Preserving Spatial-Temporal Relationship with Adaptive Node Sampling in Hierarchical Dynamic Graph Transformers

LINH THI HOANG[1], TUAN DUNG PHAM[2], SON T. MAI[2], VIET CUONG TA[3]
1. Singapore Management University, 2. Queen's University Belfast, 3. VNU – University of Engineering and Technology
Contact information: htlinh83@gmail.com

MOTIVATIONS

- Dynamic Graph Transformers (DGTs) have demonstrated remarkable performance in various applications, such as social networks, traffic forecasting, and recommendation systems.
- Training DGTs for large graphs remains a challenge. Mini-batch training is usually used to alleviate this challenge but this approach often fails to capture complex dependencies or sacrifice performance.
- We propose the **A**daptive **N**ode **S**ampling in **H**ierarchical **D**ynamic **G**raph **T**ransformers (ASH-DGT) architecture that focuses on sampling the set of suitable nodes preserving spatial-temporal relationships in the dynamic graph for training DGTs.
- Unlike previous methods that use random sampling or structural sampling, our motivation is that the contribution of nodes to learning performance can be time-sensitive, while we still care about spatial correlation in the dynamic graph with consideration to the global and local structure of the graph.

METHOD

$$\text{Maximize} \quad R(\mathcal{T}) = \sum_{t\in\mathcal{T}, v\in\mathcal{V}} R_v^t$$

$$\text{Subject to} \quad \sum_{i\in\mathcal{N}(v)} p_i = 1, \qquad 0 < \gamma < 1$$

$$R_v^t = \frac{1}{N} \sum_{v\in\mathcal{S}(v)} A_v^t \cdot ||h_v^t||$$

Algorithm 1: Extent EXP3 Algorithm for node-wise sampling

Input : K: number of chosen nodes, η: number of neighbor nodes, γ: exploration rate
Output: : p: policy distribution
Initialize $w \leftarrow (1, 1, \dots, 1)$;
for *each iteration* **do**
  Compute $p = (p_1, p_2, \dots, p_\eta)$ with

$$p_i = (1-\gamma) \cdot \frac{w_i}{\sum_{i=1}^{\eta} w_i} + \frac{\gamma}{\eta}$$

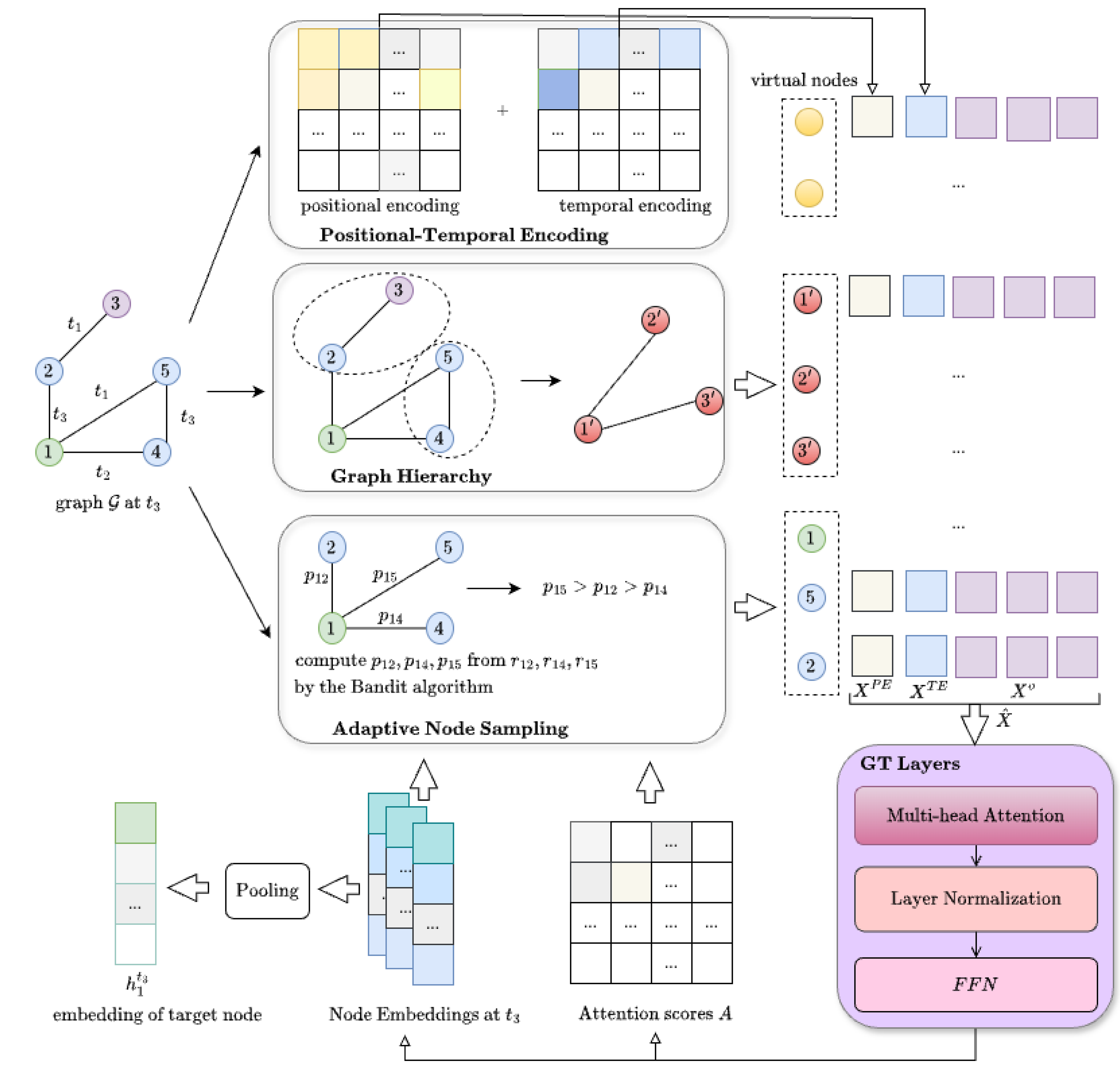
  Choose K nodes based on p; Compute reward r following eq.5,
  Update the weights using the EXP3 update:

$$w_i = w_i \cdot e^{-\eta \cdot \frac{r}{K \cdot p_i}}$$

end

Algorithm 2: Adaptive Sampling Iterative Process

Input : $G = (V, E, T)$: dynamic graph, H_0: initial node embeddings, t: maximum of time steps
Output: H^t: the updated node embedding
Initialize: $i \leftarrow 1$, $H \leftarrow H_0$;
while $i \leq numEpoch$ **do**
  for *each node* v *in* V **do**
    Train the policy p following the Algorithm 1; Sample neighbor nodes from p;
  end
  Training DGT;
  Updated node embeddings H^t for time step t; $i \leftarrow i + 1$
end



EXPERIMENTS & RESULTS

Dataset

Dataset	Wikipedia	Reddit	SocialEvo	ENRON	UCI	MOOC
# Nodes	9227	10984	74	184	1899	7145
# Edges	157474	672447	352180	2099520	59835	411749
# Edge feature dim	172	172	0	0	0	4
# Timespan	30 days	30 days	30 days	1316 days	193days	30 days

We split the data following the baseline TGN [1] with ratio of 70%/15%/15% for training, validation and test set. We compare our method with baselines on the node classification and link prediction tasks.

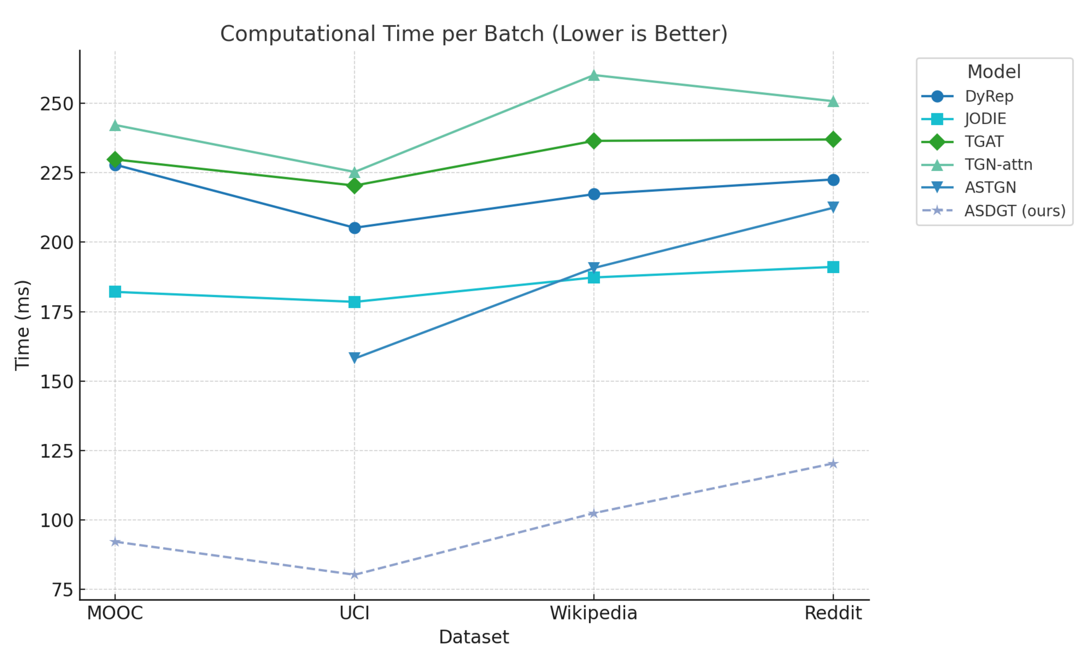
Node Classification Task

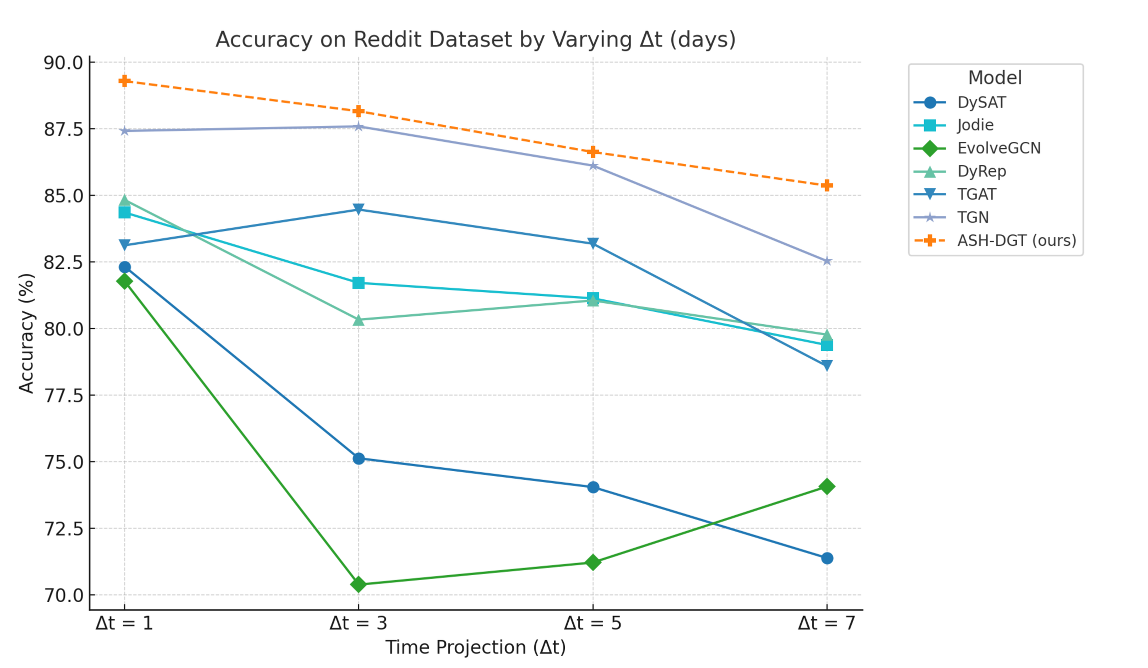
Model	MOOC	Wikipedia	Reddit
DySAT	72.11 ± 0.5	61.79 ± 0.3	74.82 ± 1.2
Jodie	73.39 ± 2.1	61.23 ± 2.5	84.35 ± 1.2
EvolveGCN	70.26 ± 0.5	63.41 ± 0.3	81.77 ± 1.2
TGAT	74.23 ± 1.2	65.43 ± 0.7	83.12 ± 0.7
DyRep	75.12 ± 0.7	62.79 ± 2.3	84.82 ± 2.2
TGN	77.47 ± 0.8	67.11 ± 0.9	87.41 ± 0.3
ASH-DGT (ours)	$\mathbf{79.08 \pm 0.5}$	$\mathbf{69.74 \pm 1.3}$	$\mathbf{89.12 \pm 0.3}$

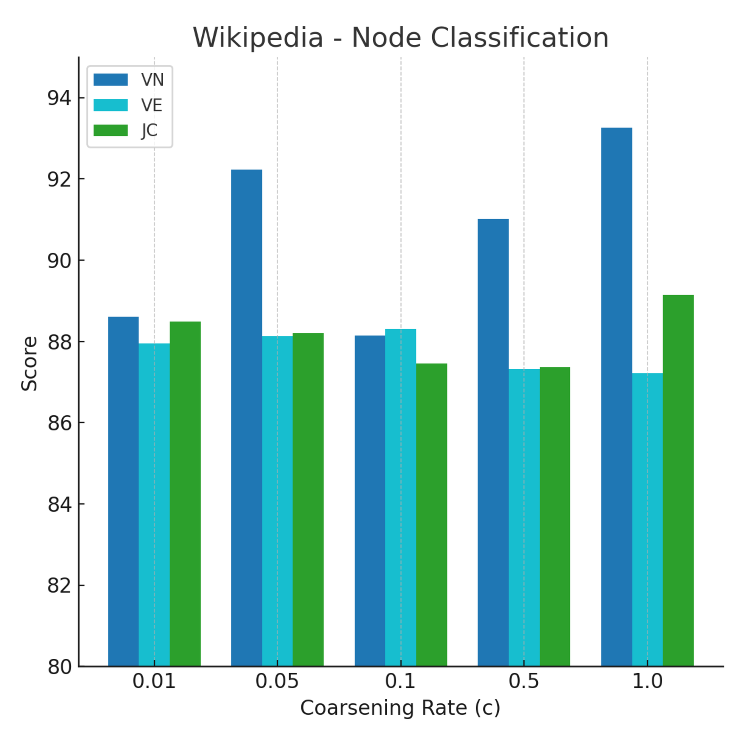
Link Prediction Task

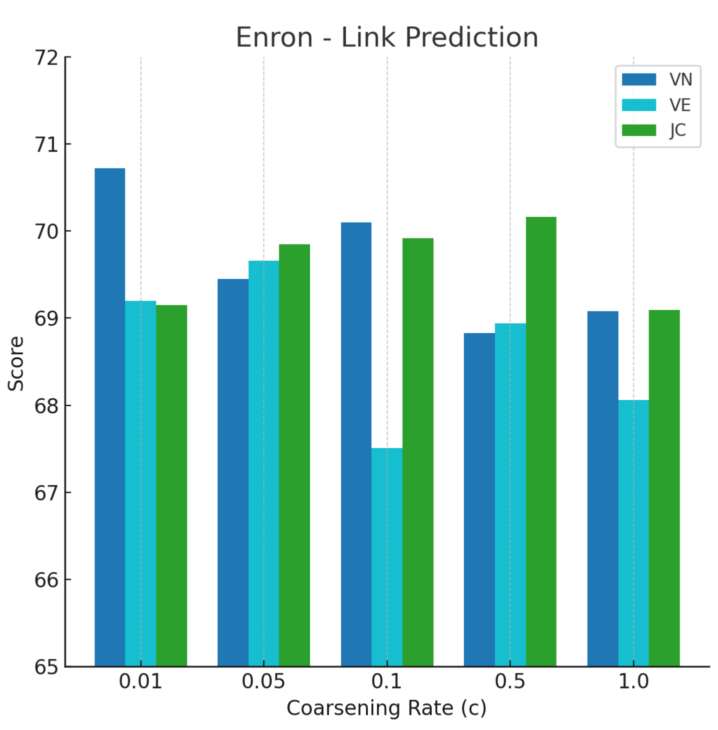
Dataset	Wikipedia		Reddit		MOOC	
Metric	AUC (%) ↑	AP (%) ↑	AUC (%) ↑	AP (%) ↑	AUC (%) ↑	AP (%) ↑
DyRep	77.40 ± 1.2	75.20 ± 1.1	67.36 ± 1.1	66.12 ± 1.0	90.49 ± 0.3	89.23 ± 0.4
JODIE	88.43 ± 0.7	87.95 ± 0.6	87.71 ± 0.4	87.43 ± 0.5	90.50 ± 0.1	90.20 ± 0.2
VGRNN	72.20 ± 0.6	71.00 ± 0.7	51.89 ± 0.2	50.56 ± 0.3	90.03 ± 0.4	89.00 ± 0.3
EvolveGCN	60.48 ± 0.5	58.74 ± 0.4	58.42 ± 0.4	57.33 ± 0.3	50.36 ± 0.8	49.76 ± 0.6
DynAERNN	71.00 ± 1.1	70.12 ± 1.0	83.37 ± 1.4	82.54 ± 1.3	89.34 ± 0.2	88.23 ± 0.3
TGAT	95.47 ± 0.1	94.89 ± 0.2	96.65 ± 0.1	95.98 ± 0.1	72.09 ± 0.3	70.95 ± 0.3
TGN-attn	95.72 ± 2.1	94.67 ± 1.9	96.12 ± 1.1	95.65 ± 1.0	81.64 ± 1.3	80.23 ± 1.1
ASTGN	96.92 ± 0.2	95.87 ± 0.3	98.97 ± 0.1	98.45 ± 0.1	92.75 ± 0.8	91.45 ± 0.7
ASH-DGT (ours)	$\mathbf{97.45 \pm 0.3}$	$\mathbf{96.50 \pm 0.2}$	$\mathbf{99.02 \pm 0.1}$	$\mathbf{98.76 \pm 0.1}$	$\mathbf{93.23 \pm 0.4}$	$\mathbf{92.10 \pm 0.3}$

Dataset	UCI		SocialEvo		Enron	
Metric	AP (%) ↑	AUC (%) ↑	AP (%) ↑	AUC (%) ↑	AP (%) ↑	AUC (%) ↑
DyRep	79.76 ± 0.1	81.02 ± 0.3	62.02 ± 1.7	64.35 ± 1.1	67.28 ± 1.3	68.55 ± 1.4
JODIE	78.02 ± 0.2	79.53 ± 0.5	60.01 ± 1.1	61.56 ± 1.2	63.10 ± 1.3	64.72 ± 1.0
EvolveGCN	76.63 ± 0.2	78.12 ± 0.4	56.90 ± 0.6	57.89 ± 0.5	57.37 ± 0.2	58.62 ± 0.3
TGAT	60.25 ± 0.2	61.32 ± 0.6	57.37 ± 0.6	58.91 ± 0.5	60.36 ± 0.7	61.67 ± 0.9
TGN-attn	64.21 ± 0.2	65.55 ± 0.4	58.11 ± 1.2	59.74 ± 0.8	62.47 ± 1.7	63.53 ± 1.3
ASTGN	81.52 ± 0.2	82.10 ± 0.3	62.41 ± 1.0	63.65 ± 1.2	69.12 ± 1.0	70.56 ± 0.9
ASH-DGT (ours)	$\mathbf{82.34 \pm 0.2}$	$\mathbf{83.21 \pm 0.2}$	$\mathbf{65.12 \pm 0.3}$	$\mathbf{66.45 \pm 0.4}$	$\mathbf{71.33 \pm 0.5}$	$\mathbf{72.67 \pm 0.4}$









CONCLUSION

We introduce an adaptive node sampling technique to preserve spatial-temporal relationships in hierarchical dynamic graph transformers. Through experiments, our approach demonstrates enhanced representation learning, leading to improved model performance and accuracy.

References:
[1] Emanuele Rossi et al, Temporal graph networks for for deep learning on dynamic graphs.

To see the full version of paper, you can reach out by scan the QR code beside.

Acknowledgements

This research was funded by the research project QG.23.32 of Vietnam National University, Hanoi. This project is also supported by QUB Agility+ funding, DfE project (R3648ECS) and UKRI/EU funded project RELAX (R1142ECS).