Live Graph Lab: Towards Open, Dynamic and Real Transaction Graphs with NFT

Zhen Zhang, Bingqiao Luo, Shengliang Lu, Bingsheng He

National University of Singapore

01

Background

- Graph is a widely used data structure to represent the complex relationships among different entities
- Examples include social networks, etc.

- It's usually impractical for us to obtain the whole real-time graphs due to privacy concerns and technical limitations

02

Background

- In the current literature, studies are usually conducted on a set of outdated and incomplete graphs

Categories	Datasets	Open	Timely Evolving	Complete Structure	Timestamp
Social Network	ego-Twitter [45]	✓	✗	✗	✗
Citation Network	DBLP [70]	✓	✗	✗	✓
The Web	web-Google [44]	✓	✗	✗	✗
Blockchain	Live Graph Lab	✓	✓	✓	✓

- We need open graph datasets that evolve dynamically and are easily accessible in a timely manner

03 Proposed Datasets

- To bridge this gap, we propose the concept of Live Graph Lab, which provides live graphs according to blockchain transactions

- We offer a set of tools for ***downloading, parsing, cleaning, and analyzing*** blockchain transactions to empower the analyses of transaction graphs
  - Getting rid of accessing massive raw transaction data
  - Providing considerable opportunities for temporal graph studies

04

Proposed Datasets

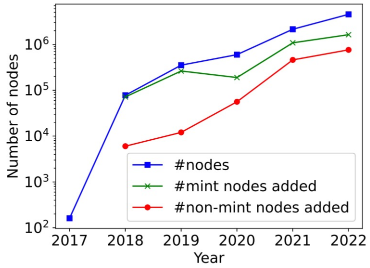
- We synchronize a full Ethereum node to continuously keep up with the latest Ethereum block

- We instantiate a live graph with NFT transaction network in the Ethereum blockchain

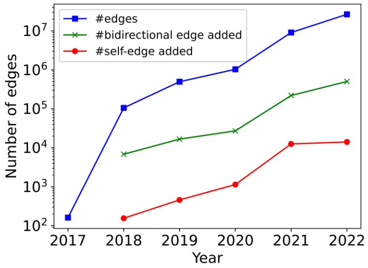
Descriptions	Statistics
Start date (mm-dd-yyyy, UTC)	07-12-2017 13:49
End date (mm-dd-yyyy, UTC)	08-01-2022 06:50
Number of NFT collections	97,667
Number of NFT tokens	77,991,885
Number of account addresses	4,531,020
Number of transactions	124,660,813

05 Observations and Analyses

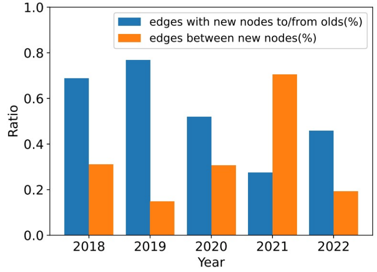
- The NFT transaction network is highly ***active*** and ***growing*** at a fast speed
- Most of the addresses remain active as the NFT ecosystem becomes mature



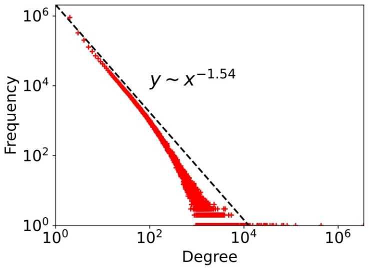
(a) Nodes



(b) Edges



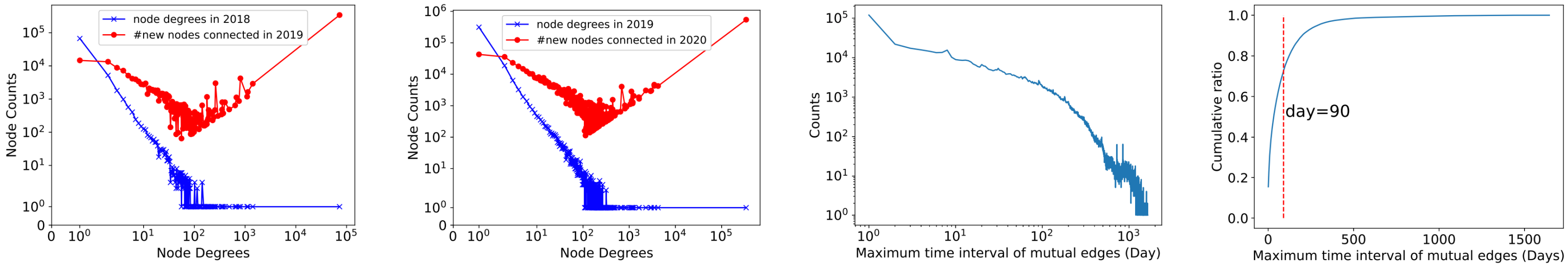
(c) Edge percentages



(d) Degree distribution

06 Observations and Analyses

- It follows ***preferential attachment growth model***
- The reciprocity of the transaction network is relatively low
- It shares similar yet distinct characteristics when compared to traditional networks like social networks, citation networks, etc.



(a) Hub nodes in year 2018 (b) Hub nodes in year 2019 (c) Mutual edges in days (d) Ratio of mutual edges

07 Experiments and Lessons Learnt

- Temporal link prediction

- ***Fixed-split***
- ***Live update***
- ***Three granularities***

Models	Snapshot Days AUC	Snapshot Days MRR	Snapshot Weeks AUC	Snapshot Weeks MRR	Snapshot Months AUC	Snapshot Months MRR
Fixed Split						
Dyngraph2vec	OOM	OOM	OOM	OOM	OOM	OOM
TGCN	53.64±1.60	14.24±2.60	61.55±6.24	36.16±6.60	74.87±4.99	45.97±4.46
EvolveGCN	OOM	OOM	OOM	OOM	OOM	OOM
GCRN-GRU	95.86±0.03	**71.48±0.49**	93.14±0.18	**68.44±0.05**	**86.74±0.80**	58.23±1.23
GCRN-LSTM	94.12±0.92	68.51±2.36	92.90±0.43	67.66±0.31	86.44±0.92	58.71±0.84
DynGEM	OOM	OOM	OOM	OOM	OOM	OOM
Roland-MA	**95.93±0.15**	66.34±0.23	**93.53±0.13**	65.06±0.43	86.23±0.85	54.93±1.42
Roland-MLP	65.46±6.10	43.76±5.94	73.34±7.87	42.04±16.8	85.88±2.22	57.58±4.12
Roland-GRU	73.33±11.5	49.45±8.91	91.48±1.67	66.28±1.86	86.59±0.88	**59.37±1.54**
Live Update						
Dyngraph2vec	OOM	OOM	OOM	OOM	OOM	OOM
TGCN	58.22±5.76	17.77±10.7	59.94±8.44	22.67±19.3	75.01±2.66	43.16±0.75
EvolveGCN	OOM	OOM	OOM	OOM	OOM	OOM
GCRN-GRU	80.95±1.92	39.13±0.39	85.34±0.26	46.08±1.43	81.40±0.34	43.68±0.39
GCRN-LSTM	79.12±2.14	37.83±1.16	84.73±0.34	42.89±3.34	81.24±0.80	41.47±3.00
DynGEM	OOM	OOM	OOM	OOM	OOM	OOM
Roland-MA	**90.47±0.66**	**49.79±0.95**	**88.74±0.37**	**50.77±1.13**	83.93±0.95	47.11±1.16
Roland-MLP	56.32±8.06	22.16±15.4	70.88±10.5	40.91±10.8	79.38±4.47	46.51±2.61
Roland-GRU	60.04±5.08	27.29±6.26	75.93±19.2	48.66±13.7	**84.36±0.46**	**50.75±0.83**

- Active graph should be modeled by fine-grained time granularity

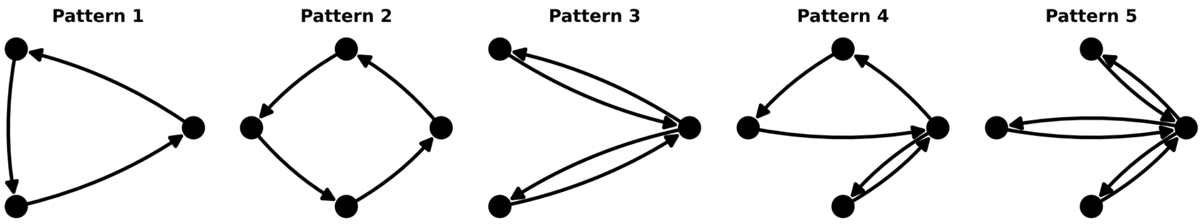
08 Experiments and Lessons Learnt

- Temporal node classification
- Patterns may be sparse at finer time granularities, making it challenging to discern meaningful features for classification

Models	Snapshot Days		Snapshot Weeks		Snapshot Months	
	Accuracy	Recall	Accuracy	Recall	Accuracy	Recall
Dyngraph2vec	OOM	OOM	OOM	OOM	OOM	OOM
TGCN	18.48±2.66	31.15±3.16	43.97±4.57	32.99±2.34	47.45±3.49	**32.53±2.95**
EvolveGCN	OOM	OOM	OOM	OOM	OOM	OOM
GCRN-GRU	41.06±3.30	34.75±2.93	46.78±0.72	**34.79±0.42**	47.42±2.16	28.97±3.22
GCRN-LSTM	46.14±3.29	**35.19±3.61**	48.04±2.37	31.58±1.75	49.32±2.01	35.39±1.49
DynGEM	OOM	OOM	OOM	OOM	OOM	OOM
Roland-MA	**51.02±2.01**	28.77±3.23	**50.39±0.45**	26.33±3.95	47.96±2.69	22.33±3.07
Roland-MLP	48.46±3.18	30.62±3.94	47.59±3.39	31.67±3.62	45.74±4.75	35.04±3.47
Roland-GRU	49.88±2.15	33.38±3.91	46.63±3.07	33.85±1.48	**50.04±0.37**	32.17±2.72

09 Experiments and Lessons Learnt

- Continuous subgraph matching



Query Patterns		Model Query Time (ms)			
Queries	Counts	SymBi	Graphflow	TurboFlux	RapidFlow
p1	19,338	1.22×10^4	1.11×10^4	1.11×10^4	$\mathbf{5.93\times10^2}$
p2	2,243,232	1.19×10^4	1.44×10^4	1.51×10^4	$\mathbf{6.13\times10^2}$
p3	3,012,738	1.11×10^4	1.13×10^4	1.08×10^4	$\mathbf{5.83\times10^2}$
p4	9,472,960	1.24×10^4	1.43×10^4	6.06×10^4	$\mathbf{6.45\times10^2}$
p5	3,154,355,868	1.34×10^5	4.24×10^5	7.72×10^4	$\mathbf{5.84\times10^2}$

10 Conclusion

- We propose the concept of ***Live Graph Lab***, which includes blockchain based temporal graphs that are openly accessible, fully recorded, and dynamically evolving over time

- **Future Tasks**
  - Detecting anomalies activities
  - Providing personalized services like recommendation
  - Informative decision-making
  - …

THANK YOU