Rethinking the Memory Staleness Problem in Dynamic Graphs

- Based on TGN for DL on Dynamic Graphs
  (Rossi et al)*

Advanced Topics in Deep Learning 236605

Dekel Brav, Hadas Ben Atya, Mor Ventura

April 2022

* arxiv: 2006.10637

TECHNION
Israel Institute of Technology

Introduction

- Accurate representations of dynamic systems using graphs has to consist temporal evolution within them.
- Dynamic graphs:
  - Discrete-time dynamic graphs (DTDG)
  - Continuous-time dynamic graphs (CTDG)
- TGN – Temporal Graph Networks
  - Encoder on CTDG – seq. time stamped events $\longrightarrow$ time stamped embeddings
- Memory staleness problem

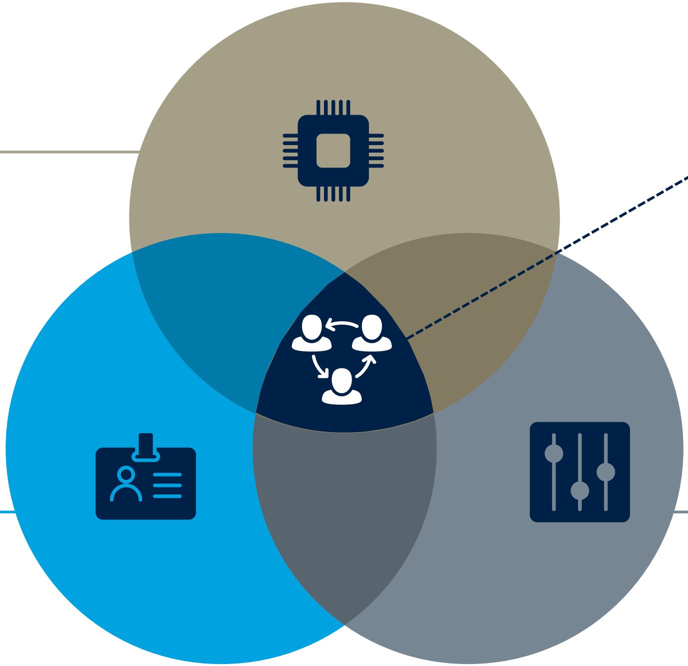
TGN – Core modules

Memory

State of the model – updated after an event. It represents the node's history

Message

Is computed foe each event involving source node and target node at time t



TGN

Embedding

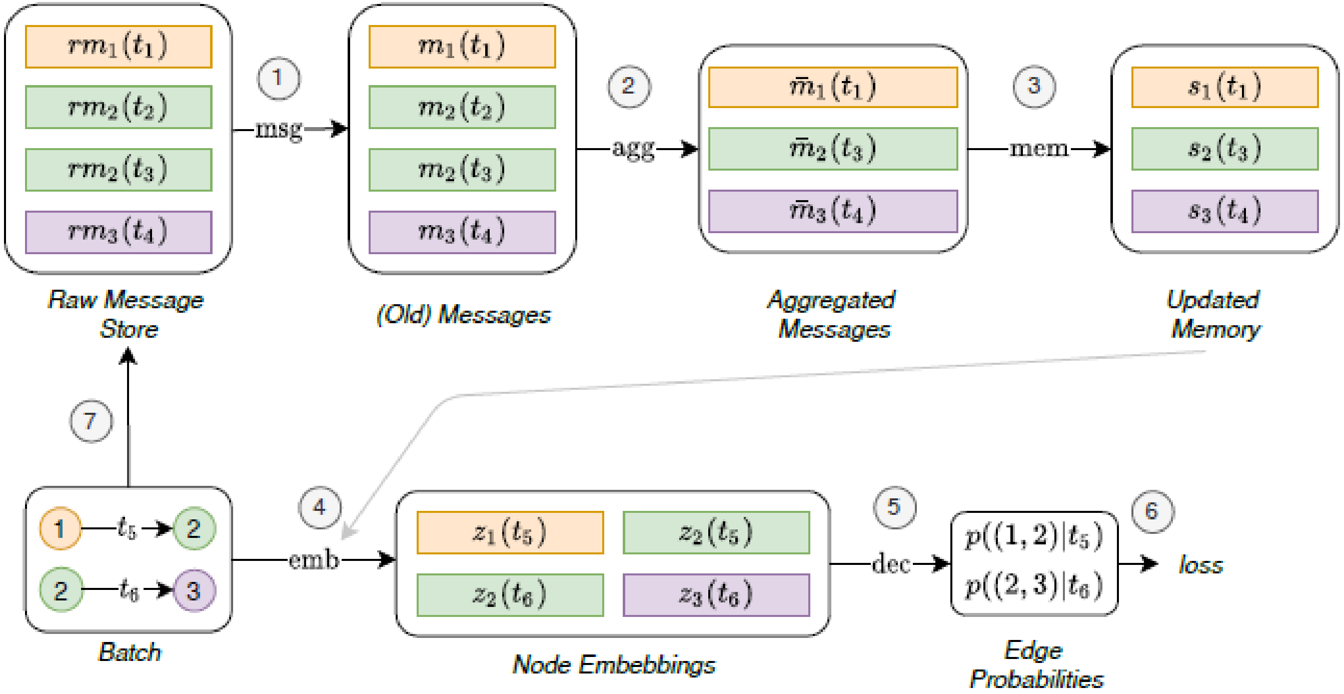
Of each node for any time. Staleness problem may occur.
Using temporal attention graph

The Staleness Problem

- Node's embedding is based on its memory
- Lack of events (messages)
- Stale memory
- Node's embedding is not up to date
- Edge prediction performance ↓

- User's state
- Temporal deactivating
- User's information is getting not relevant
- User's recommendation system

TGN - Flow of Operations



Temporal graph attention – embedding for staleness

- Temporal Neighborhood
- 10 recent
- Attention mechanism
- 1 hop

Temporal Graph Attention (attn): A series of L graph attention layers compute i's embedding by aggregating information from its L-hop temporal neighborhood.

The input to the l-th layer is i's representation $\mathbf{h}_i^{(l-1)}(t)$, the current timestamp t, i's neighborhood representation $\{\mathbf{h}_1^{(l-1)}(t), \ldots, \mathbf{h}_N^{(l-1)}(t)\}$ together with timestamps $t_1, \ldots, t_N$ and features $\mathbf{e}_{i1}(t_1), \ldots, \mathbf{e}_{iN}(t_N)$ for each of the considered interactions which form an edge in i's temporal neighborhood:

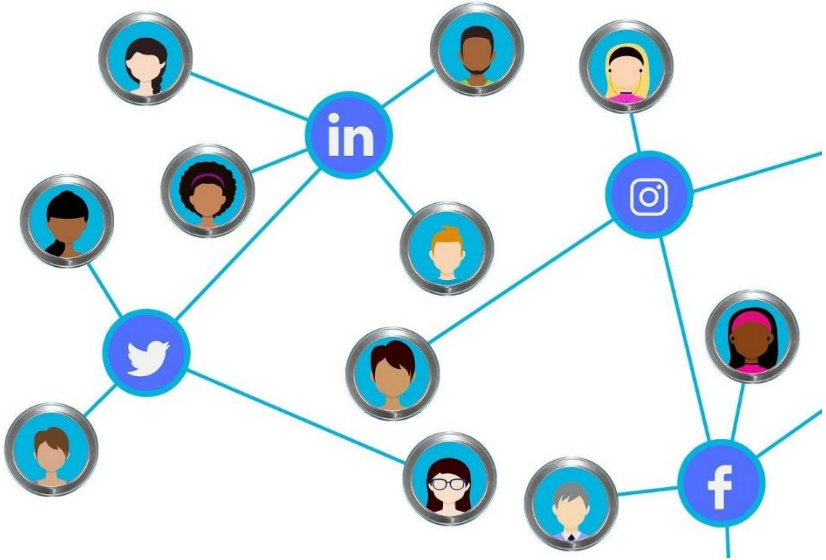
$$\mathbf{h}_i^{(l)}(t) = \mathrm{MLP}^{(l)}(\mathbf{h}_i^{(l-1)}(t) \,\|\, \tilde{\mathbf{h}}_i^{(l)}(t)), \tag{5}$$

$$\tilde{\mathbf{h}}_i^{(l)}(t) = \mathrm{MultiHeadAttention}^{(l)}(\mathbf{q}^{(l)}(t), \mathbf{K}^{(l)}(t), \mathbf{V}^{(l)}(t)), \tag{6}$$

$$\mathbf{q}^{(l)}(t) = \mathbf{h}_i^{(l-1)}(t) \,\|\, \phi(0), \tag{7}$$

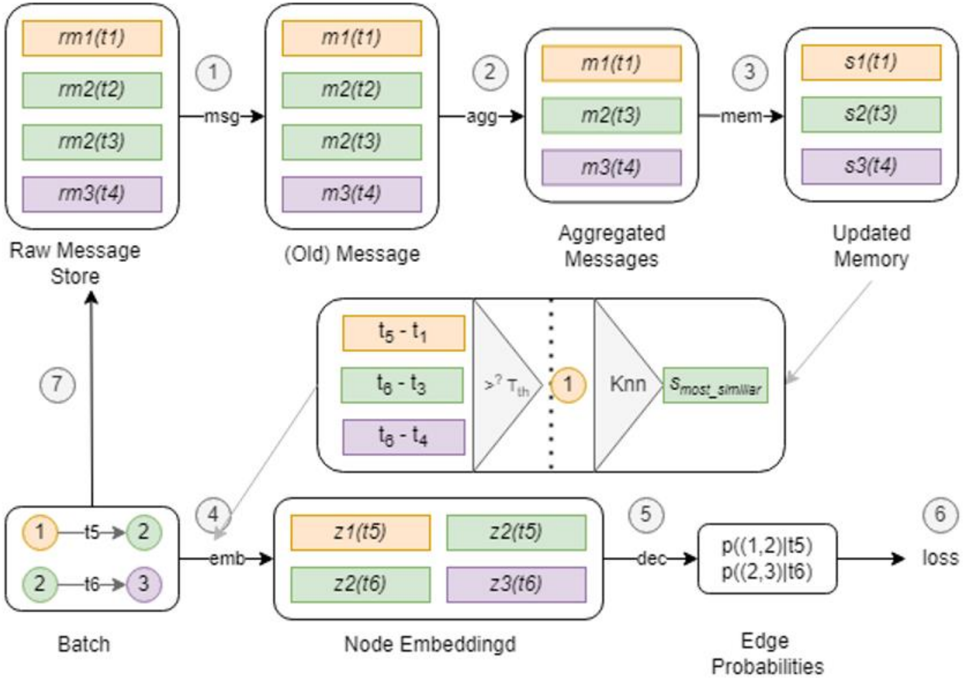
$$\mathbf{K}^{(l)}(t) = \mathbf{V}^{(l)}(t) = \mathbf{C}^{(l)}(t), \tag{8}$$

$$\mathbf{C}^{(l)}(t) = [\mathbf{h}_1^{(l-1)}(t) \,\|\, \mathbf{e}_{i1}(t_1) \,\|\, \phi(t - t_1), \ldots, \mathbf{h}_N^{(l-1)}(t) \,\|\, \mathbf{e}_{iN}(t_N) \,\|\, \phi(t - t_N)]. \tag{9}$$

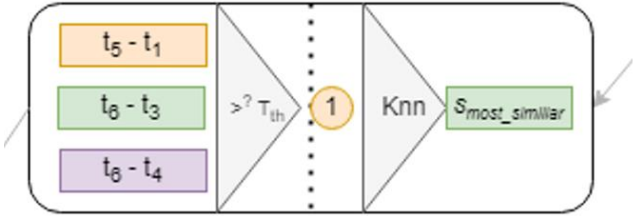


Tell me **who your friends are**, and I will tell you **who you are**

TGN - Flow of Operations + **Our Contribution**



Our Contribution



1

Time Difference Threshold
(Quantile, relative)

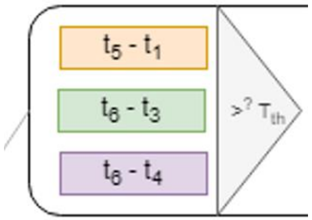
2

Similarity Metric

3

Embedding update

Time Difference



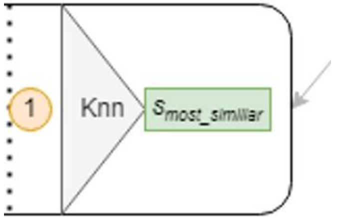
1

Time Difference Threshold
(Quantile, relative)

- **time differences vector** - current event & last updated event – for current batch nodes
- **Relative threshold**
- **Different quantiles**

$$\Delta T_{th} = Q(p) = \inf\{t \in \Re : p(t)\}$$

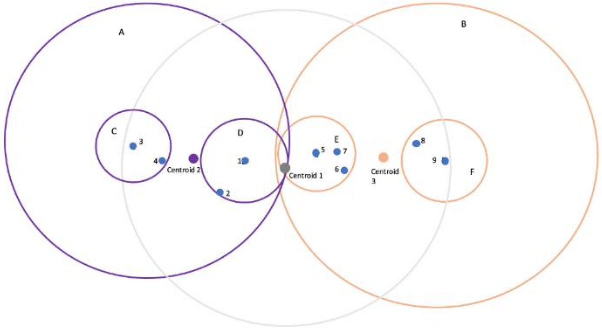
Similarity Metric



2

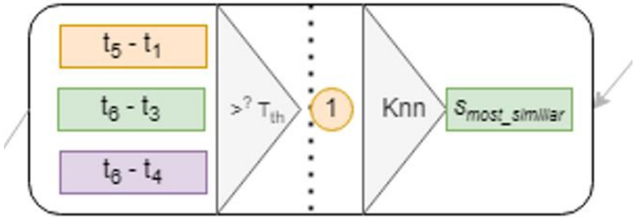
Similarity Metric

- **KNN**- K-nearest Neighbors
  - **Ball-Tree**
  - **Brute Force**
- **K = 1**
- **Only others with history (mem)**



TECHNION
Israel Institute of Technology

Embedding Update



3

Embedding update

- **Stales source nodes update**
- **Sum with memory**
- **Sum with most similar node**

- **Temporal attention**

Datasets

Wikipedia

Reddit

Results

Table : q =0.975

model	dataset	AUC	precision
TGN	**wikipedia**	**0.963**	**0.967**
Ball-Tree	wikipedia	0.961	0.966
Brute-force	**wikipedia**	**0.963**	**0.967**
TGN	reddit	0.953	0.958
Ball-Tree	**reddit**	**0.957**	**0.960**
Brute-force	reddit	0.953	0.957

Table: Different quantiles

model	quantile	AUC	precision
TGN	-	0.963	0.967
Ball-Tree	0.975	0.961	0.966
Ball-Tree	**0.8**	**0.966**	**0.970**
Ball-Tree	0.7	0.958	0.963

Discussion & Future Steps

- Staleness problem importance
- Temporal neighborhood
- Future work:
  - K > 1
  - Quantiles thresholds ↓
  - Learnable nodes similarity model
  - Specific evaluation metric
  - Hyperparamters and fintuning