Detecting Illicit Behavior

In Crypto Transaction Networks

Howard Wang
Yazhe Huang

01

Introduction

02

Dataset Description

03

Graph Construction

04

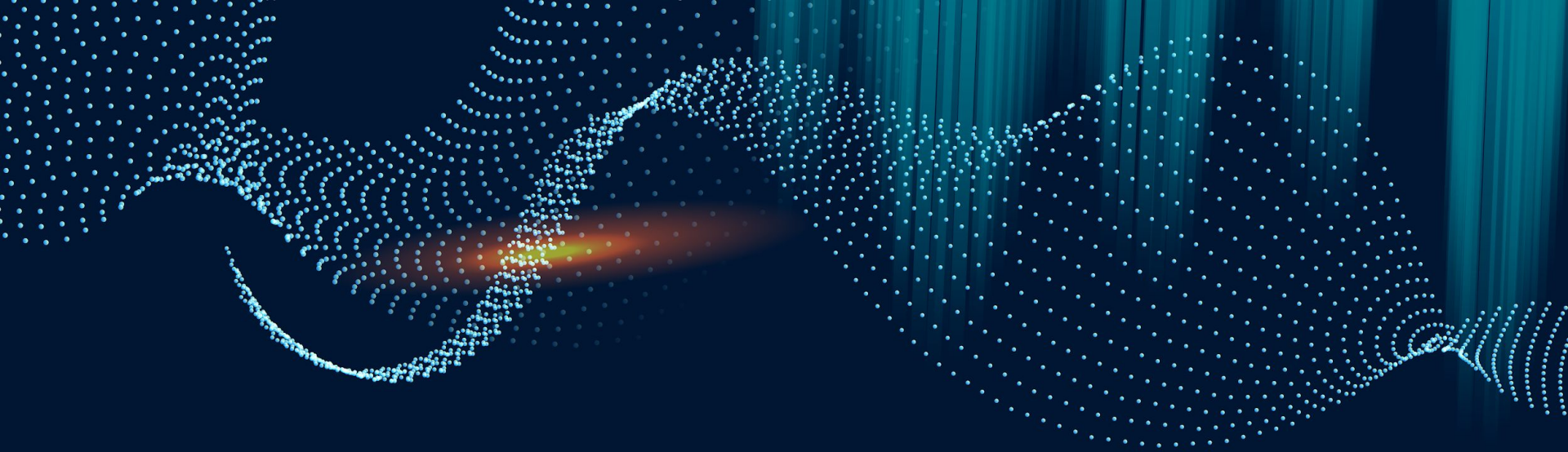
Behavior-based Detection

05

The ML Phase

06

Conclusion



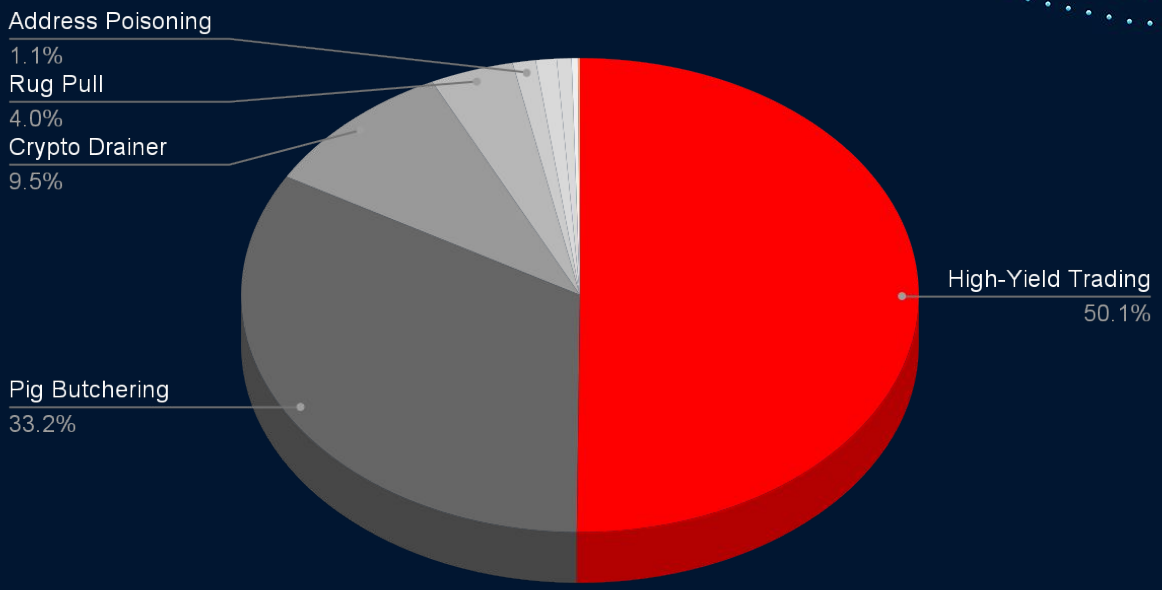
01 Introduction

Research Motivation &
Project Objective

$3.01B

Total value of assets stolen in crypto-related scams in 2024, according to Chainalysis.

Crypto Fraud in 2024 By Type

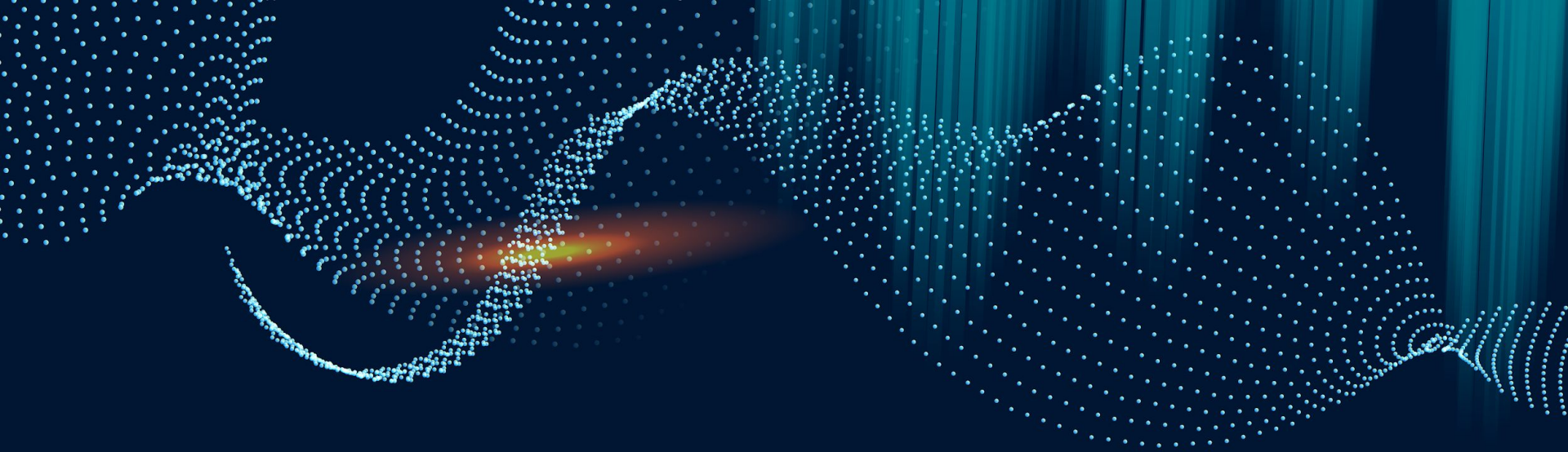


Project Objective

Applying:

- ***Graph-based methods*** (Degree Centrality, PageRank)
- ***Community Detection Algorithms*** (Louvain)
- ***GCNs*** for fraud nodes detection

To detect high-risk or scam-related wallet addresses in Ethereum transaction network.



02 Dataset Description

Transaction Data &
Behavioral and Structural Features

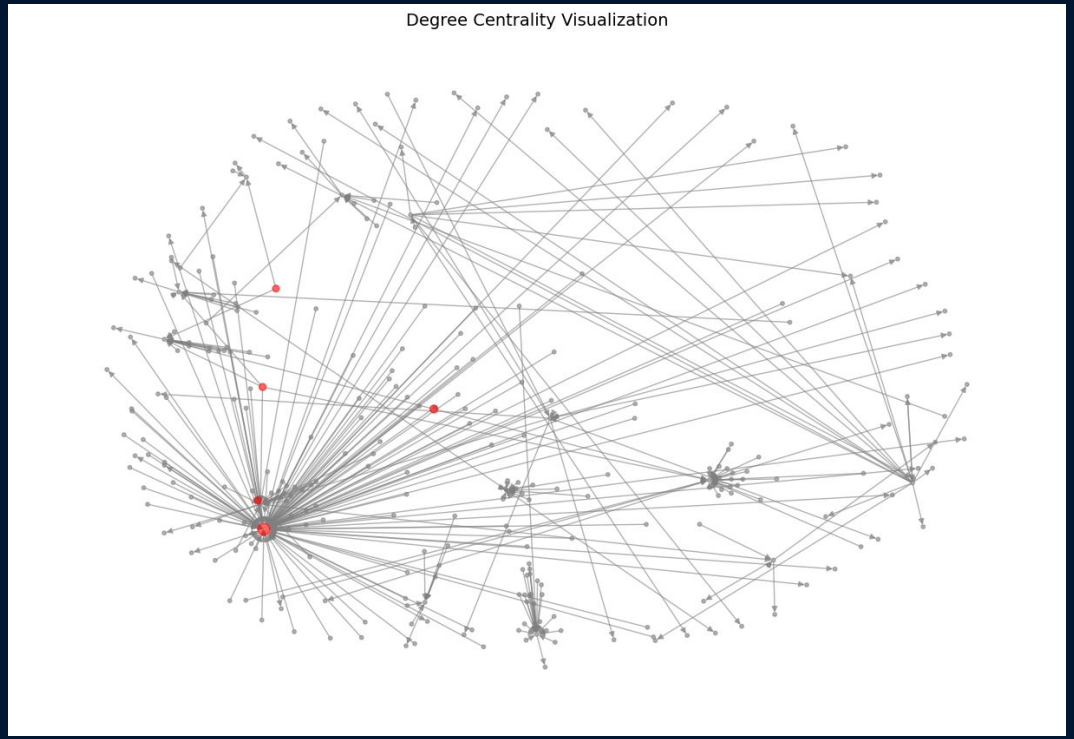
Data Source & Network Construction

Source: Ethereum Blockchain

- From/To: Wallet Addresses (Nodes)
- Value: Transaction Amount (Edge Weight)
- Timestamp/isError: Metadata

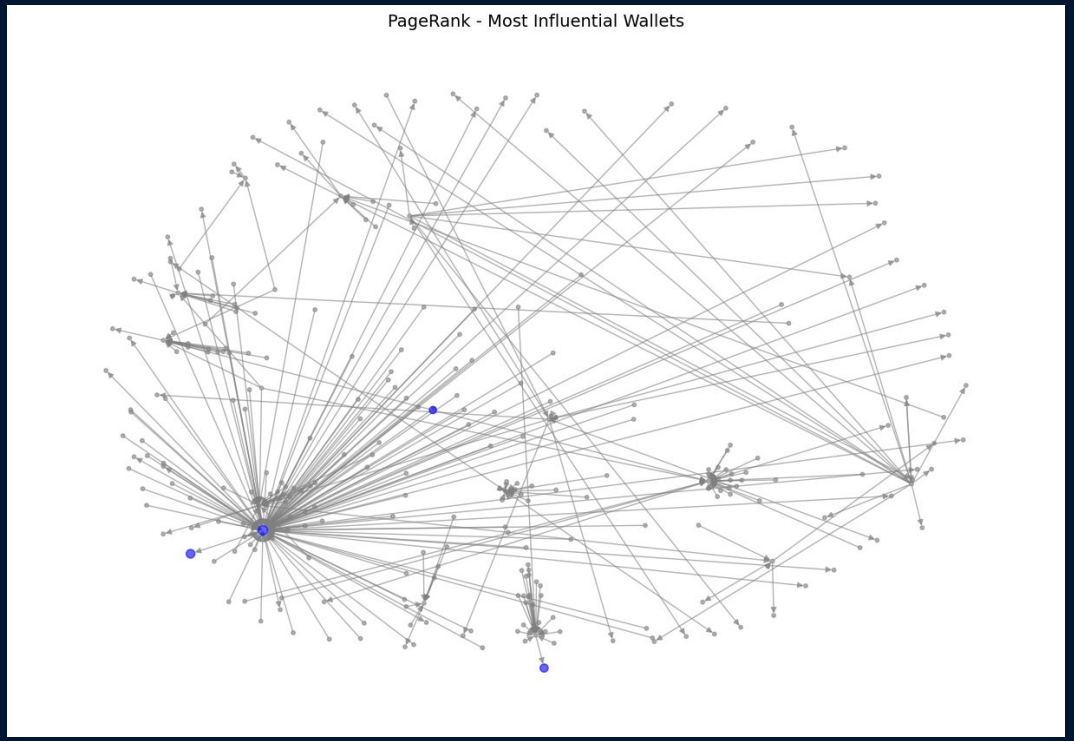
Graph

- Nodes: 5,689 Unique Wallets
- Edges: 6,553 Transactions
- Subset: Sampled from larger Ethereum transaction dataset



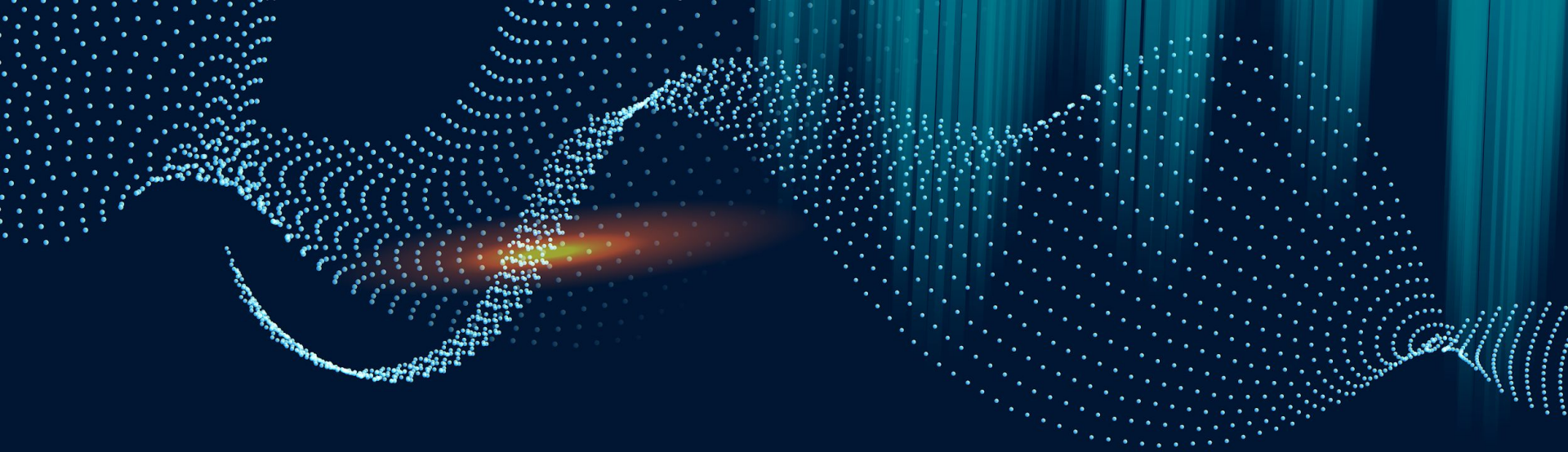
Top 50

Of the active wallets within a subgraph to highlight their central roles in the network.



Top 50

PageRank wallets to highlight the most influential addresses in the transaction network.



03

Graph Construction

Graph Creation & Centrality Metrics
& Community Detection

Why Louvain?

SCALE

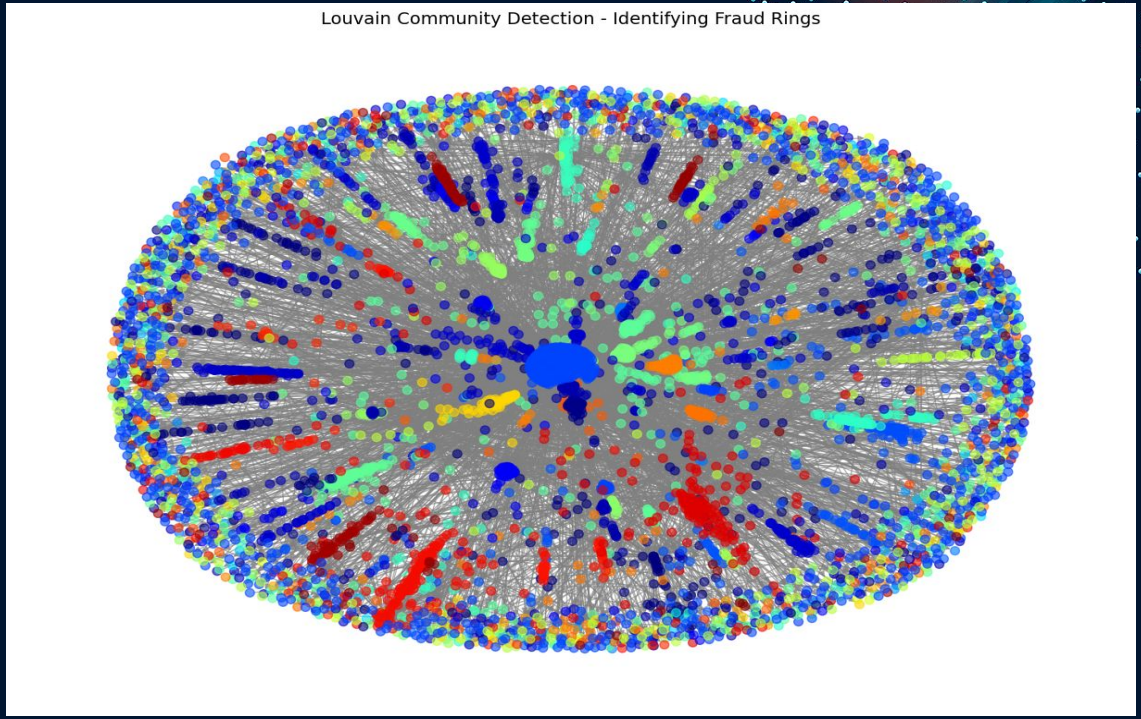
Scales well to large transaction graphs.

MODULARITY

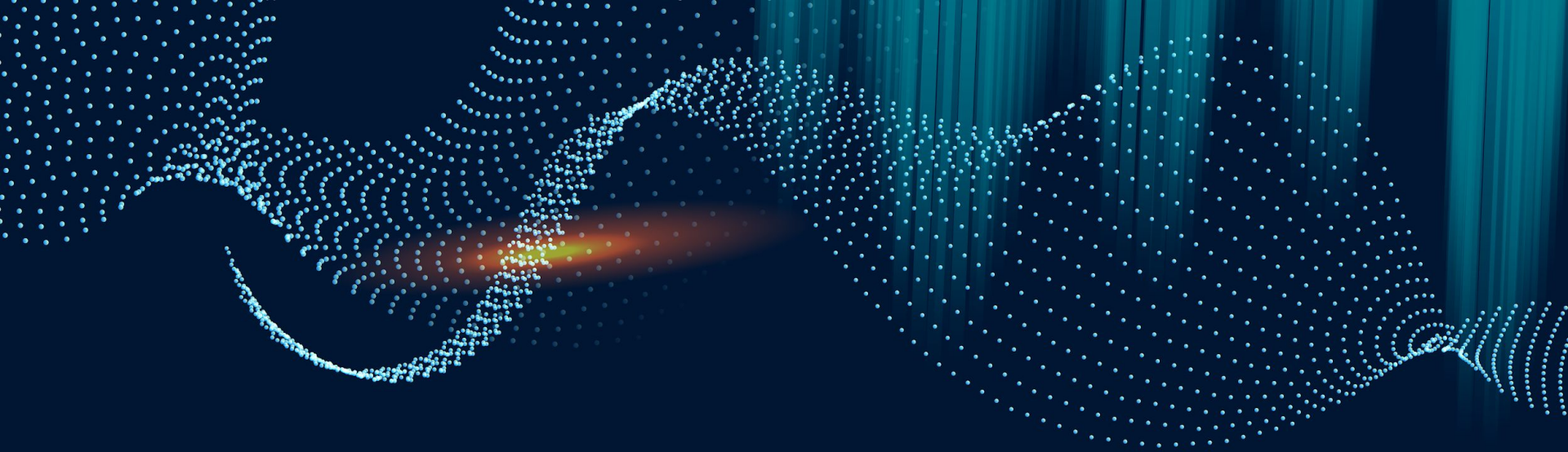
Optimizes modularity to detect dense fraud communities.

INSIGHTS

Well-suited for uncovering tightly-knit scam rings.



- **Colors** represent different detected communities
- Nodes in the **same color** cluster = tightly connected wallets
- **Center Nodes** = highly connected
- **Top 5 largest communities** flagged as possible fraud rings

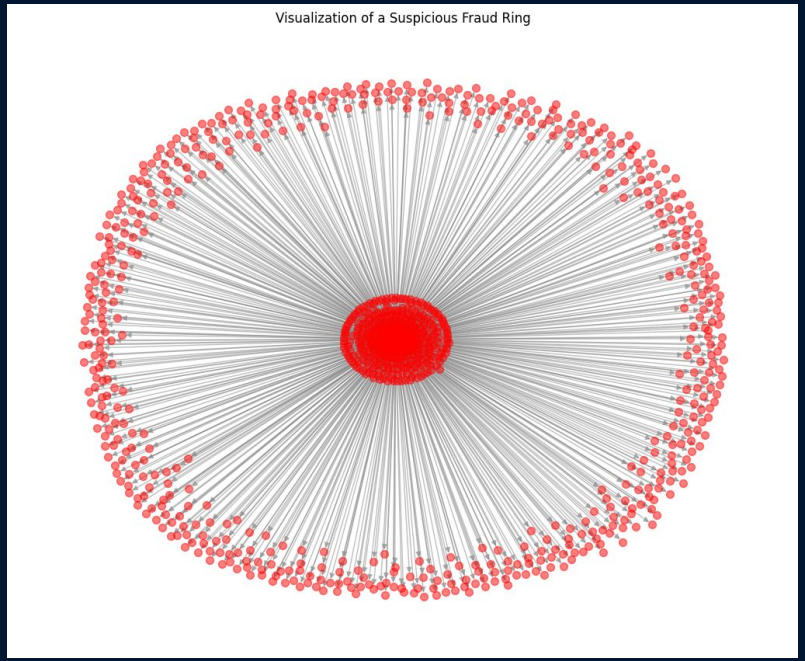


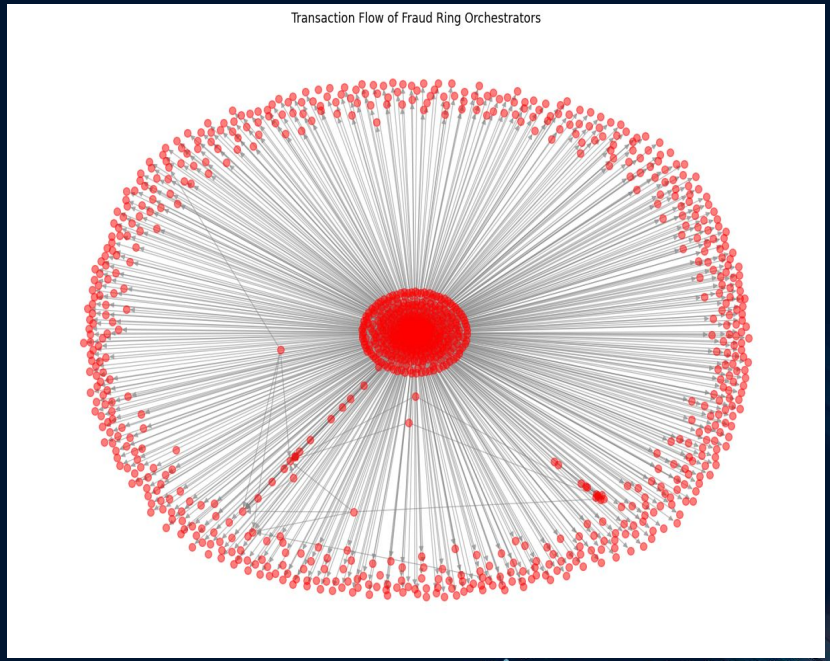
04 Behavior-based Detection

Heuristic Rules & Mixer-like Patterns and Cyclical Flows

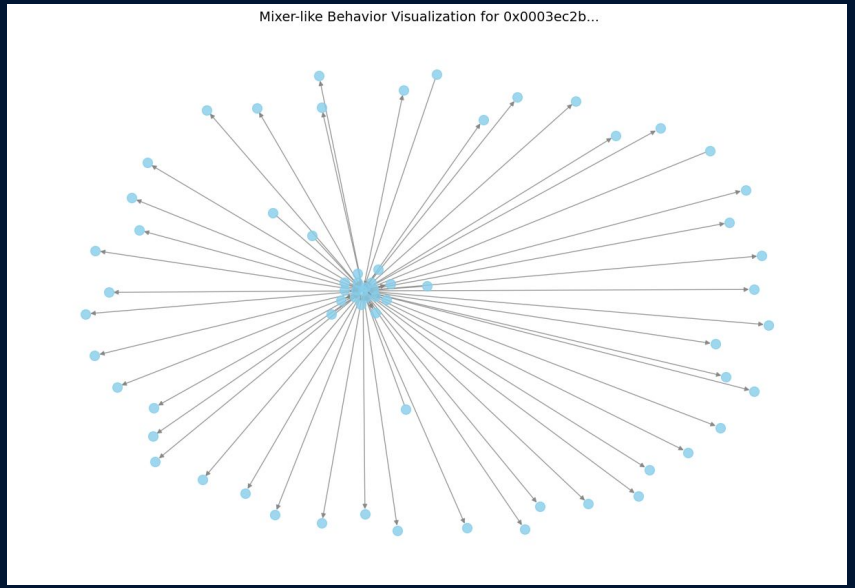
Rule-Based Detection Overview

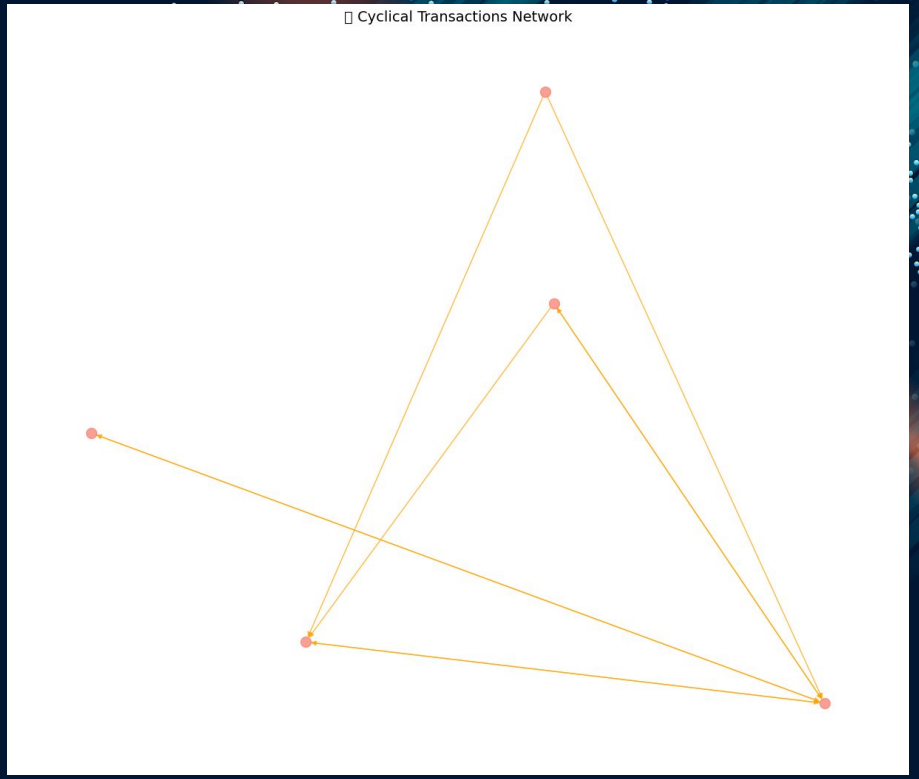
- Outgoing-Only Wallets
- Mixer-Like Patterns
- Cyclical Transactions (e.g., A → B → C → A)
- Fraud Ring Heuristics
- Used as handcrafted features for machine learning and GCN detection



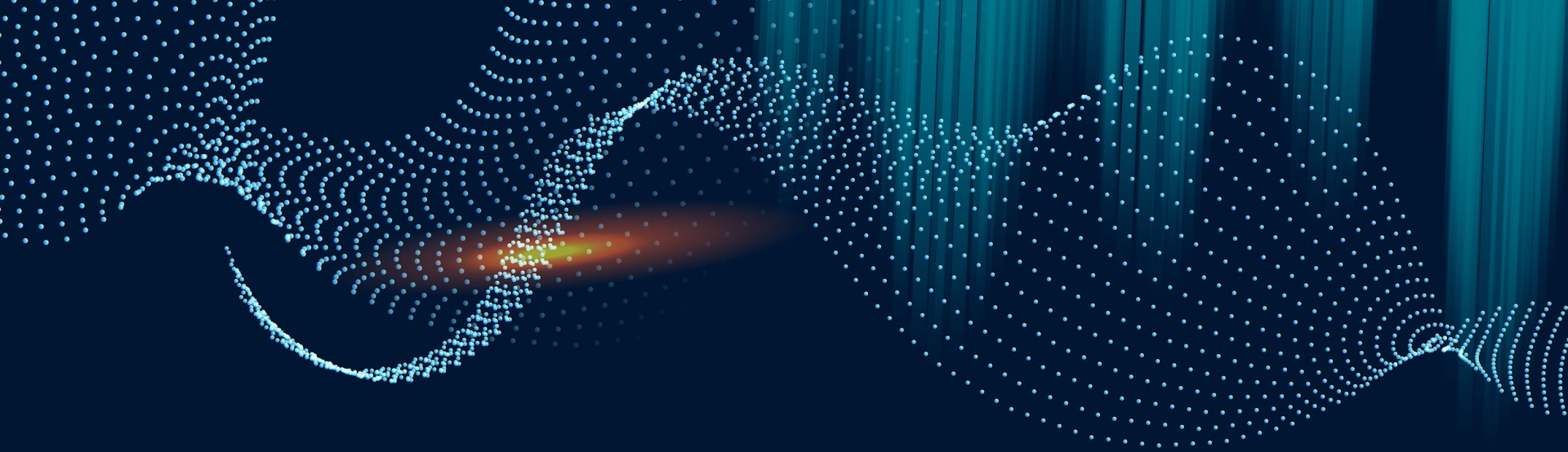


- Flagged largest fraud communities after Louvain Clustering
- Extracted suspicious addresses and visualized hub-and-spoke structures
- Identified top orchestrators to trace outward fund flows





- Detected mixer-like wallets based on transaction dispersion patterns
- Identified **346** cyclical transactions suggestive of obfuscation
- Constructed subgraphs for directed flow tracing



05 The Machine Learning Phase

GCN Model Design & Training and Evaluation & Visualizations

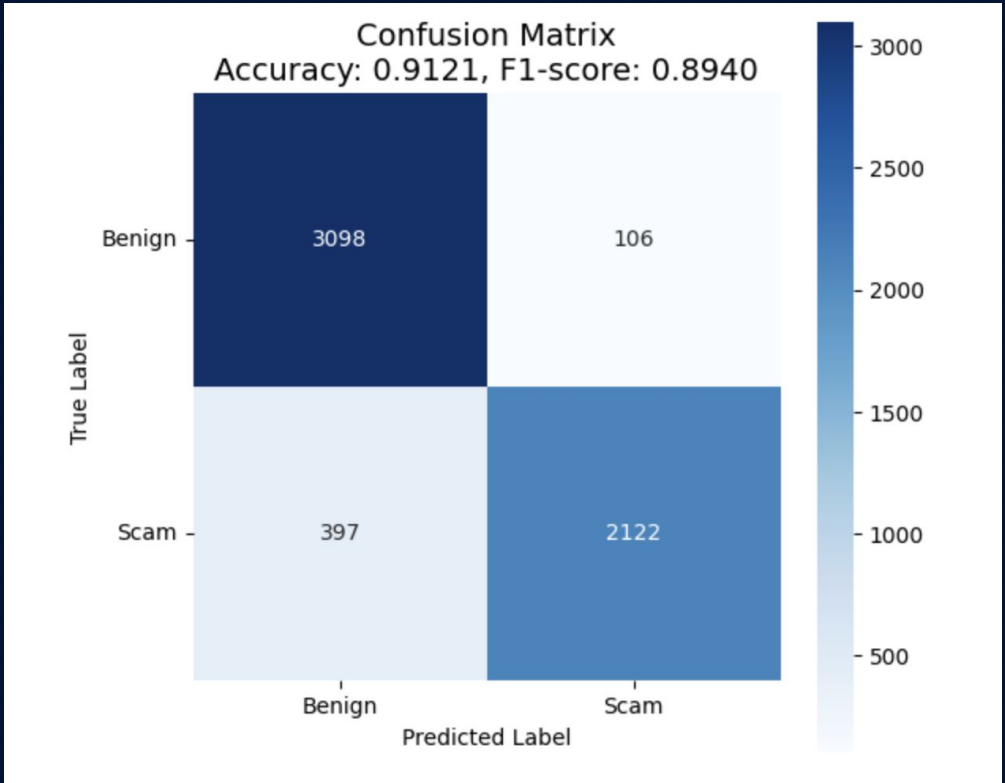
Dataset & Features

- **~20,000** labeled wallets, **9.8M** transaction edges
- Features Include:
  - Transaction Volume
  - PageRank, behavioral flags
- Graph-based learning → Considers both **individual** and **neighbor** behavior

GCN Model Architecture

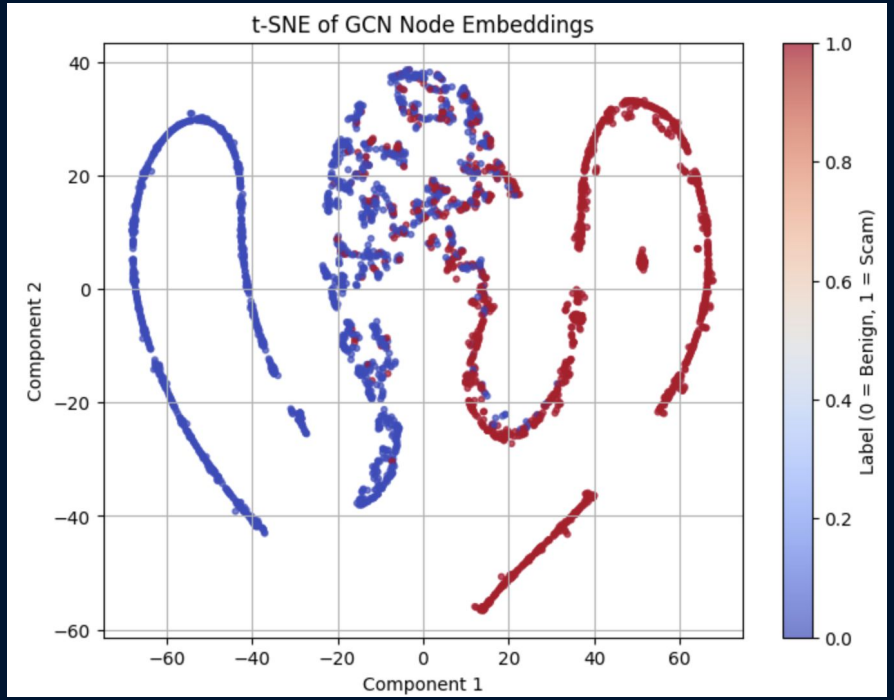
- 2-layer GCN using GCNConv (PyTorch Geometric)
- Input → GCN (ReLU) → GCN → Output (binary classification)
- Trained for 100 epochs on GPU (CUDA), 80/20 split
- Loss Function: Cross-entropy | Optimizer: Adam (lr=0.01)

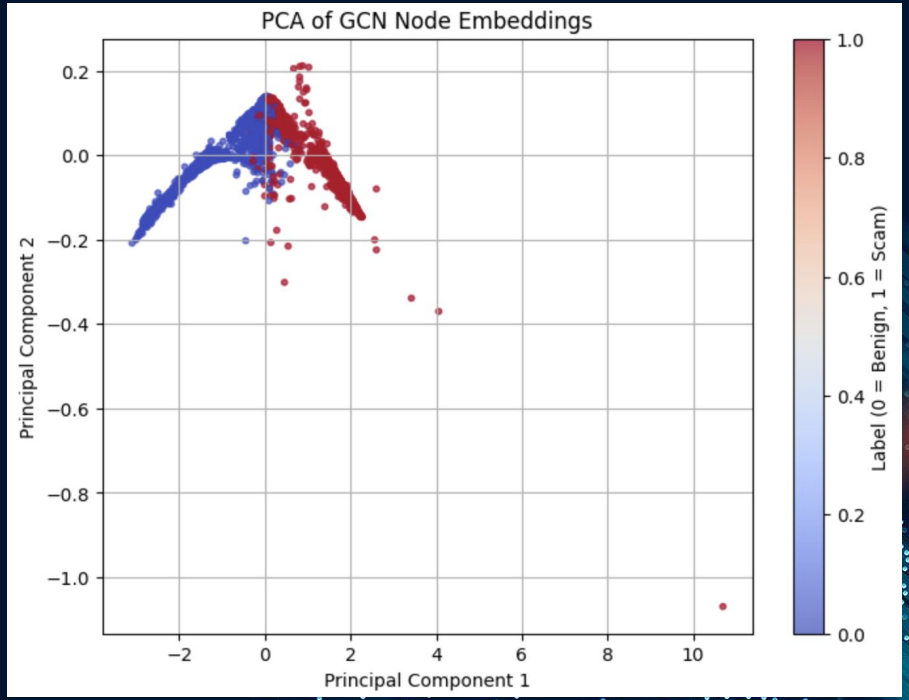
Performance Results

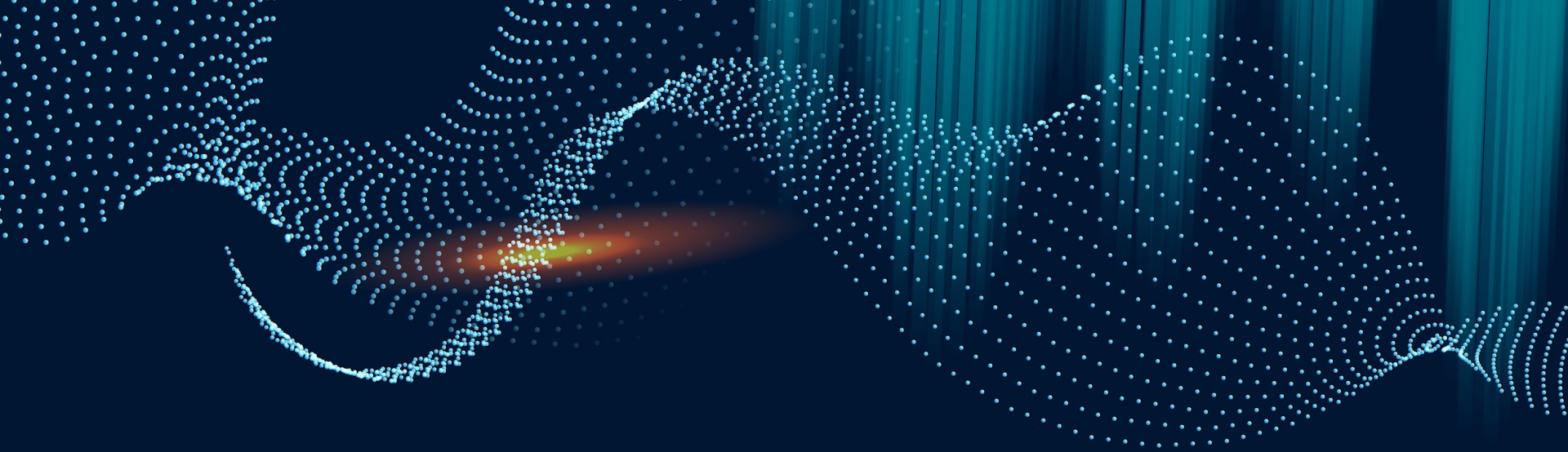


- Validation Accuracy: 91.21%
- F1 Score: 0.894
- Confusion Matrix:
  - TP: 2,122/TN: 3,098
  - FP: 106/FN: 397
- Outperforms baseline rule-based detection

Our graph-based learning approach not only detects scams accurately, but also reveals hidden patterns in the network.


t-SNE of GCN Node Embeddings
Component 2
Component 1
Label (0 = Benign, 1 = Scam)
40
20
0
−20
−40
−60
−60
−40
−20
0
20
40
60
1.0
0.8
0.6
0.4
0.2
0.0


PCA of GCN Node Embeddings
Principal Component 2
Principal Component 1
Label (0 = Benign, 1 = Scam)
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
−2
0
2
4
6
8
10
1.0
0.8
0.6
0.4
0.2
0.0



06 Conclusion

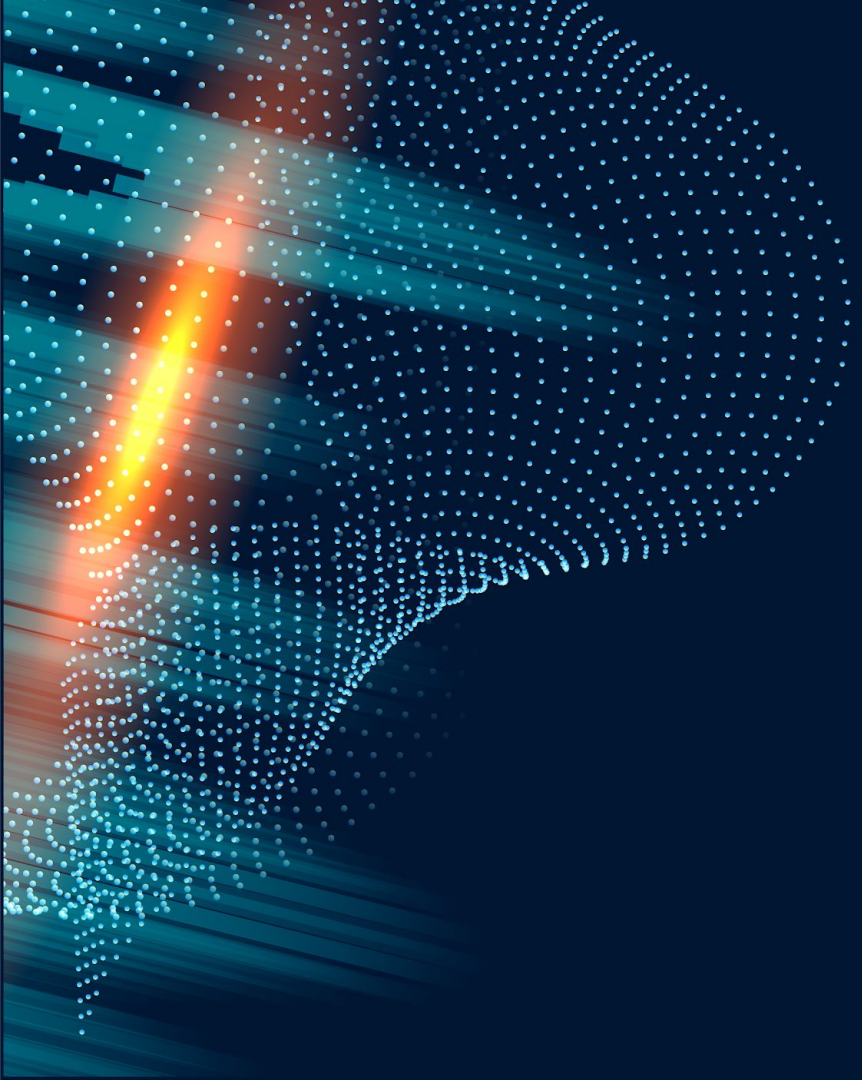
Summary & The Future

Our Vision


1
Apply to other blockchains
2
Incorporate time-series data
3
Real-time wallet risk scoring system
Next...?

“A graph is worth a thousand words.”

—Frank Harary



Thanks!

Howard Wang (sw1235)
Yazhe Huang (hy550)