



Activity Prediction using Dynamic Graph Embeddings

Jens Petit



1 Dynamic Graph Representation Learning

2 GDELT Dataset

3 Research Idea





Dynamic Graph Representation Learning



2020-10-29

Dynamic Graph Representation Learning

●○○○○○○○○○

GDELT Dataset

○○○○○

Research Idea

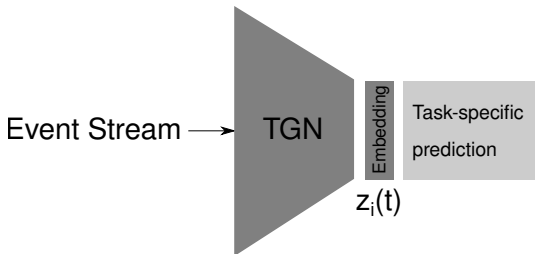
○○○○○○○

Discrete-time dynamic graphs (DTDG)

Sequences of static graph snapshots.

Continuous-time dynamic graphs (CTDG)

Timed list of events, including node addition, deletion and edge addition and deletion.



General encoder-decoder framework

Current decoders:

- future edge ('link') prediction
- dynamic node classification

Problem settings:

- Transductive: only nodes which have been used in training
- Inductive: additionally nodes which have *not* been used in training

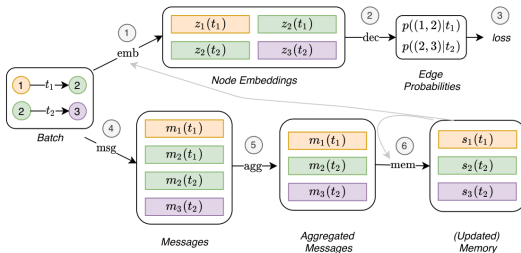
Datasets:

- Reddit, Wikipedia: Bipartite interaction graphs with users and subreddits/pages as nodes.
- Twitter: Users are nodes and retweets are interactions.

All interaction events carry text features (tweets, edits, posts) and 70%-15%-15% (train-valid-test) chronological split is used.

Exemplary decoder: simple MLP decoder mapping from the concatenation of two node embeddings to the probability of the edge

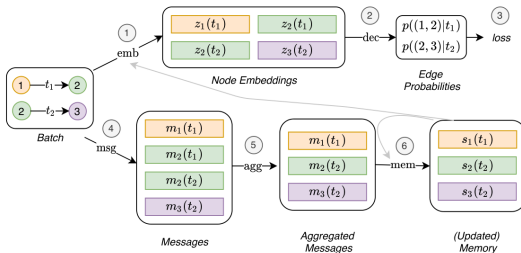
¹“Temporal Graph Networks For Deep Learning on Dynamic Graphs, *Rossi et al.*”



TGN computations on a single batch of time-stamped interactions².

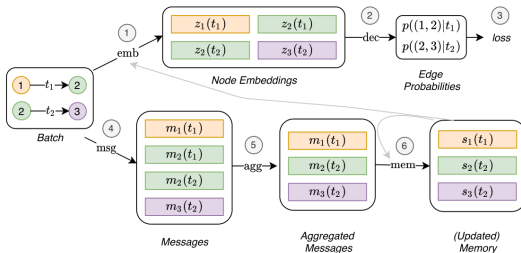
Core idea: combining memory module with graph-based operators

²Figure taken from "Temporal Graph Networks For Deep Learning on Dynamic Graphs, Rossi *et al.*"



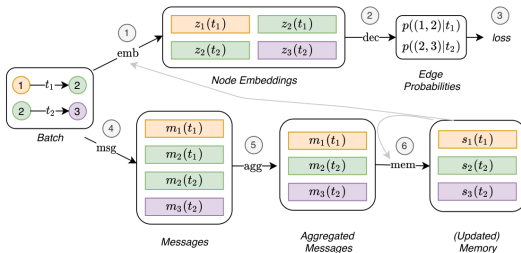
TGN computations on a single batch of time-stamped interactions.

$$\mathbf{m}_i(t) = \text{msg}(\mathbf{s}_i(t^-), \mathbf{s}_j(t^-), \Delta t, \mathbf{e}_{ij}(t))$$



Combine all messages in a single batch for a specific node:

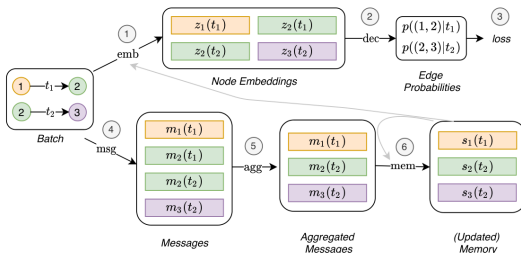
$$\bar{\mathbf{m}}_i(t) = \text{agg}(\mathbf{m}_i(t_1), \dots, \mathbf{m}_i(t_b))$$



TGN computations on a single batch of time-stamped interactions.

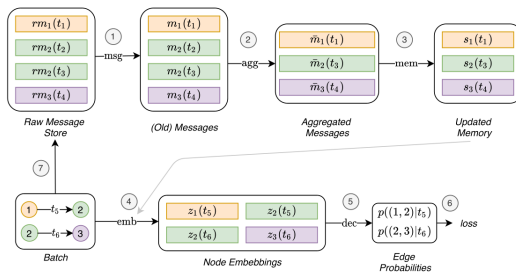
Using a Recurrent Neural Network:

$$\mathbf{s}_i(t) = \text{mem}(\bar{\mathbf{m}}_i(t), \mathbf{s}_i(t^-))$$



$$\mathbf{z}_i(t) = \text{emb}(i, t) = \sum_{j \in \mathcal{N}_i^k([0, t])} h(\mathbf{s}_i(t), \mathbf{s}_j(t), \mathbf{e}_{ij}, \mathbf{v}_i(t), \mathbf{v}_j(t)),$$

Includes specific cases like: memory directly, time projection (JODIE), Temporal Graph Attention (TGAT), Temporal Graph Sum



TGN training ³

Problem: memory-related modules (Message function, Message aggregator, and Memory updater) do not directly influence the loss and therefore do not receive a gradient -> memory update before predictions

³Figure taken from "Temporal Graph Networks For Deep Learning on Dynamic Graphs, Rossi *et al.*"



GDELT Dataset



2020-10-29

Dynamic Graph Representation Learning

oooooooooooo

GDELT Dataset

●ooooo

Research Idea

oooooooo

"The GDELT Project monitors the world's broadcast, print, and web news from nearly every corner of every country in over 100 languages and identifies the people, locations, organizations, themes, sources, emotions, counts, quotes, images and events driving our global society every second of every day, creating a free open platform for computing on the entire world."

- Global Knowledge Graph
- Global Event Database
- Global Entity Graph
- Global Frontpage Graph

Random sample of news articles every 15 minutes (roughly 100k per day)

Google NLP API extracts entities from each article

```
{
  "url": "https://chicago.suntimes.com/news/washington-state-ends-racially-biased",
  "lang": "en",
  "date": "2018-10-12T00:15:00Z",
  "score": -0.2,
  "magnitude": 12.3,
  "entities": [
    {
      "name": "Supreme Court",
      "type": "ORGANIZATION",
      "numMentions": 1,
      "avgSalience": 0.04405
    },
    ...
  ]
}
```

Each pair of entities occurring in a single article correspond to an edge event with timestamp:

Nathan Trott	RB Leipzig
Manchester United	RB Leipzig
West Ham	RB Leipzig
Timo Werner	RB Leipzig
Ralf Rangnick	RB Leipzig
Bundesliga	RB Leipzig
Patrick Dempsey	Leipzig
Leipzig	Germany
Patrick Dempsey	Leipzig
Leipzig	Germany
Patrick Dempsey	Leipzig
Leipzig	Germany
...	

Restricting to the 4 most salient entities gives roughly 200k edges per day

ScaDS.AI Data Example: IPCC

DRE!

SZ Süddeutsche.de
Klimawandel - Reaktionen auf den IPCC-Klimabericht
 In dem Bericht des Weltklimarates IPCC heißt es jetzt, dass die angestrebte Begrenzung der Erderwärmung auf 1,5 Grad im Vergleich zum vorindustriellen Niveau ...
 08.10.2018



S SPIEGEL ONLINE
Sonderbericht des Weltklimarats: Die Welt gerät aus den Fugen - fragt sich nur, wie sehr
 Klimawandel: IPCC Bericht zum 1,5-Grad-Ziel vorgestellt.
 Ausführlich - 08.10.2018



BR Bayerischer Rundfunk
Weltklimarat IPCC veröffentlicht Sonderbericht zum 1,5-Grad-Ziel
 Die Erde erwärmt sich schneller und mit schwereren Folgen als bisher angenommen, ist ein Ergebnis des IPCC-Sonderberichts zum 1,5-Grad-Ziel. Erforderlich ...
 09.10.2018



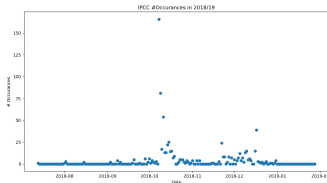
klimareporter*
Politik muss Ergebnisse des IPCC aufgreifen – klimareporter*
 Politik muss Ergebnisse des IPCC aufgreifen. In wenigen Tagen erscheint der Sonderbericht zum 1,5-Grad-Ziel. Der Weltklimarat hat alle wichtigen ...
 30.09.2018



Solarify - Energie für die Zukunft
IPCC-Sonderbericht kommt – SOLARIFY
 Sie laden den Weltklimarat ein, zur angestrebten 1,5-Grad-Grenze einen Sonderbericht zu verfassen. Am 08.10.2018 wird dieser Sonderbericht des IPCC ...
 26.09.2018



DIE WELT
IPCC-Bericht: Für das Pariser Klimaziel braucht es negative Emissionen
 IPCC-Bericht: Für das Pariser Klimaziel braucht es negative Emissionen. Wissen - IQ-Test Weltraum Natur & Umwelt Gesundheit Psychologie Biowetter.
 08.10.2018

Occurrences IPCC

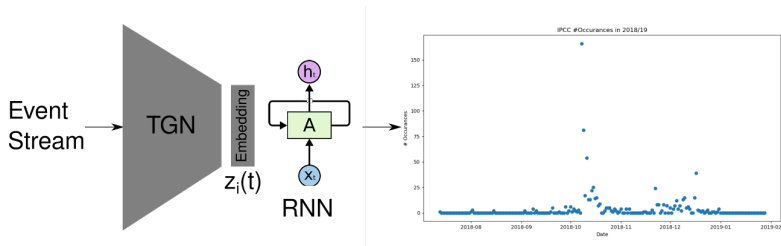
entity_1	entity_2	count
IPCC	India	18
Michael McCormack	IPCC	15
Scott Morrison	IPCC	15
India	IPCC	15
Donald Trump	IPCC	15
United States	IPCC	14
European Union	IPCC	14
Hoesung Lee	IPCC	14
ottish Government	IPCC	12
IPCC	US	12



Research Idea



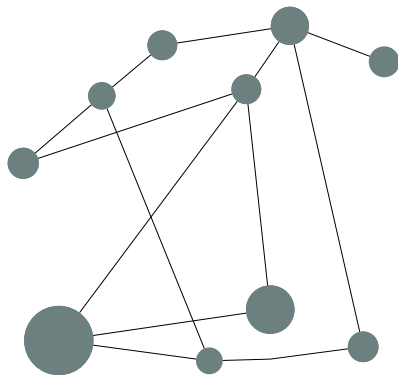
How can we identify entities with similar temporal dynamics, e.g. “hot” topics?



Replace decoder with a RNN which predicts the future #Occurrences per day for a given entity and time horizon

Why is the graph information relevant?

The neighborhood should be strong indicator for future behavior: If all my neighbors are getting popular, then it is very likely that I will too.





- before (link prediction): similarity predicts future link between two entities
- now (activity prediction): embedding space represents temporal dynamics
 - clustering
 - split relative and absolute dynamics
 - duplicate detection





- Current state:
 - preparing dataset
- Baseline: time series prediction for number of occurrences (no neighborhood info)
- Open questions:
 - What is a single data point?
 - How to batch the data?

