

Analysis of User Behaviors in Blockchain

A Network Science Perspective

Natkamon Tovanich

TU Wien

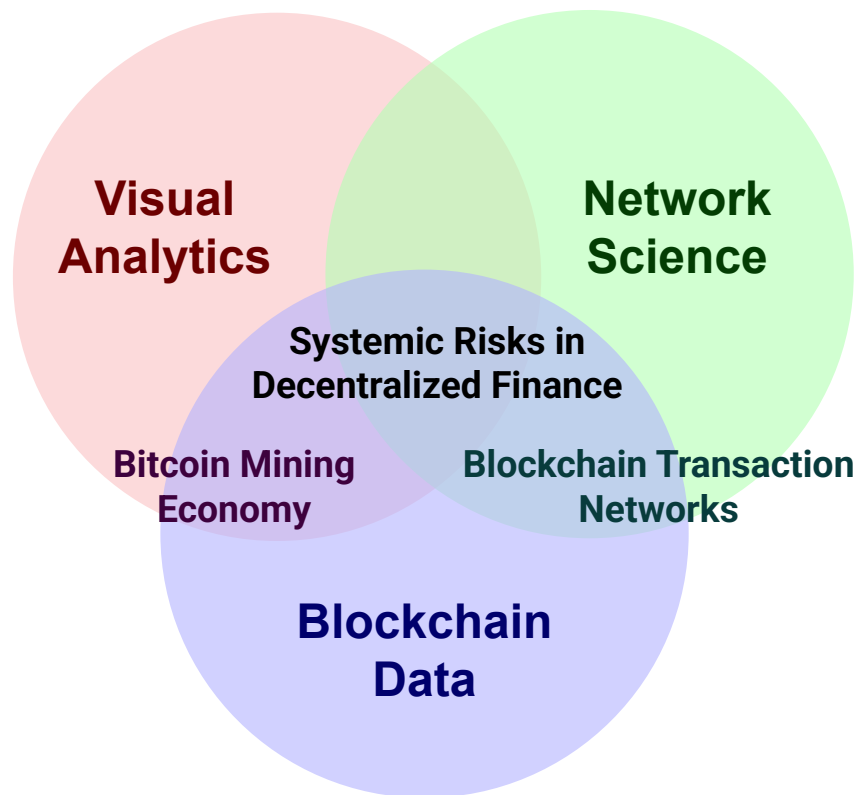
My Research Overview

2026-2030: Post-Doc 
Visual Analytics Research Unit
Faculty of Informatics, TU Wien



Laura Bassi Centre of Expertise
Centre for Visual Analytics
Science & Technology

2018-2022: Ph.D. Student 
IRT SystemX & Aviz team
Inria, Université Paris-Saclay



2022-2026: Post-Doc 
Blockchain@X, CREST,
École Polytechnique



BLOCKCHAIN
@POLYTECHNIQUE

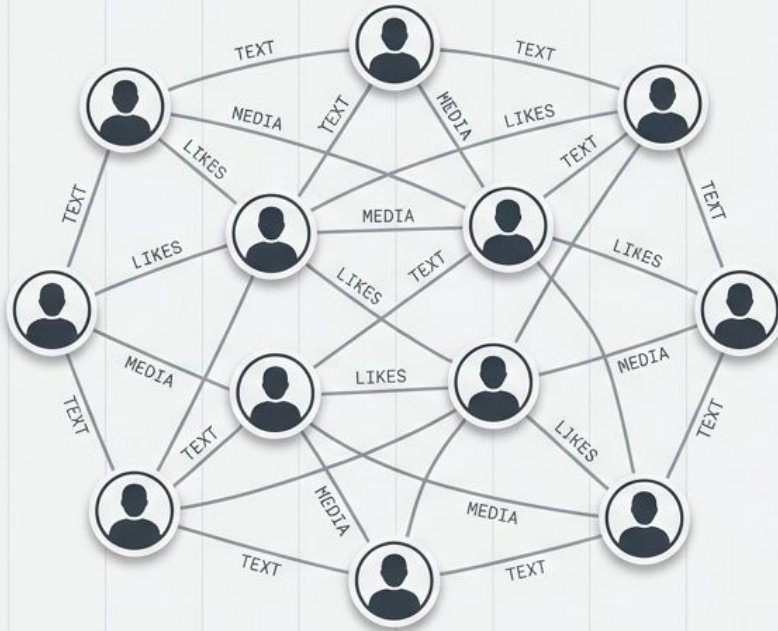


IP PARIS

Cryptocurrency Network Analysis

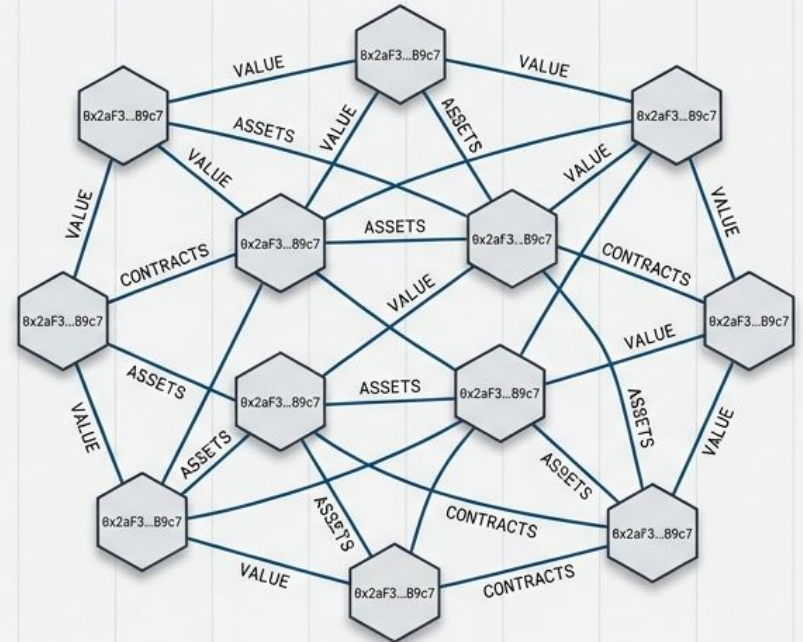
- Natkamon Tovanich, Célestin Coquidé, Rémy Cazabet. **Cryptocurrency Network Analysis**. In: *Encyclopedia of Social Network Analysis and Mining (3rd ed.)*, Springer. (Forthcoming)

Information Exchange



Online social networks map the flow of social interactions and textual content.

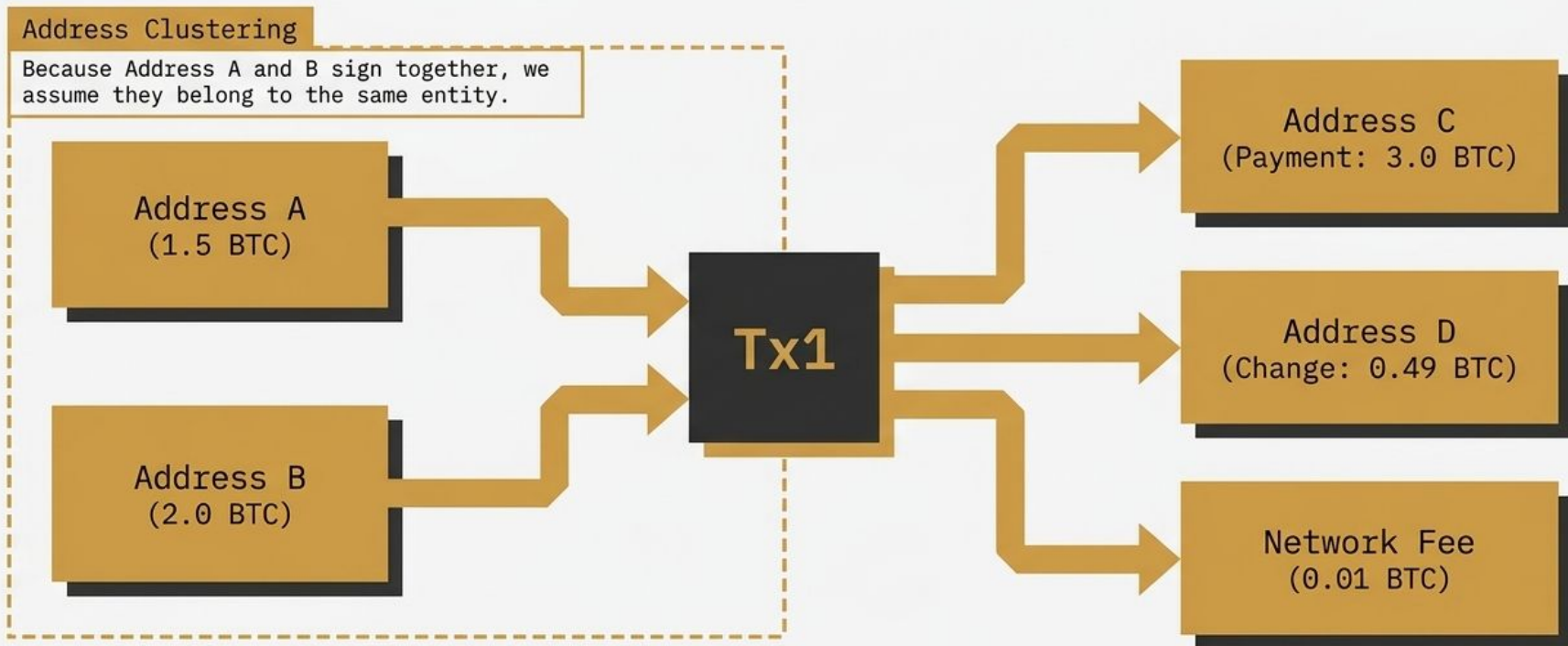
Value Exchange



Cryptocurrency networks apply the exact same network science tools, but map the flow of economic transactions and smart contract interactions.

We have the data. The challenge is no longer access—it is structure.

Bitcoin: The Anatomy of UTXO



To map a UTXO network, raw data must be statistically transformed into a 'User Network' using heuristics and change-address detection.

Ethereum: The World Computer

Protocol Layer (DeFi)

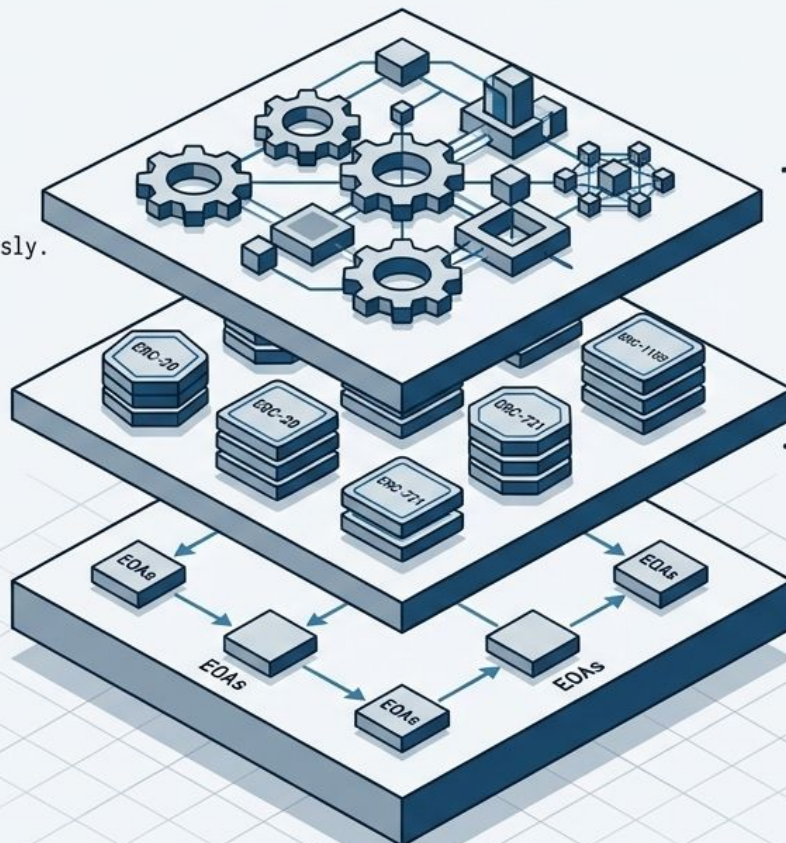
Complex decentralized finance structures (DEXs, Lending platforms) utilizing multiple tokens simultaneously.

Asset Layer (Tokens)

Smart Contracts issuing fungible (ERC-20) and non-fungible (ERC-721/1155) tokens.

Settlement Layer (Base)

Native protocol asset transfers (ETH) between Externally Owned Accounts (EOAs).



Internal State Obfuscation

Unlike native ETH transfers, ERC-20 token movements are executed inside smart contract logic. They do not appear as explicit asset transactions on the base ledger, requiring complex Event Log decoding.

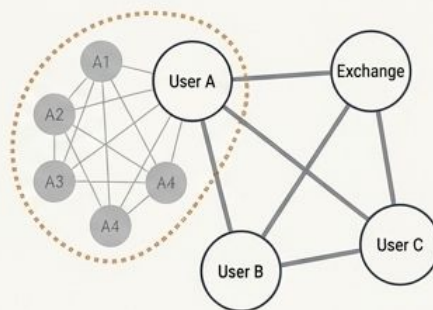
From Raw Data to Actionable Graphs

Raw Transaction
from: 0x1a...3c to: 0x5d...e8 value: 2.5 ETH timestamp: 2023-10-27 10:38:00 UTC



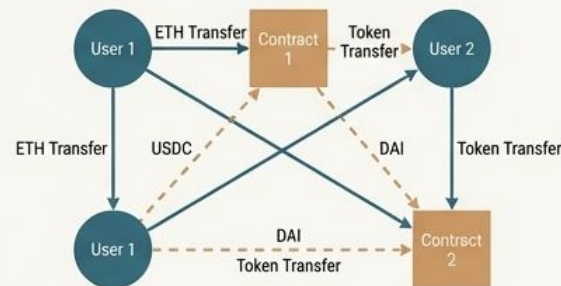
Address Network

The most direct representation. Simple but noisy, as one real-world transaction can create a large number of edges.



User Network (Post-Clustering)

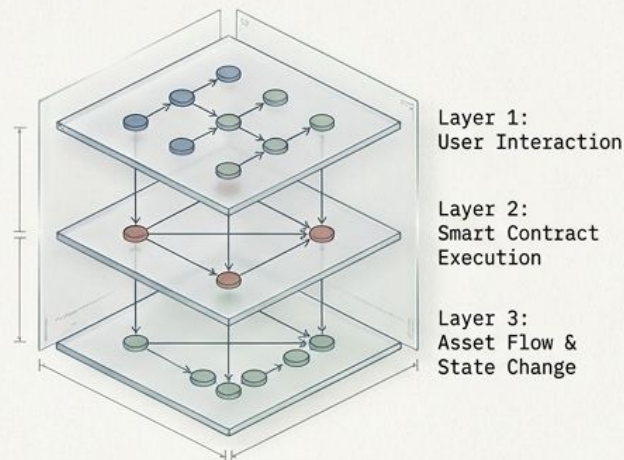
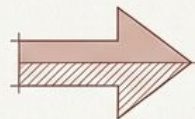
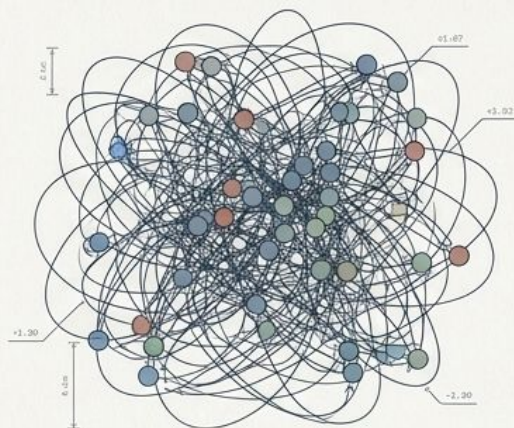
A more meaningful representation where nodes are actors. Requires accurate address clustering heuristics or ML models to build.



Token Transfer Network (Ethereum)

Captures the full richness of the Ethereum ecosystem, representing the flow of diverse digital assets. This is the foundation for advanced DeFi analysis.

The Decentralized Finance Observability Challenge



Composability

Smart contracts automatically interact with other smart contracts within a single transaction, obscuring the primary user's intent.

Missing ABIs

Application Binary Interfaces (ABIs) are required to read human-readable method labels. Over 45% of transactions lack these labels, rendering them opaque.

Asset Transformation

A user can borrow ETH, swap it for USDC, and deposit it into a liquidity pool—all within one transaction block. Standard 2D tracing loses the trail immediately.

Cryptocurrency network analysis has three major areas of impact.



Financial Analysis & Intelligence

- Understanding market structure, liquidity flows, and systemic risk.
- Identifying novel trading strategies (e.g., Maximal Extractable Value - MEV).
- Informing quantitative models with on-chain flow data.



Forensics & Compliance

- Tracing the flow of illicit funds from hacks, scams, and ransomware attacks.
- Providing evidence for law enforcement investigations.
- Helping exchanges comply with Anti-Money Laundering (AML) regulations.

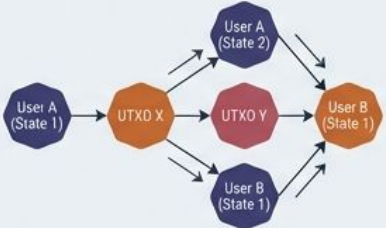
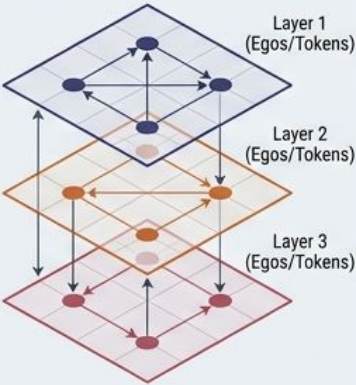
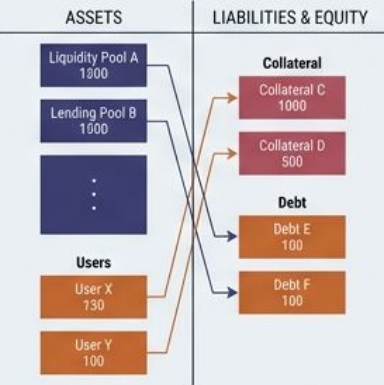


Socio-Technical Understanding

- Studying the emergence and evolution of a decentralized economy.
- Characterizing user behavior and protocol adoption patterns.
- Identifying trends like centralization (e.g., the dominance of exchanges).

The Unifying Framework

A scale-based methodology for blockchain forensic and economic analysis

MICRO	MESO	MACRO	SYSTEMIC
Decoding Opaque Semantics 	Fingerprinting Entities & Strategies 	Mapping Market Structures 	Quantifying Protocol Fragility 
• Nodes: Accounts & Contracts • Edges: Token Transfers • Methodology: Graphlet & Motif Mining	• Nodes: User States / UTXOs • Edges: Temporal Flows • Methodology: Taint Analysis & Path Tracing	• Nodes: Egos & Tokens • Edges: Intra/Inter-layer Transfers • Methodology: Multilayer PageRank / CheiRank	• Nodes: Pools & Users • Edges: Collateral & Debt • Methodology: DebtRank & Contagion Simulation

Micro-Level Analysis

The Topology of DeFi Transactions

- [Natkamon Tovanich](#), Célestin Coquidé, Rémy Cazabet. **Decoding Decentralized Finance Transactions through Ego Network Motif Mining**. *The 13th International Conference on Complex Networks and their Applications*, Dec 2024, Istanbul, Türkiye.
- Célestin Coquidé, Alessio Brini, [Natkamon Tovanich](#). **Graphlet-based Decentralized Finance Transaction Method Classification**. *The 15th International Conference on Complex Networks and their Applications*, Dec 2026, Granada, Spain. (Forthcoming)

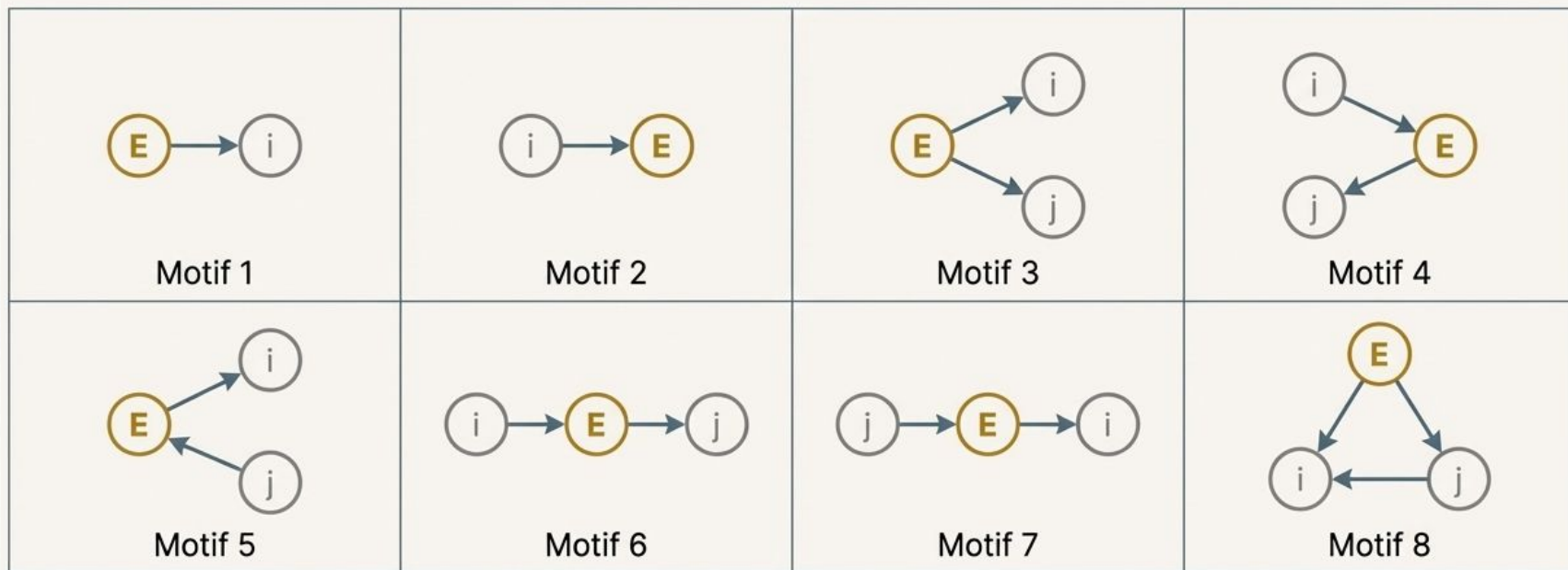
The Blind Spot of On-Chain Forensics

DeFi interactions are complex, but public contract interfaces (ABIs) are frequently incomplete, missing, or deliberately obfuscated. Our baseline dataset reveals that barely 55% of transactions have reliable human-readable labels. Without them, standard forensic tools go blind.



The Elemental Alphabet of DeFi Transactions

The average DeFi transaction involves just 2.09 accounts. By extracting the Ego Transfer Network, we can decompose millions of chaotic transfers into a fundamental alphabet of just 8 directed, 2-to-3-node structural motifs.



From Motifs to Meaning: Interpretable Signatures of DeFi Methods

By analyzing the most frequent motifs in transactions with known labels, we can extract “signatures” for specific DeFi methods. A pruned decision tree model helps identify the most discriminative patterns, making the classification highly interpretable.

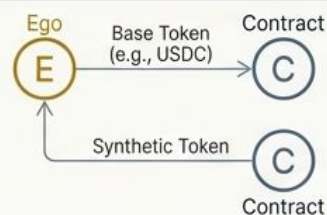
Swap (Leaf 31)



Signature Motif

A simple, direct exchange of tokens between the Ego and a Contract.

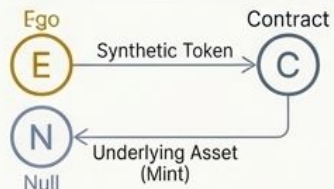
Deposit (Leaf 28)



Signature Motif

Ego sends a token to a Contract and receives a synthetic representative token in return.

Withdraw (Leaf 17)



Signature Motif

Ego returns a synthetic token to a Contract to retrieve the original underlying asset.

Borrow (Leaf 16)

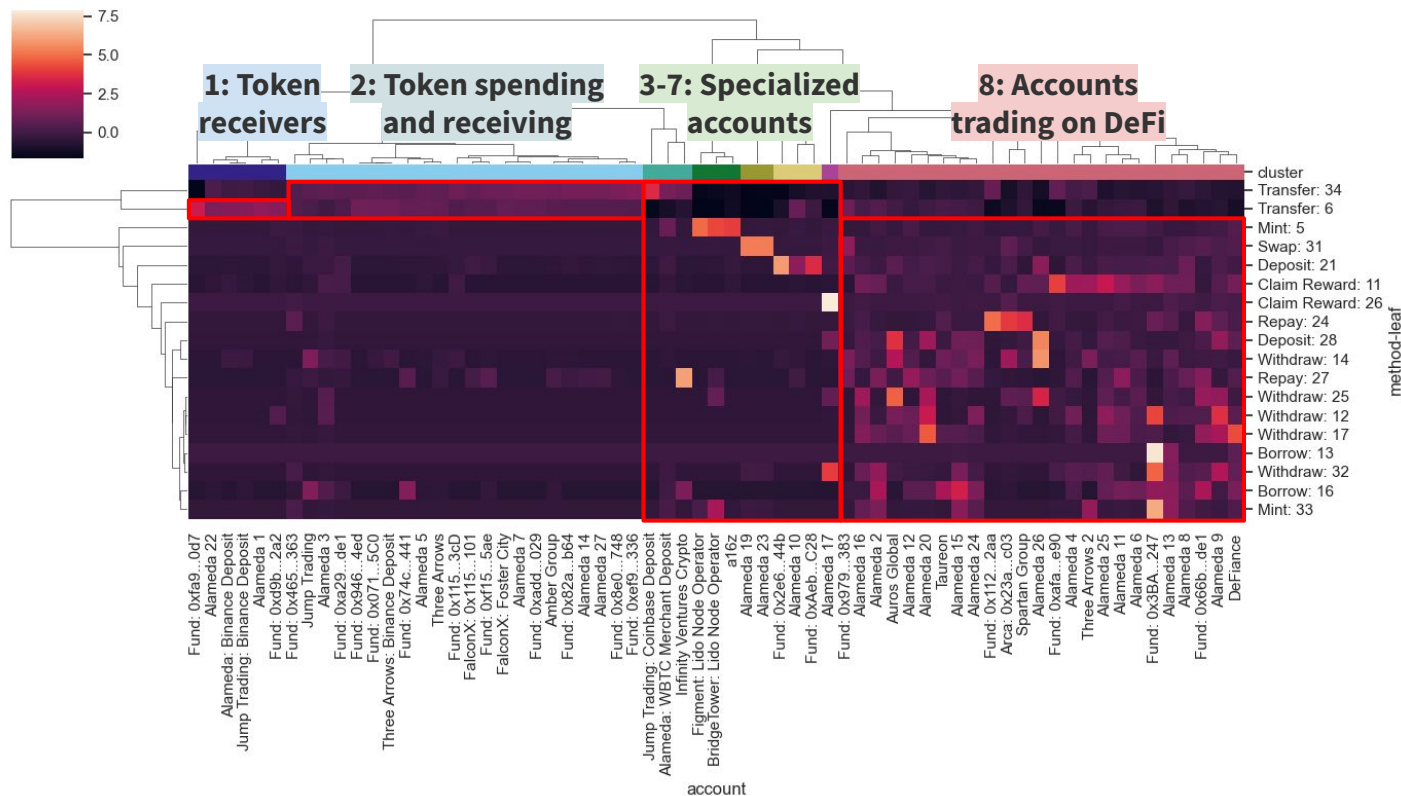


Signature Motif

Ego receives a stablecoin from a lending contract, typically after posting collateral in a prior action.

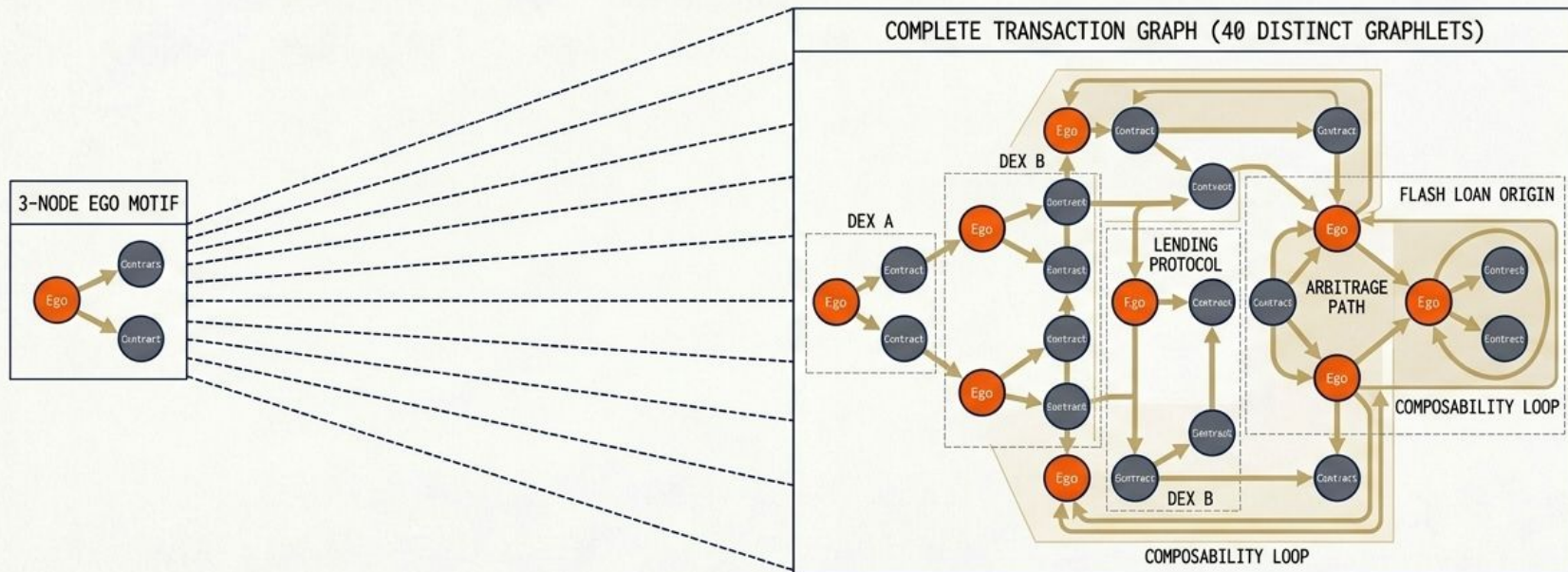
Fingerprinting On-Chain Entities

Behavioral Profiling via Motifs: By counting and normalizing structural signatures across transactions, we can cluster entities by behavioral identity, such as liquidity providers, DEX traders, and off-chain exchanges.

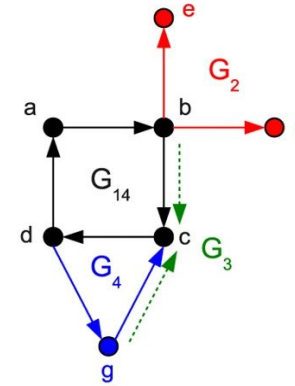
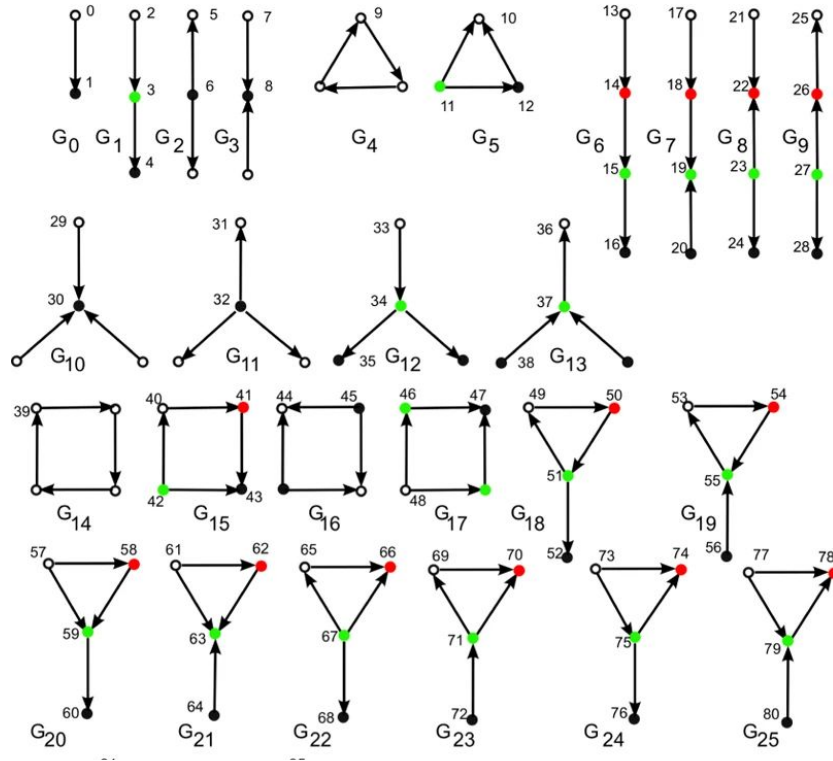


Scaling the Lens: From Ego to Ecosystem

As DeFi evolves—introducing multi-DEX aggregators, flash loans, and high-composability strategies—3-node motifs lose resolution. We must expand our topological X-ray to **Complete Graphs**, utilizing 40 distinct directed Graphlets (G0-G39) to map complex multi-step execution.



Directed Graphlets and Structural Positions



40 patterns on 2-4 nodes

Node numbers identify structural roles.
A transaction can participate
in several overlapping graphlets.

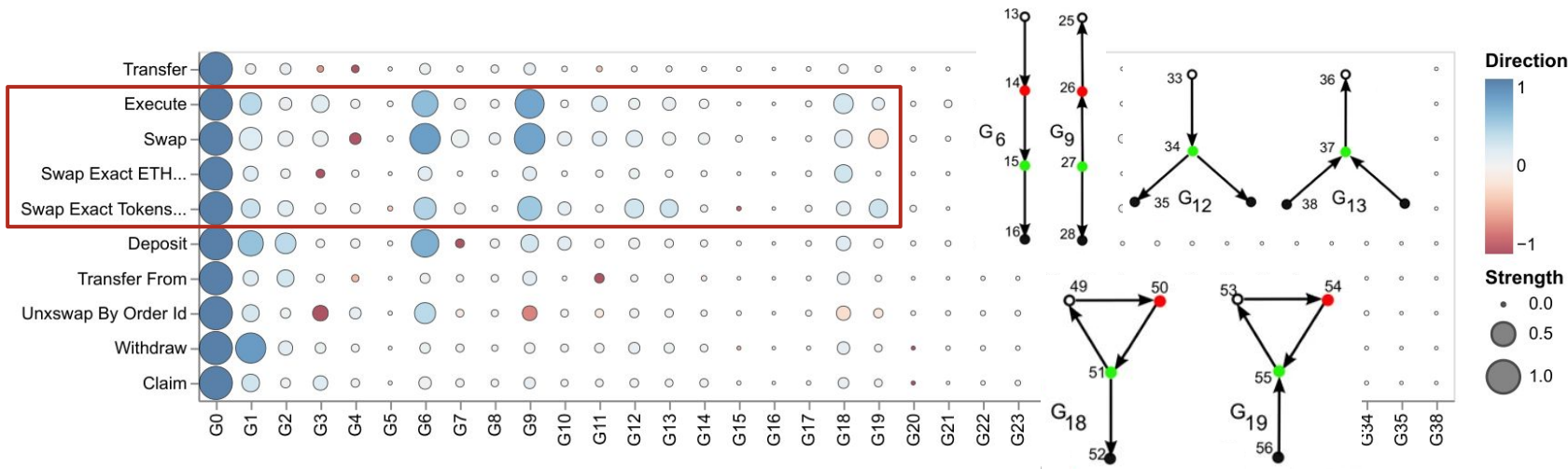
Evaluating Predictive Methodologies

While topological motifs independently establish a strong baseline for classification, integrating **complete graphlets** with **smart contract event logs** results in near-perfect predictive capabilities.

Methodology	Features Extracted	Best Model	F1-Score
Ego Motifs (2024)	Shape + Node Types	Logistic Regression	0.710
Graphlets Only (2026)	Complete Graph + Node/Token Types	WeightedEnsemble L2	0.693
All Event Logs	All Emitted Logs	WeightedEnsemble L2	0.793
Graphlets + Event Logs (2026)	Topology + Top 1,000 Emitted Logs	WeightedEnsemble L2	0.918

Method-specific Graphlet Signatures

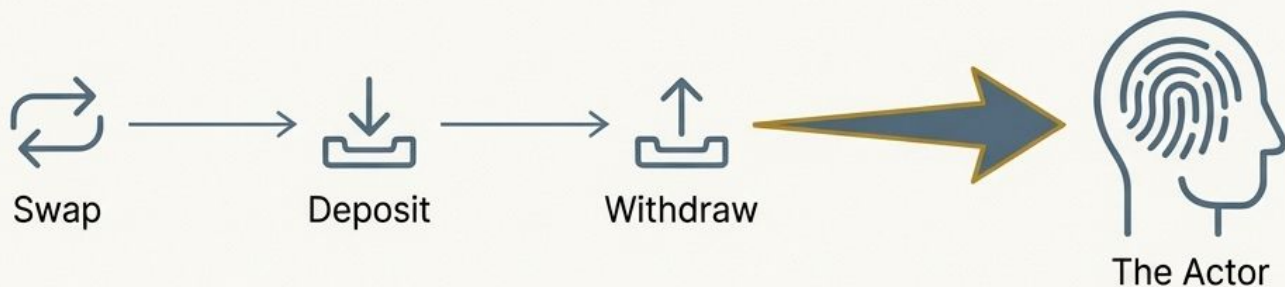
By applying **SHapley Additive exPlanation (SHAP)** and aggregating contribution at the graphlet level, we can extract human-interpretable structural rules that drive the model's predictions



G0 (simple transfer) contribute across all top 10 methods and dominates Transfer.

Swap and Execute use larger patterns, i.e., G6, G9, G12, G13, G18 and G19.

From 'What' to 'Who': A Series of Actions Creates a Behavioral Pattern



Recap

We have demonstrated that the structure of a single transaction reveals its function—the **'what'**.

The Next Question

But actors don't perform actions in isolation. They execute sequences of transactions over time, creating a unique behavioral pattern.

Posing the Challenge for Part II

Can we analyze the flow of funds originating from an entity to create a unique 'fingerprint' that allows us to identify the **'who'**?

Meso-Level Analysis

Cryptocurrency Money Flows

- [Natkamon Tovanich](#), Rémy Cazabet. **Fingerprinting Bitcoin Entities Using Money Flow Representation Learning**. *Applied Network Science*, 2023, 8 (1), pp.1–22
- [Natkamon Tovanich](#), Célestin Coquidé. **Tracing Money Flows in the Multi-Token Cryptocurrency Network**. *The 15th International Conference on Complex Networks and their Applications*, Dec 2026, Granada, Spain. (Forthcoming)

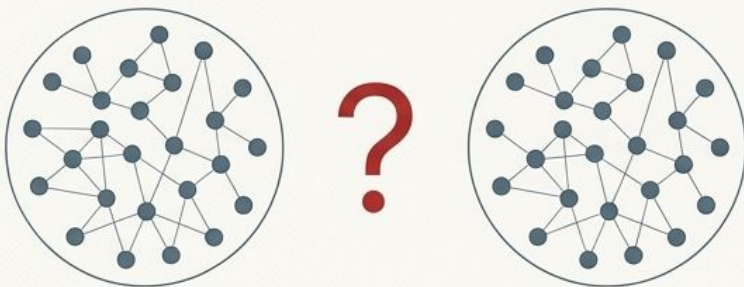
Part II: The Fingerprint of an Actor

The Challenge: Identifying Entities Across Disparate Address Clusters

How can we identify that multiple, unlinked clusters of addresses belong to the same entity?

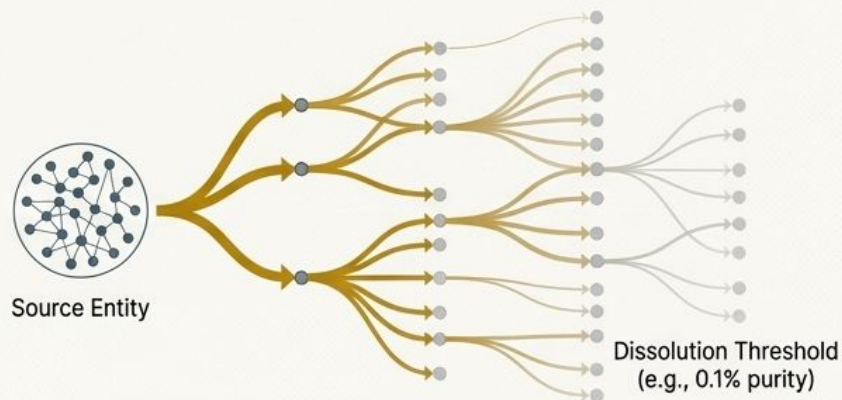
Problem Statement

Standard address clustering heuristics (e.g., common-input) are effective but incomplete. A single entity can operate multiple, separate “wallets” or address pools that never interact directly. These methods cannot match sub-clusters if they never make a transaction together.



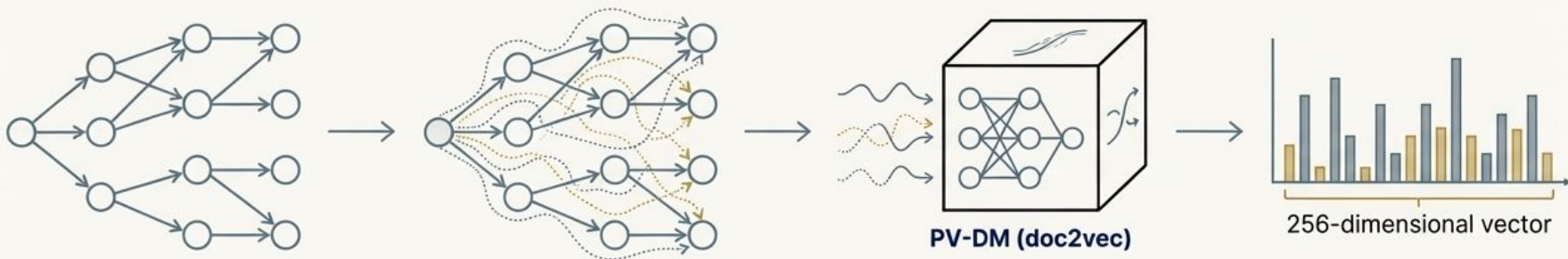
Proposed Solution

We can characterize an entity by its downstream money flow. We introduce the concept of a **Taint Flow**: the directed acyclic graph (DAG) of all subsequent transactions involving coins originating from a source entity on a given day. The flow is tracked until the “taint” is dissolved (e.g., purity drops below a threshold of 0.1%).



The Solution: Learning a Fingerprint from Taint Flow

Our method synthesizes each complex taint flow graph into a single 256-dimensional vector—a “fingerprint”—that captures its unique topological and temporal characteristics.



1. Extract Taint Flow

Start with an entity's transactions on a given day. Show a small DAG representing the flow.

2. Generate Walks

Generate 10,000 random walks from the source to dissolution. This process translates the graph structure into a collection of "sentences".

3. Embed Walks

Feed these "sentences" into a Distributed Memory Model of Paragraph Vectors (PV-DM), a `doc2vec` model. The model learns a vector representation for the entire document (the taint flow).

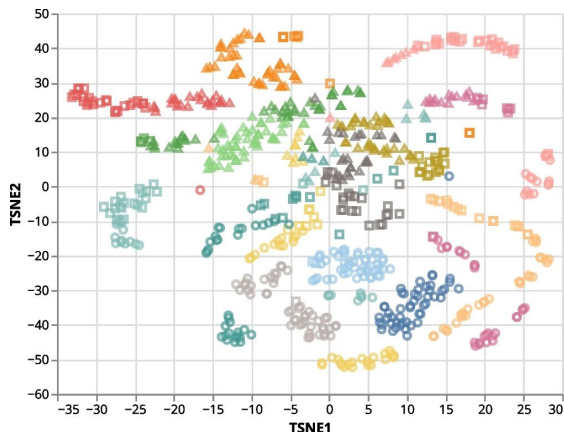
4. The Fingerprint

The output is a single vector that serves as the fingerprint for that specific flow.

Taint Flow Form Distinct, Classifiable Clusters

By projecting 256-dimensional fingerprints into 2D space using t-SNE, we can visualize the separation between entities. The results show that taint flows are a valid basis for recognizing and classifying actors.

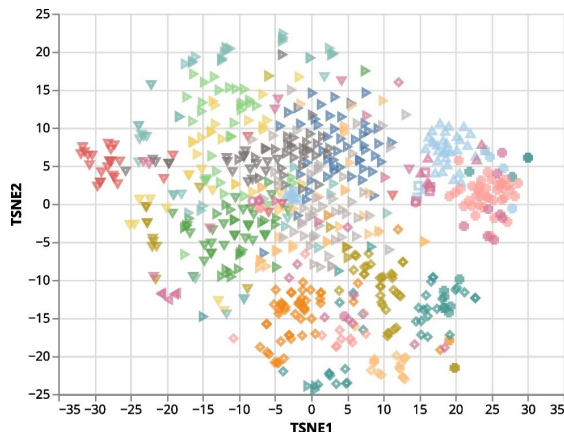
Mining Pools



Classification accuracy 96%

Clear separation between pools.
Within-pool shifts reflect changes
in payout flows over time.

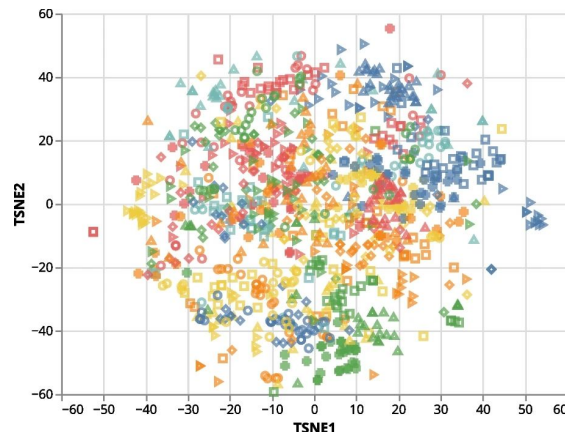
Ransomware Families



Classification accuracy 70%

More overlap, but distinct clusters are
still visible for major families.

WalletExplorer Entities



Classification accuracy: 63%

Highly complex, but flows from the
same entity (shape) and entity type
(color) tend to be closer

Key Insight: High accuracy is achieved using very short path lengths. Tracking flows just **2-3 steps** from the source is often sufficient to identify the entity, implying that near-neighbor interactions are most characteristic.

Following value across DeFi is hard



New tokens, new protocols

Smart-contract chains let anyone issue tokens and build trading, lending and liquidity protocols.



Complex trading strategies

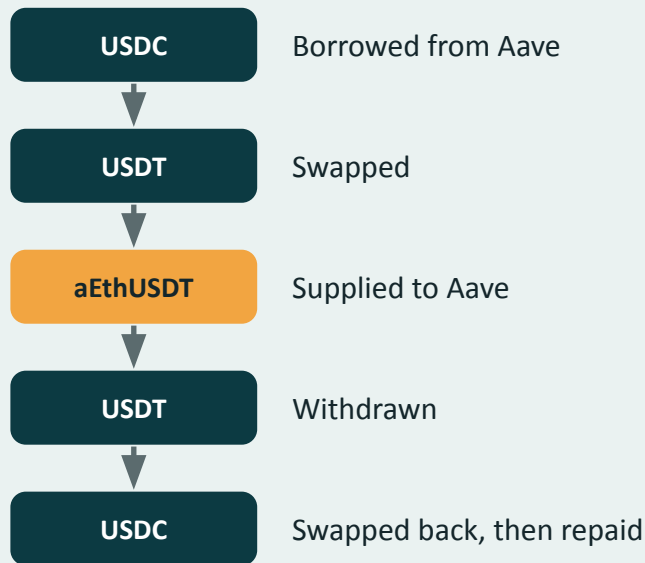
Flash loans, MEV and arbitrage combine several protocols and assets in one go.



Account graphs fall short

Accounts as nodes hide which token moved, in which transaction, and when.

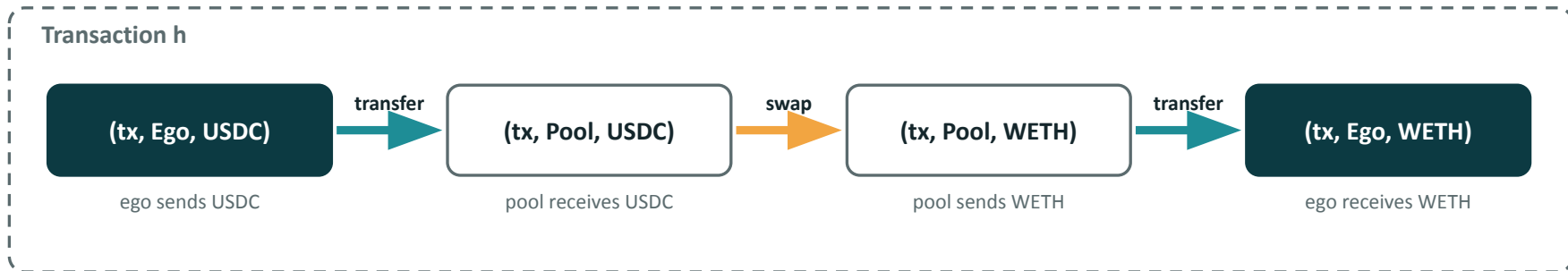
Where does a borrowed asset go?



Example pattern from our Aave case study

A Transaction-Aware Multi-Token Graph

Node $v = (tx, a, \tau)$ — transaction hash, account, asset. One account is split across transactions and assets, keeping time and token information.



Transfer edges

Observed token movement, including mints and burns.



Swap edges

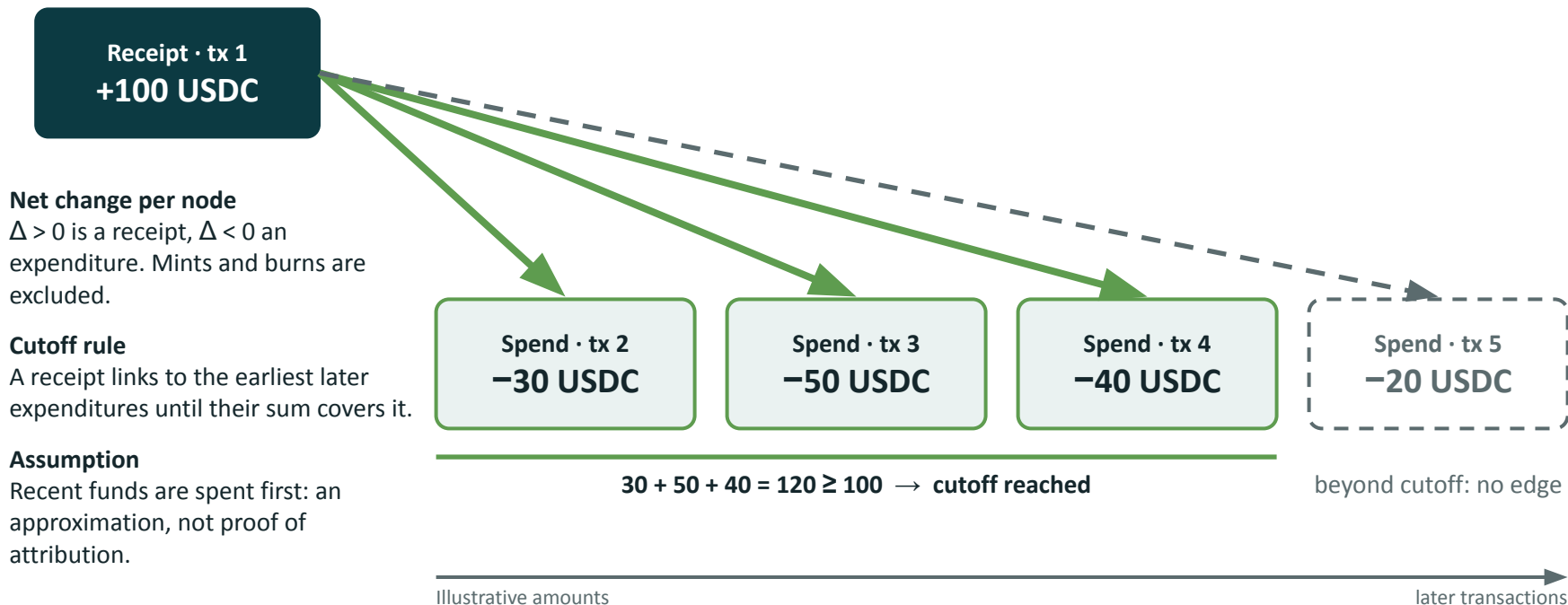
Inferred conversion at a counterparty (never the ego): it receives one asset and sends another.



Teleport edges

Fallback link between related token operations when no transfer or swap path exists. Lower confidence.

Linking Inter-Transactions with a LIFO Cutoff



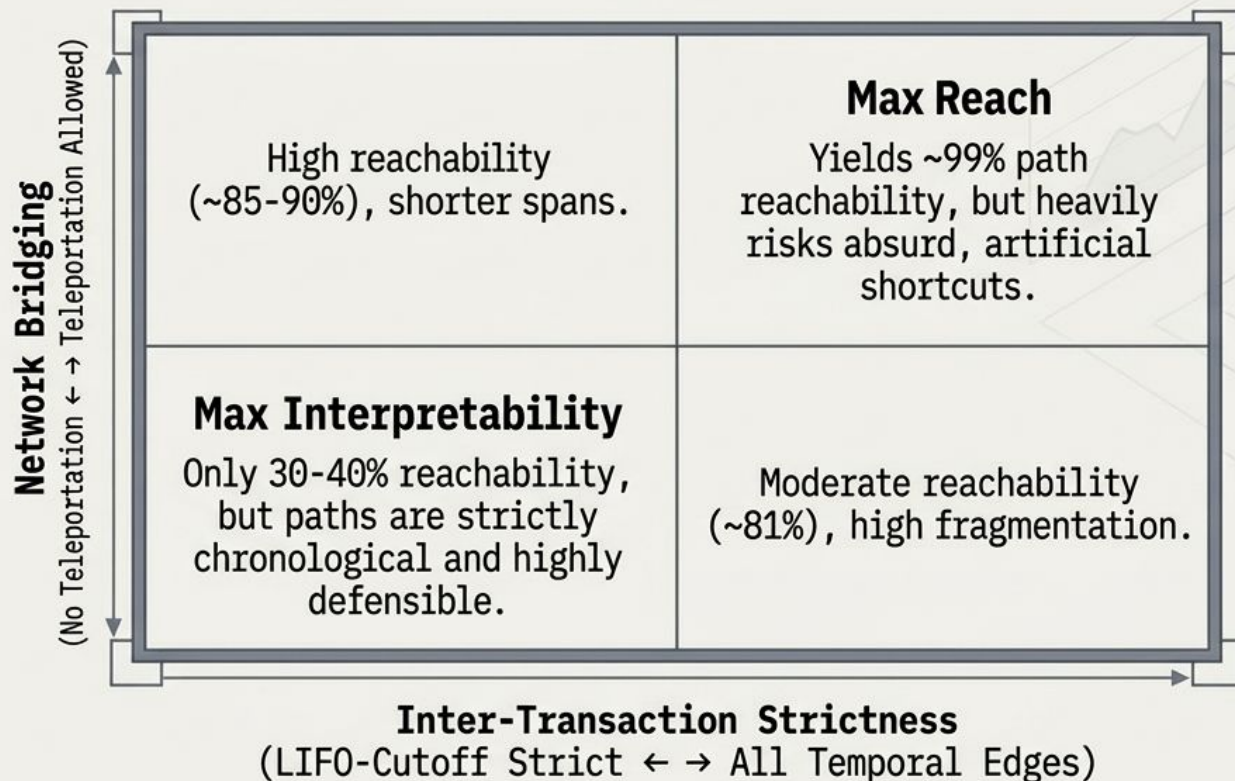
The Case Study: Aave V3 Borrow-Repay Flows



The Objective: Navigate the analytical fog to find the shortest, most economically accurate path connecting the Borrow node to the Repay node.

The Core Trade-Off Matrix

Reachability vs. Interpretability



Insight: There is no “perfect” configuration. Analysts must choose between maximum evidentiary strictness and total network coverage.

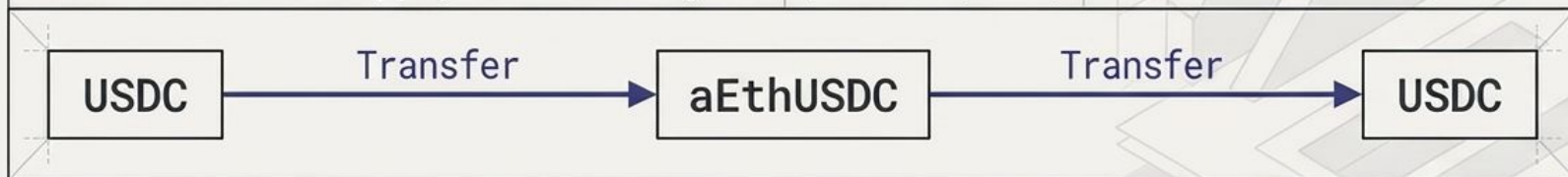
Real-World Activity: Uncovering Economic Strategies

Pattern 1: The Round-Trip Swap (3.4% of paths)



Borrowing to trade a stablecoin peg for arbitrage, then returning to repay.

Pattern 2: The Supply-Withdraw Cycle (3.1% of paths)



Supplying the borrowed asset straight back into an Aave liquidity pool to farm yield or increase leverage.

Real-World Activity: Exposing Network Artifacts

The Artifact Pattern: The Debt-Token Shortcut (8.8% of paths)

WETH

variableDebtEthWETH

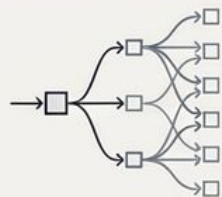
WETH

This sequence is a mathematical artifact, not an economic strategy. Because no continuous transfer/swap path exists, the graph uses the underlying debt token as an artificial intermediary bridge to force a connection.

These anomalies highlight the frontier of DeFi forensics—areas where multi-contract interactions still obscure the true flow of funds.

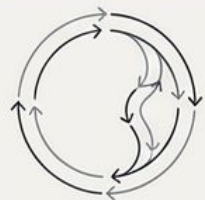
Next Frontiers

Advancing the tracing framework.



Hypergraph
Expansion

Moving beyond pairwise edges to map single transactions that involve multiple simultaneous inputs and outputs across vast protocol arrays.



Cycle
Detection

Upgrading from simple "shortest paths" to capturing complex flows that split, merge, or loop continuously.



Cross-Chain
Tracing

Expanding the (h, a, τ) framework across multiple blockchains simultaneously for a universal, chain-agnostic tracing tool.

Macro-Level Analysis

Multi-Layer Cryptocurrency Network

- Célestin Coquidé, Rémy Cazabet, [Natkamon Tovanich](#). **Analysis of Ego Multi-Token Transfer Networks: A Multilayer Approach**. *Applied Network Science*, 2025, 10 (1), pp.37.
- Célestin Coquidé, Rémy Cazabet, Natkamon Tovanich. **Inside Alameda Research: A Multi-Token Network Analysis**. *The 13th International Conference on Complex Networks and their Applications*, Dec 2024, Istanbul, Türkiye.

Part III: The Blueprint of a Strategy

The Challenge: Modeling Cross-Token Interactions in DeFi

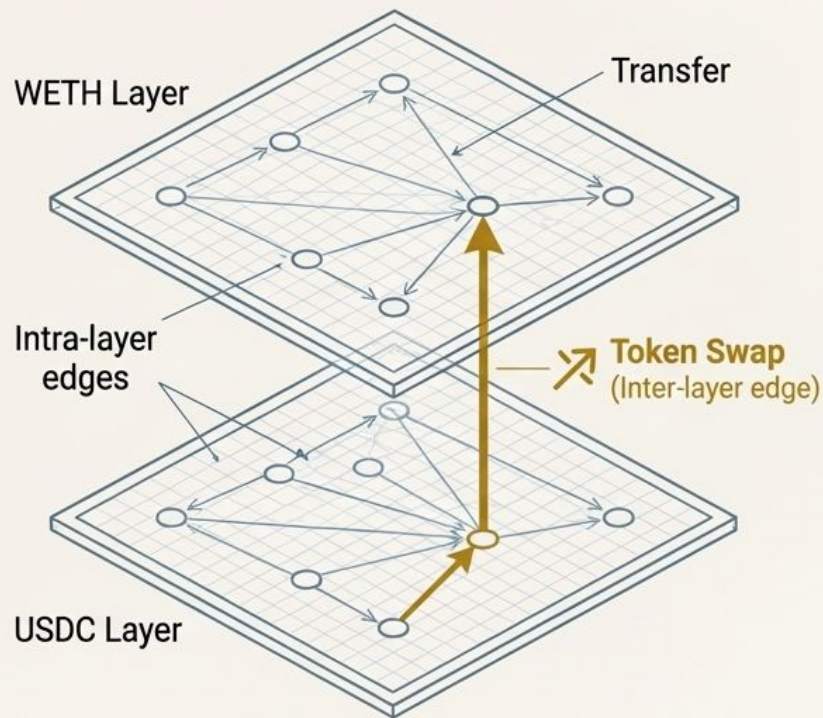
How do we capture inter-token transformations crucial to DeFi, such as token swaps?

Problem Statement

Standard network models treat each token's transfer graph in isolation. They fail to capture the critical moment when one asset is transformed into another within the same transaction (e.g., swapping ETH for USDC).

Proposed Solution: The Multilayer Token Network (MLTN)

- **Framework:** A network where nodes are addresses, and each token exists on its own "layer".
- **Intra-layer edges:** Represent transfers of the same token between two addresses.
- **Inter-layer edges:** The key innovation. A directed edge is created between layers for the same address if, within one transaction, it sends one token and receives another. This explicitly models token transformation.



The Solution: Quantifying Strategy with Multilayer Centrality

To measure influence and behavior in the MLTN, we adapt standard network centralities and introduce a new metric to quantify trading strategy.

1. Biased PageRank & CheiRank

We use a specialized Biased PageRank model. Unlike standard PageRank, its “teleportation” is biased to prefer jumps along inter-layer edges. This gives greater weight to actors who frequently engage in token transformations.

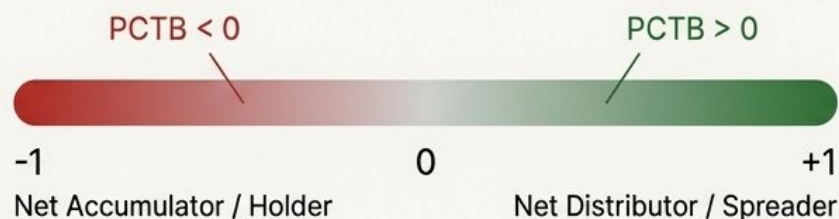
- **PageRank** measures an address's capacity to accumulate tokens (an influence sink).
- **CheiRank** (PageRank on the reversed graph) measures its capacity to distribute tokens (an influence source).



2. PageRank-CheiRank Trade Balance (PCTB)

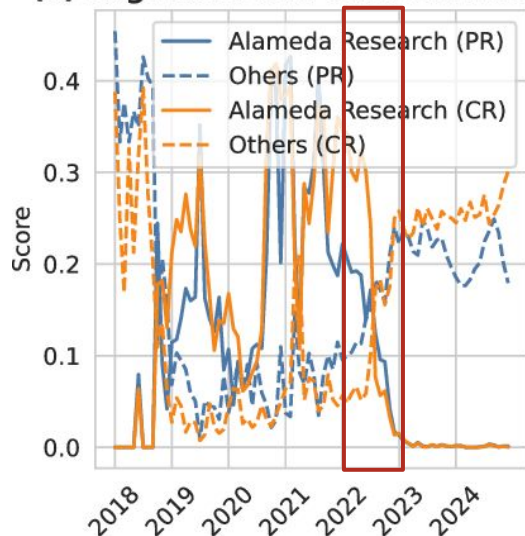
We define a single score to capture net behavior:

$$\text{PCTB} = \frac{\text{CheiRank} - \text{PageRank}}{\text{CheiRank} + \text{PageRank}}$$

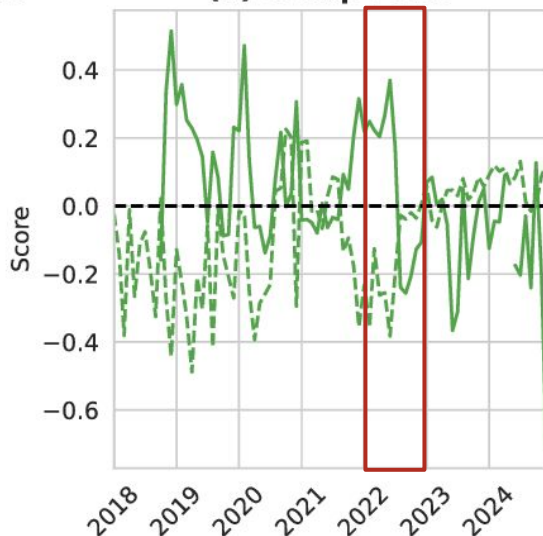


Alameda Research's Trading Behavior

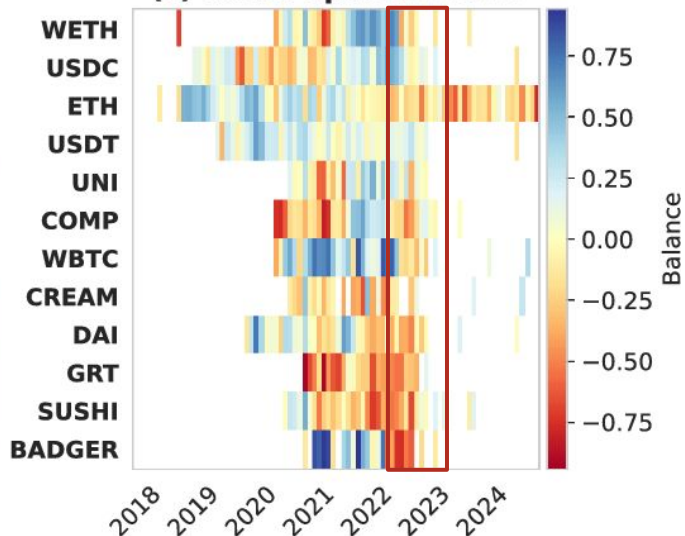
(a) PageRank and CheiRank Scores



(b) Group PCTB



(c) Token-specific PCTB



Network importance

Alameda's PageRank and CheiRank fluctuate strongly, then fall sharply after its November 2022 bankruptcy.

Accumulation and dispersal

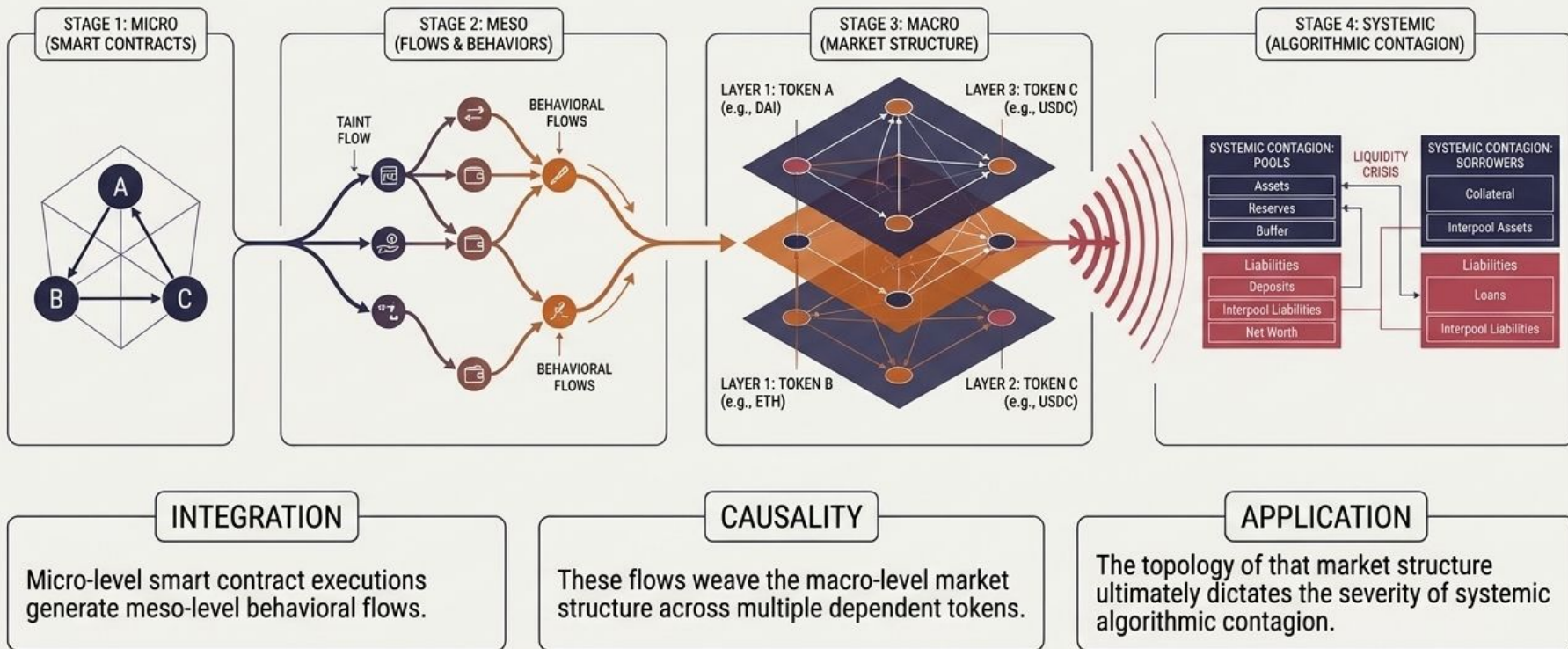
Alameda often exhibits an opposite PCTB pattern to other funds, alternating between token accumulation and dispersal.

Token-specific behavior

USDC shifts from dispersal toward accumulation, while USDT remains dispersive. ETH shows sustained accumulation after bankruptcy.

The Connected Blueprint

DeFi cannot be understood at a single scale.



More to Talk About My Work

- **Systemic Risks in DeFi Lending Protocols**

- Contagion in Decentralized Lending Protocols: A Case Study of Compound <https://dl.acm.org/doi/10.1145/3605768.3623544>
- **My talk at EthereumZuri.ch 2025** https://www.youtube.com/watch?v=vFa9ZZd_W5M
- Shock Propagation in Decentralized Lending Networks: Evidence from the Compound Protocol <https://ssrn.com/abstract=5380164> (Coming Soon)

- **FairOnChain** (November 2023 – July 2026) <https://www.faironchain.org/>

- *Project collaboration with EPFL, HEG Genève, École Polytechnique, and Imperial College London*
- SoK: Unified Blockchain Data <https://hal.science/hal-05185296v1/>

- **Visualization Tools**

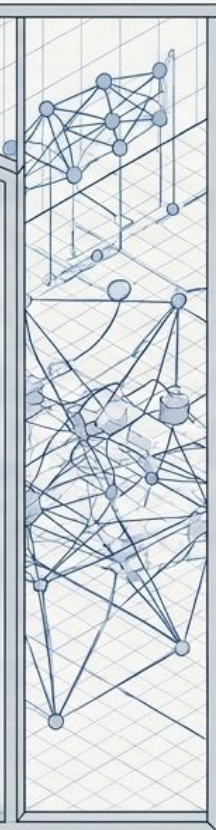
- MiningVis (Bitcoin Mining Economy) <https://miningvis.fr/>
- **MoneroVis (Monero Exploration and Traceability)** <https://monerovis.com/>
- DeFi Lending Risk Analysis Dashboard (Coming Soon)

Future Frontiers in On-Chain Observability

Frontier 1

AI & Graph Neural Networks (GNNs)

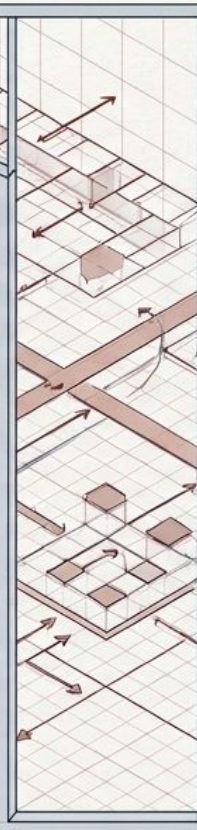
Scaling ML models for real-time MEV discovery, wash-trading identification, and automated fraud detection without relying on smart contract ABIs.



Frontier 2

Cross-Chain Forensics

Expanding Multilayer Token Networks (MLTN) to track multi-asset teleportation across different Layer 1 blockchains and bridging protocols.



Frontier 3

Real-Time Regulatory Stress Testing

Implementing automated DebtRank simulations as a live compliance standard, shifting DeFi regulation from post-mortem audits to predictive topological risk management.

