A Review on Dynamic Graph Neural Networks

Alberto Gonzalo Rodriguez Salgado

Seminar- Selected Topics in Machine Learnnin Research ,
WiSe 21/22

Outline

- What are Dynamic Graphs ?
- Tasks in Dynamic Graphs
- Graph Neural Networks (GNNs) Overview
- From standard GNNs to Dynamic GNNs:
  - Temporal Graph Networks (TGN) approach
  - TGAT, DYREP approaches
- Datasets and performance
- Future Research

What are Dynamic Graphs?

What are dynamics Graphs

- Graph consists of a set of nodes G and set of edges E

- Till now, most Graph papers cover static Graphs i.e, the sets do not change over time

- However, in many applications as social networks, the sets of nodes and edges change over time

↓

Event types: a user messaging a friend (new edge), a new user logging in (new node) or the user changing its profile information (changing node feature)

Tasks in Dynamic Graphs?

Presented by:
Alberto Rodriguez

Tasks in Dynamic Graphs

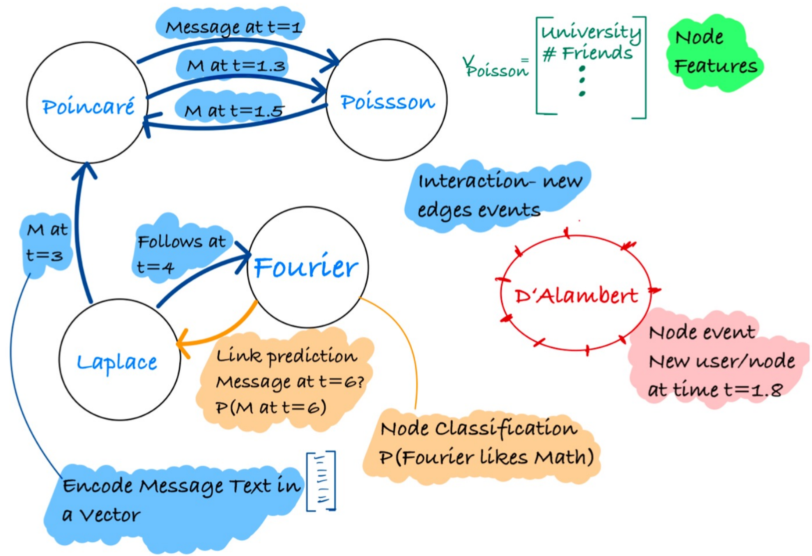


Figure 1. Social Network Dynamic Graph Example showing two event types: 1) Edge creation and 2) Node creation

- Node Classification
- Link prediction: will two nodes. n_i, n_j interact with each other time t_i $i.e, p(n_i, n_j | t_i)$
- Predict **which** type of event will happen
- Predict **when** the event will happen
- Solve inductive task
- Scalable to large graphs

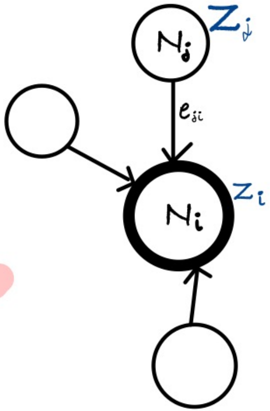
Presented by:
Alberto Rodriguez

Graph Neural Networks (GNNs) Overview

Presented by:
Alberto Rodriguez

Graph Neural Networks Overview

- Find learnable function in a message passing network to compute node embeddings



$$\underbrace{\mathbf{m_i}^{(k)}}_{\text{Message}} = \underbrace{\sum_{u\in N_i}}_{\text{Aggregate over neighbours}} h\left(\underbrace{\mathbf{z}_i^{(k-1)}, \mathbf{z}_u^{(k-1)}}_{\text{Own and neighbours embeddings}}, \underbrace{\mathbf{e}_{ui}}_{\text{Edge representation}}\right)$$

$$\underbrace{\mathbf{z}_i^{(k)}}_{\text{New embedding}} = F\left(\underbrace{\mathbf{z}_i^{(k-1)}}_{\text{Old embedding}}, \underbrace{\mathbf{m}_i^{(k)}}_{\text{Message}}\right)$$

From static GNNs to dynamic GNNs :

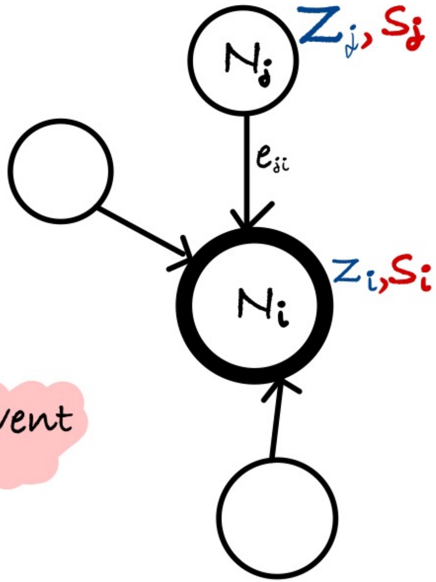
Temporal Graph Networks (TGN)

Presented by:
Alberto Rodriguez

Temporal Graph Networks (TGN)

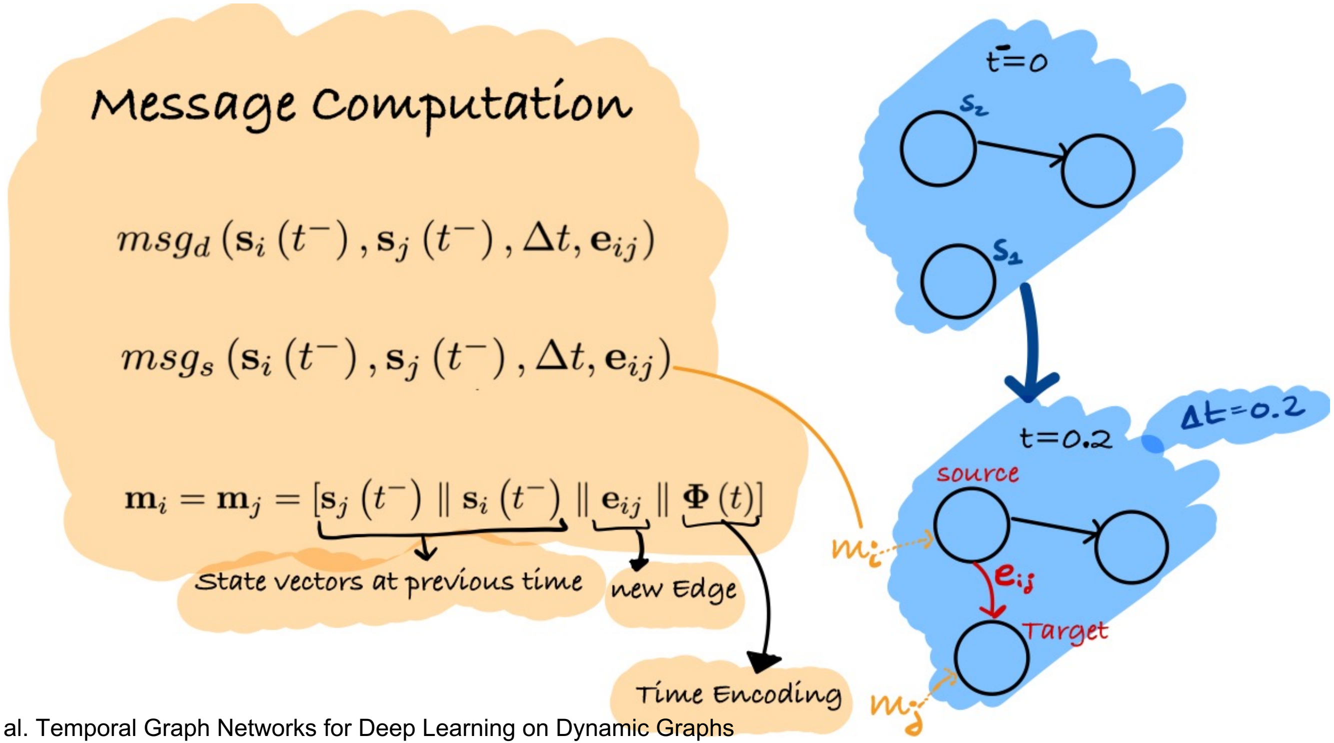
Define Memory state vector s_i
that keeps track over time about
what is happening to node n_i

Update state vector s every time an event
involving node n happens



ICML 2020,E.Rossi et al. Temporal Graph Networks for Deep Learning on Dynamic Graphs

TGN-Message Computation



ICML 2020,E.Rossi et al. Temporal Graph Networks for Deep Learning on Dynamic Graphs

TGN-Message Batching

For Computational reasons it would be too expensive to compute a message and update the state every time an event occurs

Batching : From set of messages only use the most recent one

$$\{\mathbf{m}_i\,(t_1)\ldots\mathbf{m}_i\,(t_n)\}$$

Potential Problem: in some cases other messages may be more important than the Most recent one

ICML 2020,E.Rossi et al. Temporal Graph Networks for Deep Learning on Dynamic Graphs

TGN-Memory State Updater

Once the message is computed, the state is updated as a combination of the old state and the message

$$\underbrace{\mathbf{s}_i(t)}_{\text{New state}} = mem\left(\underbrace{\tilde{\mathbf{m}}_i(t)}_{\text{Message}}, \underbrace{\mathbf{s}_i(t^-)}_{\text{Old state}}\right)$$

LSTM or GRU to allow learning long term dependencies

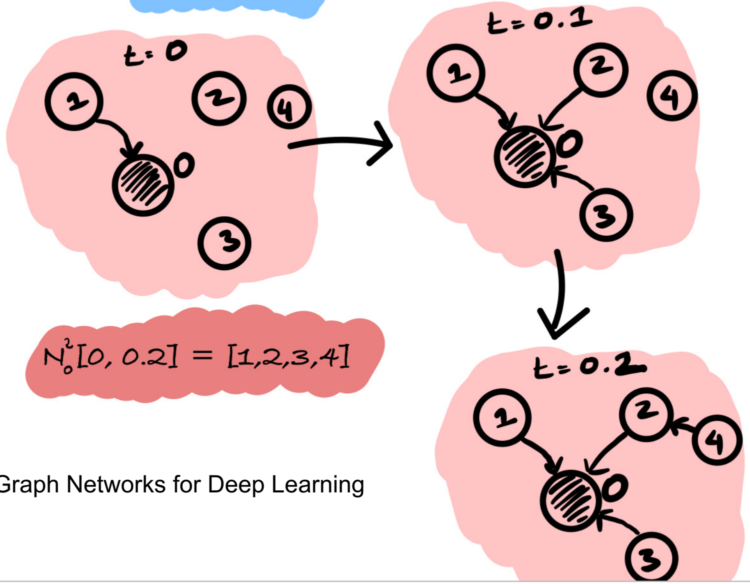
ICML 2020,E.Rossi et al. Temporal Graph Networks for Deep Learning on Dynamic Graphs

Presented by:
Alberto Rodriguez

TGN-Computing Embedding

The embedding vector is computed by aggregating over the temporal neighbourhood of the node

$$\underbrace{\mathbf{z}_i(t)}_{\text{Embedding}} = \sum_{\underbrace{j\in N_i^k[0,t]}_{\text{Temporal neighbourhood}}} \overbrace{h}^{\text{Learnable function e.g, Attention}}\left(\underbrace{\mathbf{s}_i(t), \mathbf{s}_j(t)}_{\text{state vectors}}, \underbrace{\mathbf{e}_{ij}}_{\text{Edge information}}, \underbrace{\mathbf{v}_i(t), \mathbf{v}_j(t)}_{\text{Feature vectors}}\right)$$



ICML 2020,E.Rossi et al. Temporal Graph Networks for Deep Learning on Dynamic Graphs

Presented by:

Alberto Rodriguez

Temporal Graph Attention

Idea: replace positional encoding by a time encoding

$$\underbrace{\Phi(t)}_{\text{Time encoding}} = [\omega_0 + \phi_0, \underbrace{\Psi}_{\text{Periodic function}} (\omega_1 t + \phi_1) \ldots \Psi (\omega_d t + \phi_d)]$$

Time encoding

Periodic function

Frequencies encoding a periodicity e.g. "how often does a user message other user"

Frequencies are learned during training as the parameters of a single layer Neural Network!

Input to attention layer

$$\mathbf{Z} = [\mathbf{z}_i \parallel \underbrace{\Phi(t_i)}_{\text{Time encoding}}, \mathbf{z}_1 \parallel \Phi(t_1) \ldots \mathbf{z}_N \parallel \Phi(t_N)] .$$

Time encoding

State vector, Edge vector, Node feature etc.

ICML 2020,E.Rossi et al. Temporal Graph Networks for Deep Learning on Dynamic Graphs

Presented by:

Alberto Rodriguez

Other Approaches : TGAT and DyRep

Other approaches: TGAT and DyRep

- The *Temporal Graph Attention* (TGAT) model [16] can be seen as a special case of TGN where no node's memory representations states are used (1)

- DyRep aims to capture the graphs dynamics with temporal point processes (TPP) (2)

(1) ICLR 2020, Kumar et al., Inductive representation learning on temporal graphs
(2) ICLR 2019, Trivedi et al., DyRep: Learning Representations over Dynamic Graphs

Other approaches: DyRep

Nodes

Type of event (new node or new edge)

$$\lambda_k^{u,v}(t) = f_k\left(g_k^{u,v}(\bar{t})\right)$$

TPP conditional Distribution

Learnable similarity function between zwo nodes

intensity function modeled by a softplus function with learnable parameters

→ With this it is possible to predict when an event is likely to occur between two nodes

(1) ICLR 2019, Trivedi et al., DyRep: Learning Representations over Dynamic Graphs

Datasets

Datasets

- Only a few datasets available e.g., Wikipedia and Reddit bipartite graphs which are open sourced

- Twitter bipartite data set is not open sourced

- None of these datasets contain node creation or deletion events, neither node features!

Performance

Performance

Table 2: Average Precision (%) for future edge prediction task in transductive and inductive settings. **First**, **Second**, **Third** best performing method. *Static graph method. †Does not support inductive.

	Wikipedia		Reddit		Twitter	
	Transductive	Inductive	Transductive	Inductive	Transductive	Inductive
GAE*	91.44 ± 0.1	†	93.23 ± 0.3	†	—	†
VAGE*	91.34 ± 0.3	†	92.92 ± 0.2	†	—	†
DeepWalk*	90.71 ± 0.6	†	83.10 ± 0.5	†	—	†
Node2Vec*	91.48 ± 0.3	†	84.58 ± 0.5	†	—	†
GAT*	**94.73** ± 0.2	91.27 ± 0.4	97.33 ± 0.2	95.37 ± 0.3	67.57 ± 0.4	62.32 ± 0.5
GraphSAGE*	93.56 ± 0.3	91.09 ± 0.3	97.65 ± 0.2	**96.27** ± 0.2	65.79 ± 0.6	60.13 ± 0.6
CTDNE	92.17 ± 0.5	†	91.41 ± 0.3	†	—	†
Jodie	94.62 ± 0.5	**93.11** ± 0.4	97.11 ± 0.3	94.36 ± 1.1	**85.20** ± 2.4	**79.83** ± 2.5
TGAT	**95.34** ± 0.1	**93.99** ± 0.3	**98.12** ± 0.2	**96.62** ± 0.3	70.02 ± 0.6	66.35 ± 0.8
DyRep	94.59 ± 0.2	92.05 ± 0.3	**97.98** ± 0.1	95.68 ± 0.2	**83.52** ± 3.0	**78.38** ± 4.0
TGN-attn	**98.46** ± 0.1	**97.81** ± 0.1	**98.70** ± 0.1	**97.55** ± 0.1	**94.52** ± 0.5	**91.37** ± 1.1

Long story short: TGN accomplishes best scores both on inductive and transductive tasks on all data sets

Table from ICML 2020,E.Rossi et al. Temporal Graph Networks for Deep Learning on Dynamic Graphs

Future Research

Future Research

- Lack of data: publish data sets that contain node creation and deletion events.
- Message batching not with the most recent message but by attending over the messages
- Predicting **when** with a neural TPP using a LSTM instead of a softplus function as in. DyRep.

Thank you

for your attention

Presented by:
Alberto Rodriguez