



Activity Prediction using Dynamic Graph Embeddings

Jens Petit



1 Graph Representation Learning

2 GDELT Dataset

3 Research Idea: Activity Prediction





Graph Representation Learning



2020-12-16

Graph Representation Learning

● ○ ○ ○ ○ ○ ○

GDELT Dataset

○ ○ ○ ○ ○

Research Idea: Activity Prediction

○ ○ ○ ○ ○ ○ ○ ○ ○

Project nodes into a latent vector space where geometric relations correspond to relationships in the original graph.

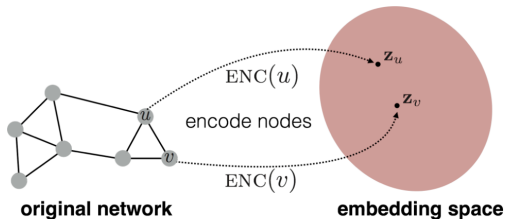


Figure 1: The node embedding problem¹.

¹Graph Representation Learning, *Hamilton 2020*.

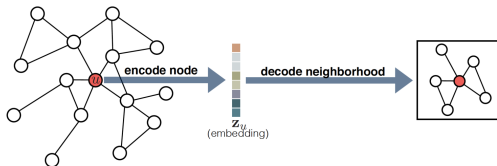


Figure 2: Overview of the encoder-decoder perspective².

²Graph Representation Learning, *Hamilton 2020*.

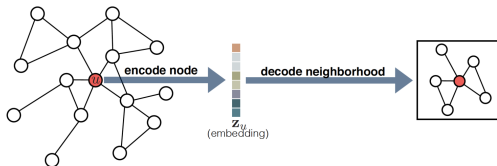


Figure 2: Overview of the encoder-decoder perspective².

$$\text{DEC}(\text{ENC}(u), \text{ENC}(v)) = \text{DEC}(z_u, z_v) \approx \mathbf{S}[u, v]$$

²Graph Representation Learning, *Hamilton 2020*.

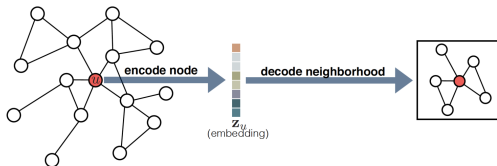


Figure 2: Overview of the encoder-decoder perspective².

$$\text{DEC}(\text{ENC}(u), \text{ENC}(v)) = \text{DEC}(\mathbf{z}_u, \mathbf{z}_v) \approx \mathbf{S}[u, v]$$

$$\mathcal{L} = \sum_{(u,v) \in \mathcal{D}} \ell(\text{DEC}(\mathbf{z}_u, \mathbf{z}_v), \mathbf{S}[u, v])$$

²Graph Representation Learning, *Hamilton 2020*.

Discrete-Time Dynamic Graphs

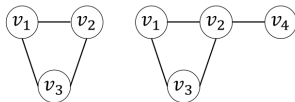


Figure 3: Two graph snapshots.

Continuous-Time Dynamic Graphs

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

```

{
  AddNode(v_4, t_1),
  AddEdge((v_2, v_4), t_2),
  AddEdge((v_2, v_3), t_3),
  ...
}

```

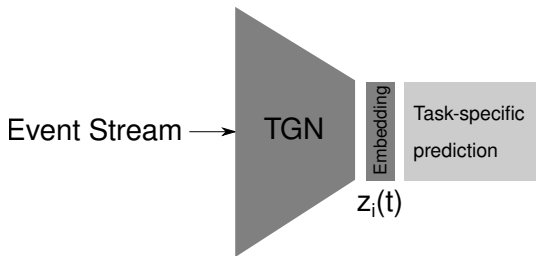
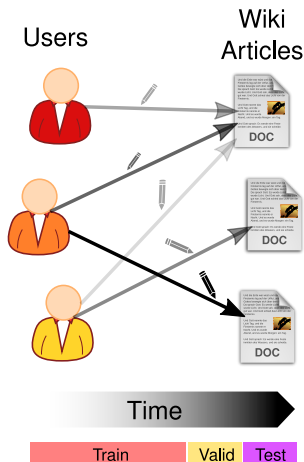



Figure 4: TGN architecture.

³“Temporal Graph Networks For Deep Learning on Dynamic Graphs, *Rossi et al.*”



Dataset:

- Wikipedia with bipartite interaction graphs (users and pages as nodes)
- interaction events carry text features (edits) and chronological split

Decoder:

- predict future edits -> concatenate two node embeddings into simple MLP

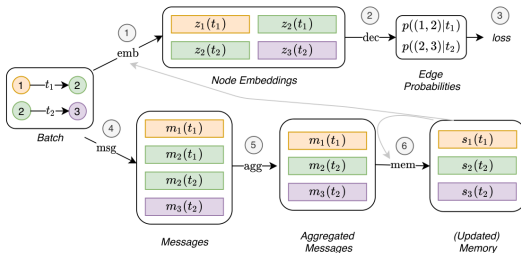


Figure 5: TGN computations on a single batch of time-stamped interactions⁴.

Core idea: combining memory module with graph-based operators

⁴Figure taken from "Temporal Graph Networks For Deep Learning on Dynamic Graphs, Rossi *et al.*"



GDELT Dataset



2020-12-16

Graph Representation Learning

○○○○○○○

GDELT Dataset

●○○○○

Research Idea: Activity Prediction

○○○○○○○

"... monitors the world's broadcast, print, and web news."

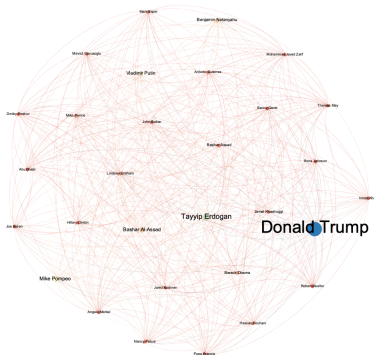


Figure 6: Person co-occurrence in articles covering Syria and Turkey.⁵

⁵<https://medium.com/@atakunguey94/a-brief-introduction-into-gdelt-gtc64d03a>

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!



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






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






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



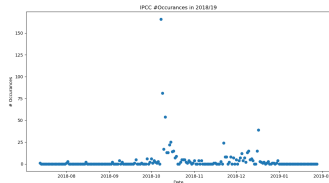


Figure 7: # Occurrences IPCC

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: Activity Prediction



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

Standard setting:

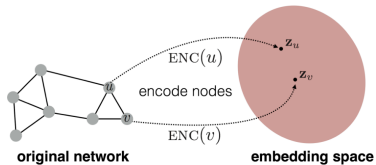


Figure 8: Closeness in embedding space predicts neighborhood.

Proposed approach:

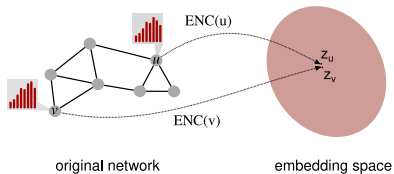
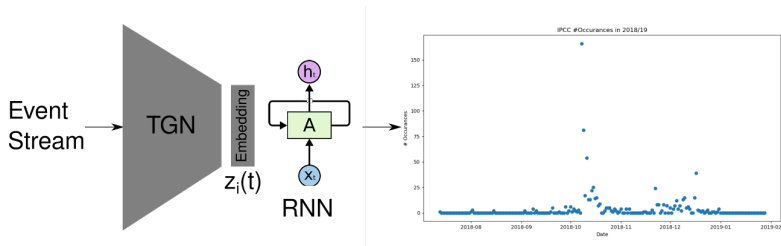
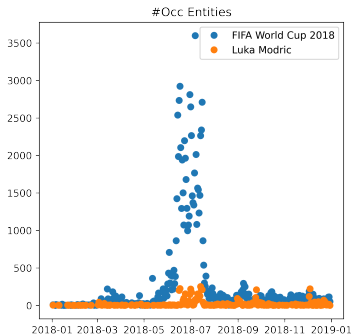


Figure 9: Closeness in embedding space predicts temporal activity.



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.

Dataset preparation:

- implemented data fetching, processing and interim representation
- large dataset -> setup streaming scenario



Dataset preparation:

- implemented data fetching, processing and interim representation
- large dataset -> setup streaming scenario

Running baseline experiments:

- time series prediction for number of occurrences (no neighborhood info)



Limitations:

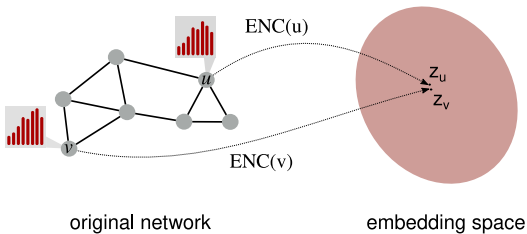
- cannot predict one time events

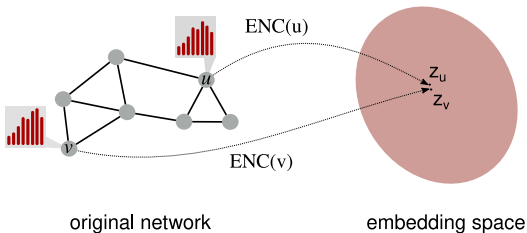
Limitations:

- cannot predict one time events

Open questions:

- Can we split relative and absolute dynamics?
- What should be the prediction horizon?





Thank you! Questions?