black hat ASIA 2025

APRIL 3-4, 2025
BRIEFINGS

Sweeping the Blockchain: Unmasking Illicit Accounts in Web3 Scams

Speaker: Wenkai Li

Hainan University, China

Collaborators: Zhijie Liu (ShanghaiTech University),

Xiaoqi Li* (Hainan University)

*Corresponding author: csxqli@ieee.org

The Team

Li Wenkai

— PhD Student, Hainan University, China

— cswkli@hainanu.edu.cn

— https://cswkli.github.io/

Liu Zhijie

— Msc Student, ShanghaiTech University, China

— liuzhj2022@shanghaitech.edu.cn

— https://rroscha.github.io/

Li Xiaoqi

— Associate Professor, Hainan University, China

— csxqli@ieee.org

— https://csxqli.github.io/

About Us

Security Research

- 9 year-experience (since 2016) of Ethereum (Born in 2015) Blockchain Security.
- Blockchain/Software/System Security and Privacy, Ethereum/Smart Contract, Malware Detection, and etc.
- 40+ papers including ASE、INFOCOM、ICSE、WWW、AAAI、TSE, etc. within 5 years
- 30+ CVE/CNVD Vulnerabilities identified within 5 years
- 3700+ citations within 5 years
- Best Paper from INFOCOM、ISPEC、CCF, etc.
- SV Insight Annual Global Top-50 Blockchain Research Paper
- ESI Hot (Top 0.1%)、Highly Cited Paper (Top 1%)

Overview

Agenda

Introduction

Motivation

ScamSweeper

Experiments

Case Study

Introduction

The 3rd Generation Internet – Web 3.0

- Many ways for crypto users to engage with Web3.0:

NFT

DECENTRALAND

HORIZON WORLDS

META

- The most used Web3.0 Services:

DEX

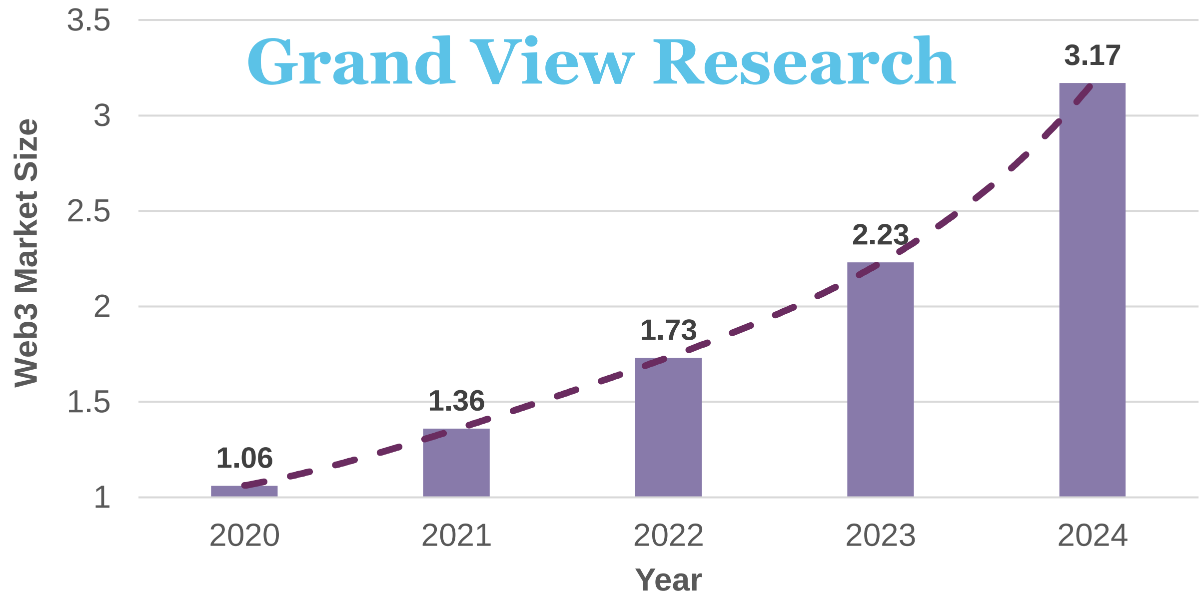
CEX

CryptoKitties
Axie INFINITY
THE SANDBOX
GODS UNCHAINED
Happy LAND
ALIEN WORLDS

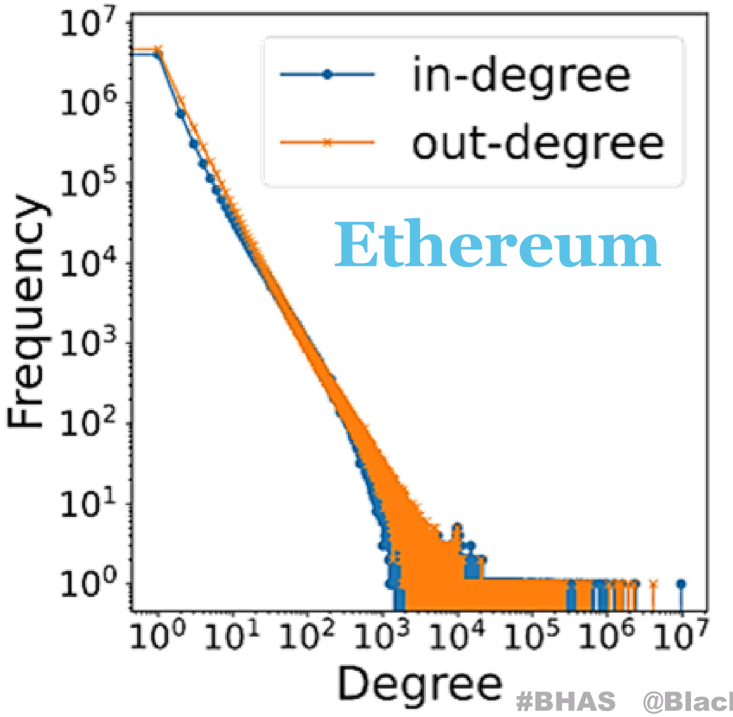
CRYPTO GAMING

The 3rd Generation Internet – Web 3.0

- What is the scale of Web3.0 tech market?
  - ➢ A growing trend.
  - ➢ The accelerating growth rate.
  - ➢ USD 3.17 billion in 2024.



- Web3.0 applications
  - ➢ DApp, DeFi protocol, DID, and etc. based on blockchain.
  - ➢ The blockchain node network follows a power-law distribution.
  - ➢ A minority of accounts appear at majority of Txs.



The Web3 environment comes with ***scam risks*** …

Motivation

Motivation: Web3 Scams

- The situation of Web3 scams:
  - ➢ Phishing, Rug Pulls, Harmful Airdrops, Giveaway Scams…
  - ➢ Crypto Drainer, Pig Butchering, Address Poisoning Scams…

NEWS 6 JAN 2025 www.infosecurity-magazine.com

Scammers Drain $500m from Crypto Wallets in a Year

- The scams on Web3 ecosystem can be catastrophic

Crypto Phishing Report 2024

Annual Report 2024 by ScamSniffer

$494M+ ↑67% TOTAL LOSS	332K+ ↑3.7% VICTIMS
$55.4M ↑130% LARGEST SINGLE LOSS	30 ↑58% LARGE LOSS CASES

ScamSniffer

NEWS 22 DEC 2023
Crypto Drainer Steals $59m Via Google and X Ads

NEWS 12 MAR 2024
Victims Lose $47m to Crypto Phishing Scams in February

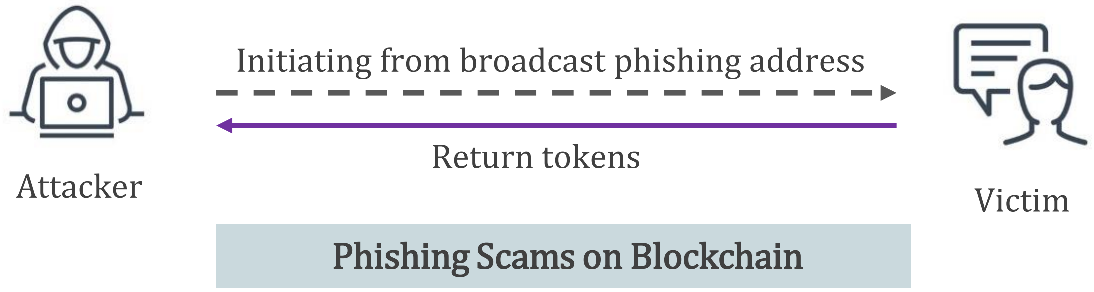
NEWS 16 JAN 2024
Inferno Drainer Spoofs Over 100 Crypto Brands to Steal $80m+

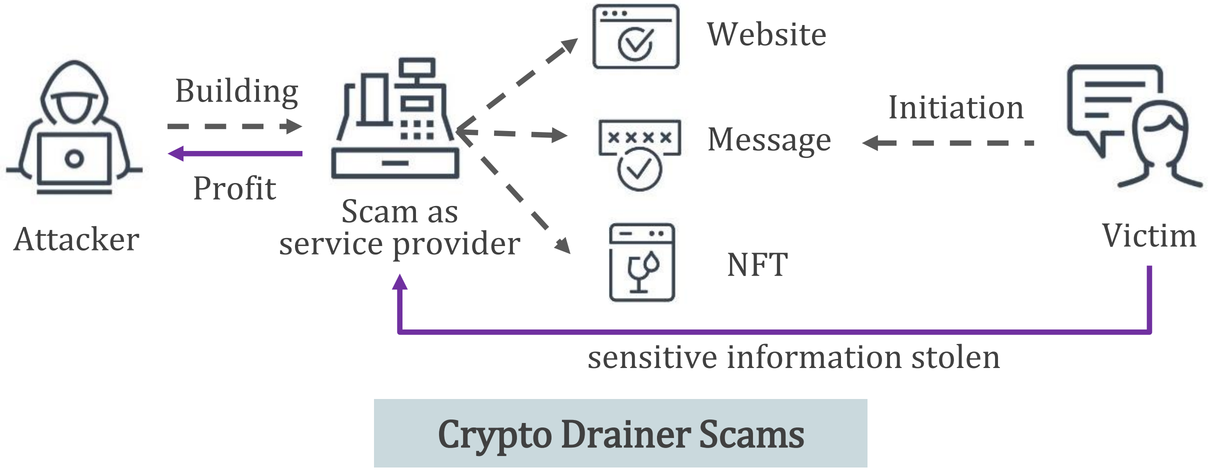
NEWS 8 JAN 2024
Security Firm Certik's Account Hijacked to Spread Crypto Drainer

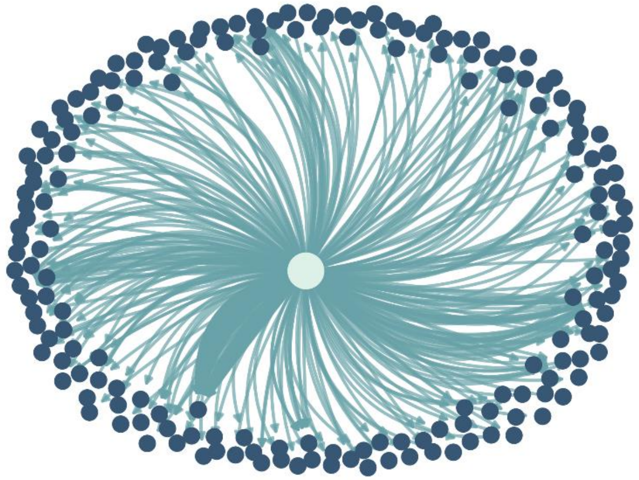
NEWS 3 JAN 2025
Web3 Attacks Result in $2.3Bn in Cryptocurrency Losses

Motivation: Web3 Scams

- What do the **Web3 Scams** on blockchain look like?
  - ➢ e.g., crypto drainers often masquerade as web3 projects, enticing victims into the drainer and getting the control access.

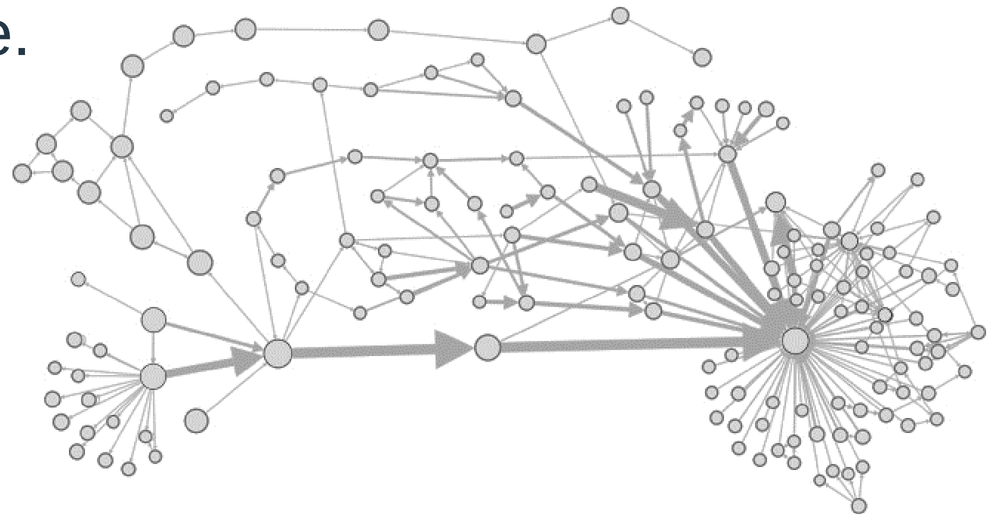


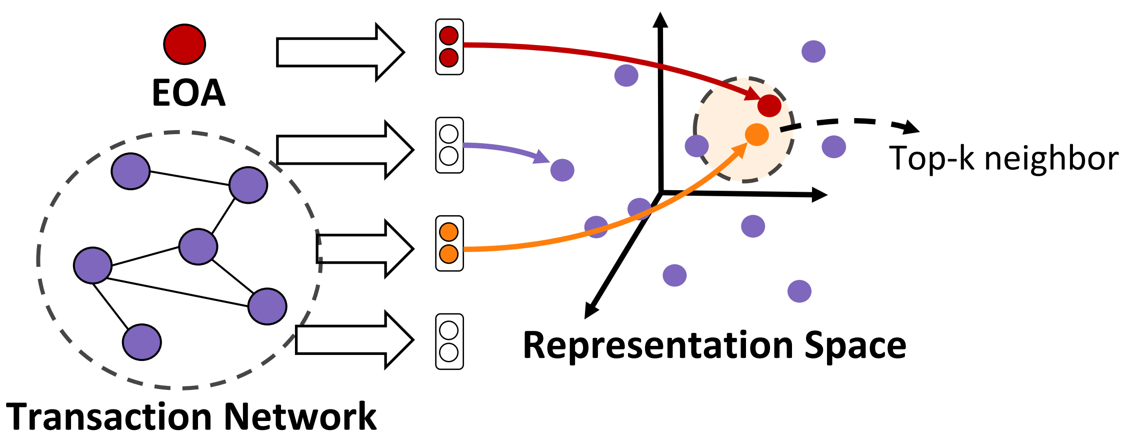




Motivation: previous research

- Graph Learning Methods
  - ➢ Intuitive to represent interactions of the topology structure.
  - ➢ Account as node, transaction as edge.
  - ➢ Top-k algorithm.
  - ➢ Power-law distribution leads lots of noise.





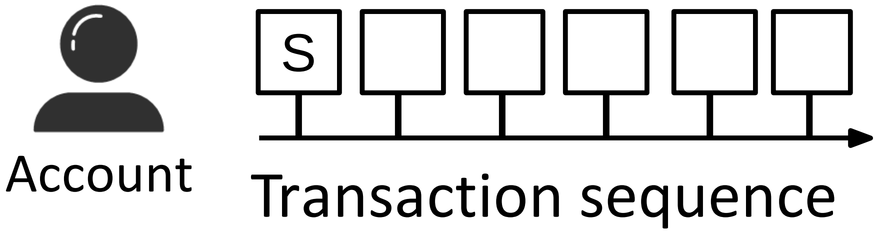
Top-k Algorithm

Random Walk

[1] Li, Shucheng and et al. "SIEGE: Self-Supervised Incremental Deep Graph Learning for Ethereum Phishing Scam Detection." in *Proc. of MM*. 2023.
[2] Wu, Zhiying and et al. "TRacer: Scalable graph-based transaction tracing for account-based blockchain trading systems. " *TIFS*. 2023.
[3] Li, Sijia and et al. "TTAGN: Temporal transaction aggregation graph network for Ethereum phishing scams detection." in *Proc. of WWW*. 2022.

Motivation: previous research

- Sequence Learning Methods
  - ➢ Transductive to learn the logic of account behavior feature.
  - ➢ Analyzing an account is related to its length.
  - ➢ Large-scale transactions, e.g., **2.7 billion txs** on Ethereum.



Tab.1 – The statistical information of some accounts on Ethereum.

No.	Account Address	Tx Cnt
1	0xC02aaA39b223FE8D0A0e5C4F27eAD9083C756Cc2	16,514,200
2	0x28C6c06298d514Db089934071355E5743bf21d60	18,921,592
3	0x267be1C1D684F78cb4F6a176C4911b741E4Ffdc0	3,832,284
4	0x32400084C286CF3E17e7B677ea9583e60a000324	3,094,481
5	0xf7858Da8a6617f7C6d0fF2bcAFDb6D2eeDF64840	1,588,678
6	0xA7EFAe728D2936e78BDA97dc267687568dD593f3	3,482,451
7	0xBf94F0AC752C739F623C463b5210a7fb2cbb420B	1,611,882
8	0xae0Ee0A63A2cE6BaeEFFE56e7714FB4EFE48D419	1,798,762
9	0x0D0707963952f2fBA59dD06f2b425ace40b492Fe	7,527,833
10	0x6262998Ced04146fA42253a5C0AF90CA02dfd2A3	1,183,120

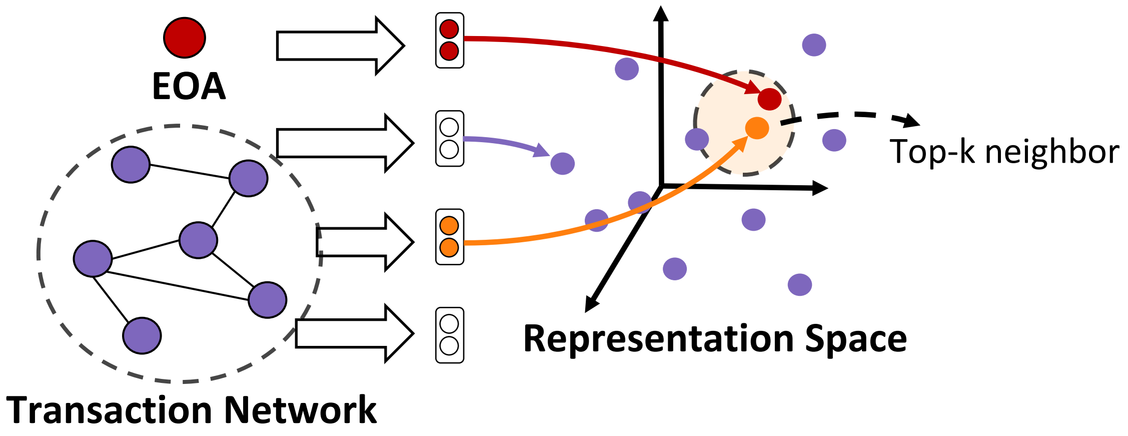
[4] Hu, Sihao and et al. "BERT4ETH: A Pre-trained Transformer for Ethereum Fraud Detection." in *Proc. of WWW*. 2023.

Motivation: previous research

- Graph Learning Methods
  - ➢ Not suitable to capture dynamic information.

    Merging multiple edges into one for graph computation

    e.g., graph convolution or random walk

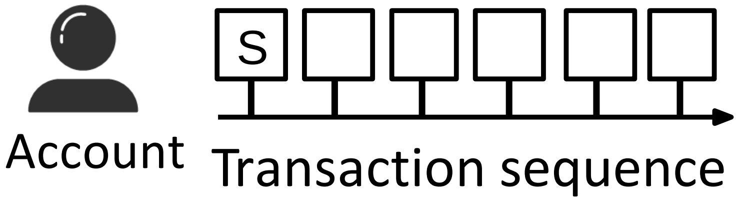


  - ➢ Not suitable for power law distribution.

    Introducing noise when multi-hop convolution,

    In GRU, Model capability is limited (# of GNN layers = # of hop)

- Sequence Learning Methods



  - ➢ Not suitable to large-scale transactions.

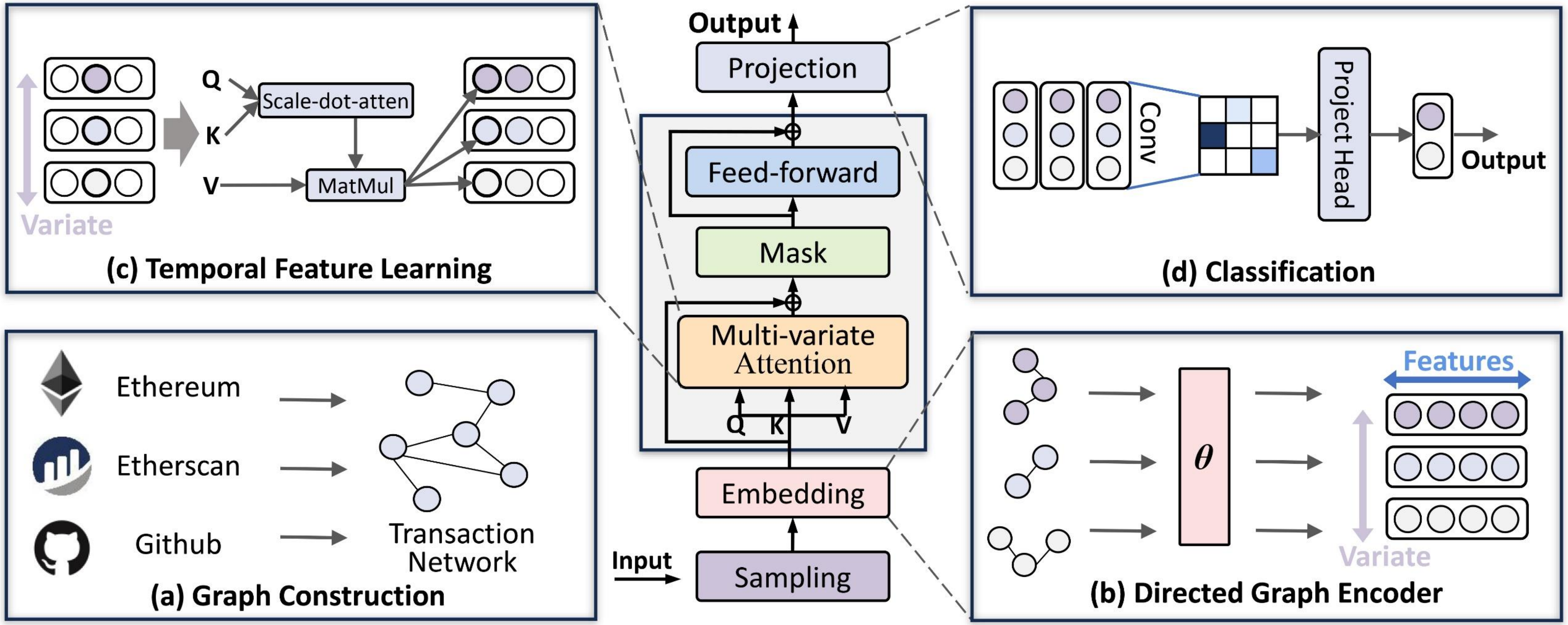
    Analyzing an account is related to the length of its transaction sequence.

[4] Hu, Sihao and et al. "BERT4ETH: A Pre-trained Transformer for Ethereum Fraud Detection." in *Proc. of WWW*. 2023.

ScamSweeper

ScamSweeper

- Learning the dynamic evolution of transaction graph, and applying to account detection
  - ➢ Sequence learning from the graph structure.



ScamSweeper (1)

- (a) Graph Construction
  - ➢ Most previous works used the **random walk** to sample the transaction network.
  - ➢ Random walk is like a ***dice game!***

Motivation:

To lower the computing consumption, and learn features from temporal sequence and topology structure.

We designed a new walk-sampling method:

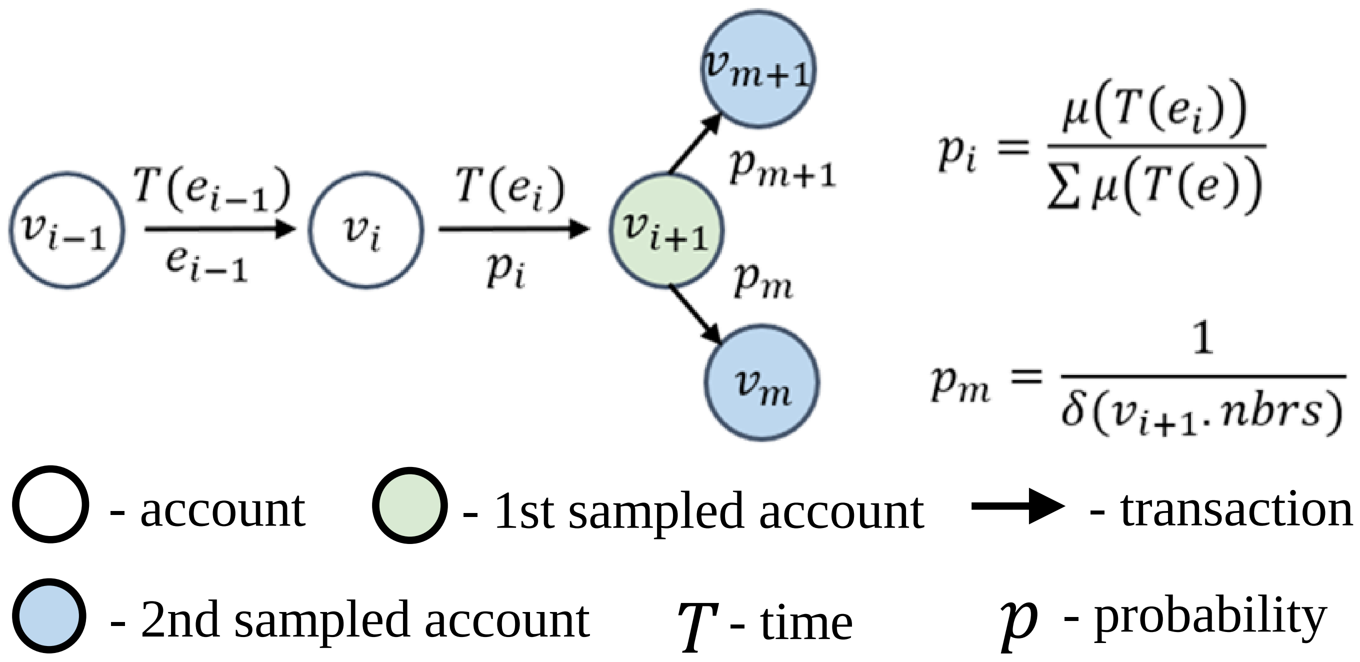
Struct-Temporal Random walk (STRWalk)

ScamSweeper (1)

- (a) Graph Construction
  - current node is v_i, next node is v_{i+1},
  - the edge is e_i
  - $\mu(T(e_i))=T(e_i)-minTime$,
  - $\delta(v)$ represents the number of nodes that are in the same interval with v

With P_i and p_m, Struct-Temporal Random walk (STRWalk)

With P_i, Temporal Random Walk (TRWalk)

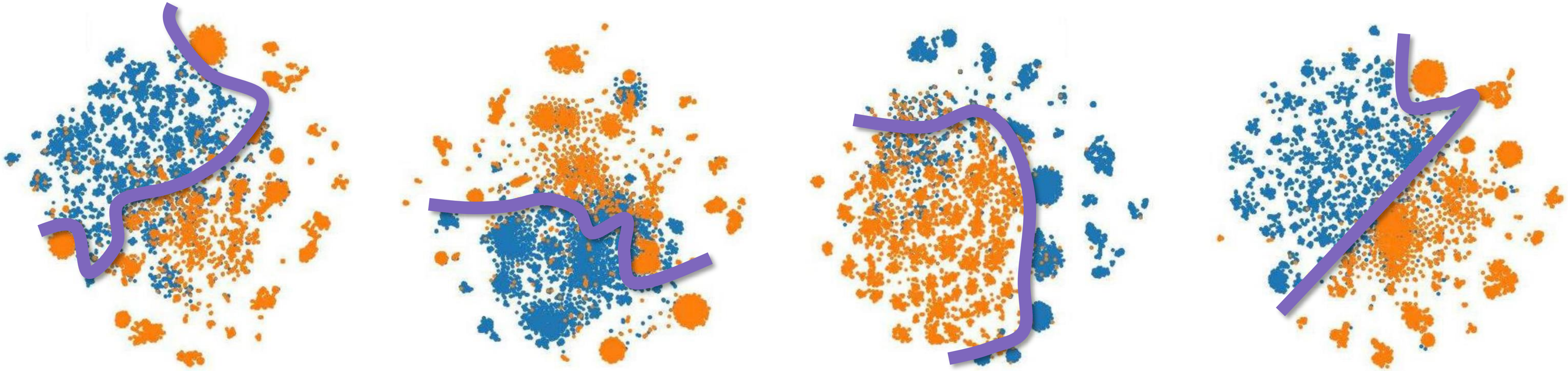


The 1st sampled node selected by the alias sample algorithm with the **probability p_i**.

The 2nd sampled node selected by the alias sample algorithm with the **probability p_m**.

ScamSweeper (1)

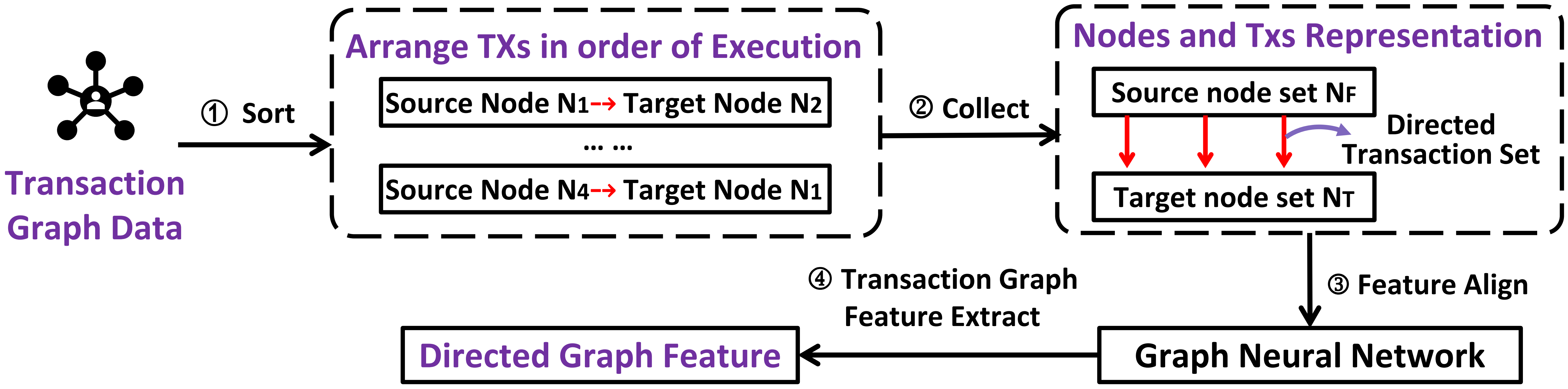
- (a) Graph Construction
  - ➢ Walk length: 20, the window size: 4, and the embedding dimension: 128
  - ➢ Phishing dataset, 1165 **malicious** nodes and 636 **normal** nodes.
  - ➢ T-SNE Visualization



Random Walk Deep Walk TRWalk STRWalk

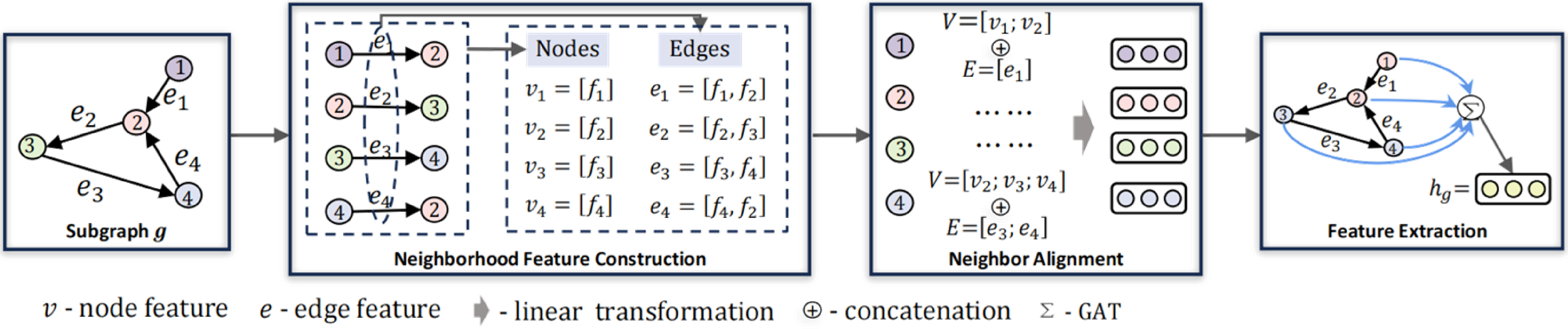
ScamSweeper (2)

- (b) Directed Graph Encoder
  - Split the whole graph according to the interval, generating several sub-graphs
  - Learning the feature of each subgraph in time sequence



ScamSweeper (2)

- (b) Directed Graph Encoder



$$V = \{X_f;\ X_t \mid (X_f^1; X_t^1, X_f^2; X_t^2, \ldots, X_f^n; X_t^n)\}$$
$$E = \{X_f \rightarrow X_t \mid (e_1, e_2, \ldots, e_n)\}$$

Θ - linear transformation layer

h - hidden feature of nodes

$$\hat{v} = LeakyRelu(\Theta_v \cdot [v \mid e]) \quad (1)$$
$$e_{ij} = LeakyRelu(\Theta_n \cdot [h_i \mid h_j]) \quad (2)$$
$$\alpha_{ij} = \frac{exp(e_{ij})}{\sum_{x \in N(i)} exp(e_{ix})} \quad (3)$$
$$h_g = Elu(\alpha_{ij} \cdot \Theta \cdot h_i + \sum_{x \in N(i)} \alpha_{ix} \cdot \Theta \cdot h_x) \quad (4)$$

ScamSweeper (3)

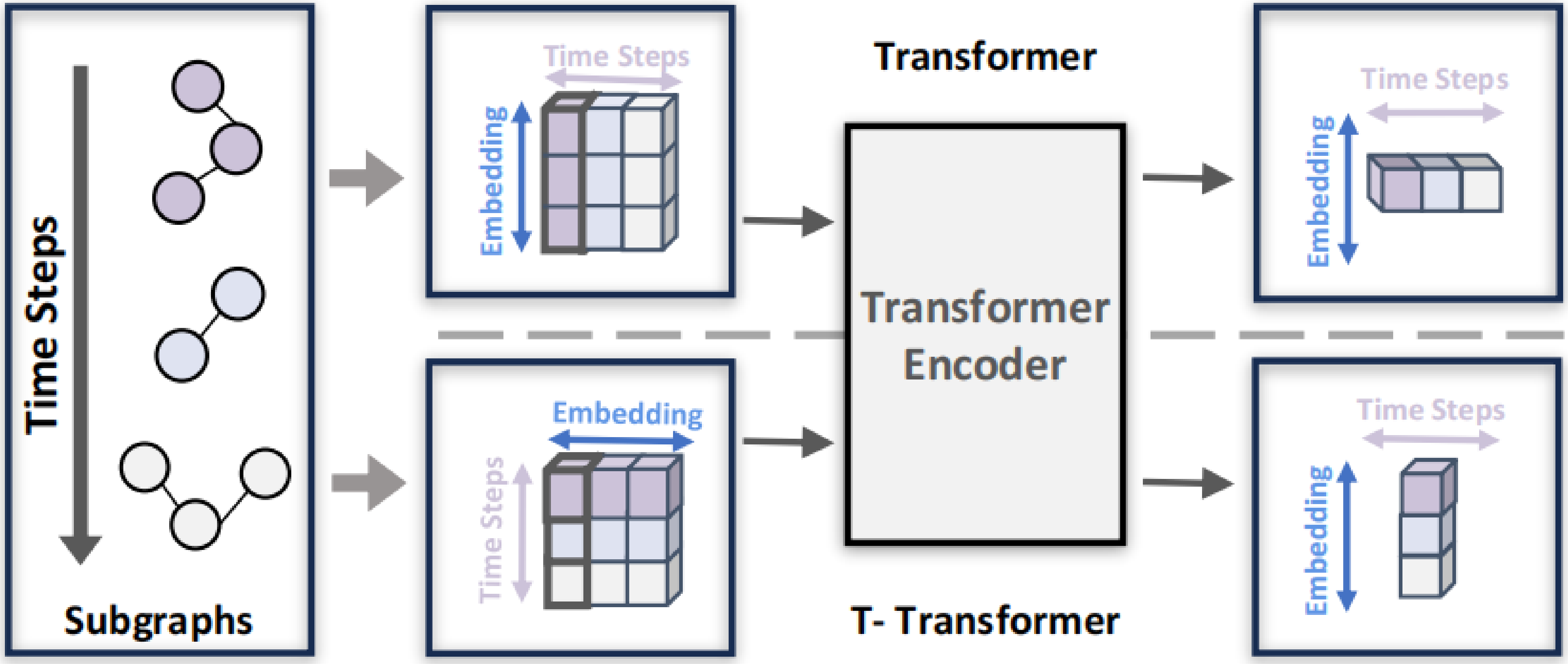
- (c) Temporal Feature Learning
  - ➢ Leveraging the ability of Transformer

$$H^{(l+1)} = Attention({H^{(l)}}^T \Theta_Q, {H^{(l)}}^T \Theta_K, {H^{(l)}}^T \Theta_V) \quad (5)$$

$$h = Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}} V) \quad (6)$$

$$H^{(l+1)} = FFN(h) \quad (7)$$

$$FFN(x) = Sigmiod\left(xW_1^{(l)} + b_1^{(l)}\right) W_2^{(l)} + b_2^{(l)} \quad (8)$$



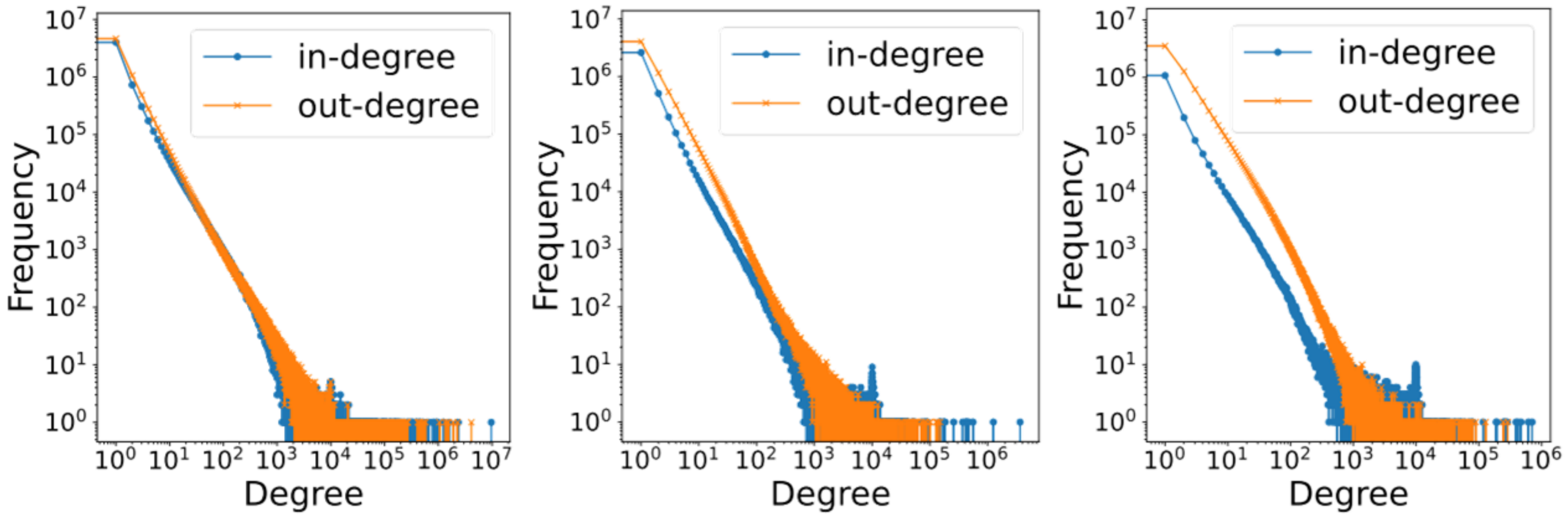
Experiments

Experiments: large-scale data

- Data & distribution
    - ➢ Crawling the first 18 million block height on Ethereum
    - ➢ Phishing labels from Etherscan
    - ➢ Web3 scams from [5]
    - ➢ Normal nodes contains 4 types: exchange, mining, ICO wallet, and gambling.

Tab.2 – The statistical information of dataset.

Datasets	#Nodes	#Labeled	#Edges	#Std Degree
Normal	12,042,066	636	142,750,370	3555.35
Phishing	10,159,847	4,905	62,011,219	1285.25
Web3 Scams	8,736,430	3,125	64,265,586	541.10

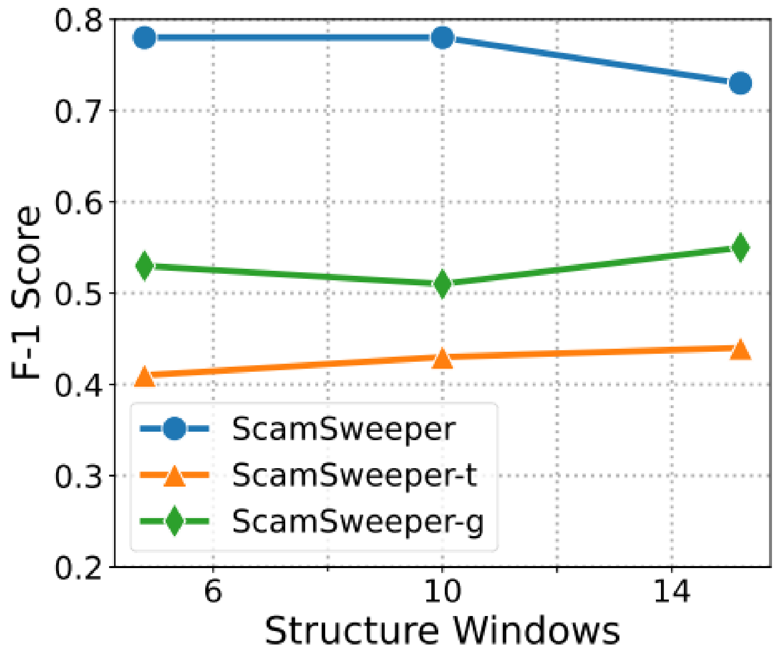


(a) Normal (b) Phishing (c) Web3 Scams

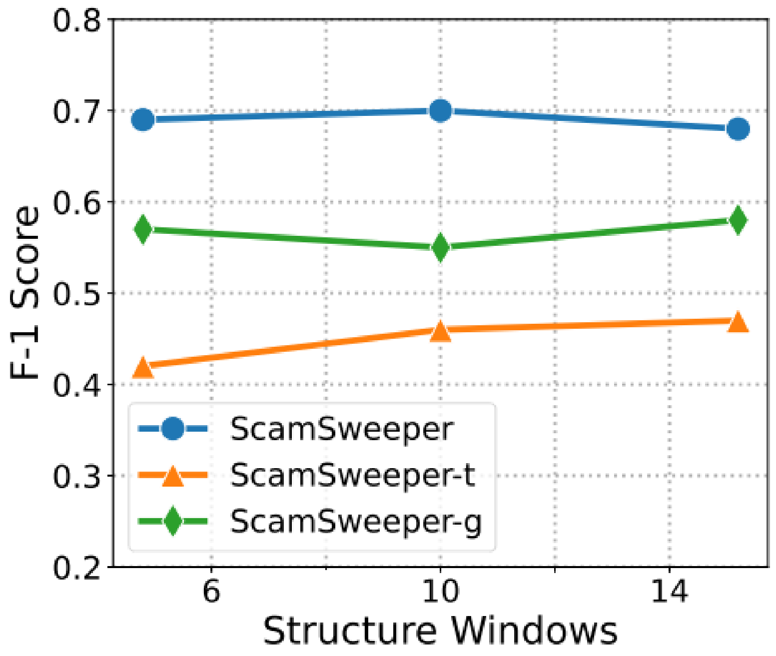
[5] https://github.com/scamsniffer/scam-database..“, 2024.

Experiments: Ablation

- How well do the components work?
  - ➢ the importance of graph encoder and T-Transformer



(a) F1-scores



(b) Weighted F1-scores

ScamSweeper with all components,

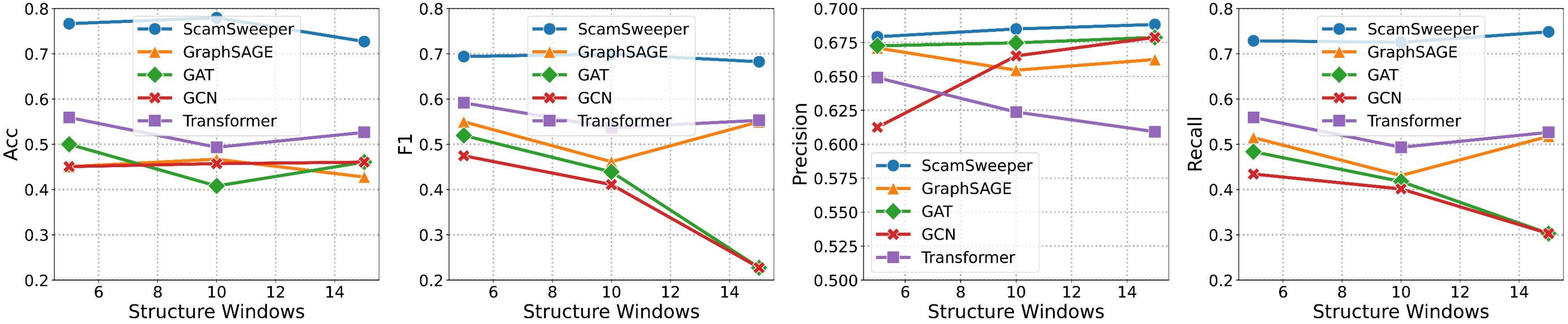
ScamSweeper-t without the T-Transformer,

ScamSweeper-g without the graph encoder.

ScamSweeper > Graph encoder > T-Transformer

Experiments: Comparison

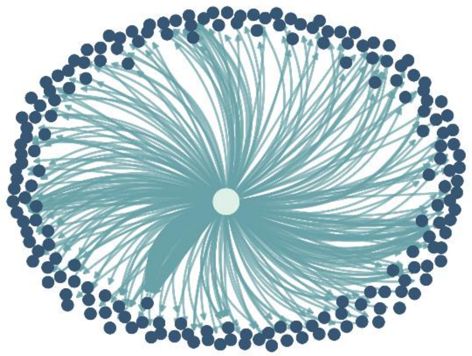
- How well do the ScamSweeper work?
  - ➢ Compared with Graph methods and Transformer
  - ➢ Structure window: {5,10,15}, Adam weight decay rate: $5e-4$.
  - ➢ Training: 70%, Validation: 20%, Test:10%



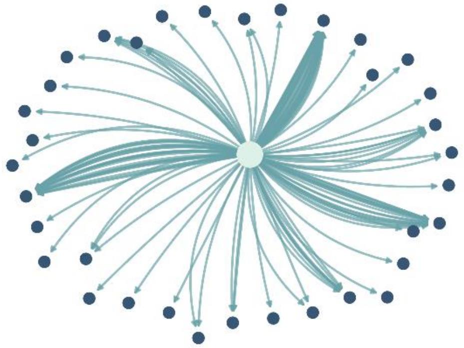
Case Study

Case Study: Web3 Scam

- Dynamic Evolution
  - τ is a time interval.

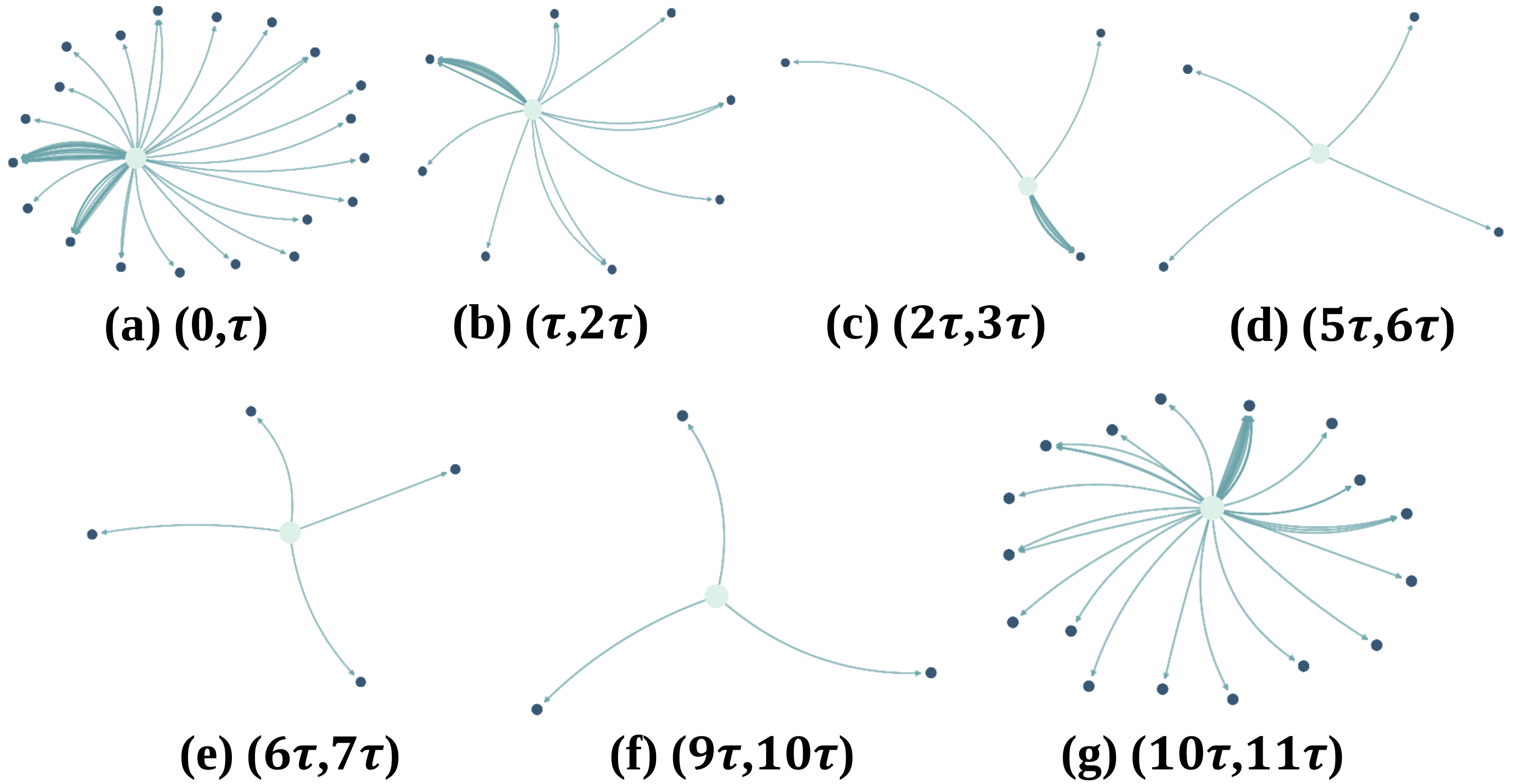


Transaction Graph



Transaction Pattern

Malicious Obfuscation



Takeaways

- Summary & key takeaways

  - **Web3 Scams Proliferation:** Web3 applications are increasingly targeted by scammers who mimic legitimate transactions to deceive users, highlighting a critical gap in current detection methods.

  - **Research Gap:** Prior studies focus on de-anonymization and phishing nodes, neglecting the unique temporal and structural patterns of web3 scams, while existing detection tools struggle with power-law distributed transaction networks.

  - **ScamSweeper Framework:** A novel approach that combines structure-temporal random walks for efficient transaction network sampling and variational transformers for dynamic pattern analysis, capturing both temporal and structural evolution of scams.

  - **Practical Insights:** Large-scale dataset collection, cost-effective data sampling, and dynamic evolution analysis, enabling real-world application in Ethereum transaction monitoring.