GENTI: GPU-powered Walk-based Subgraph Extraction for Scalable Representation Learning on Dynamic Graphs

Zihao Yu*, Ningyi Liao*, Siqiang Luo

Nanyang Technological University, Singapore

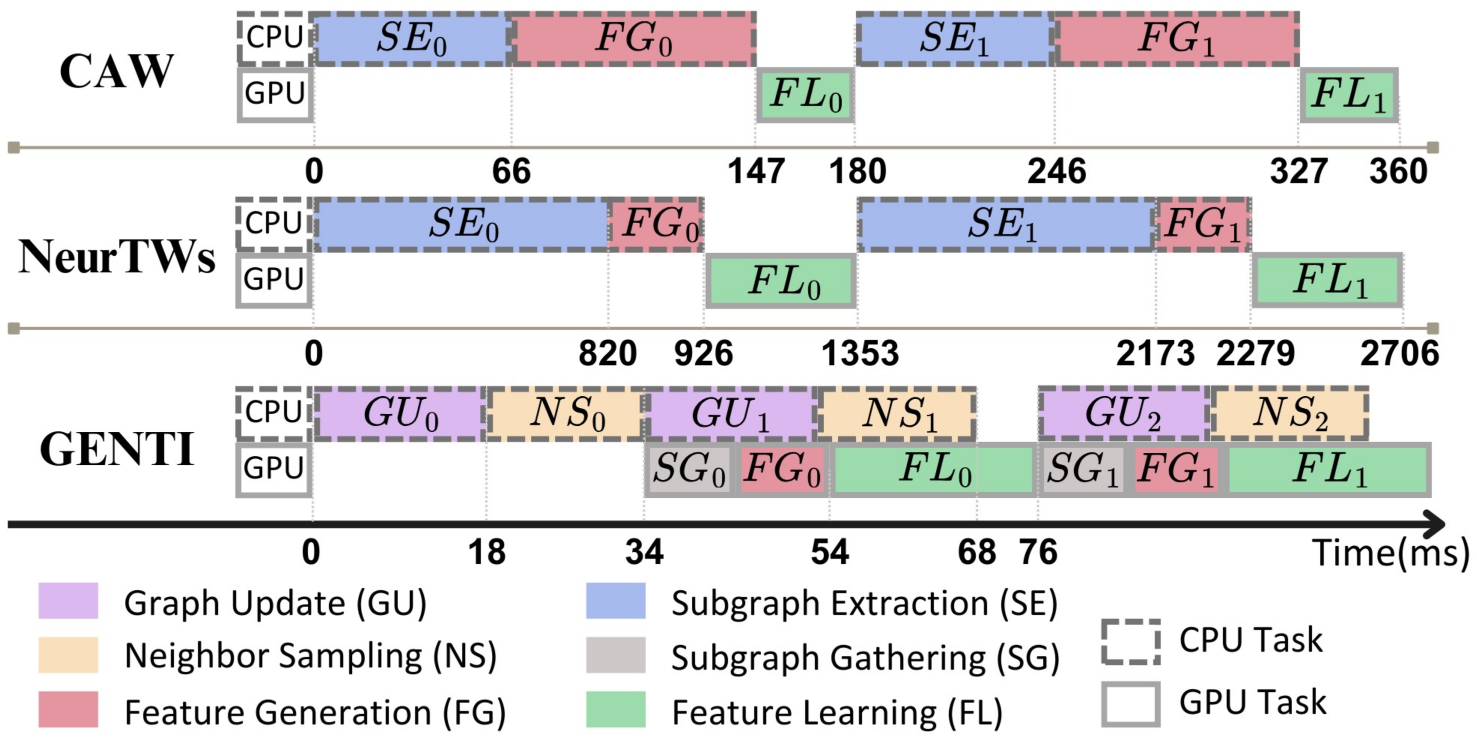
INTRODUCTION

Subgraph-based Graph Representation Learning (SGRL) achieve strong performance on many graph understanding tasks due to its ability to capture complex graph motifs. However, existing SGRL methods ignore the dynamic property of real-world graphs, and the time-consuming subgraph extraction stage limits the scalability of SGRL.

Key Ideas:

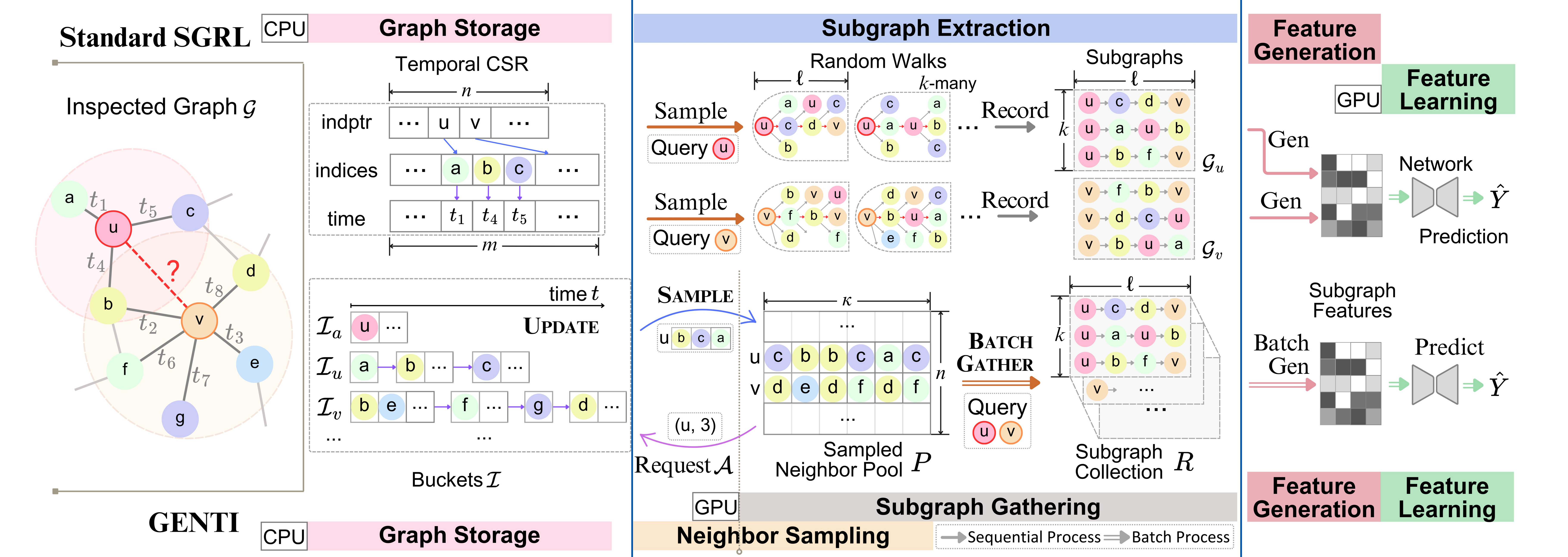
1. We decouple the resource-intensive subgraph extraction stage to be separately conducted on CPUs and GPUs, which enables full GPU utilization and offers improved efficiency.
2. We design the data structure for maintaining subgraphs on GPU with improved memory complexity and fast batch processing ability.
3. We upgrade the graph storage scheme on CPU to support frequent graph update and sampling.

RELATED WORK & Challenges



1. Imbalanced workloads and poor parallelism.
2. Moving the subgraph extraction stage to the GPU is challenging because large-scale graphs are hard to store on the GPU.
3. Online graph updates and sampling are not supported.

PROPOSED FRAMEWORK



① **Bucket-based data structure** supports streaming graph updates and neighbor sampling, while existing methods that rely on temporal CSR, which cannot accommodate streaming updates.

② **Sampled Neighbor Pool (SNP)** is an N×κ ($\kappa = O(k\sqrt{l})$) matrix stored on GPU, which is the key design to decouple the subgraph extraction into two parallel stages.

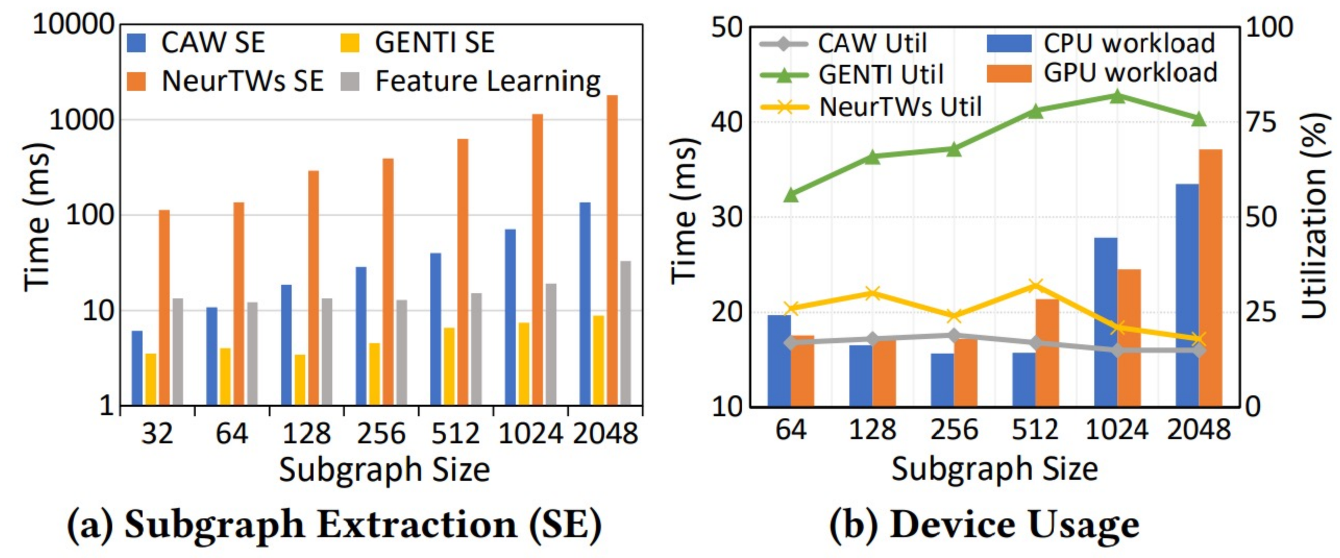
③ **Batch Processing:** Thanks to the SNP, the subsequent stages can be fully executed on GPU, maximizing its computational power.

EXPERIMENTS

Model	UCI-MSG Trans AP	UCI-MSG Induct AP	UCI-MSG Time (#)	Wikipedia Trans AP	Wikipedia Induct AP	Wikipedia Time (#)	Reddit Trans AP	Reddit Induct AP	Reddit Time (#)
JODIE	80.27 ± 0.1	71.64 ± 0.6	431(12)	95.16 ± 0.4	93.13 ± 0.5	1985(18)	95.83 ± 0.3	93.20 ± 0.4	8320.4(12)
TGAT	60.25 ± 0.3	75.27 ± 2.3	689(25)	94.26 ± 0.1	92.88 ± 0.3	2428(29)	97.80 ± 0.2	96.08 ± 0.3	11138(24)
TGN	78.91 ± 0.1	75.47 ± 0.1	507(21)	98.58 ± 0.1	98.05 ± 0.1	1839(26)	98.66 ± 0.1	97.55 ± 0.1	8152(26)
APAN	84.02 ± 0.3	83.14 ± 0.5	266(25)	96.41 ± 0.5	96.06 ± 0.4	1352(21)	98.50 ± 0.2	97.62 ± 0.7	7728(9)
Zebra	92.74 ± 0.2	91.16 ± 0.3	483(31)	98.63 ± 0.1	98.65 ± 0.1	1329(32)	98.73 ± 0.1	98.42 ± 0.1	6207(25)
D-DGNN	90.41 ± 0.1	89.72 ± 0.1	14467(30)	99.16 ± 0.3	98.54 ± 0.2	15173(30)	98.93 ± 0.2	98.56 ± 0.1	50342(30)
CAW	95.33 ± 0.3	95.19 ± 0.2	1488(8)	99.18 ± 0.1	99.34 ± 0.1	3720(5)	98.80 ± 0.1	98.99 ± 0.1	30912(8)
NeurTWs	95.46 ± 0.3	95.70 ± 0.2	44064(12)	99.17 ± 0.1	99.32 ± 0.1	65448(9)	98.32 ± 0.2	98.05 ± 0.1	TLE
GENTI	95.36 ± 0.3	**95.82 ± 0.3**	394(8)	**99.18 ± 0.1**	**99.37 ± 0.1**	**739(8)**	98.87 ± 0.1	**99.18 ± 0.1**	**5890(8)**

Model	SuperUser Trans AP	SuperUser Induct AP	SuperUser Time (#)	Wiki-Talk Trans AP	Wiki-Talk Induct AP	Wiki-Talk Time (#)	Tgbl-Comment Trans AP	Tgbl-Comment Induct AP	Tgbl-Comment Time (#)	MAG Trans AP	MAG Induct AP	MAG Time (#)
APAN	89.63 ± 0.5	86.32 ± 0.4	5.18(16)	TLE	TLE	TLE	TLE	TLE	TLE	TLE	TLE	TLE
Zebra	93.34 ± 0.3	97.63 ± 0.2	0.92(21)	95.25 ± 0.1	97.56 ± 0.3	4.65(18)	91.32 ± 0.5	95.15 ± 0.3	18.4(10)	OOM	OOM	OOM
CAW	93.12 ± 0.2	97.36 ± 0.4	10.9(5)	95.37 ± 0.1	98.07 ± 0.3	73.3(10)	TLE	TLE	TLE	TLE	TLE	TLE
GENTI	**94.05 ± 0.2**	**98.27 ± 0.3**	**0.66(8)**	**95.53 ± 0.1**	**98.58 ± 0.2**	**3.70(8)**	**93.51 ± 0.3**	**95.61 ± 0.2**	**12.3(6)**	**94.88 ± 0.4**	**99.45 ± 0.1**	**192(10)**

- Achieves comparable prediction accuracy with 3 ~ 26× faster overall learning time compared to state-of-the-art walk-based methods like CAW [1] and NeurTWs [2].



(a) Subgraph Extraction (SE) (b) Device Usage

- Achieves up to 30× faster subgraph extraction with balanced and concurrent workloads, even as subgraph size increases.
- Average device utilization reaches 80% when extracting large subgraphs in batches.

REFERENCES

[1] Yanbang Wang, Yen-Yu Chang, Yunyu Liu, Jure Leskovec, and Pan Li. 2021. Inductive representation learning in temporal networks via causal anonymous walks. International Conference on Learning Representations (2021).

[2] Ming Jin, Yuan-Fang Li, and Shirui Pan. 2022. Neural temporal walks: Motif-aware representation learning on continuous-time dynamic graphs. Advances in Neural Information Processing Systems 35 (2022), 19874–19886.

ACKNOWLEDGMENTS

This research is supported by Singapore MOE AcRF Tier-2 funding (MOE-T2EP20122-0003) and NTU NAP startup grant (022029-00001). Ningyi Liao is supported by the Joint NTU-WeBank Research Centre on FinTech, Nanyang Technological University, Singapore.

* Both authors contributed equally to this research.