DeFi Transaction Tracing

Jacky Jang
Rami Younes
Helen Samara Dos Santos
Philippe Béliveau
Lingyi Yang
Kiyan Karimi Nemch
Odile Marcotte
Jean-François Plante
Frédéric Dupont-Marillia

Problem #3

14th Montreal Industrial Problem Solving Workshop
May 13-17, 2024

Autorité des marchés financiers

LAMBDA
Laboratoire analytique, modélisation et IIi de l'AUTORITÉ DES MARCHÉS FINANCIERS

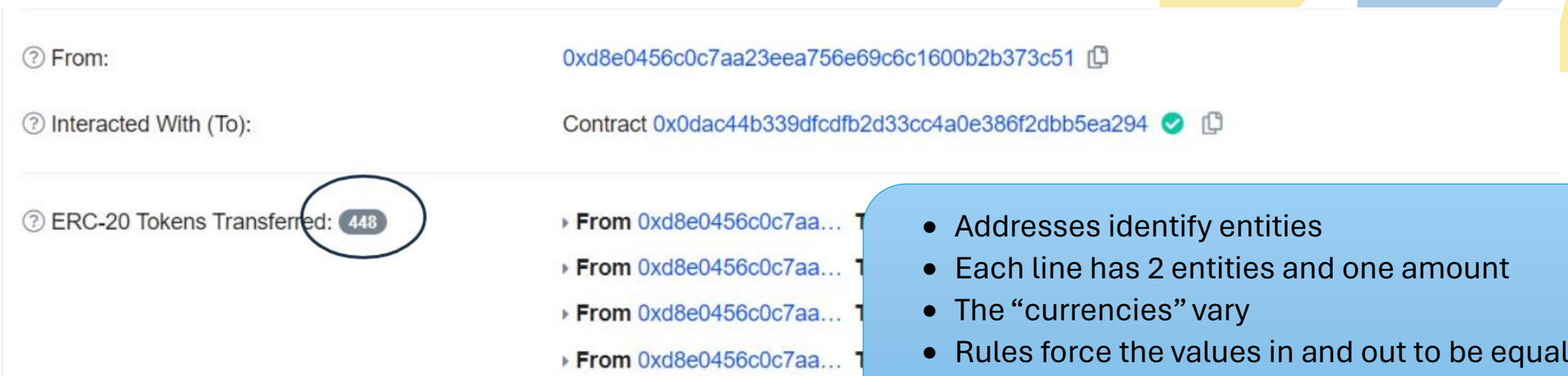
Plan

- The problem
- Sub-transactions interpreted as a graph
- Algorithm to extract the total value expressed in WETH
- Large Language Models
- Scoping Supervised Learning Approaches
- Conclusion

The problem

Crypto-currency transactions are complex.
One transaction can be made from many sub-transactions (sometimes hundreds of them!).

Key problem: from these sub-transactions, how can we make sense of what the overall purpose of the main transaction is?

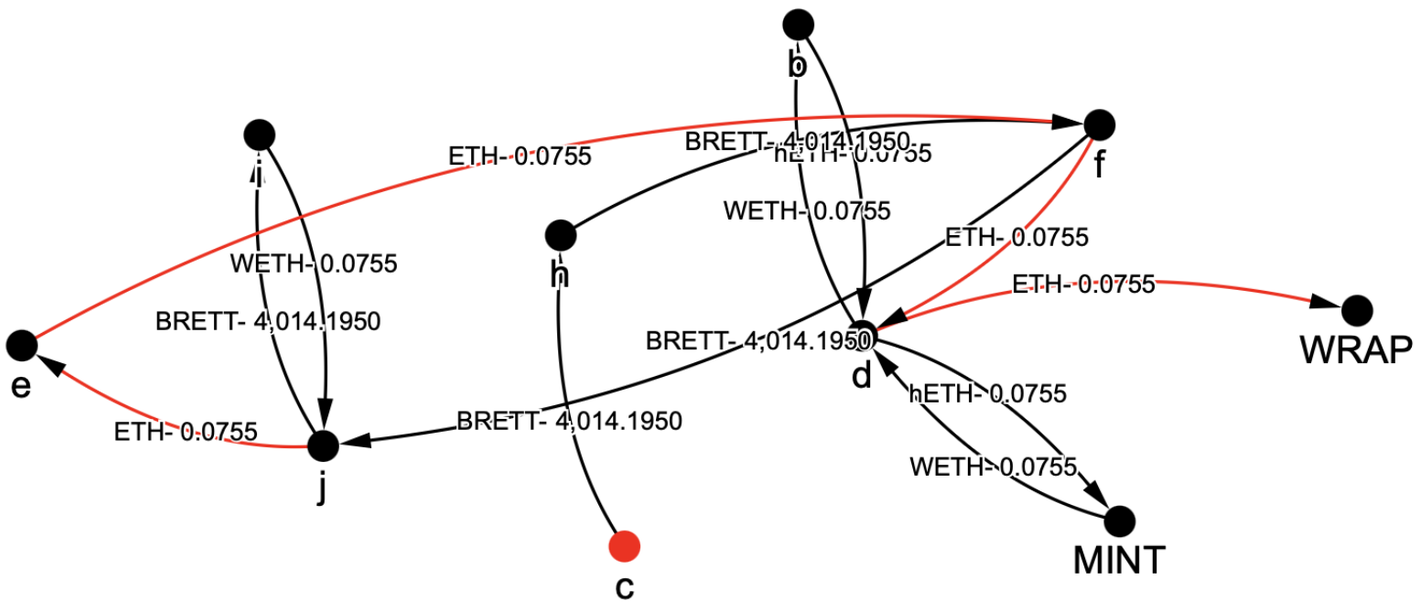


- Addresses identify entities
- Each line has 2 entities and one amount
- The “currencies” vary
- Rules force the values in and out to be equal
- Info defines a network!

Additional info

- Thousands of currencies – anyone can create a currency in only 15 minutes!
- Wrapped Ether (WETH) is the basic currency (or reference currency)
- The value of transactions can be understood through tracing the flow of WETH

How to extract the WETH value?



The Breadth-First Search (BFS) Algorithm

BFS is a well-known and linear-time algorithm for searching graphs.
It proceeds by marking a source, then its children, grandchildren, and so on.
We adapted it to the case of multiple sources and sinks and made sure that there was at least one source and one sink by splitting the “initiator” if necessary.

At the end of the search the algorithm either

(a) determines that there is a path from a source to a sink not including any WETH arc, or
(b) returns an arc cut separating the set of sources from the set of sinks.

At the end of the search the algorithm either

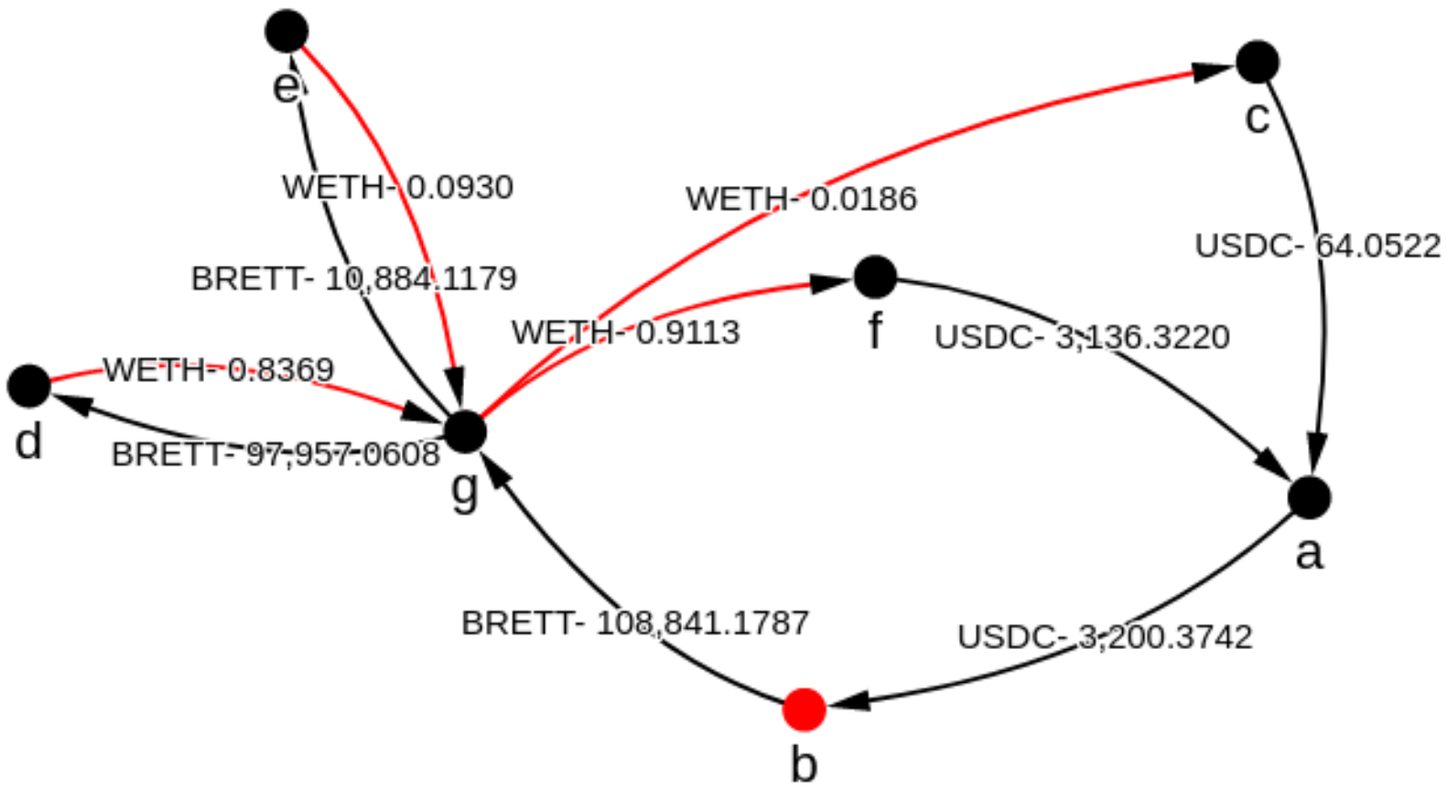
(a) determines that there is a path from a source to a sink not including any WETH arc, or

(b) returns an arc cut separating the set of sources from the set of sinks.

In the first case the transaction is not traceable because there is a path whose “value” is not known.
In the second case the transaction has been traced and its value is the sum of the values of the arcs contained in the cut.

In theory the total value of fees (expressed in WETH) should be deducted from the value of the transaction.
BFS should probably be modified so that some sinks are marked after the other nodes (if at all).

An example


e
WETH- 0.0930
BRETT- 10,884.1179
WETH- 0.0186
c
USDC- 64.0522
WETH- 0.9113
f
USDC- 3,136.3220
WETH- 0.8369
d
BRETT- 97,957.0608
g
a
BRETT- 108,841.1787
USDC- 3,200.3742
b

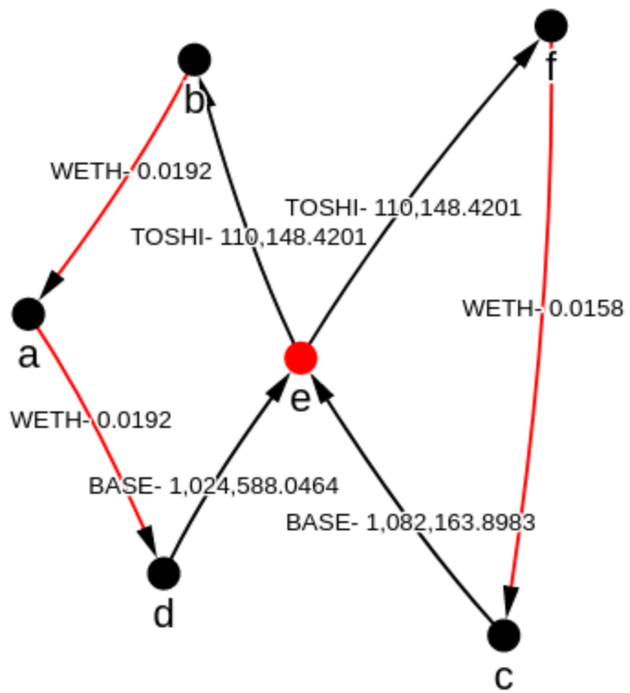
Performance of the Proposed Method

Key Statistics

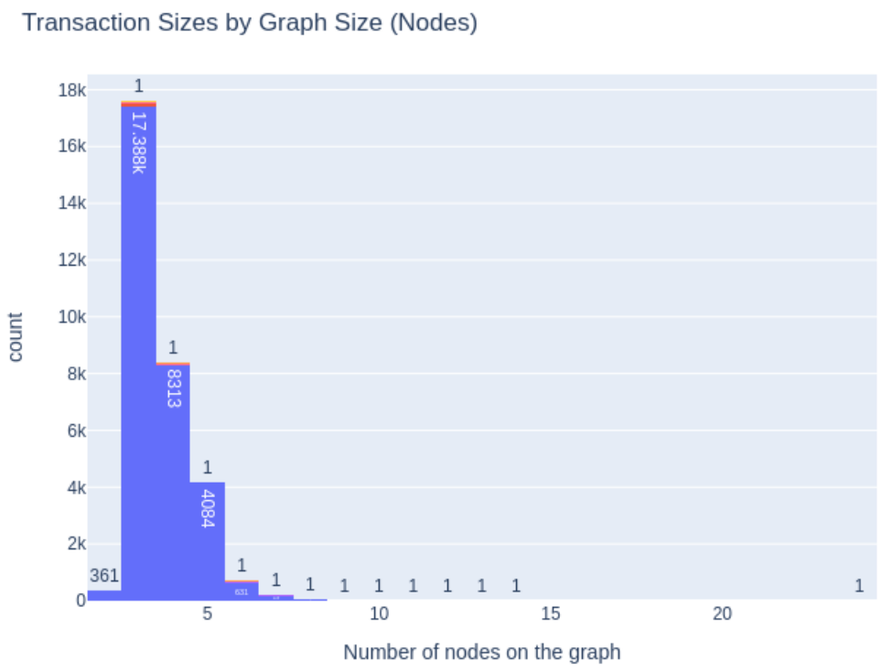
- 32 091 transactions processed with the AMF previous method.
  - 21 631 were “swaps”
  - 7 866 were “transfers”
  - 2 594 were too complex for the existing algorithm to be certain
- Overall 97.36% accuracy for our new method.
  - 98.72% for swaps
  - 95.52% for transfers
  - 91.60% accuracy on complex cases

- Specific performance breakdown:
  - Few false positives:
    - Most inaccurate results were due to the algorithm treating the graph as untraceable
  - Even when the algorithm was deemed wrong **it might not have been**:
    - Hash: 0xeaeb0684213931f114bfc978caf68d012bf0e598de9aa8ca21c4cc84e2e7e758
      - Instigator node received 0.0192 + 0.0158 = 0.035 WETH

type	h_est	truth	h_diff
str	f64	f64	f64
swap	0.034944	-0.019158	-0.015786



The proposed method does not depend on transaction type or graph size



ML-Approach - Large Language Model

We tried testing LLMs' ability to make sense of the log of sub transactions

Input

```
main_from : 0x12A6f5d83E0c643c2056EB47c55b31fD10Be20AD
tx1 : {from: 0x12A6f5d83E0c643c2056EB47c55b31fD10Be20AD, to: 0x3fC91A3afd70395Cd496C647d5a6CC9D4B2b7FAD, value: 0.0424163, symbol: ETH}
tx2 : {from: 0x3fC91A3afd70395Cd496C647d5a6CC9D4B2b7FAD, to: 0x0000000000000000000000000000000000000000, value: 0,0424163, symbol: ETH}
tx3 : {from: 0x12A6f5d83E0c643c2056EB47c55b31fD10Be20AD, to: 0xBA3F945812a83471d709BCe9C3CA699A19FB46f7, value: 4297.313994, symbol: BRETT}
tx4 : {from: 0xBA3F945812a83471d709BCe9C3CA699A19FB46f7, to: 0x3fC91A3afd70395Cd496C647d5a6CC9D4B2b7FAD, value: 0.0424163, symbol: ETH}
tx5 : {from: 0x3fc91a3afd70395cd496c647d5a6cc9d4b2b7fad, to: 0x4200000000000000000000000000000000000006, value: 0.0424163, symbol: ETH}
```

Output

```
{
"transaction_type": "swap",
"token_1": {
"symbol": "ETH",
"amount": 0.0424163
},
"token_2": {
"symbol": "BRETT",
"amount": 4297.313994
},
"value_in_WETH": 0.0424163
}
```

Scoping Supervised Learning Approaches

- Transactions are described by multiple features
- We can extract additional features from the network:
  - Number of nodes
  - Node with most edges
  - Longest path
  - Number of sinks, etc.
- An interesting characteristic of the transaction may be determined manually on a finite number of examples
  - Transaction requiring investigation
  - Presence of a pool
- A supervised learning model is trained on those manually coded labels.
- Predictions are then made on a large number of transactions

Illustration of potential use

Imagine a long list of transactions:

- The network algorithm extracts the value in WETH
- We extract information on the shape of the sub-transactions network
- The ML algorithm classifies the transactions
- We can summarize the value of all types of transactions

Summary

- Crypto-currencies transactions are made of numerous sub-transactions
- Modelling the structure of sub-transactions as a graph allows us to:
  - Easily identify important information (WETH value)
  - Leverage graph tools such as breadth-first search algorithms
  - Summarize features that can become predictors of alternative models
- Additional exploration and research will be necessary to fully assess the performance and robustness of the proposed graph algorithm, but the proof of concept is promising
- Exciting developments in machine learning may lead to further enhancement