Transaction Hypergraph Models for Pattern Identification in the Bitcoin Blockchain

Cliff Joslyn[1,⋆], Chase Dowling[1,2], Sean Kreyling[3], Stephen Ranshous[1,4], Curtis West[1], and Amanda White[5]

[1] Data Sciences and Analytics, PNNL, Seattle, WA
[2] Electrical Engineering, University of Washington
[3] Global Security and Technology Policy, PNNL, Seattle, WA
[4] Computer Science, North Carolina State University
[5] Computational Mathematics, PNNL, Richland, WA

This poster reports on the recent efforts of our research group to explore the potential of advanced graph models, including directed, weighted, hypergraphs, to represent transactions within the Bitcoin blockchain to identify patterns of illicit activity. Few groups have attempted whole-graph analysis of the Bitcoin block chain. Our interest is to do so in the context of the range of mathematical structures most appropriate for this data object, in order to understand the potential to identify patterns of money laundering. More generally, we seek methods to analyze Bitcoin as an exemplar of these burgeoning "social machines": distributed, autonomous, cryptographic databases poised to automate a range of social functions, from trade regulation to contracts and public records. We exploit multiple mathematical representations of the block chain, including:

Transaction Graph: Modeling transactions within blocks as nodes, the result is an address-labeled weighted directed graph.

Transaction Bipartate Multigraph: Gathering common addresses into a distinct node type yields potentially multiple weighted edges between them.

Transaction Hypergraph: Contracting parallel edges then results in a directed hypergraph, with transactions as hyperedges joining multiple addresses as inputs and outputs.

While additional structures like user graphs are commonly analyzed, they depend on additional information and/or heuristic methods. In contrast, we seek mathematical and computational methodologies to identify common situations such as change-making, aggregation, and mixing services as hypergraph patterns based on their inherent structure and content.

We delineate each of these mathematical representations with examples, including descriptive summary statistics and visualizations. We finally report on our additional work in exploitation of high-performance computation capabilities to handle graphs of $O(100M)$ size; the use of integer partition measures to generalize graph degree measurement; and the use of higher order graph invariants such as graph spectra to identify localization of transaction network pivot points and central entities.

⋆ Corresponding author, Pacific Northwest National Laboratory, 1100 Dexter Ave. #400, Seattle, WA 98109, cliff.joslyn@pnnl.gov