ChronoSpike: An Adaptive Spiking Graph Neural Network for Dynamic Graphs

Md Abrar Jahin[1] Taufikur Rahman Fuad[2] Jay Pujara[1] Craig A. Knoblock[1]

[1]University of Southern California [2]Islamic University of Technology

LoG CONFERENCE

Code & paper — scan me

THE PROBLEM: PICK TWO?

Graphs that change over time (who cites, buys, or talks to whom) need a model that sees **structure** *and* **history**. Today's options each give something up:

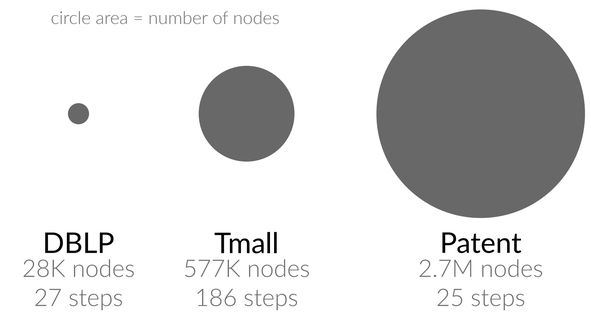
Recurrent	Attention	Spiking
JODIE, EvolveGCN	TGAT, TGN	SpikeNet, Dy-SIGN, Delay-DSGN
memory of the past fades	every step × every step	sparse 0/1 events
remembers step by step	expressive, long-range	cheap, event-driven
dense state per node gradients vanish	cost grows as T^2 and with neighbours	binary detail lost short-sighted, local

Can we keep the efficiency of spikes and still look far back in time?

ChronoSpike: yes, by letting each part do what it is best at.

The Task & Data

Temporal node classification: given T snapshots of a graph, predict each node's label.

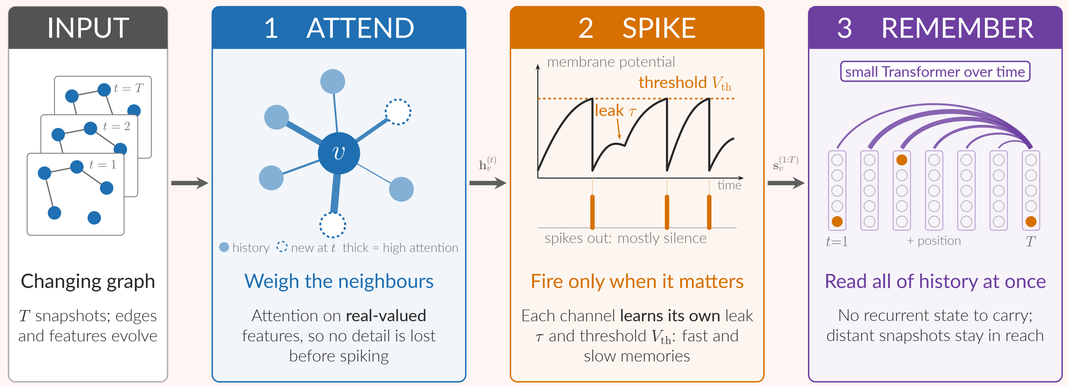


Compared against **12 baselines** (static, recurrent, attention, spiking) at 40 / 60 / 80% training labels, 5 runs each.

TAKEAWAYS

- (1) (2) (3) Attend on real values, spike adaptively, remember with a Transformer: **accuracy without giving up efficiency**
- **Limits**: discrete snapshots and full history required; hand-made contrastive augmentations
- **Next**: continuous-time / streaming graphs, neuromorphic hardware

HOW IT WORKS: ATTEND → SPIKE → REMEMBER



Output: embedding $\mathbf{z}_v \rightarrow$ classifier

Training: cross-entropy + $\lambda \cdot$ contrastive loss; surrogate gradients pass through spikes

The Result in Five Numbers

+2.4 Micro-F1 points over best prior model (avg.)

+2.0 Macro-F1 points over best prior model (avg.)

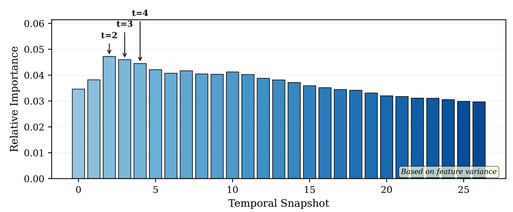
3–10× faster training than recurrent models

105K parameters, whether 28K or 2.7M nodes

83–88% of neurons silent at any moment

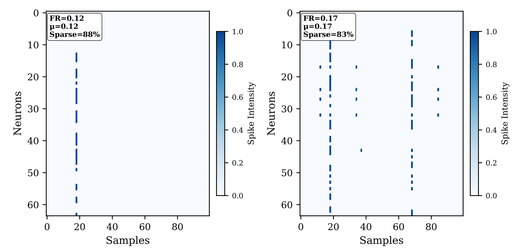
What Did It Learn? (DBLP, 1,000 test nodes)

It reaches back to the beginning



Attention peaks at **early snapshots** (t=2 to 4), then decays slowly: a **primacy effect**, not the recency bias of recurrent models.

It stays quiet until needed



Layer 1 (left) and layer 2 (right): **88%** and **83%** of neurons are silent. Sparse events keep compute and memory low.

Stable by Design: Three Guarantees



Potential stays bounded
for any learned leak $\tau_i > \frac{1}{2}$

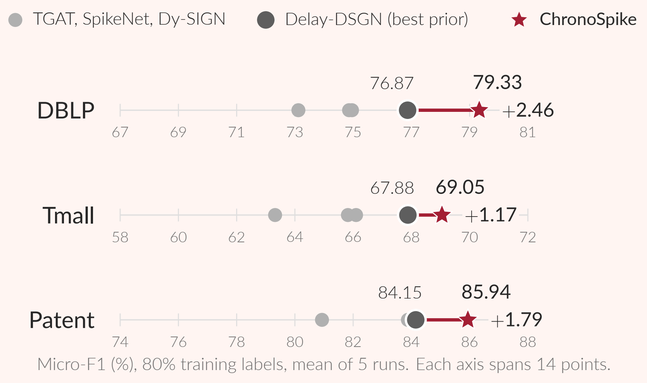


Gradients cannot explode
$\|\Delta(t)\|_1 \le \rho^{t'-t}\|\Delta(t')\|_1,\ \rho < 1$



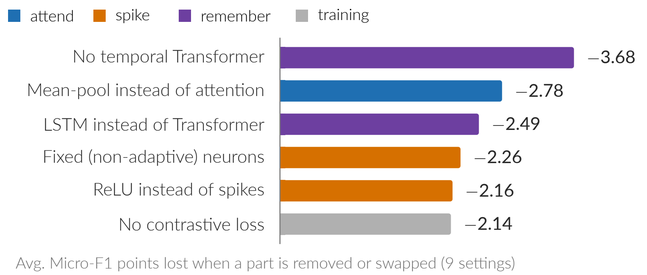
Bounded in ⇒ bounded out
BIBO-stable as a whole network

RESULTS: BEST ON EVERY DATASET



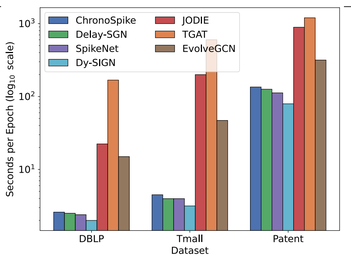
- Wins **all 18 settings** (3 datasets × 3 label ratios × 2 metrics)
- Biggest margin when labels are scarce: **+4.2** Micro-F1 on DBLP at 40%

Every Part Earns Its Place



The Transformer matters most; and ChronoSpike still beats *every* variant everywhere, so the three parts help each other.

Fast and Small



Seconds per epoch
(log scale, lower is better)

- On par with other spiking models
- Far below JODIE, TGAT, EvolveGCN
- 2.6 s / 4.5 s / 135 s per epoch