefs

相关工作



- | [PHYSA] Jialan Chen, Dan Lin, **Jiajing Wu***, “Do cryptocurrency exchanges fake trading volumes? An empirical analysis of wash trading based on data mining” , Physica A
- | [CHAOS] Qing Han, **Jiajing Wu***, Zibin Zheng, “Long-range dependence, multi-fractality and volume-return causality of Ether market “ , Chaos: An Interdisciplinary Journal of Nonlinear Science
- | [ISCAS] Haixian Wen, Junyuan Fang, **Jiajing Wu*** and Zibin Zheng, “Transaction-based hidden strategies against general phishing detection framework on Ethereum” , in Proc. IEEE International Symposium on Circuits and Systems,
- | [BlockSys] Zhuoming Gu, Dan Lin, Jiatao Zheng, **Jiajing Wu*** and Chaoxin Hu, “Deep learning-based transaction prediction in Ethereum” , in Proc. International Conference on Blockchain and Trustworthy Systems
- | [BlockSys] Yijun Xia, Jieli Liu, Jiatao Zheng, **Jiajing Wu***, and Xiaokang Su, “Portraits of Typical Accounts in Ethereum Transaction Network” , in Proc. International Conference on Blockchain and Trustworthy Systems



谢谢!