University of Pittsburgh

MemFreezing: A Novel Adversarial Attack on Temporal Graph Neural Networks under Limited Future Knowledge

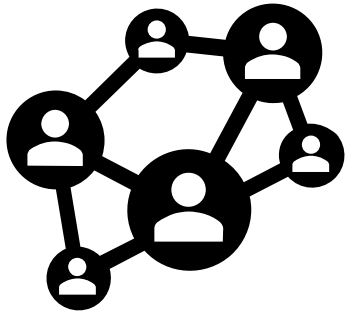
Yue Dai, Liang Liu*, Xulong Tang, Youtao Zhang, Jun Yang*

University of Pittsburgh

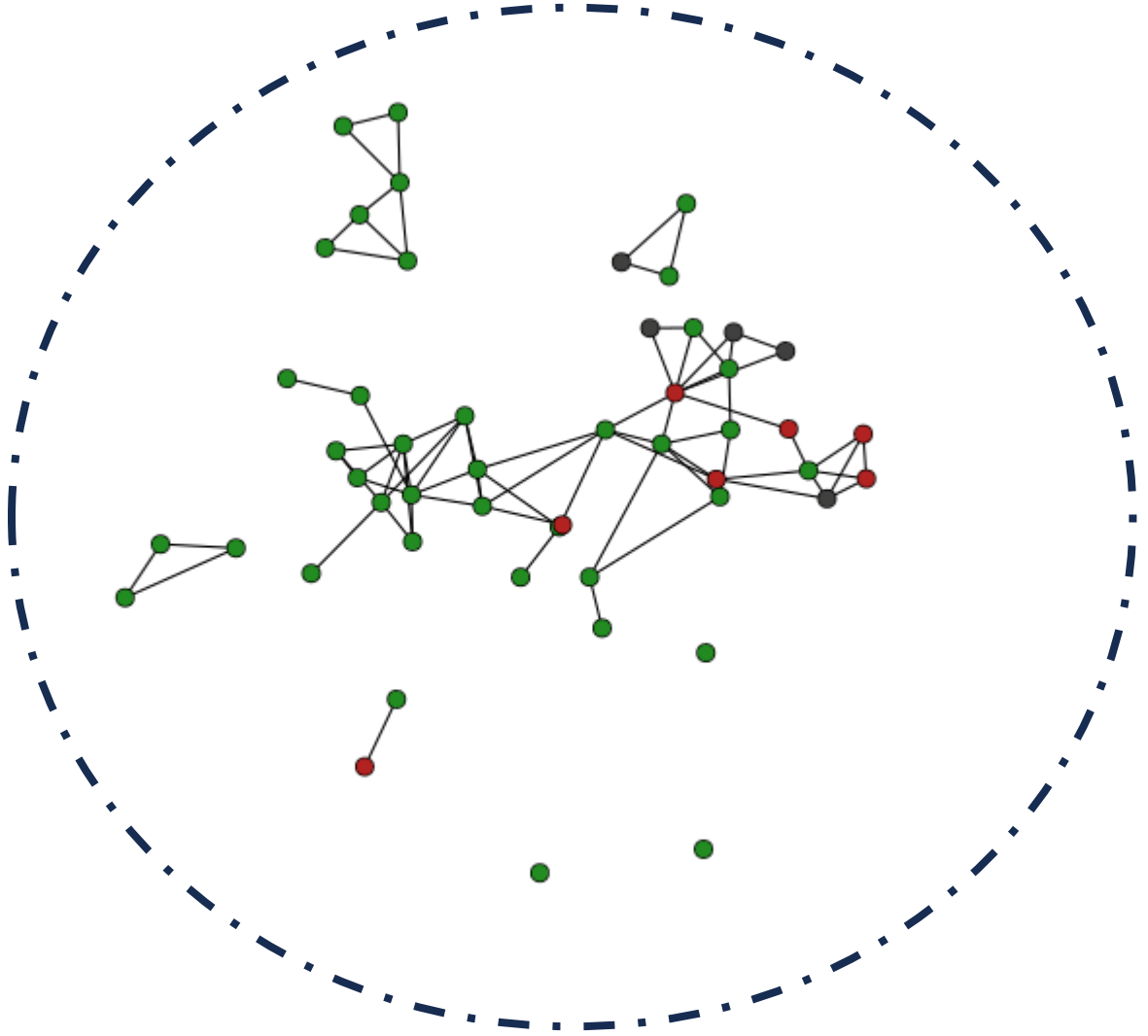
Forty-second International Conference on Machine Learning(ICML-2025)

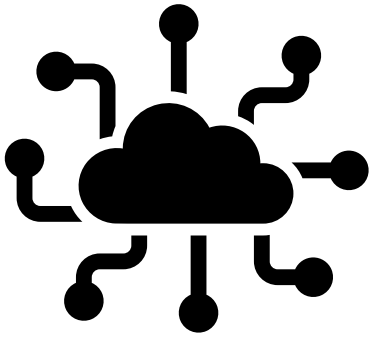
Dynamic Graphs in Real-World Applications

E-Commerce:
Changing Shopping History



Social Media:
Changing Relationship





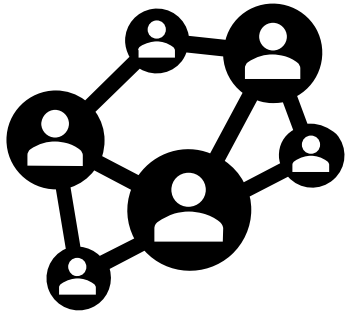
Internet-of-things:
Changing connectivity

Navigating:
Changing traffic

University of Pittsburgh

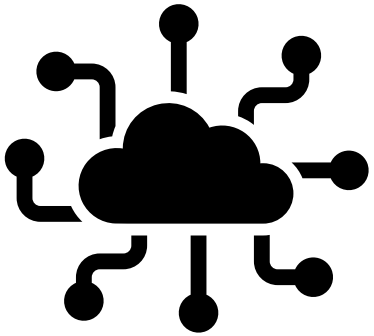
Dynamic Graphs in Real-World Applications

E-Commerce:
Changing Shopping History



Social Media:
Changing Relationship

Temporal Graph Neural Networks



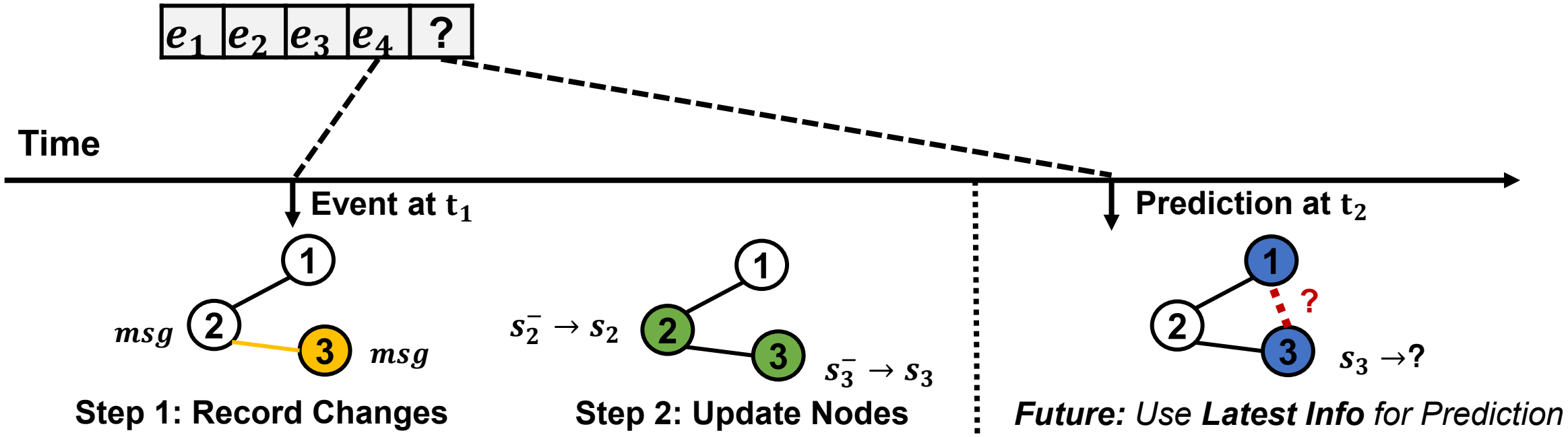
Internet-of-things:
Changing connectivity

Navigating:
Changing traffic

University of Pittsburgh

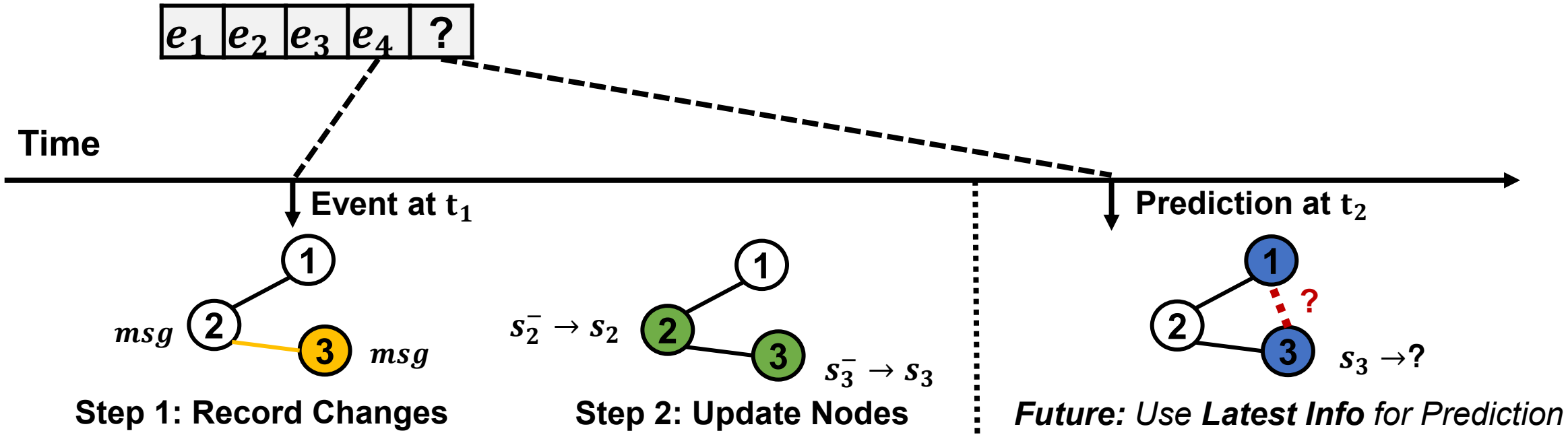
Temporal Graph Neural Networks

TGNN Record Events (changes) and **Update** Affected Nodes (memories).



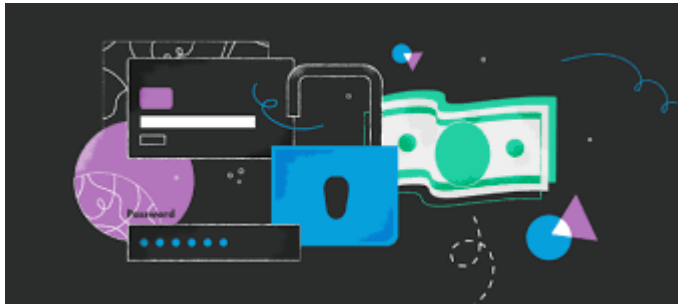
Temporal Graph Neural Networks

TGNN Record Events (changes) and **Update** Affected Nodes (memories).

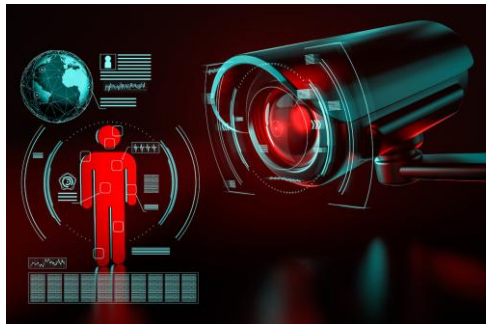


ADVANTAGE over static GNNs:
can capture temporal history information---
achieving **SOTA in dynamic graph tasks**

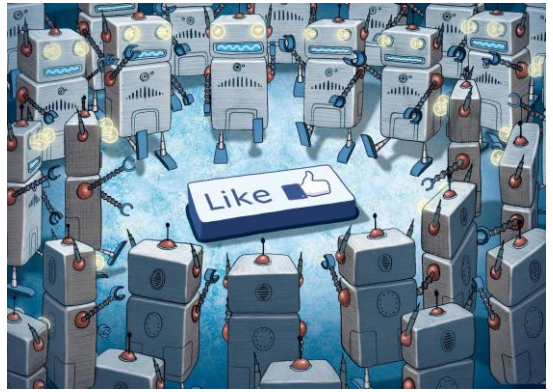
Robustness is A General Concern



Fraud Detection

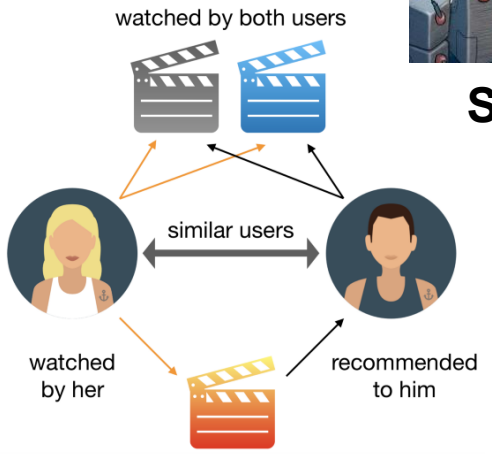


Intrusion Detection



Spam Bot Filtering

Drug Discovery



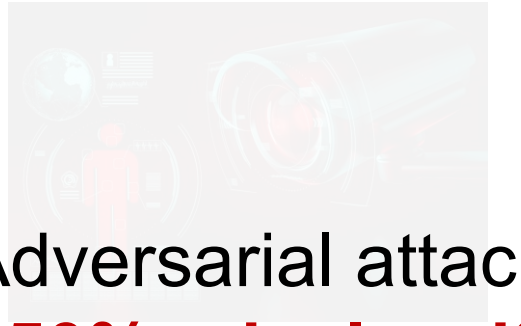
Personal Recommendation

Severe financial or safety losses

Privacy Concerns

Robustness is A General Concern

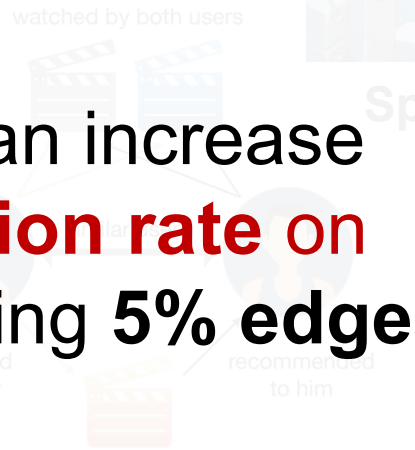
Fraud Detection



Intrusion Detection

Spam Bot Filtering

Drug Discovery



Personal Recommendation

Adversarial attacks can increase **~50% misclassification rate** on static GNNs by affecting **5% edges**.

Severe financial or safety losses

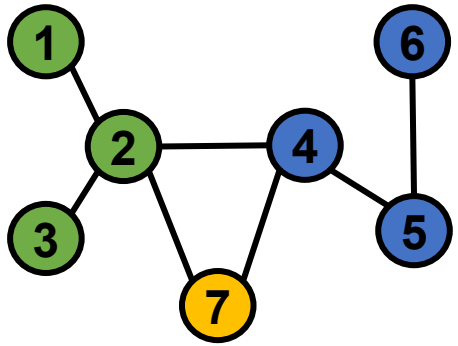
Privacy Concerns

Robustness is A General Concern

What happened when adversarial attacks meet dynamic graphs?

Adversarial Attacks on Graph Learning Models

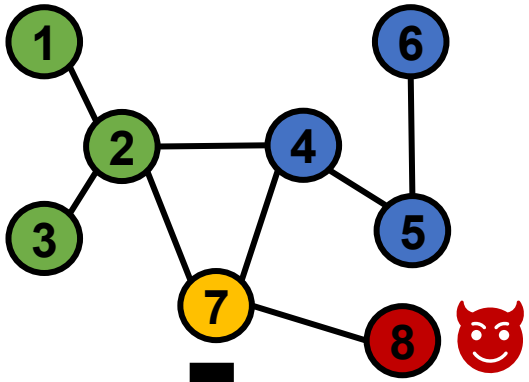
Original:



GNN

User 7 is suspicious!
Do not trust!

Adversarial:



GNN

User 7 is valid and trustworthy!

Adversarial Attacks on Graph Learning Models

Adversarial Attacks on TGNNs can be very **different and unexpected**

Attacker's Capacity:
Modify a small set of graph

~~**Attacker's Knowledge:**
Attacker can observe entire input~~

Static GNN

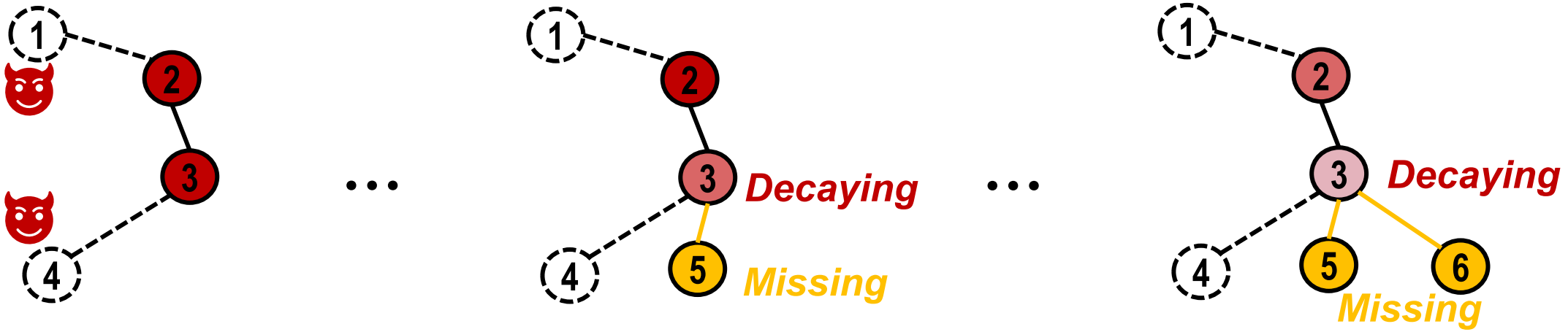
Attacker's Capacity:
Modify a small set of graph

Attacker's Knowledge:
Attacker can observe up-to-attack input

TGNN

University of Pittsburgh

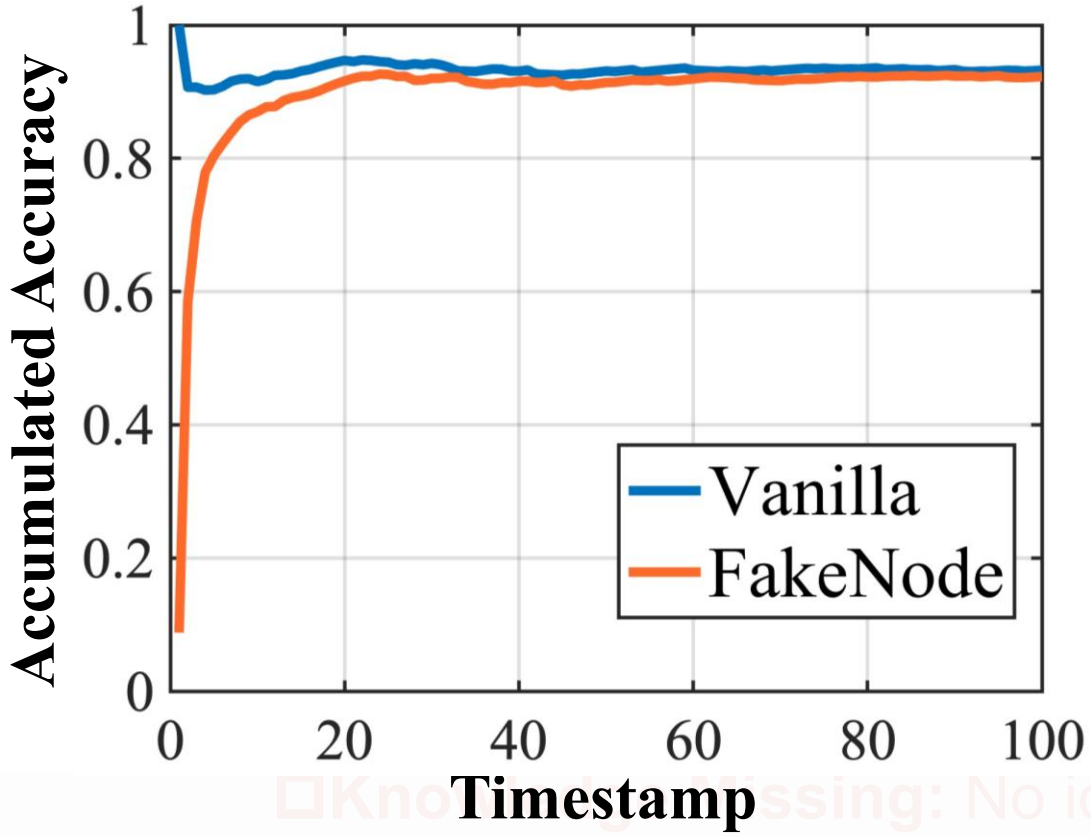
Challenge: Limited Knowledge due to Changing Graphs



Graph Continue evolving after attack

- **Noise Decaying:** Future changes can dilute noises.
- **Knowledge Missing:** No idea about unseen edges/nodes.

Challenge: Limited Knowledge due to Changing Graphs

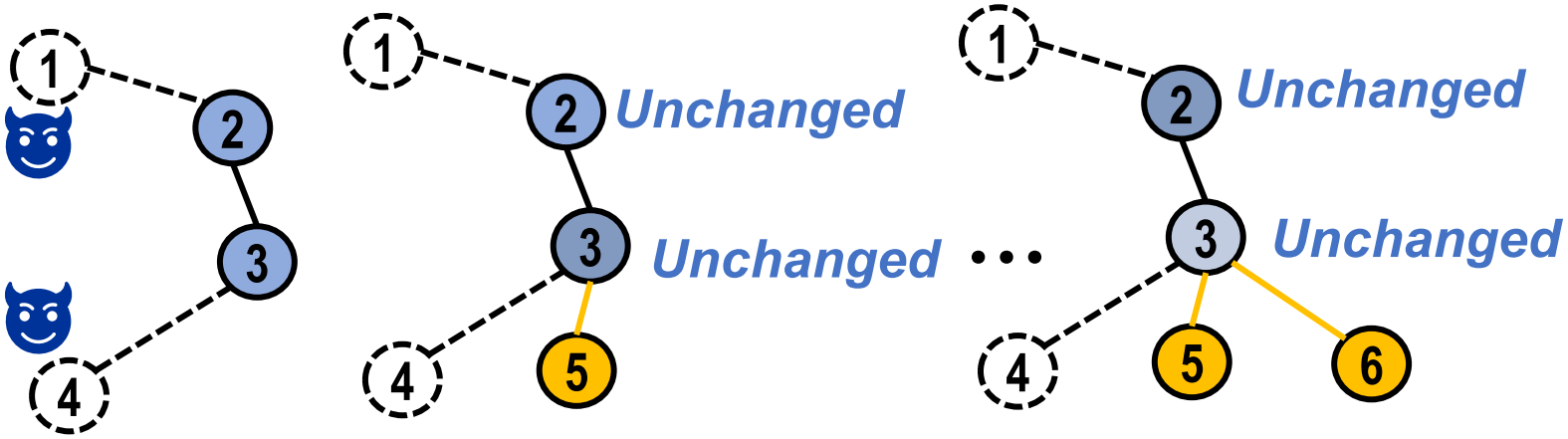


- TGNNs soon recover after the attack time (timestamp=0)!
- Impossible to solve noises maximizing unknown losses.

Alternative Objective: Memory Freezing

Fact: Stale information may hurt accuracy.

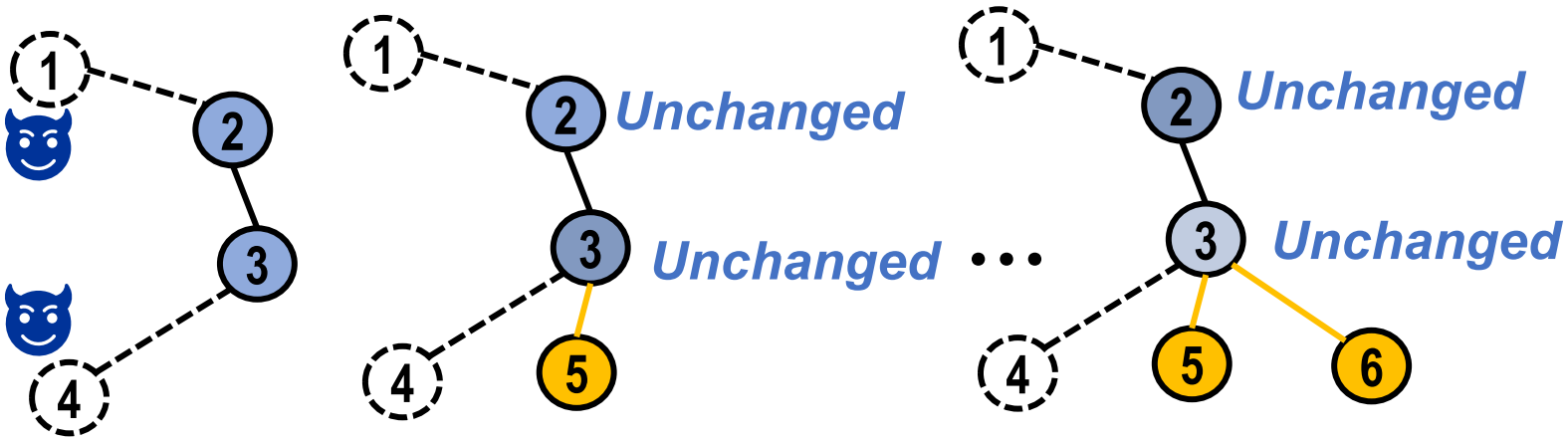
Q: What if node memory **no longer changes?**

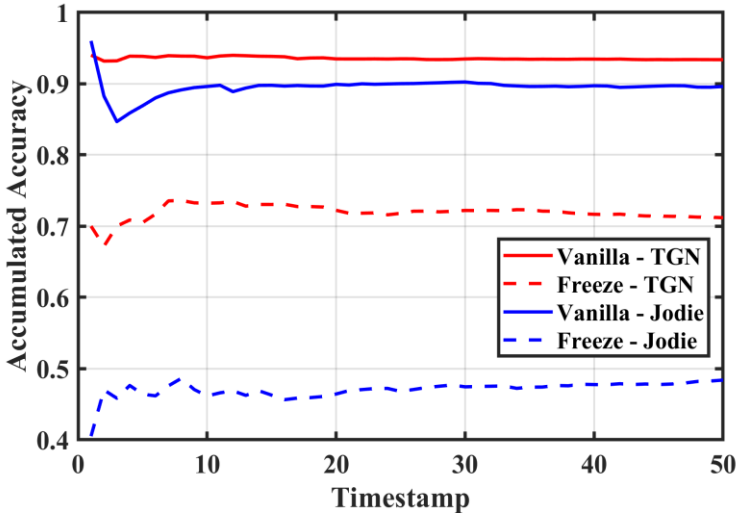


Alternative Objective: Memory Freezing

Fact: Stale information may hurt accuracy.

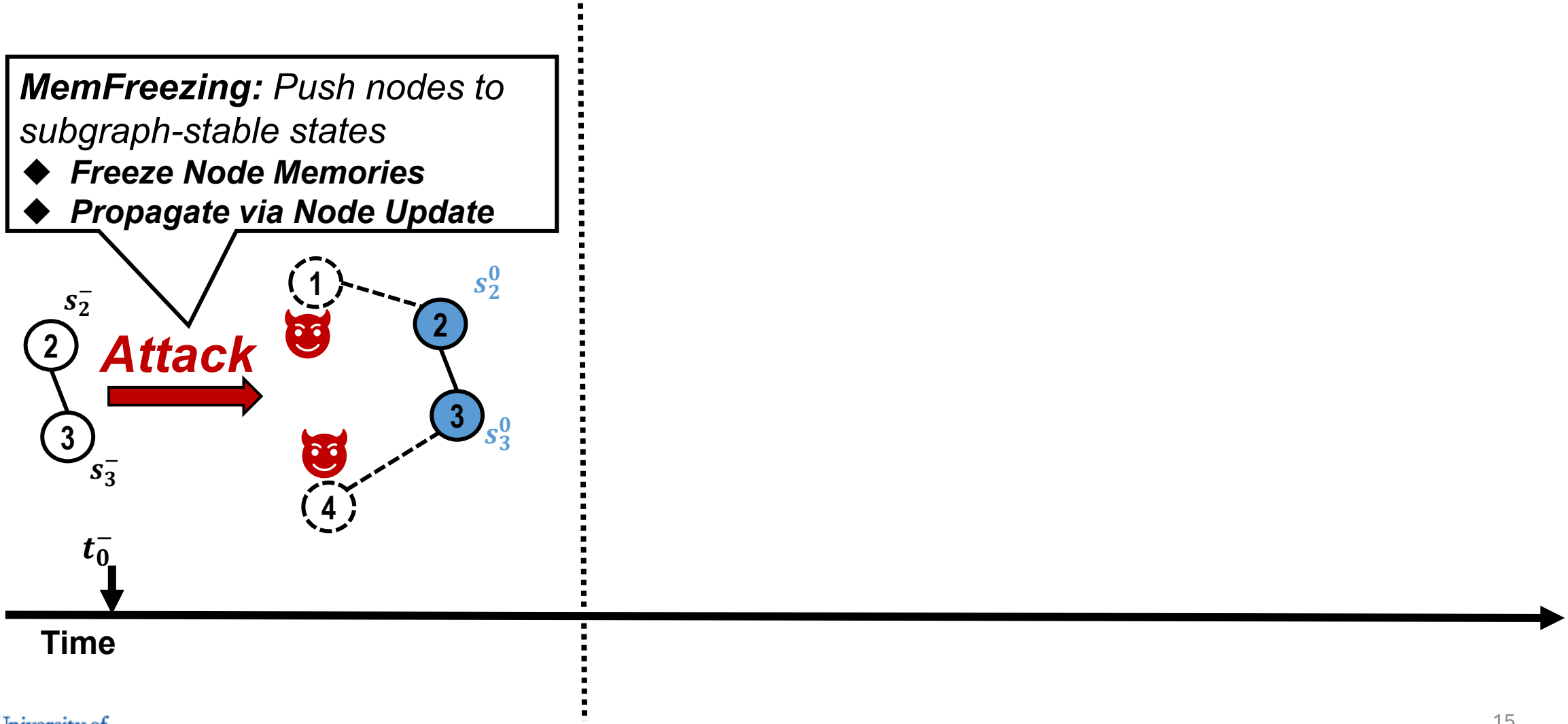
Q: What if node memory **no longer changes?**



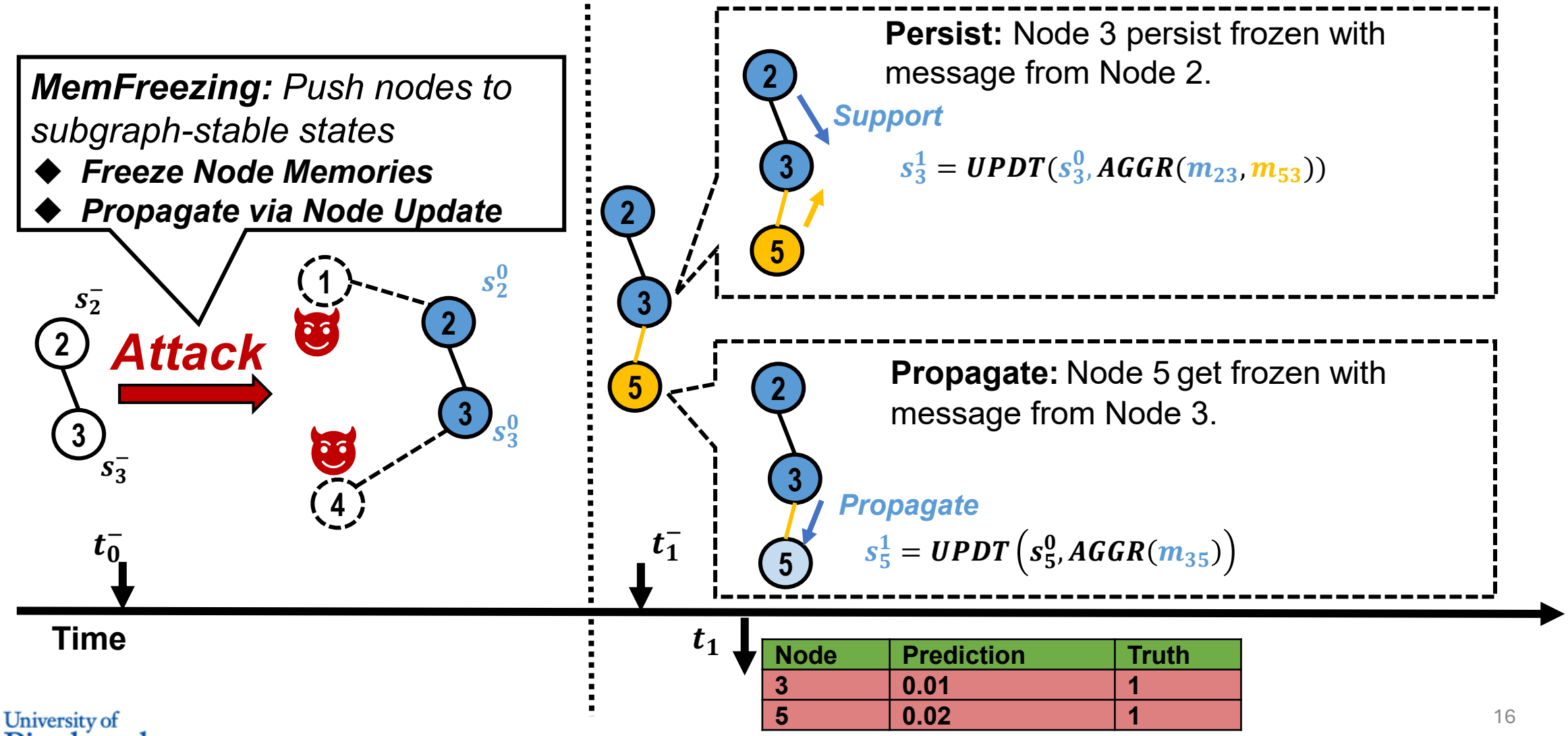


A: Leading to >30% Accuracy Drop!

Memfreezing: Mislead TGNNs by Preventing Node Updates



Memfreezing: Mislead TGNNs by Preventing Node Updates



Node	Prediction	Truth
3	0.01	1
5	0.02	1

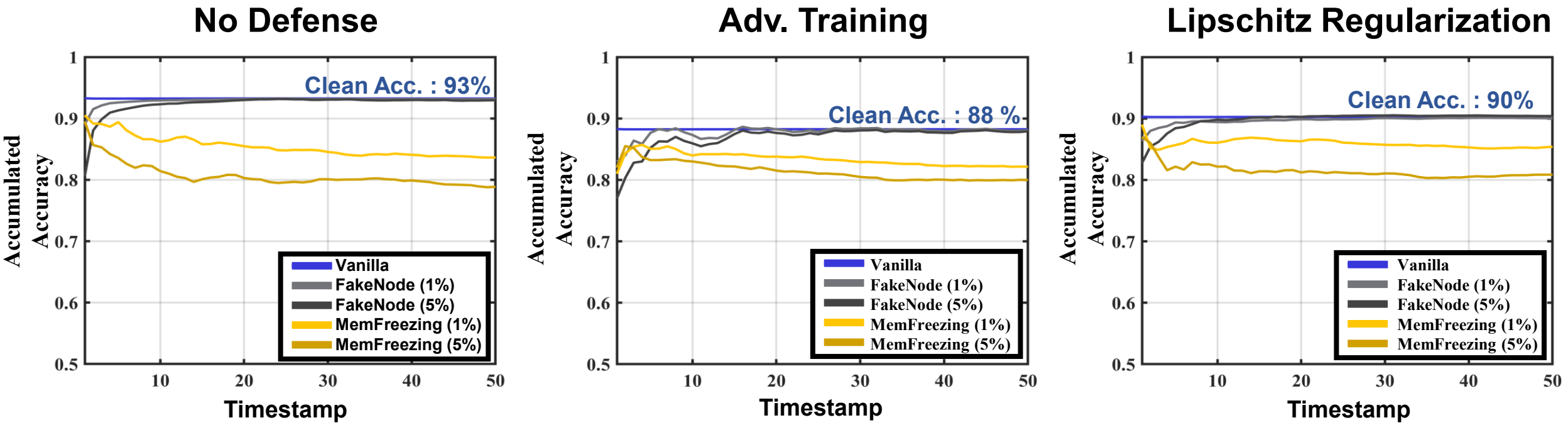
Memfreezing persistently mislead TGNNs

Dataset		WIKI				REDDIT				REDDIT-BODY			
Model		TGN	JODIE	Dyrep	Roland	TGN	JODIE	Dyrep	Roland	TGN	JODIE	Dyrep	Roland
Vanilla		0.93	0.87	0.86	0.94	0.97	0.98	0.96	0.95	0.90	0.87	0.90	0.88
t_0	FN	0.81	0.74	0.74	0.82	0.84	0.83	0.84	0.83	0.76	0.82	0.77	0.79
	Meta-h	0.90	0.83	0.81	0.85	0.93	0.95	0.90	0.92	0.86	0.83	0.88	0.85
	TDGIA	**0.77**	**0.72**	**0.71**	**0.80**	**0.74**	**0.80**	**0.81**	**0.74**	**0.72**	**0.81**	**0.74**	**0.76**
	Ours	0.89	0.78	0.83	0.87	0.75	0.84	0.94	0.82	0.84	0.85	0.81	0.78
t_{25}	FN	0.92	0.87	0.85	0.94	0.97	0.97	0.96	0.93	0.90	0.86	0.89	0.88
	Meta-h	0.93	0.87	0.84	0.93	0.96	0.98	0.94	0.96	0.89	0.86	0.90	0.87
	TDGIA	0.93	0.81	0.84	0.92	0.94	0.95	0.95	0.90	0.89	0.85	0.89	0.88
	Ours	**0.80**	**0.75**	**0.77**	**0.85**	**0.81**	**0.84**	**0.91**	**0.80**	**0.81**	**0.84**	**0.76**	**0.80**
t_{50}	FN	0.94	0.87	0.86	0.94	0.97	0.97	0.96	0.95	0.90	0.86	0.90	0.88
	Meta-h	0.93	0.87	0.85	0.93	0.97	0.98	0.94	0.95	0.90	0.86	0.90	0.88
	TDGIA	0.93	0.87	0.85	0.93	0.96	0.97	0.95	0.92	0.89	0.86	0.90	0.87
	Ours	**0.75**	**0.76**	**0.75**	**0.84**	**0.80**	**0.84**	**0.91**	**0.80**	**0.77**	**0.82**	**0.76**	**0.77**

SOTA GNN **attacks soon diminish after attack**

Memfreezing lead to **>10% accuracy drop over time**

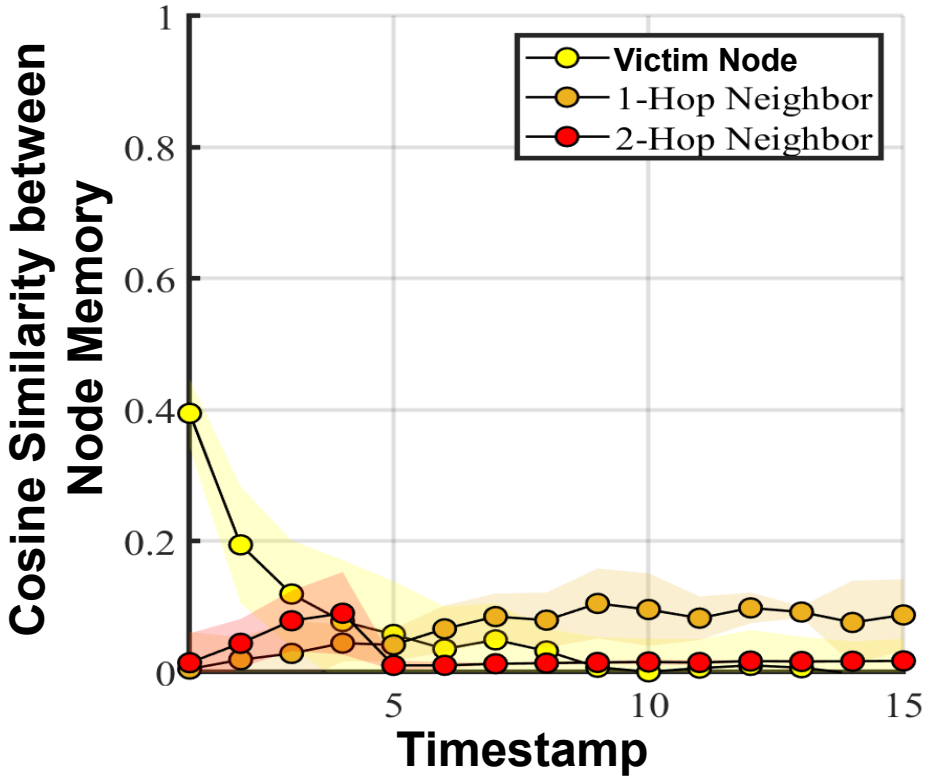
Memfreezing is effective under defenses

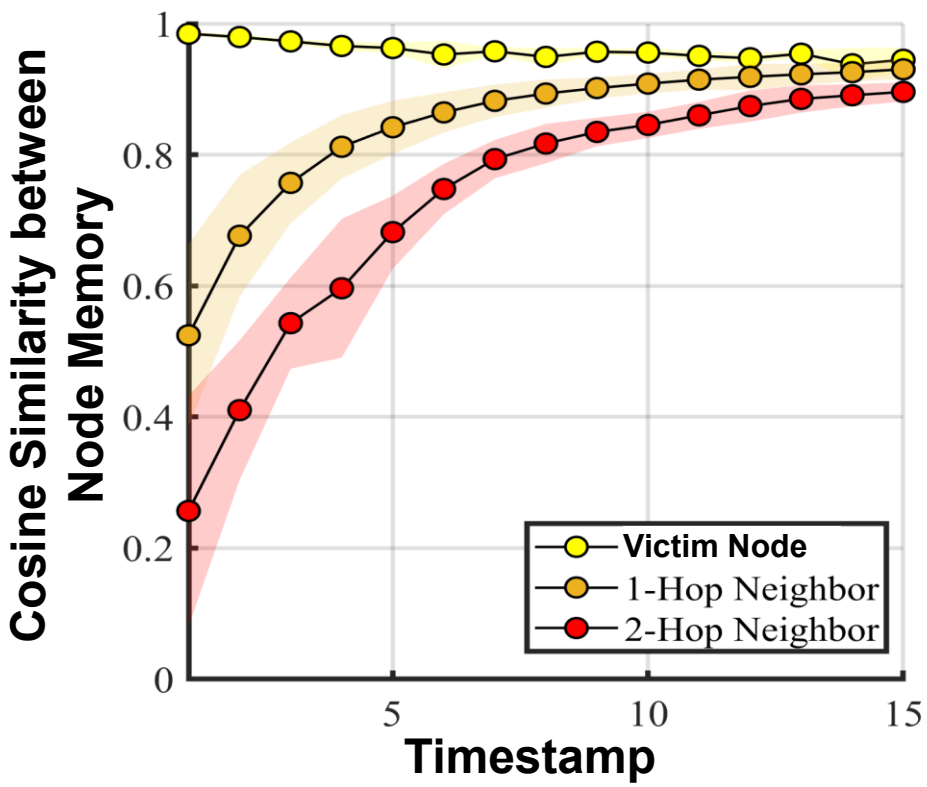


SOTA GNN **attacks soon diminish after attack** under defenses

Memfreezing lead to **~10% accuracy drop over time** under defenses

Memfreezing successfully freeze node memory





In vanilla TGNNs, victim nodes **change drastically**
Under Memfreezing attack, victim nodes **tend to stable**

More Results are Present in the Paper

- *Black-box experiments*
- *More ablation study*
- *Sensitivity study*
- *Stealthiness study*
- *Overhead*
- *Potential Defenses*

…

University of Pittsburgh

Thanks!

MemFreezing: A Novel Adversarial Attack on Temporal Graph Neural Networks under Limited Future Knowledge

For more questions: Yue Dai: yud42@pitt.edu