Multi-Chain Graphs of Graphs: A New Approach to Analyzing Blockchain Datasets

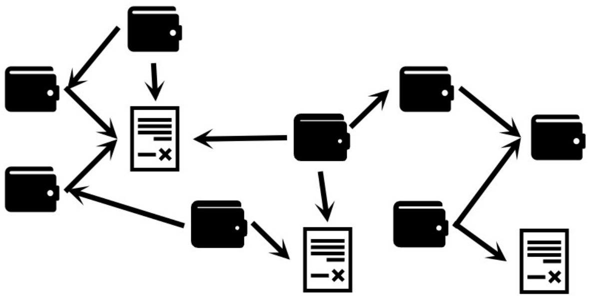
Bingqiao Luo, Zhen Zhang*, Qian Wang, Bingsheng He

National University of Singapore

Blockchain Graph

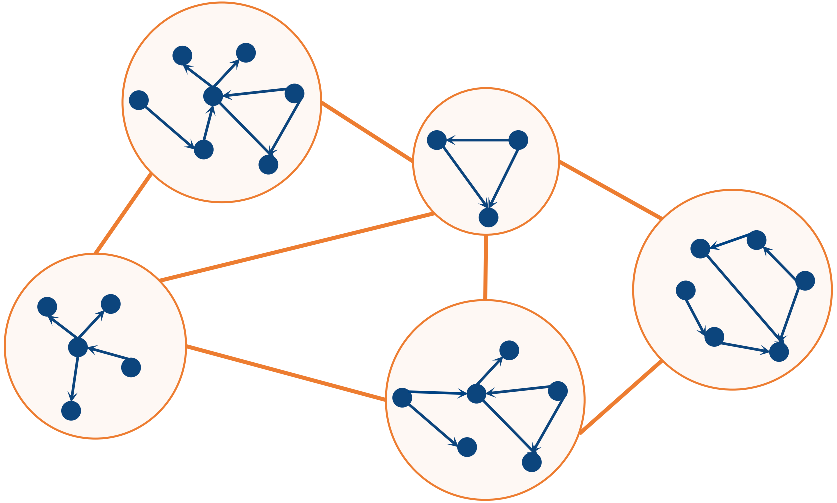
- Blockchain graph represents decentralized transaction activities, revealing patterns, key players, and tokenomics within blockchain networks.

Timestamp	From	To	Gas Fee	Amount of Ether
1629255600	A	B	30000000000	0.1
1629255603	C	C	30000080000	0.8
1629255605	C	D	30000100000	0.3
		…		



Graph of Graph

- Graph of Graph (GoG): connects individual graphs into a larger, hierarchical structure, widely used in chemical, social media and document analysis.



Problem and Solution

- Blockchain ecosystem includes diverse tokens (e.g., DeFi, MEME) that are distinct yet interconnected on the same blockchain, sharing user groups.
- Problem: Existing blockchain graph datasets often miss this interconnectivity and lack large-scale, cross-chain, hierarchical structures.
- Solution: GoG models token transactions as local graphs and their relationships as a global graph, forming a comprehensive dataset.

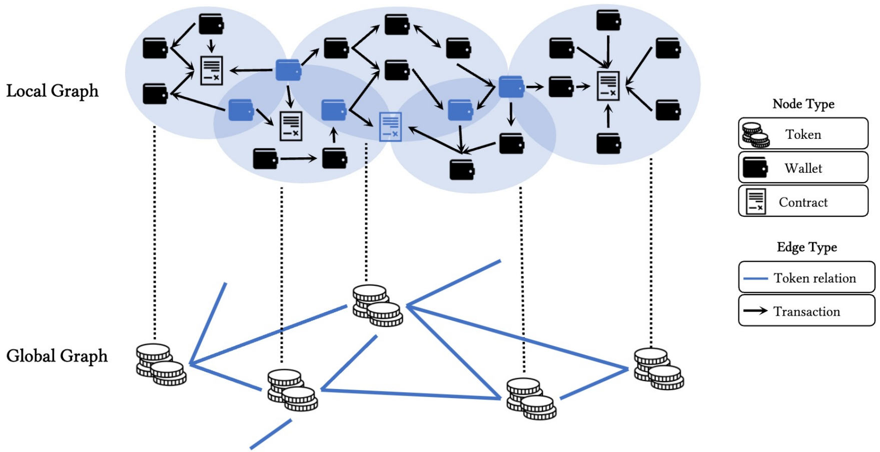
Dataset Overview

Chain	# Token	Start Month	End Month	# Transactions	# Addresses
Ethereum	14,464	2016-02	2024-02	81,788,211	10,247,767
Polygon	2,353	2020-08	2024-02	64,882,233	1,801,976
BSC	7,499	2020-09	2024-02	121,612,480	6,550,399

- Fraud cases: suspicious phishing or hack tokens.
- Other classes: category tag, such as Finance, Meme.
- Data source from prominent blockchain explorers (Etherscan, Polygonscan and BSCscan).

[1] Etherscan. https://etherscan.io/
[2] Polygonscan. https://polygonscan.com/
[3] Bscscan. https://bscscan.com/

Graph Construction



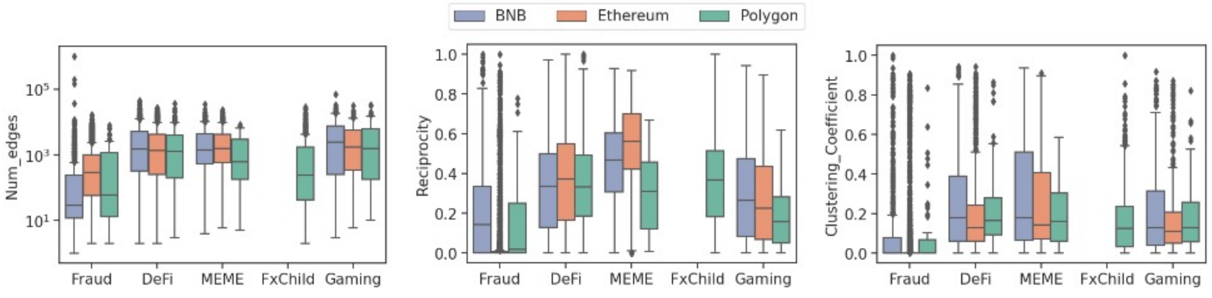
- Local graph: models the transactions of each token
- Global Graph: models the relationships between tokens, with nodes representing token graph and edges weighted by the Jaccard Coefficient to indicate shared users between tokens.

$G_{local} = (N_L, E_L).\ N_L\text{: account};\ \ E_L\text{: transaction}$

$G_{global} = (N_G, E_G).\ N_G\text{: token};\ \ E_G\text{: inter} - \text{token relationship}$

Observation 1: Local Graph Analysis

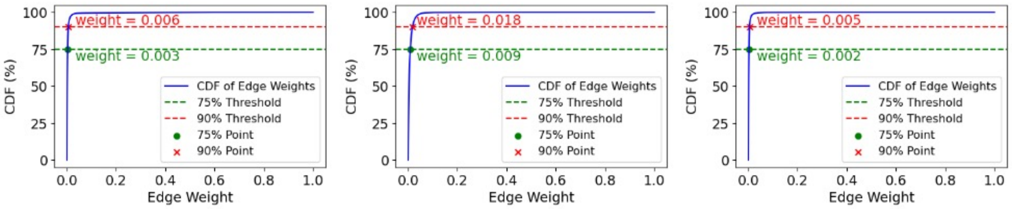
- The distribution of token categories varies across different chains. Tokens within the same class can exhibit distinct network characteristics depending on the blockchain.



(a) Number of edges (b) Reciprocity (c) Clustering coefficient

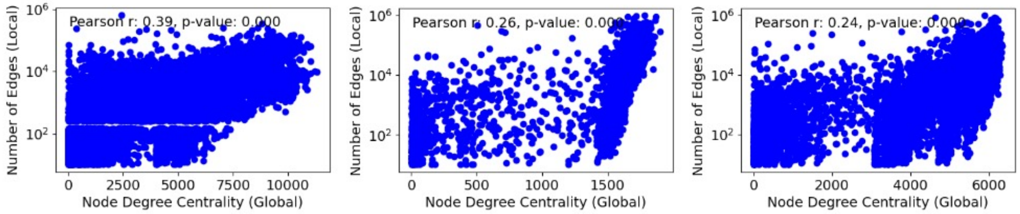
Observation 2: Global Graph Analysis

- Low edge weights suggest limited token interaction, with high weights mainly appearing within same-class local graphs, especially in fraud cases.



(a) Ethereum (b) Polygon (c) BSC

- More edges in local graphs usually indicate central roles in global graphs



(a) Ethereum (b) Polygon (c) BSC

Application 1: Node Classification

- Goal: to categorize tokens into distinct classes.

Model	Ethereum F1-macro	Ethereum F1-micro	Polygon F1-macro	Polygon F1-micro	BSC F1-macro	BSC F1-micro
3-Class Classification						
GCN	62.48±6.31	85.05±1.38	28.82±1.86	74.24±0.83	51.43±5.93	57.02±3.37
GAT	60.22±7.04	84.62±1.23	29.90±2.60	73.94±1.79	54.48±6.15	59.96±3.19
GIN	39.79±11.02	78.58±3.07	28.82±1.53	74.26±0.83	43.29±2.93	55.90±2.86
ResidualGCN	62.85±6.07	84.18±1.50	28.50±0.35	74.37±0.18	50.73±4.59	56.78±2.29
GraphSage	64.17±8.53	85.51±2.05	31.71±2.56	74.48±0.68	56.70±6.12	61.36±2.78
SEAL	**67.31** ± 3.60	**86.65** ± 1.30	29.64±1.70	**74.51** ± 0.16	**63.77** ± 0.59	**65.59** ± 0.42
GoGNN	64.20±4.29	85.89±0.47	**36.11** ± 0.50	66.09±11.02	53.98±4.55	58.03±2.90
DVGGA	37.23±10.57	77.84±4.16	28.44±0.004	74.22±0.17	41.31±8.67	47.03±7.64

- GoG superior performance compared to GNN models across most tasks for node classification.
- GoG models perform well in 3-class but struggle with minor classes in 5-class classification.

Application 2: Anomaly Detection

- Goal: to categorize tokens into distinct classes

	Ethereum (8387: 6022)		Polygon (2257: 58)		BSC (6339: 1042)	
Model	AUC	AP	AUC	AP	AUC	AP
COPOD	83.27±1.09	27.25±0.4	60.52±13.27	**11.33** ± 6.49	52.87±2.09	14.18±0.69
IForest	84.10±0.55	26.93±0.56	64.33±11.43	10.79±5.67	58.36±2.83	11.58±1.57
DIF	**84.56** ± 1.31	32.69±0.95	68.04±10.11	7.99±2.06	51.57±0.49	17.52±2.05
VAE	67.25±1.61	31.46±0.49	**72.45** ± 10.41	10.56±5.09	59.03±0.20	18.70±1.13
GAE	70.85±2.58	31.21±0.68	62.16±0.09	3.85±0.01	56.33±1.25	17.11±0.35
DONE	74.93±2.91	29.03±0.92	62.21±0.30	1.95±0.07	65.86±3.70	10.64±1.10
DOMINANT	75.18±2.69	**43.14** ± 19.69	70.45±7.93	3.55±1.48	**78.87** ± 0.23	8.49±0.03
AnomalyDAE	65.82±8.47	39.24±10.09	60.94±3.06	3.72±0.42	62.49±9.23	**22.71** ± 6.98
CoLA	65.15±7.17	35.80±7.04	54.90±2.74	3.51±0.64	60.87±3.63	19.64±6.29

- GoG models perform inconsistently across blockchains for anomaly detection, highlighting the need for adaptable, network-specific approaches.
- Both groups of methods exhibit poorer performance on the Polygon dataset.

Application 3: Link Prediction

- Goal: to forecast interactions for newly launched tokens

Model	Ethereum		Polygon		BSC	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
GCN	58.07±0.36	62.02±0.23	59.64±1.71	66.92±5.37	66.73±3.12	72.87±3.42
GAT	50.80±0.43	54.50±2.43	50.70±2.07	54.64±4.47	52.82±0.77	53.62±2.86
GIN	56.48±1.61	56.36±1.77	59.03±3.47	58.17±4.33	59.98±2.61	63.57±3.48
ResidualGCN	50.31±0.37	50.66±0.54	49.91±0.08	49.92±0.10	50.41±0.43	50.74±0.94
GraphSage	50.92±1.03	53.67±2.11	56.63±8.88	60.17±12.83	**71.02** ± 0.05	**78.07** ± 1.08
SEAL	57.09±1.64	64.74±4.83	56.98±4.93	64.62±10.34	56.52±4.62	58.05±6.04
GoGNN	**66.94** ± 2.08	**72.04** ± 2.41	57.10±5.21	56.72±4.75	58.99±2.77	66.25±1.84
DVGGA	50.40±1.79	62.93±1.73	**72.38** ± 1.36	**76.00** ± 0.32	63.63±4.94	69.11±3.95

- GoG models face challenges in link prediction, particularly on BSC datasets.
- Most current GoG models lack dynamic algorithms, indicating a need for future research.

Conclusion

- We introduced a novel Graphs of Graphs (GoG) approach for blockchain data analysis, encompassing both local token transaction graphs and global token interaction graphs across platforms.
- Our analysis shows that GoG models can enhance applications like link prediction, anomaly detection, and token classification, offering a foundation for future blockchain graph research.
- Our data and code are accessible here: