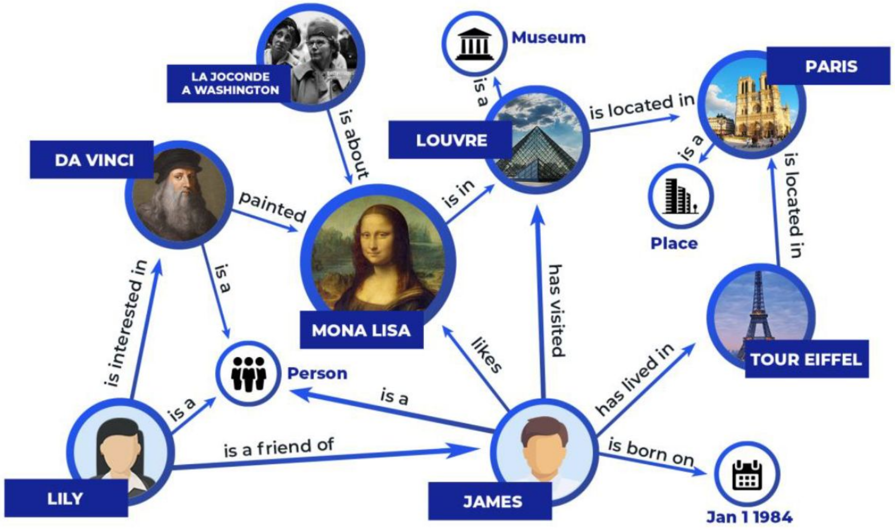
Temporal Knowledge Graph

Dong-Ho Lee

Recap: Knowledge Graph



Link Prediction

- (Lily, is a friend of, ?)
- (Lily, ? James)

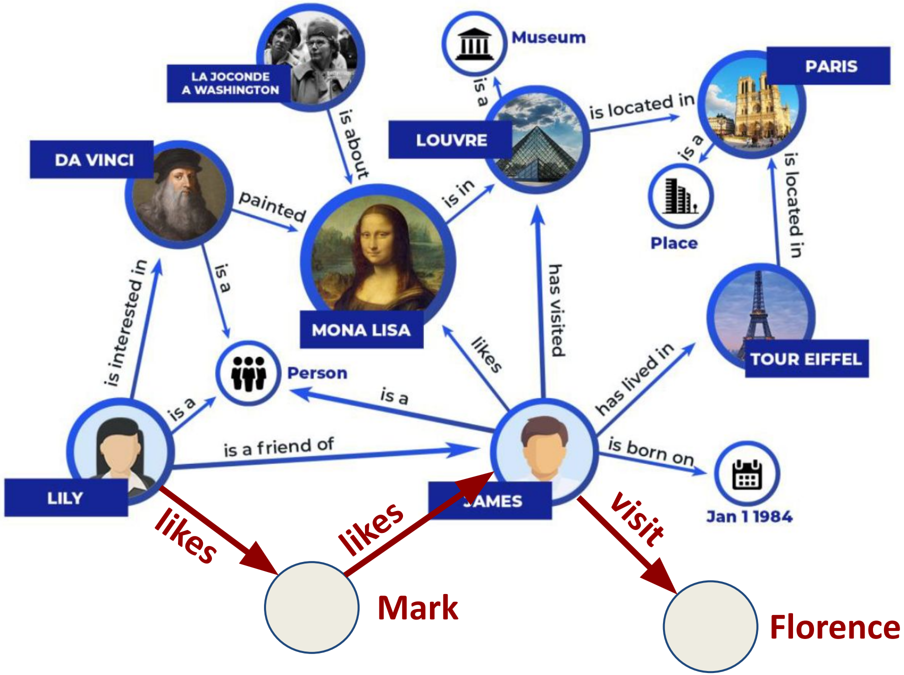
Node Classification

Which class does Paris belong to?

Graph Classification

Which class does the whole graph belong to?

Temporal Knowledge Graph



Link Prediction

- (Lily, is a friend of, ?)
- (Lily, ? James)
  - **(Lily, ?, James, 2023)**
  - **(Lily, ?, James, 2024)**
- **(Mark, ?, James, 2025)**

Node Classification

Which class does Paris belong to?

Graph Classification

Which class does the whole graph belong to?

- **KG changes over time**
  - **Relation can change over time**
  - **New node can appear over time**

Temporal Knowledge Graph

Temporal Knowledge Graph: Knowledge graphs, where facts occur, recur, or evolve over time, and each edge in the graphs has temporal information associated with it (Trivedi et al., 2017)

Goal

- Model time dependency of changes in graph
- **Infer hidden links and other information** by taking both
  - Present graph structure
  - Past graph embeddings

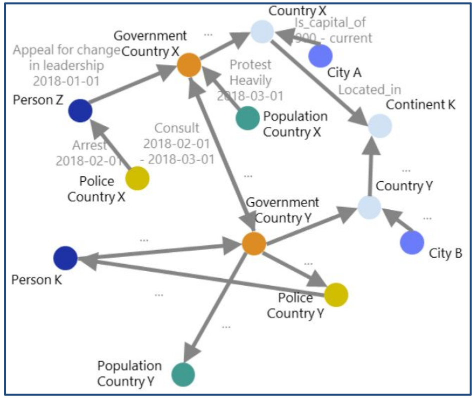
TODO

1. Extend triple to Quadruples! (s, p, o) → (s, p, o, t)
2. Learn graph embeddings over Quadruples.

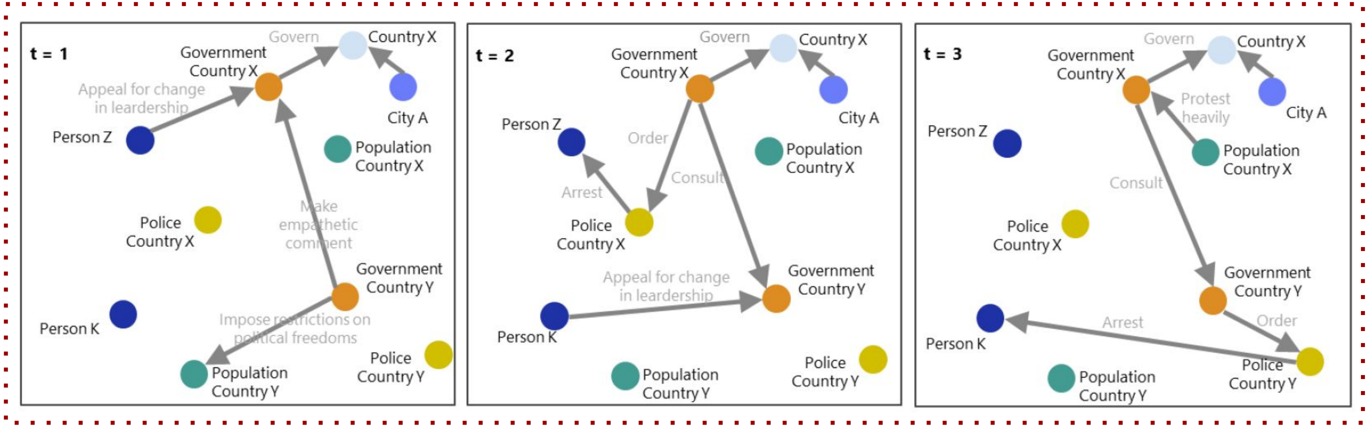
Deep Temporal Reasoning for Dynamic Knowledge Graphs., ICML 2017

TKG Approaches

1. Time in node/edge



2. Graph snapshot per time



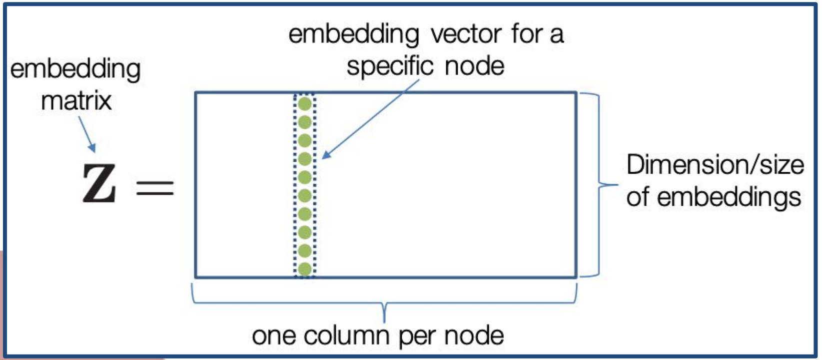
Recap: Graph Representations

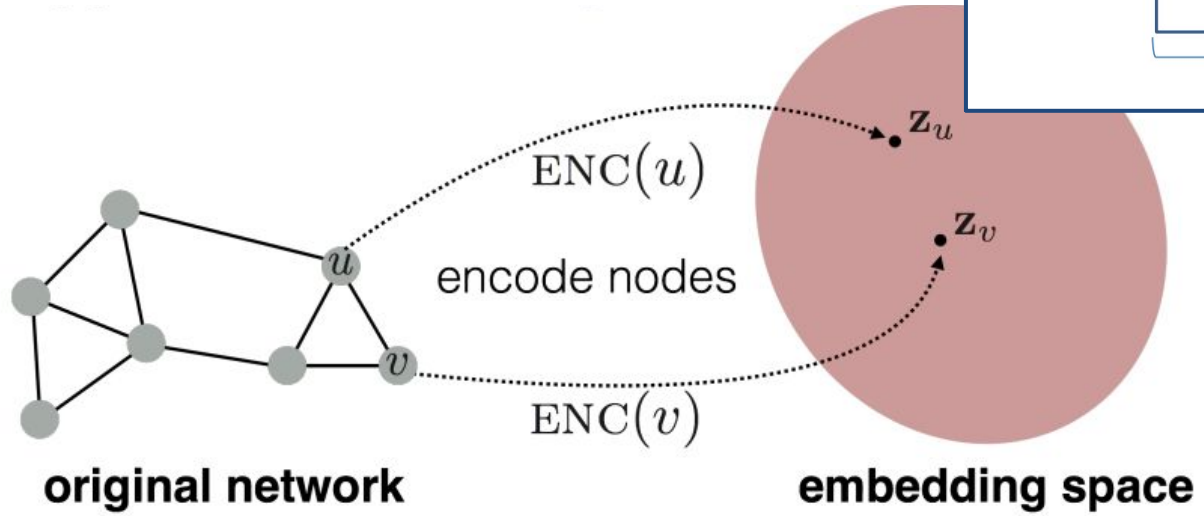
1. **Shallow KG Embeddings** (TransE, ComplEx, … etc.)
   - Lecture 9. Representation Learning

2. **Graph Neural Networks** (GraphSAGE, GCN, … etc.)
   - Lecture 16. Advanced Representation Learning

Recap: Graph Representations

Shallow KG embedding





Represent each node in the network into embedding.

Representation Learning on Networks, WWW 2018 Tutorial

Recap: Graph Representations

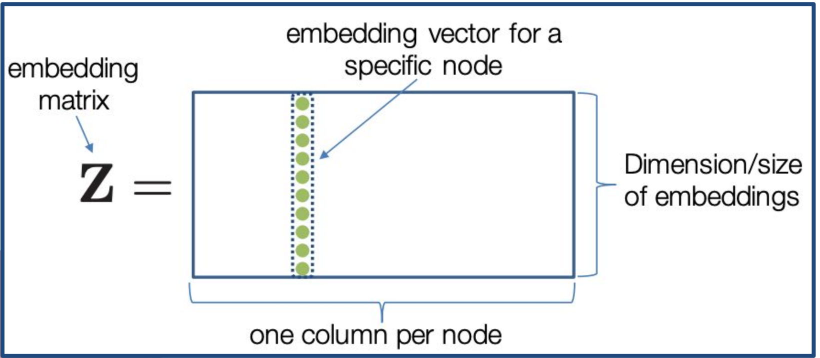
Shallow KG embedding

Score function

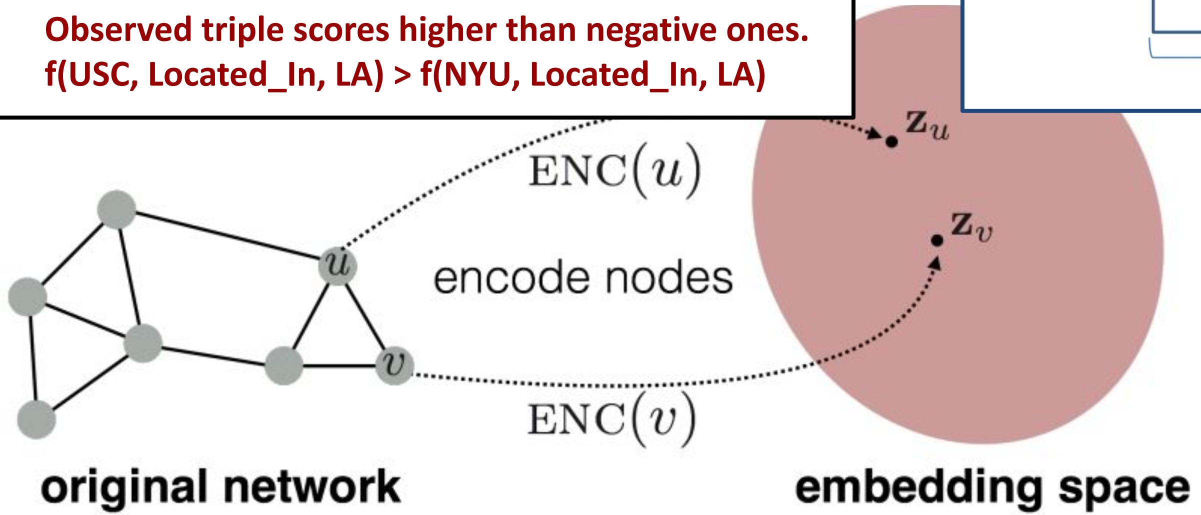
- *score = f(USC, Located_In, LA)*

Loss function

- **Observed triple scores higher than negative ones.**
- **f(USC, Located_In, LA) > f(NYU, Located_In, LA)**



Represent each node in the network into embedding.

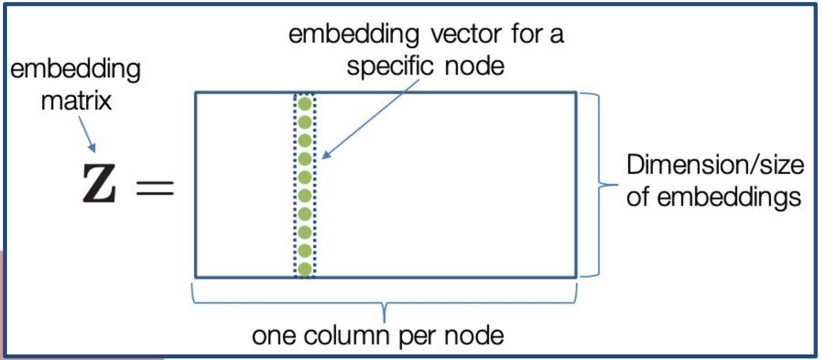


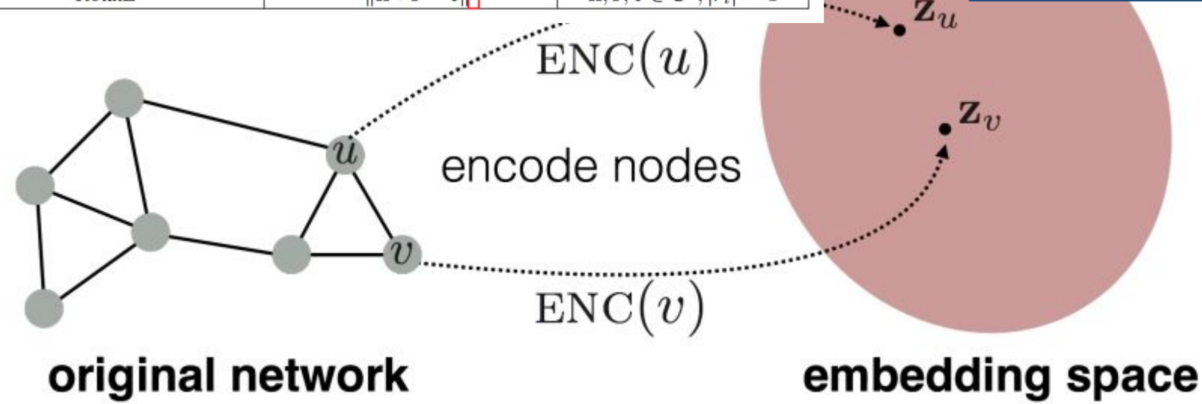
Representation Learning on Networks, WWW 2018 Tutorial

Recap: Graph Representations

Shallow KG embedding

Model	Score Function	
SE (Bordes et al., 2011)	$-\|\boldsymbol{W}_{r,1}\mathbf{h} - \boldsymbol{W}_{r,2}\mathbf{t}\|$	$\mathbf{h}, \mathbf{t} \in \mathbb{R}^k, \boldsymbol{W}_{r,\cdot} \in \mathbb{R}^{k\times k}$
TransE (Bordes et al., 2013)	$-\|\mathbf{h} + \mathbf{r} - \mathbf{t}\|$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
TransX	$-\|g_{r,1}(\mathbf{h}) + \mathbf{r} - g_{r,2}(\mathbf{t})\|$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
DistMult (Yang et al., 2014)	$\langle \mathbf{r}, \mathbf{h}, \mathbf{t} \rangle$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
ComplEx (Trouillon et al., 2016)	$\mathrm{Re}(\langle \mathbf{r}, \mathbf{h}, \overline{\mathbf{t}} \rangle)$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k$
HolE (Nickel et al., 2016)	$\langle \mathbf{r}, \mathbf{h} \otimes \mathbf{t} \rangle$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
ConvE (Dettmers et al., 2017)	$\langle \sigma(\mathrm{vec}(\sigma([\overline{\mathbf{r}}, \overline{\mathbf{h}}] * \boldsymbol{\Omega}))\boldsymbol{W}), \mathbf{t} \rangle$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{R}^k$
RotatE	$-\|\mathbf{h} \circ \mathbf{r} - \mathbf{t}\|^2$	$\mathbf{h}, \mathbf{r}, \mathbf{t} \in \mathbb{C}^k, \|r_i\| = 1$

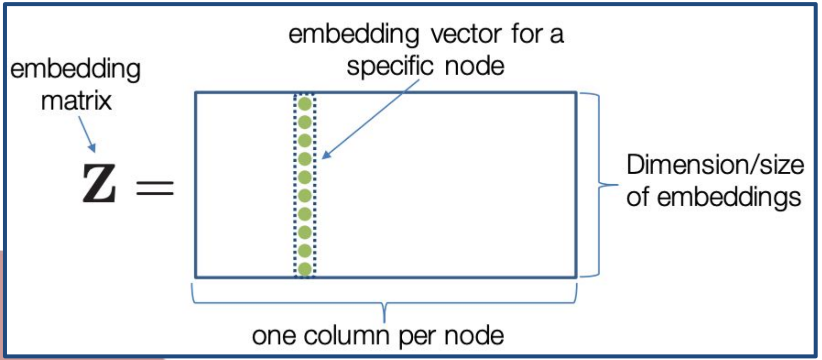




Represent each node in the network into embedding.

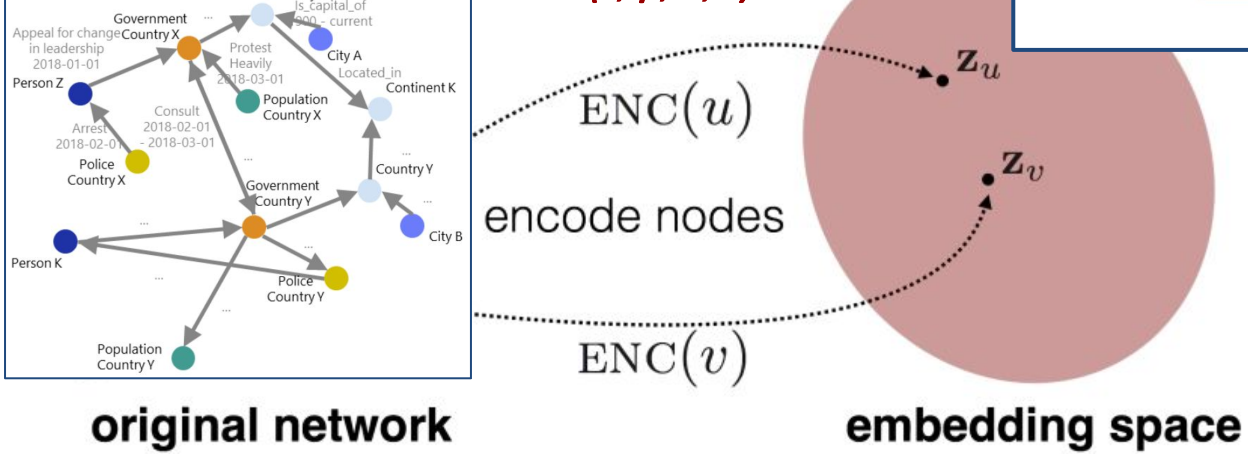
Representation Learning on Networks, WWW 2018 Tutorial

Temporal Shallow KG Embedding



1. Time in node/edge

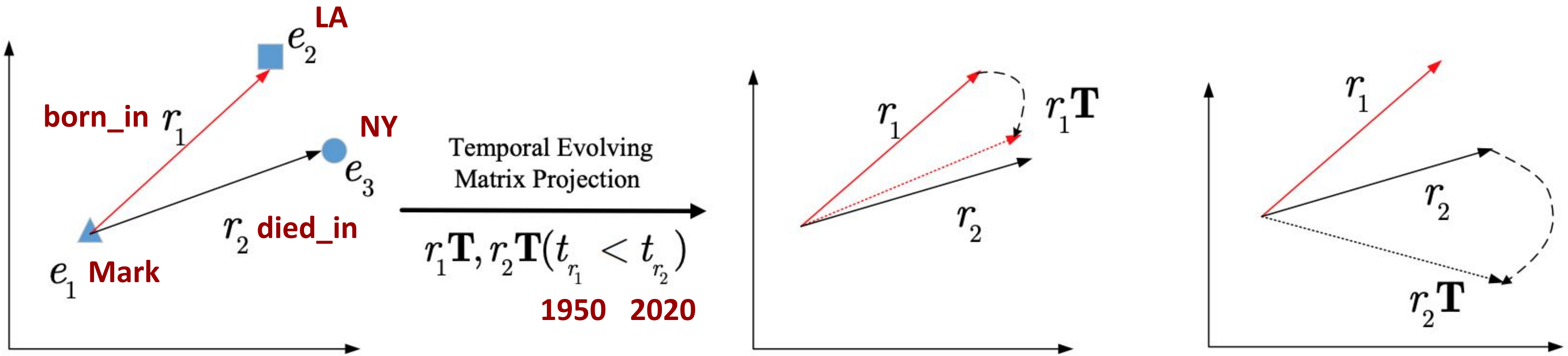
(s, p, o, t)



Represent each node in the network into time-aware embedding.

Temporal Shallow KG Embedding

TransE → TTransE



TransE Score function

- *s+r-o = 0*
- *Mark + born_in = LA*
- *Mark + died_in = NY*

TTransE Score function

- *s+r+t-o = 0*
- *Mark + born_in + 1950 = LA*
- *Mark + died_in + 2020 = NY*

Towards Time-Aware Knowledge Graph Completion, COLING 2016

Temporal Shallow KG Embedding

ComplEx → TComplEx

$$\hat{X}_{i,j,k} = \langle U_i, V_j, \overline{U}_k \rangle, U \in \mathbb{C}^{N \times d}, V \in \mathbb{C}^{P \times d}$$

$$\hat{X}_{i,j,k,t} = \langle U_i, V_j, \overline{U}_k, T_t \rangle = \langle U_i, V_j \odot T_t, \overline{U}_k \rangle$$

Time-dependent
predicate representation

Tensor Decompositions for Temporal Knowledge Base Completion., ICLR 2020

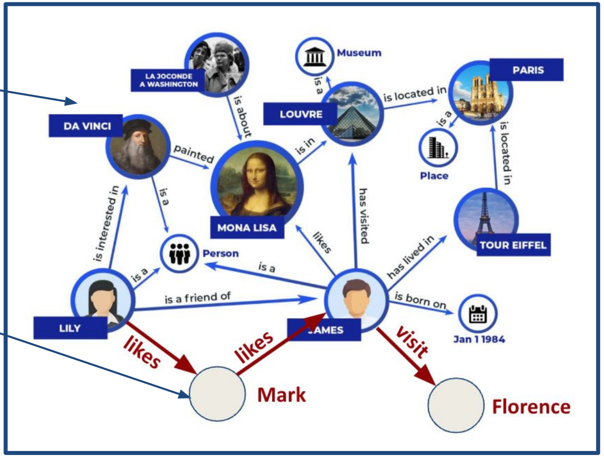
Recap: Graph Representations

Shallow KG embedding - Limitations

1. **Cannot handle node attributes**

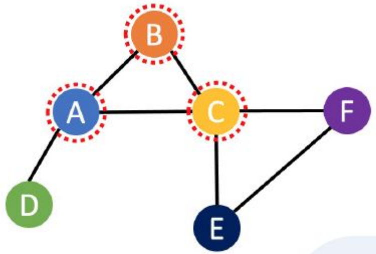
Text, Image, Categorical Attributes (e.g., age, gender, … etc.)

2. **Not inductive**
   a. Cannot handle new nodes

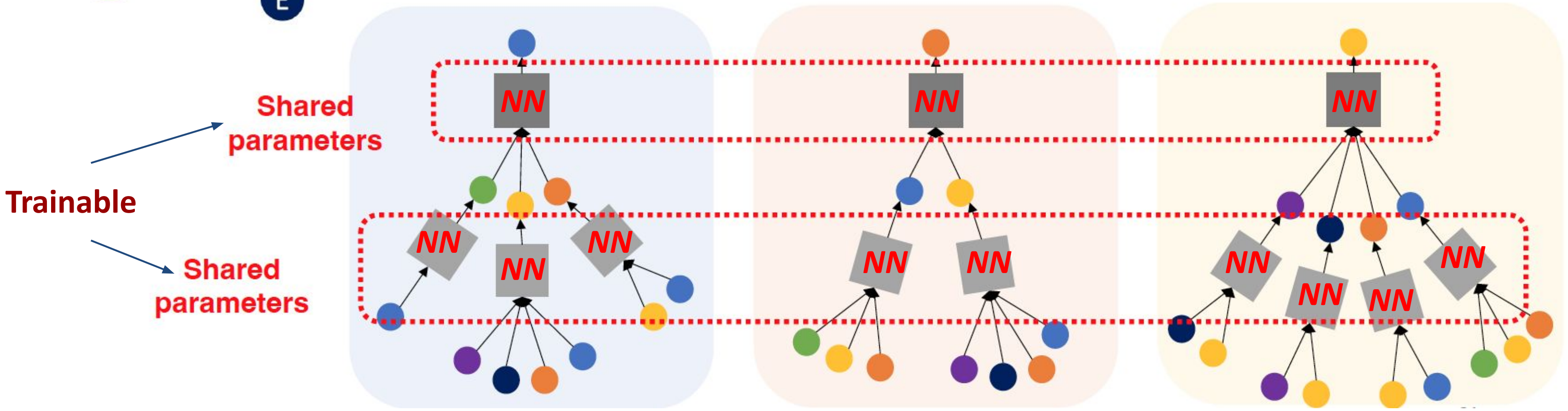


Recap: Graph Representations

Graph Neural Networks

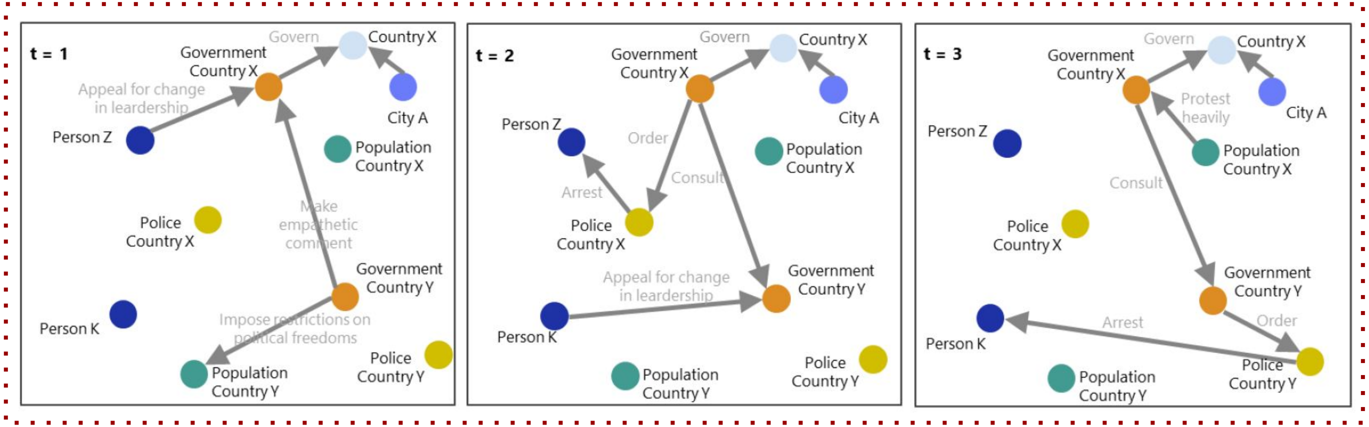


1. Node can have attribute features.
2. Available to use NN to create node representation for **unseen (NEW) nodes**.



Temporal Graph Neural Networks

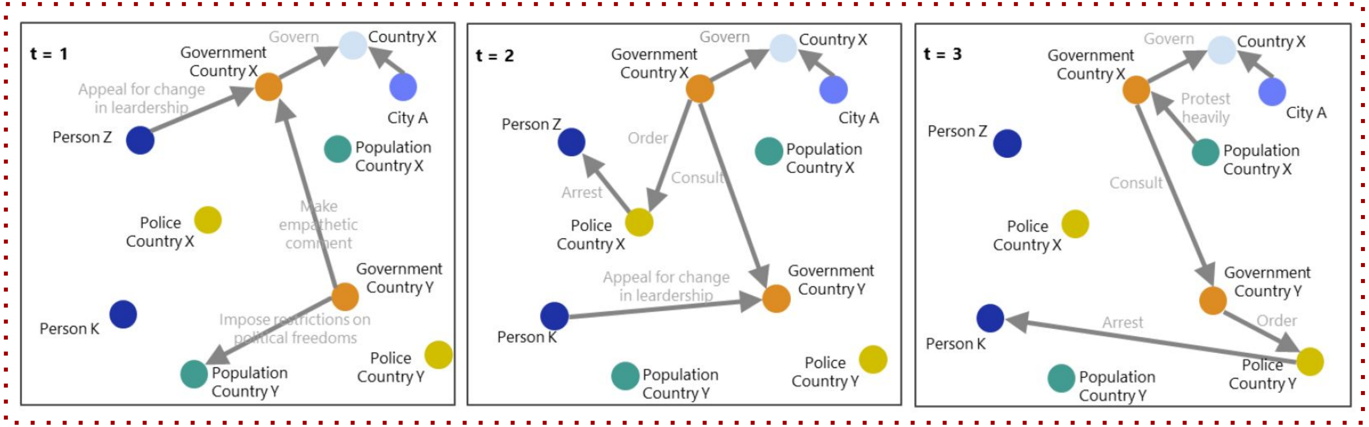
Message passing over Time



Terminology

Interpolation vs. Extrapolation

Message passing over Time



t = 4
(Person K, ?, Country Y)

Extrapolation
(Forecasting)

t = 2
(Person K, ?, Country Y)

Interpolation
(Inference on missing link in the past)

Terminology

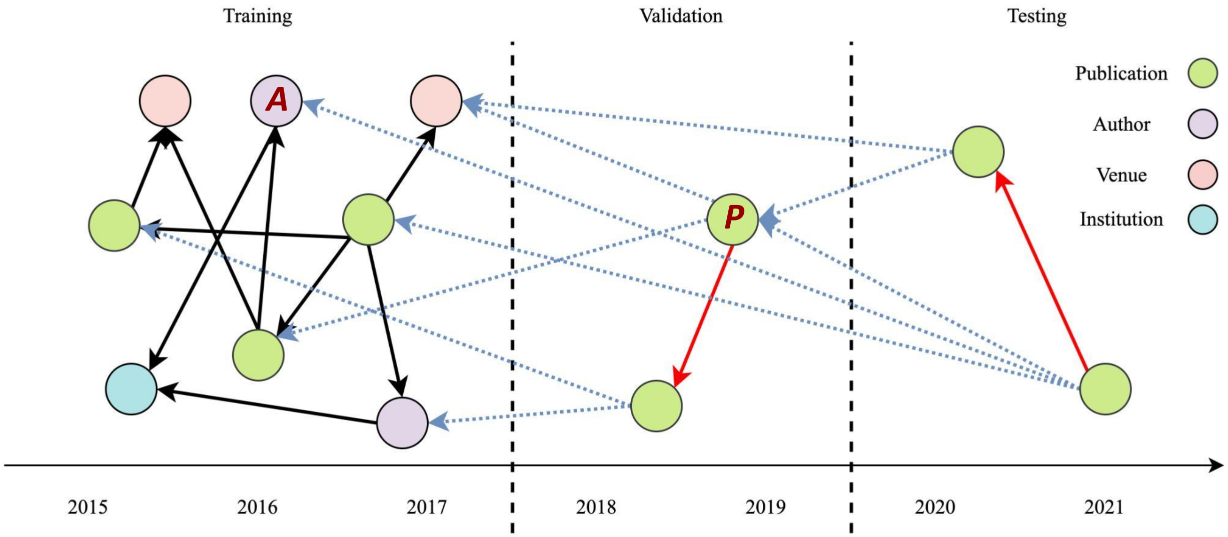
Transductive vs. Inductive

Inference with seen nodes → (Chelsea, ?, Tottenham, t4)

Terminology

Transductive vs. **Inductive**

Inference with New nodes → (Author A, ?, Publication P, 2019)



Slides by Kian Ahrabian

Temporal KG Completion

Data looks

Subject	Relation	Object	Time
DRC	Conflict:Attack	bandits	04/14/2021
leaders	Contact:Meet	Maputo	04/14/2021
north	Life:Die	people	04/15/2021
Mozambique	Movement:Transport	north	04/15/2021
province	Conflict:Attack	insurgents	04/15/2021

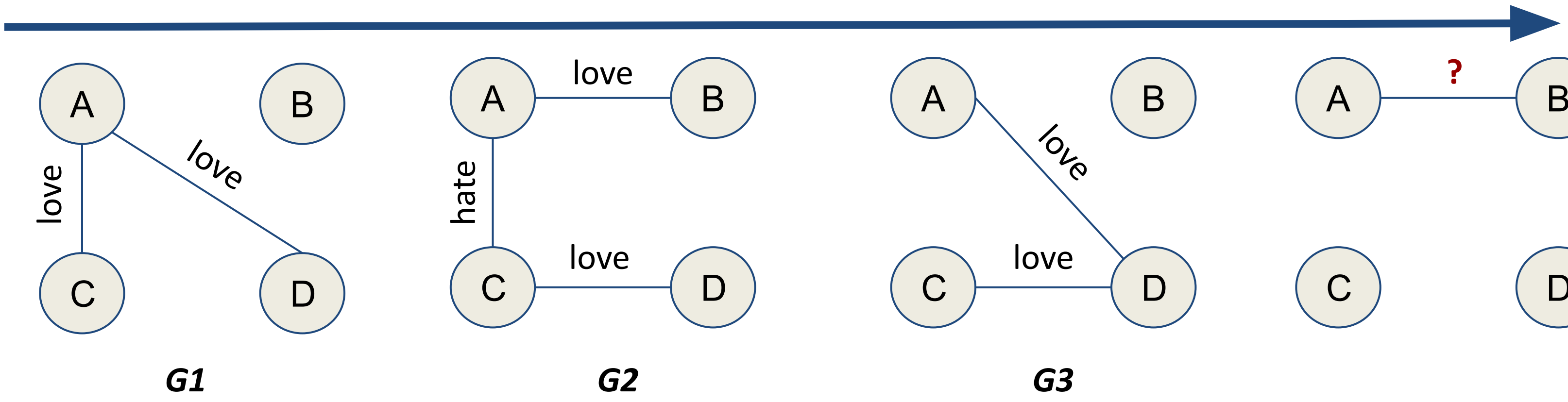
Relation_to_idx / Entity_to_idx

Subject	Relation	Object	Time
548	2	549	247
391	9	79	247
47	1	2	248
42	3	47	248
32	2	16	248

Train 80% / Valid 10% / Test 10%
Train time < Valid time < Test time

Transductive TKG Reasoning

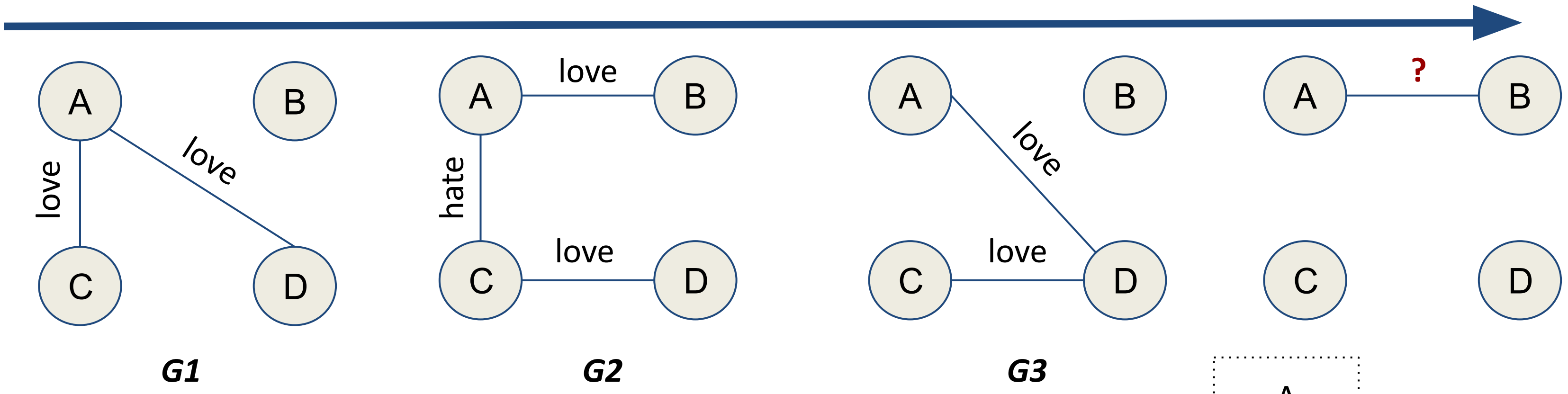
Autoregressive Model: RE-Net (2020)



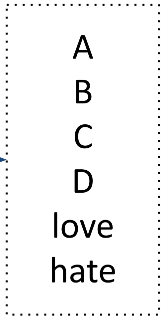
Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs., EMNLP 2020

Transductive TKG Reasoning

Autoregressive Model: RE-Net (2020)




```
self.ent = Parameter(torch.Tensor(num_nodes, hidden_channels))
self.rel = Parameter(torch.Tensor(num_rels, hidden_channels))
```

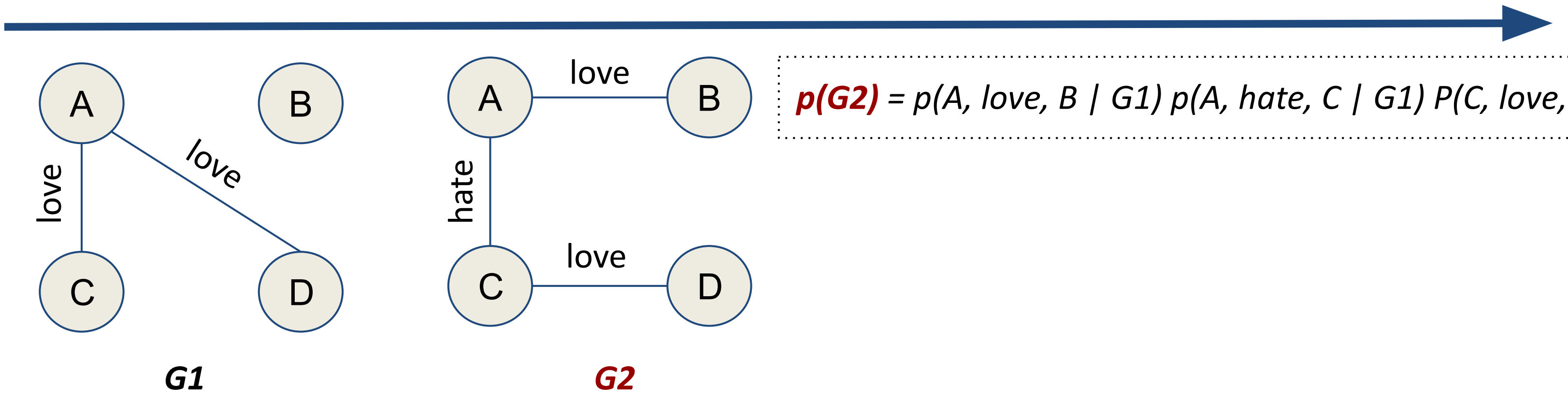


Each node & relation has randomly initialized embedding.

Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs., EMNLP 2020

Transductive TKG Reasoning

Autoregressive Model: RE-Net (2020)

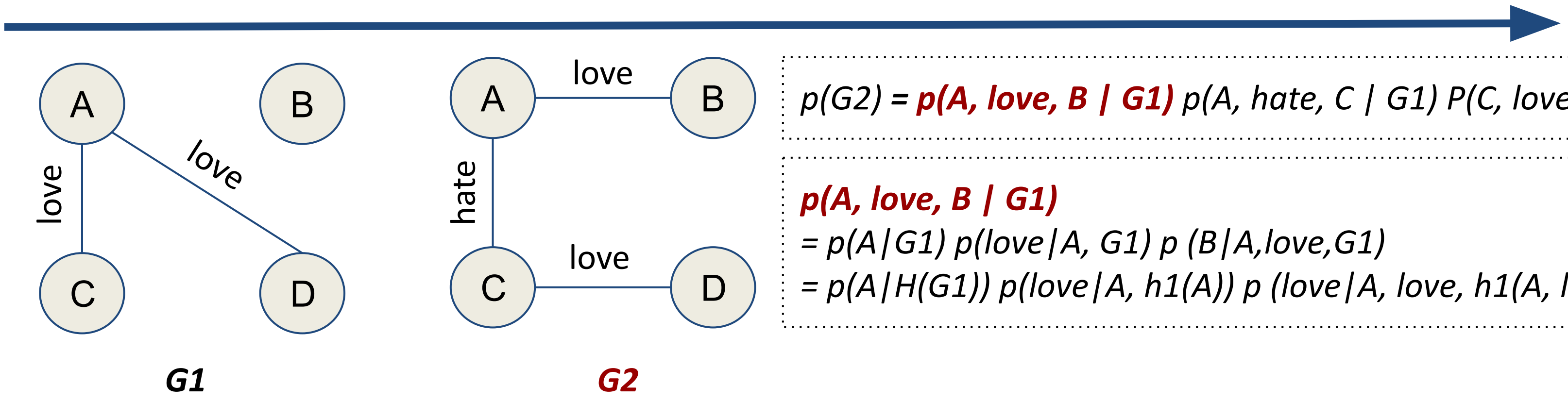


p(G2) *= p(A, love, B | G1) p(A, hate, C | G1) P(C, love, D | G1)*

Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs., EMNLP 2020

Transductive TKG Reasoning

Autoregressive Model: RE-Net (2020)



p(G2) = ***p(A, love, B | G1)*** *p(A, hate, C | G1) P(C, love, D | G1)*

p(A, love, B | G1)

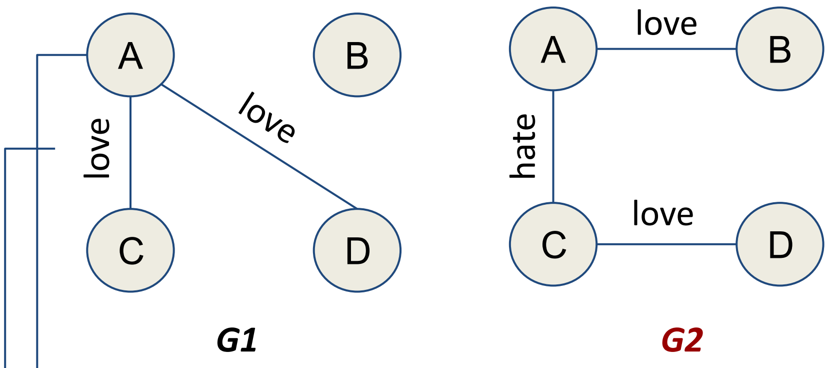
= p(A|G1) p(love|A, G1) p (B|A,love,G1)

= p(A|H(G1)) p(love|A, h1(A)) p (love|A, love, h1(A, love))

Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs., EMNLP 2020

Transductive TKG Reasoning

Autoregressive Model: RE-Net (2020)



$p(G2) = \mathbf{p(A, love, B \mid G1)}\ p(A, hate, C \mid G1)\ P(C, love, D \mid G1)$

$\mathbf{p(A, love, B \mid G1)}$

$= p(A \mid G1)\ p(love \mid A, G1)\ p\ (B \mid A, love, G1)$

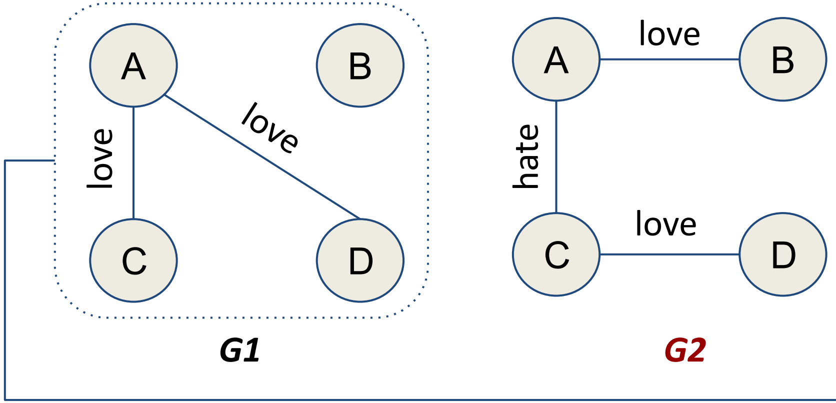
$= p(A \mid H(G1))\ p(love \mid A, \boxed{h1(A)})\ p\ (love \mid A, love, \boxed{h1(A, love)})$

Local Information

Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs., EMNLP 2020

Transductive TKG Reasoning

Autoregressive Model: RE-Net (2020)



$$p(G2) = \mathbf{p(A, love, B \mid G1)}\; p(A, hate, C \mid G1)\; P(C, love, D \mid G1)$$

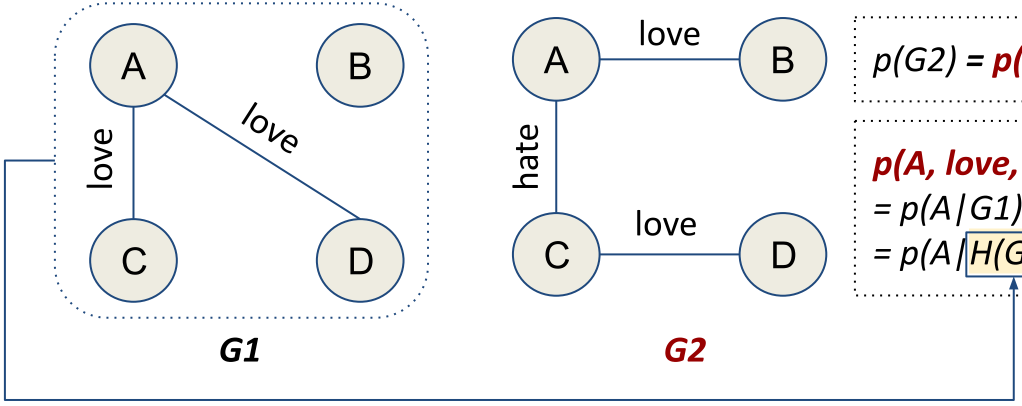
$$\mathbf{p(A, love, B \mid G1)}$$
$$= p(A \mid G1)\; p(love \mid A, G1)\; p\,(B \mid A, love, G1)$$
$$= p(A \mid \boxed{H(G1)})\; p(love \mid A, h1(A))\; p\,(love \mid A, love, h1(A, love))$$

Global Information

Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs., EMNLP 2020

Transductive TKG Reasoning

Autoregressive Model: RE-Net (2020)



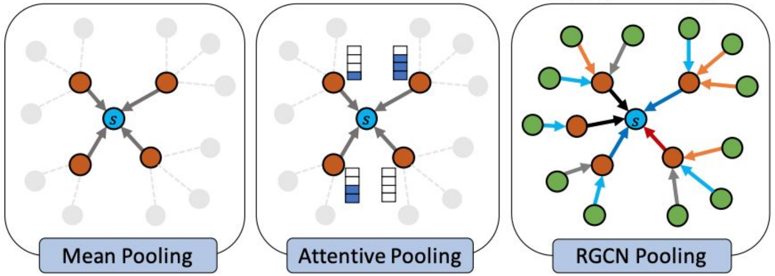
$p(G2) = \mathbf{p(A, love, B \mid G1)}\ p(A, hate, C \mid G1)\ P(C, love, D \mid G1)$

$\mathbf{p(A, love, B \mid G1)}$

$= p(A|G1)\ p(love|A, G1)\ p(B|A,love,G1)$

$= p(A|\boxed{H(G1)})\ p(love|A, h1(A))\ p(love|A, love, h1(A, love))$

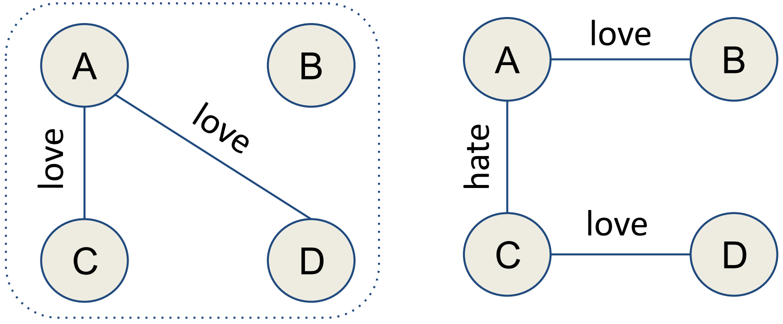
Aggregation → Global Information



Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs., EMNLP 2020

Transductive TKG Reasoning

Autoregressive Model: RE-Net (2020)



G1 ***G2***

p(G2) = ***p(A, love, B | G1)*** *p(A, hate, C | G1) P(C, love, D | G1)*

p(A, love, B | G1)

= p(A|G1) p(love|A, G1) p (B|A,love,G1)

= p(A|$\boxed{\text{H(G1)}}$) p(love|A, $\boxed{\text{h1(A)}}$) p (love|A, love, $\boxed{\text{h1(A, love)}}$)

$$\boldsymbol{H}_t = \mathrm{RNN}^1(g(G_t), \boldsymbol{H}_{t-1}),$$

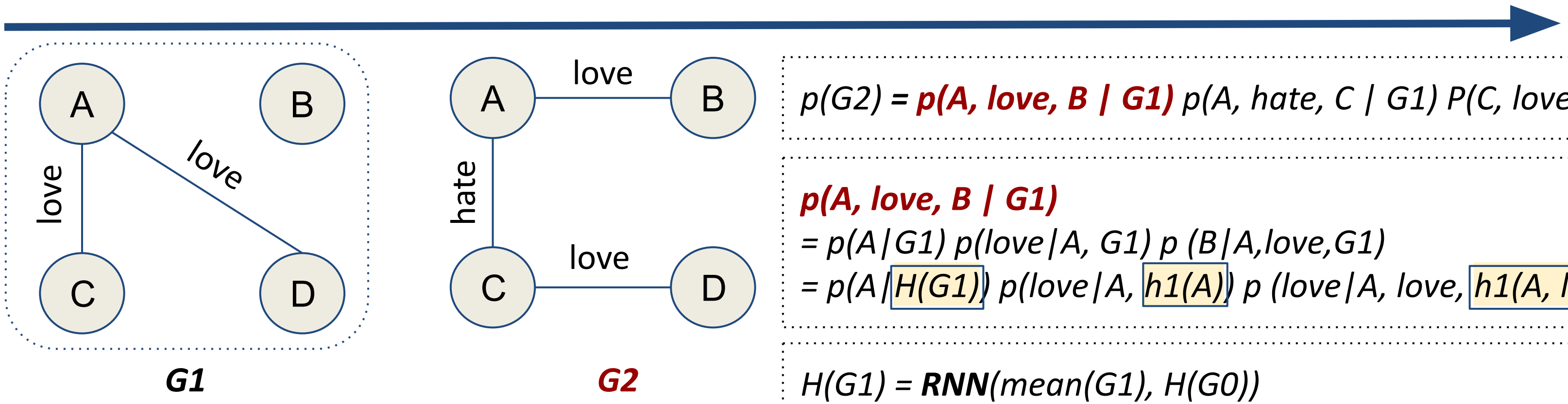
$$\boldsymbol{h}_t(\mathrm{s}, \mathrm{r}) = \mathrm{RNN}^2(g(\mathrm{N}_t^{(\mathrm{s})}), \boldsymbol{H}_t, \boldsymbol{h}_{t-1}(\mathrm{s}, \mathrm{r})),$$

$$\boldsymbol{h}_t(\mathrm{s}) = \mathrm{RNN}^3(g(\mathrm{N}_t^{(\mathrm{s})}), \boldsymbol{H}_t, \boldsymbol{h}_{t-1}(\mathrm{s})),$$

Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs., EMNLP 2020

Transductive TKG Reasoning

Autoregressive Model: RE-Net (2020)



p(G2) = ***p(A, love, B | G1)*** *p(A, hate, C | G1) P(C, love, D | G1)*

p(A, love, B | G1)

= p(A|G1) p(love|A, G1) p (B|A,love,G1)

= p(A|H(G1)) p(love|A, h1(A)) p (love|A, love, h1(A, love))

H(G1) = ***RNN****(mean(G1), H(G0))*

h1(A, love) = ***RNN****(mean(N1(A)), H(G1), h0(A, love))*

h1(A) = ***RNN****(mean(N1(A)), H(G1), h0(A))*

N1(A) = Graph of all the historical events associated with A.

Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs., EMNLP 2020

Transductive TKG Reasoning

Autoregressive Model: RE-Net (2020)

```python
def get_history(
    self,
    hist: List[int],
    node: int,
    rel: int,
) -> Tuple[Tensor, Tensor]:
    hists, ts = [], []
    for s in range(seq_len):
        h = hist[node][s]
        hists += h
        ts.append(torch.full((len(h), ), s, dtype=torch.long))
    node, r = torch.tensor(hists, dtype=torch.long).view(
        -1, 2).t().contiguous()
    node = node[r == rel]
    t = torch.cat(ts, dim=0)[r == rel]
    return node, t
```

p(G2) = ***p(A, love, B | G1)*** *p(A, hate, C | G1) P(C, love, D | G1)*

p(A, love, B | G1)

= p(A|G1) p(love|A, G1) p (B|A,love,G1)

= p(A|H(G1)) p(love|A, h1(A)) p (love|A, love, h1(A, love))

H(G1) = ***RNN****(mean(G1), H(G0))*

h1(A, love) = ***RNN****(mean(N1(A)), H(G1), h0(A, love))*

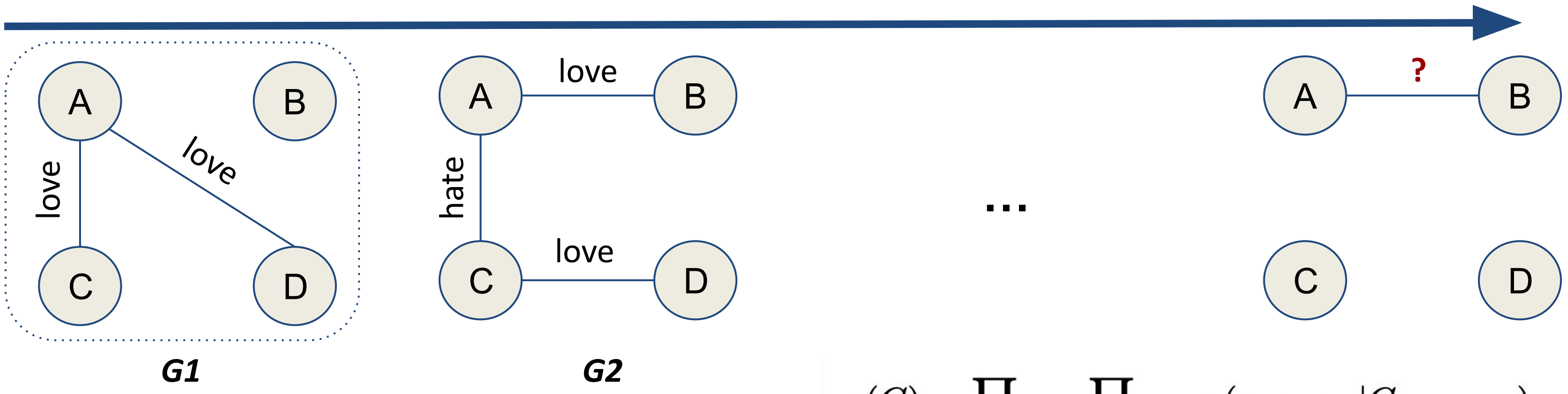
h1(A) = ***RNN****(mean(N1(A)), H(G1), h0(A))*

N1(A) = Graph of all the historical events associated with A.

Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs., EMNLP 2020

Transductive TKG Reasoning

Autoregressive Model: RE-Net (2020)



$$p(G) = \prod_t \prod_{(s_t, r_t, o_t) \in G_t} p(s_t, r_t, o_t | G_{t-m:t-1})$$
$$= \prod_t \prod_{(s_t, r_t, o_t) \in G_t} p(s_t | G_{t-m:t-1}) \cdot p(r_t | s_t, G_{t-m:t-1})$$
$$\cdot\, p(o_t | s_t, r_t, G_{t-m:t-1}). \quad (1)$$

Recurrent Event Network: Autoregressive Structure Inference over Temporal Knowledge Graphs., EMNLP 2020

Transductive TKG Reasoning

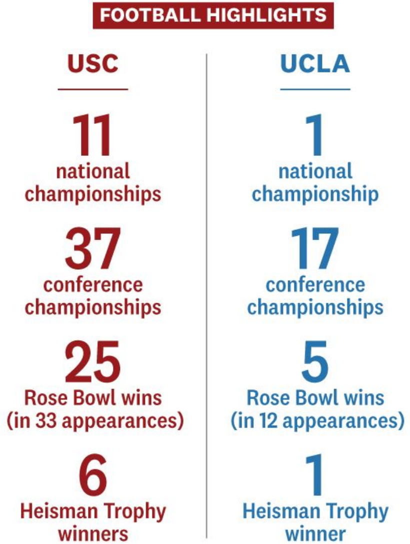
Heuristic-based Relevance: CyGNET (2021)

Motivation: Many facts occur repeatedly along the history.

→ Capture the repetition of temporal facts

Global economic crises

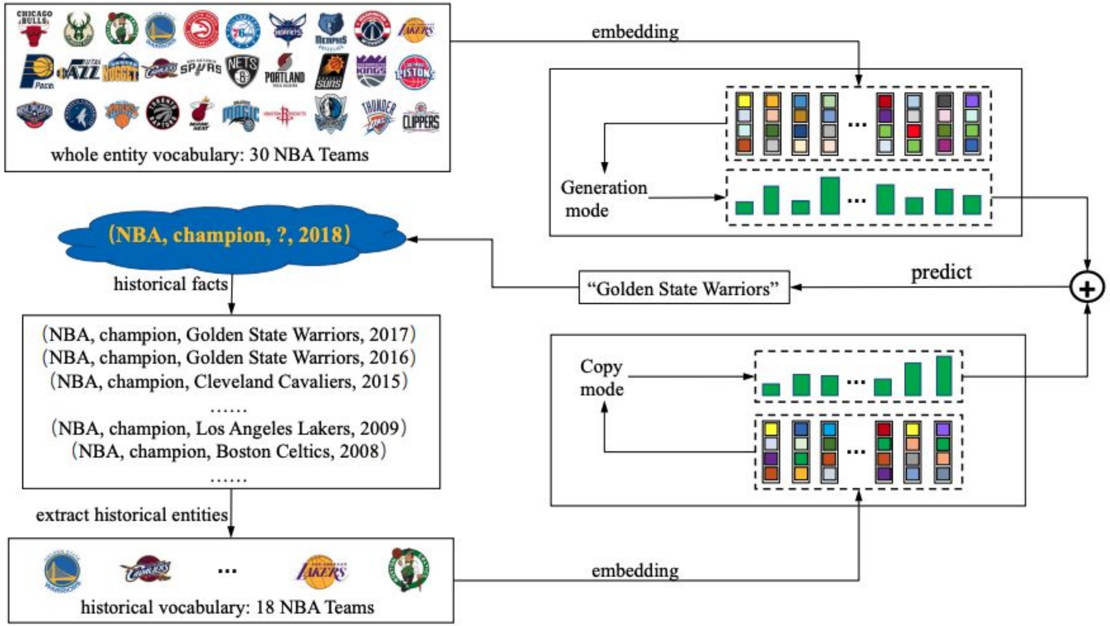
Brazil wins a lot



Learning from History: Modeling Temporal knowledge Graphs with Sequential Copy Generation Networks., AAAI 2021

Transductive TKG Reasoning

Heuristic-based Relevance: CyGNET (2021)



Generation Mode

Copy Mode

Learning from History: Modeling Temporal knowledge Graphs with Sequential Copy Generation Networks., AAAI 2021

Transductive TKG Reasoning

Heuristic-based Relevance: CyGNET (2021)

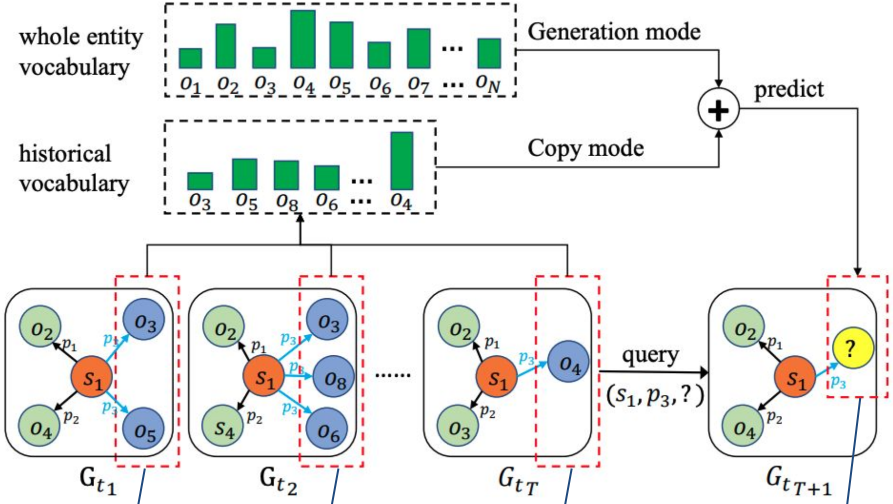
For each query *(s, p, ?, t)*,
extend historical vocabulary specific to *(s, p, t)*

t1:
- Tottenham, wins, Chelsea
- Tottenham, wins, Mancity

t2:
- Tottenham, wins, Chelsea
- Tottenham, wins, Newcastle
- Tottenham, wins, Arsenal

Ht1:
Chelsea
Mancity

Ht2:
Chelsea
Newcastle
Arsenal



Ht1(s1, p3) + Ht2(s1, p3) + HtT(s1, p3) = Ht(T+1)(s1, p3)

Learning from History: Modeling Temporal knowledge Graphs with Sequential Copy Generation Networks., AAAI 2021

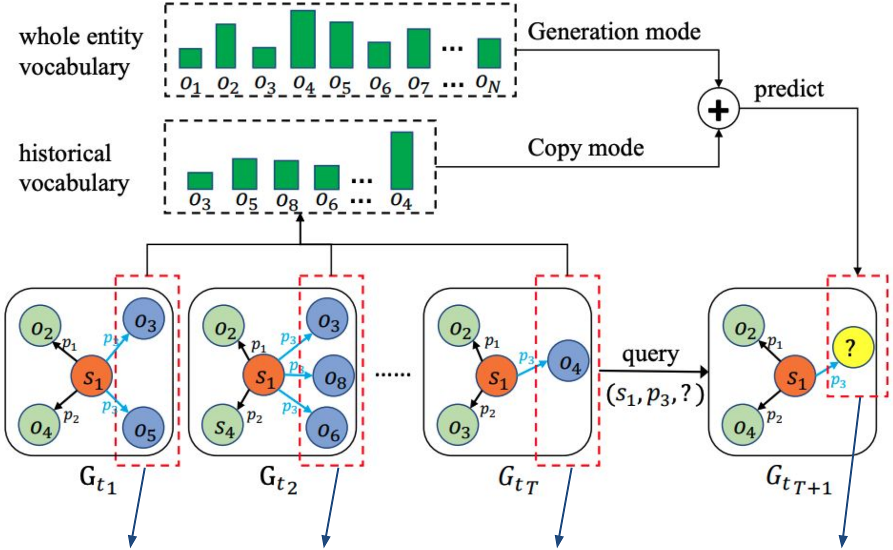
Transductive TKG Reasoning

Heuristic-based Relevance: CyGNET (2021)

For each query *(s, p, ?, t)*,
extend historical vocabulary specific to *(s, p, t)*

Ht2(Tottenham, Wins):
Chelsea
Newcastle
Arsenal

Arsenal	1
Mancity	0
Chelsea	1
Newcastle	1
…	…
Southhampton	0



Ht1(s1, p3) + Ht2(s1, p3) + HtT(s1, p3) = Ht(T+1)(s1, p3)

Learning from History: Modeling Temporal knowledge Graphs with Sequential Copy Generation Networks., AAAI 2021

Transductive TKG Reasoning

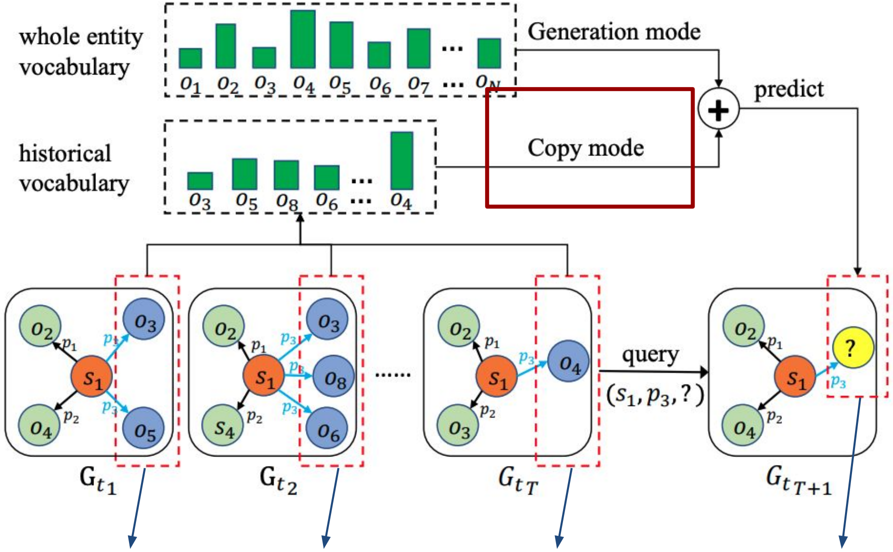
Heuristic-based Relevance: CyGNET (2021)

Copy Mode: Select known facts based on the historical vocabulary of Htk(s, p)

Tottenham, wins, t

$$\mathbf{v}_q = \tanh(\mathbf{W}_c[\mathbf{s}, \mathbf{p}, \mathbf{t_k}] + \mathbf{b}_c)$$

Trainable parameters



Ht1(s1, p3) + Ht2(s1, p3) + HtT(s1, p3) = Ht(T+1)(s1, p3)

Learning from History: Modeling Temporal knowledge Graphs with Sequential Copy Generation Networks., AAAI 2021

Transductive TKG Reasoning

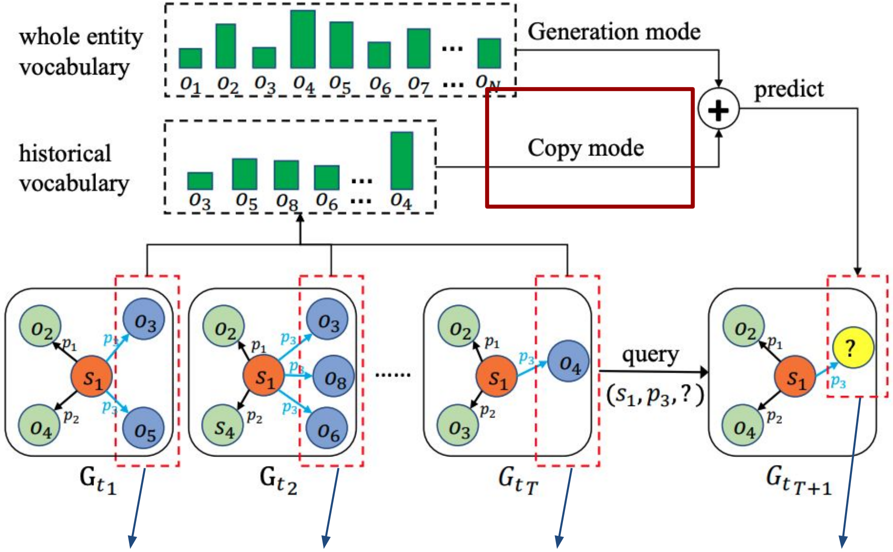
Heuristic-based Relevance: CyGNET (2021)

Copy Mode: Select known facts based on the historical vocabulary of Htk(s, p)

$$\mathbf{v}_q = \tanh(\mathbf{W}_c[\mathbf{s}, \mathbf{p}, \mathbf{t_k}] + \mathbf{b}_c)$$

$$\mathbf{c}_q = \mathbf{v}_q + \dot{\mathbf{H}}_{t_k}^{(s,p)}$$

Uninterested word
(not in history word) = -1



Ht1(s1, p3) + Ht2(s1, p3) + HtT(s1, p3) = Ht(T+1)(s1, p3)

Learning from History: Modeling Temporal knowledge Graphs with Sequential Copy Generation Networks., AAAI 2021

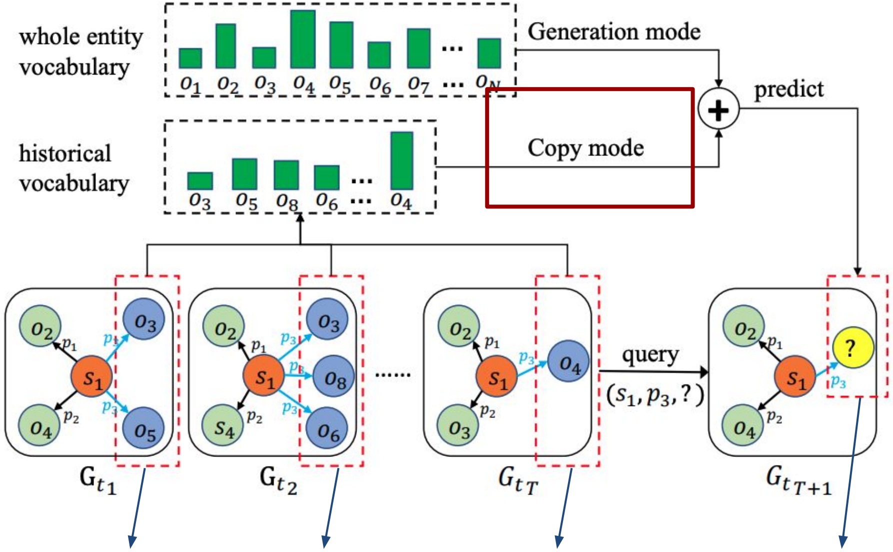
Transductive TKG Reasoning

Heuristic-based Relevance: CyGNET (2021)

Copy Mode: Select known facts based on the historical vocabulary of Htk(s, p)

Ht2'(Tottenham, Wins):
Chelsea
Newcastle
Arsenal

Arsenal	1
Mancity	-1
Chelsea	1
Newcastle	1
…	…
Southhampton	-1



Ht1(s1, p3) + Ht2(s1, p3) + HtT(s1, p3) = Ht(T+1)(s1, p3)

Learning from History: Modeling Temporal knowledge Graphs with Sequential Copy Generation Networks., AAAI 2021

Transductive TKG Reasoning

Heuristic-based Relevance: CyGNET (2021)

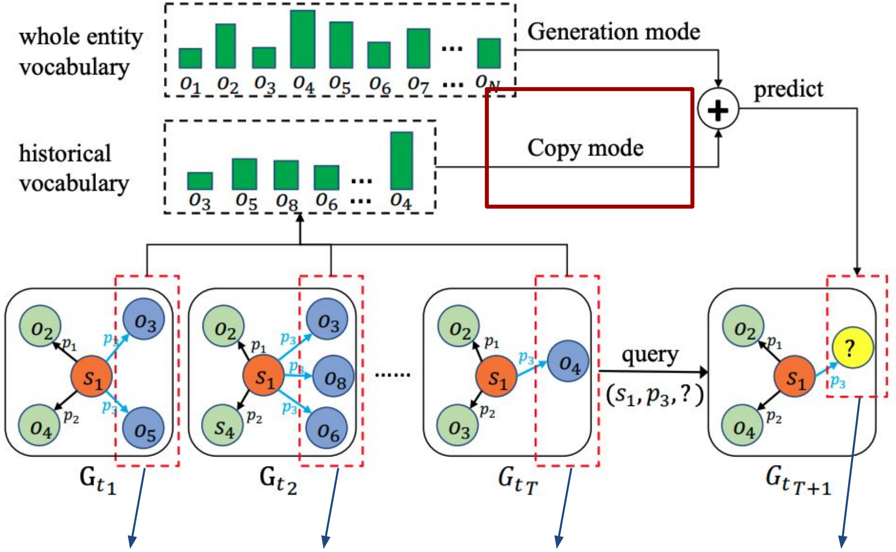
Copy Mode: Select known facts based on the historical vocabulary of Htk(s, p)

$$\mathbf{v}_q = \tanh(\mathbf{W}_c[\mathbf{s}, \mathbf{p}, \mathbf{t_k}] + \mathbf{b}_c)$$

$$\mathbf{c}_q = \mathbf{v}_q + \dot{\mathbf{H}}_{t_k}^{(s,p)}$$

$$\mathbf{p}(c) = \mathrm{softmax}(\mathbf{c}_q)$$

Probability of the whole entity vocabulary



Ht1(s1, p3) + Ht2(s1, p3) + HtT(s1, p3) = Ht(T+1)(s1, p3)

Learning from History: Modeling Temporal knowledge Graphs with Sequential Copy Generation Networks., AAAI 2021

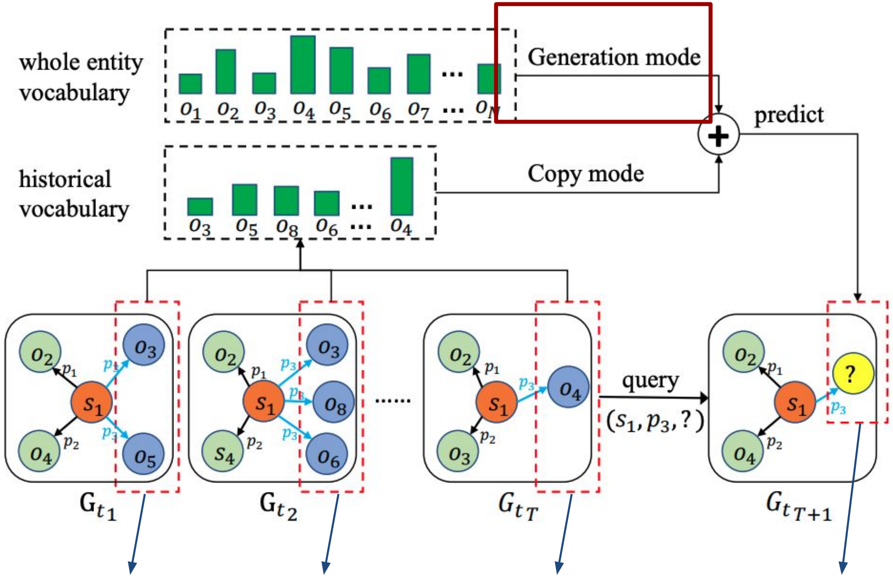
Transductive TKG Reasoning

Heuristic-based Relevance: CyGNET (2021)

Generation Mode: Predict future facts from the whole entity vocabulary

$$\mathbf{g}_q = \mathbf{W}_g[\mathbf{s}, \mathbf{p}, \mathbf{t_k}] + \mathbf{b}_g$$

$$\mathbf{p}(g) = \text{softmax}(\mathbf{g}_q)$$



Ht1(s1, p3) + Ht2(s1, p3) + HtT(s1, p3) = Ht(T+1)(s1, p3)

Learning from History: Modeling Temporal knowledge Graphs with Sequential Copy Generation Networks., AAAI 2021

Transductive TKG Reasoning

RE-Net, RE-GCN, CyGNet

Datasets	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{R}\rvert$	$\lvert\mathcal{E}_{train}\rvert$	$\lvert\mathcal{E}_{valid}\rvert$	$\lvert\mathcal{E}_{test}\rvert$	$\lvert\mathcal{E}^{s}\rvert$	$\lvert\mathcal{V}^{s}\rvert$	Time interval
ICEWS18	23,033	256	373,018	45,995	49545	29,774	8,647	24 hours
ICEWS14	6,869	230	74,845	8,514	7,371	8,442	3,499	24 hours
ICEWS05-15	10,094	251	368,868	46,302	46,159	12,392	5,179	24 hours

Model	ICE18				ICE14				ICE05-15			
	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10	MRR	H@1	H@3	H@10
DistMult	13.86	5.61	15.22	31.26	20.32	6.13	27.59	46.61	19.91	5.63	27.22	47.33
ComplEx	15.45	8.04	17.19	30.73	22.61	9.88	28.93	47.57	20.26	6.66	26.43	47.31
R-GCN	15.05	8.13	16.49	29.00	28.03	19.42	31.95	44.83	27.13	18.83	30.41	43.16
ConvE	22.81	13.63	25.83	41.43	30.30	21.30	34.42	47.89	31.40	21.56	35.70	50.96
ConvTransE	23.22	14.26	26.13	41.34	31.50	22.46	34.98	50.03	30.28	20.79	33.80	49.95
RotatE	14.53	6.47	15.78	31.86	25.71	16.41	29.01	45.16	19.01	10.42	21.35	36.92
HyTE	7.41	3.10	7.33	16.01	16.78	2.13	24.84	43.94	16.05	6.53	20.20	34.72
TTransE	8.44	1.85	8.95	22.38	12.86	3.14	15.72	33.65	16.53	5.51	20.77	39.26
TA-DistMult	16.42	8.60	18.13	32.51	26.22	16.83	29.72	45.23	27.51	17.57	31.46	47.32
RGCRN	23.46	14.24	26.62	41.96	33.31	24.08	36.55	51.54	35.93	26.23	40.02	54.63
CyGNet	24.98	15.54	28.58	43.54	34.68	25.35	38.88	53.16	35.46	25.44	40.20	54.47
RE-NET	26.17	16.43	29.89	44.37	35.77	25.99	40.10	54.87	36.86	26.24	41.85	57.60

Inductive TKG Reasoning

Data looks

Train 80% / **Valid 10% / Test 10%**
Train time < **Valid time < Test time**

Subject	Relation	Object	Time
DRC	Conflict:Attack	bandits	04/14/2021
leaders	Contact:Meet	Maputo	04/14/2021
north	Life:Die	people	04/15/2021
Mozambique	Movement:Transport	north	04/15/2021
province	Conflict:Attack	insurgents	04/15/2021

Relation_to_idx / Entity_to_idx

Subject	Relation	Object	Time
548	2	549	247
391	9	79	247
47	1	2	248
42	3	47	248
32	2	16	248

Same as Transductive TKG setting.
Only difference: Validation, Test sets contain triples with "unseen entity (Not in the train data)"

Temporal Graph Networks for Deep Learning on Dynamic Graphs., ICML workshop 2020

Inductive TKG Reasoning
Evaluation

	Wikipedia	
	Transductive	Inductive
GAE*	91.44 ± 0.1	†
VAGE*	91.34 ± 0.3	†
DeepWalk*	90.71 ± 0.6	†
Node2Vec*	91.48 ± 0.3	†
GAT*	**94.73** ± 0.2	91.27 ± 0.4
GraphSAGE*	93.56 ± 0.3	91.09 ± 0.3
CTDNE	92.17 ± 0.5	†
Jodie	94.62 ± 0.5	**93.11** ± 0.4
TGAT	**95.34** ± 0.1	**93.99** ± 0.3
DyRep	94.59 ± 0.2	92.05 ± 0.3
TGN-attn	**98.46** ± 0.1	**97.81** ± 0.1

Easily change the setting to inductive by different sampling strategy.

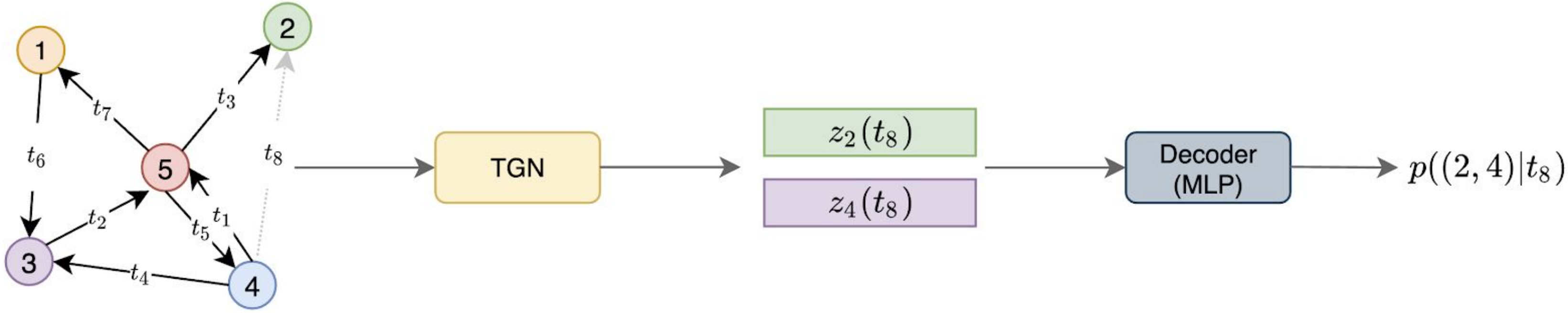
Same as Transductive TKG setting.

Only difference: Validation, Test sets contain triples with "unseen entity (Not in the train data)"

Temporal Graph Networks for Deep Learning on Dynamic Graphs., ICML workshop 2020

Inductive TKG Reasoning

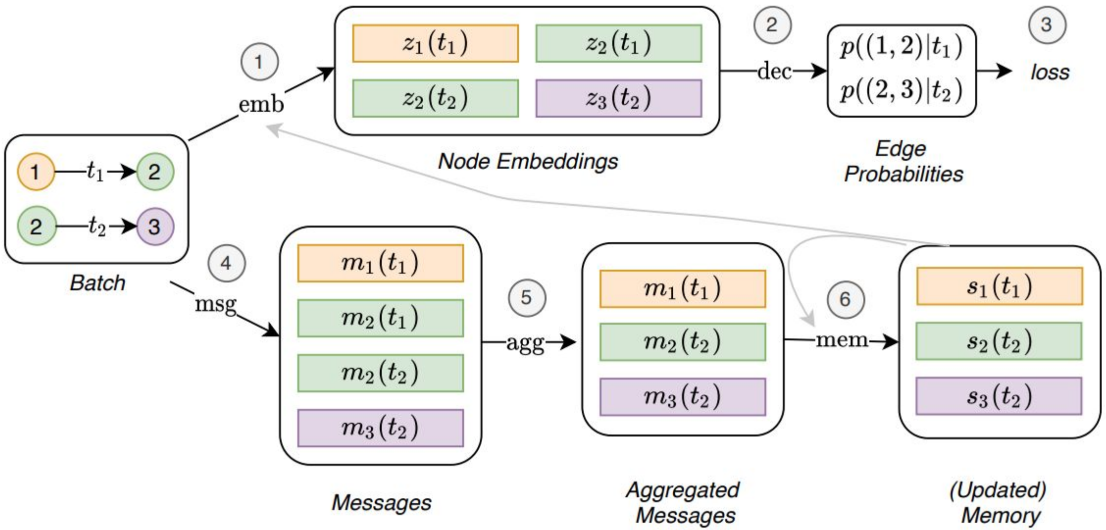
TGN (2020)



TGN is an encoder model which is able to generate temporal node embeddings ***z(t)*** for any node and time.

Temporal Graph Networks for Deep Learning on Dynamic Graphs., ICML workshop 2020

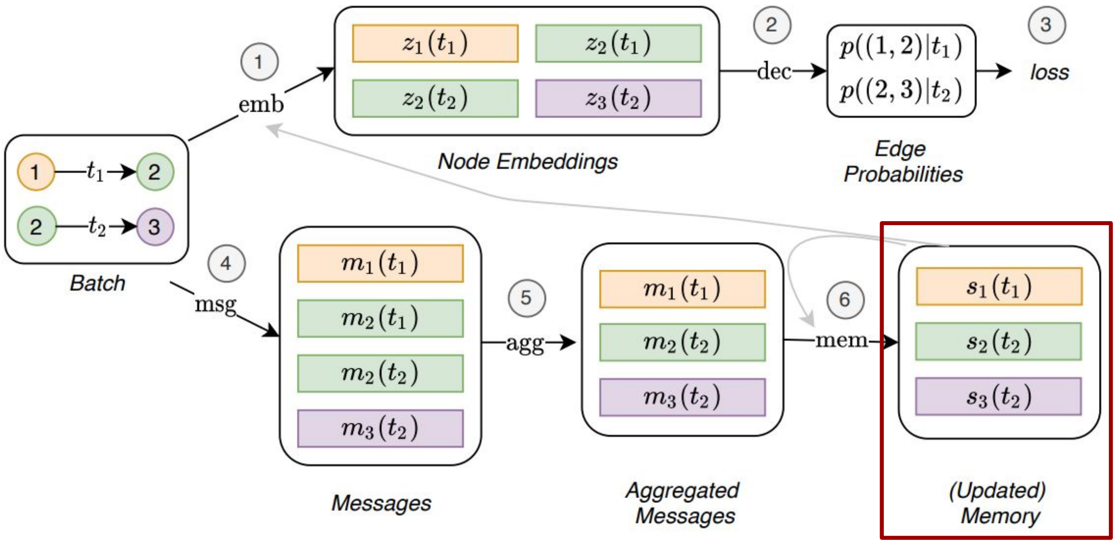
Inductive TKG Reasoning
TGN (2020) - Train



Temporal Graph Networks for Deep Learning on Dynamic Graphs., ICML workshop 2020

Inductive TKG Reasoning

TGN (2020) - Memory



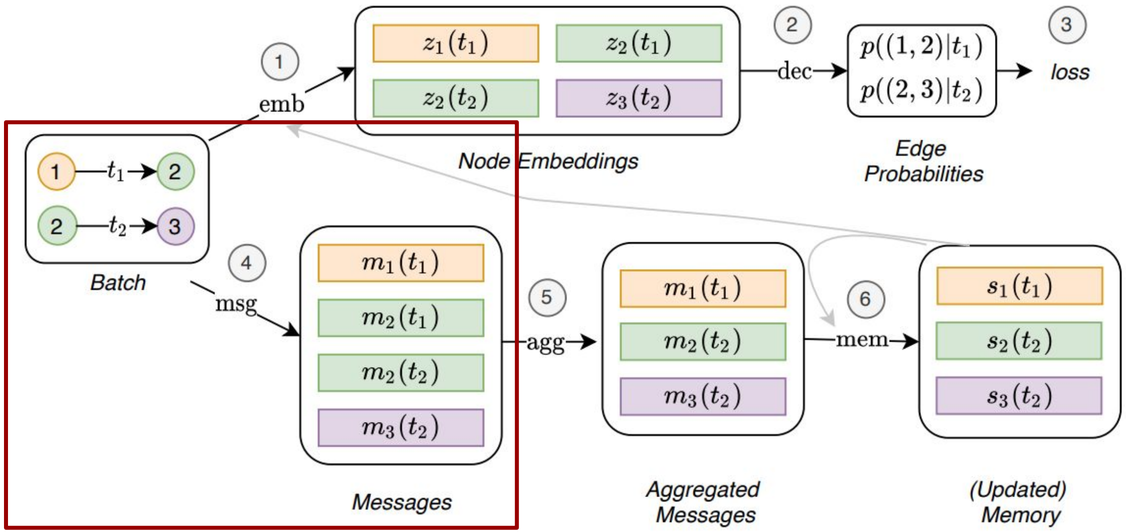
Database (Cache)

- state for each node the model has seen so far.
- Compressed representation of all past interactions of a node.
- Not a parameter, just storage.

Temporal Graph Networks for Deep Learning on Dynamic Graphs., ICML workshop 2020

Inductive TKG Reasoning

TGN (2020) - Message Function



Message Function

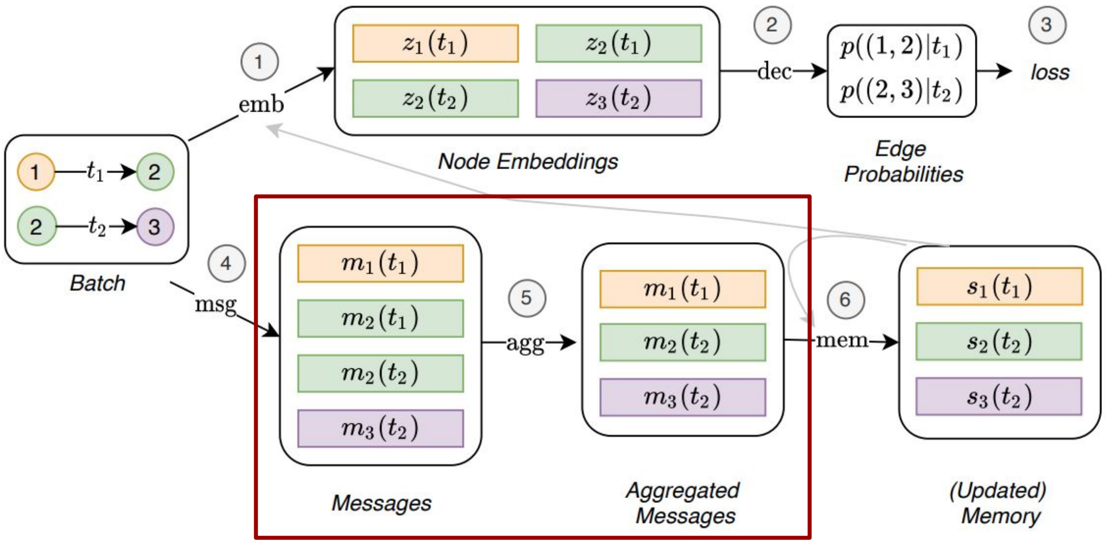
- Given an interaction between nodes ***i*** and ***j*** at the time ***t***, the message function (*MLP layer*) computes two messages one for ***i*** and one for ***j***.

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

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

Temporal Graph Networks for Deep Learning on Dynamic Graphs., ICML workshop 2020

Inductive TKG Reasoning
TGN (2020) - Aggregate Function



Aggregate Function

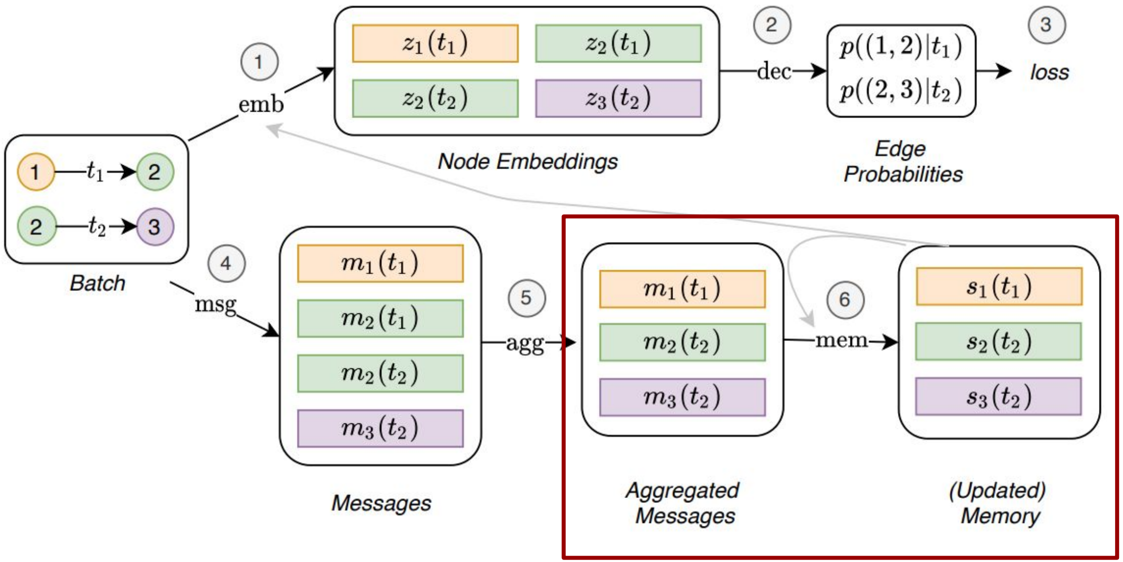
- Aggregate multiple messages for the same node in a batch.

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

Temporal Graph Networks for Deep Learning on Dynamic Graphs., ICML workshop 2020

Inductive TKG Reasoning

TGN (2020) - Update Memory



Update Memory

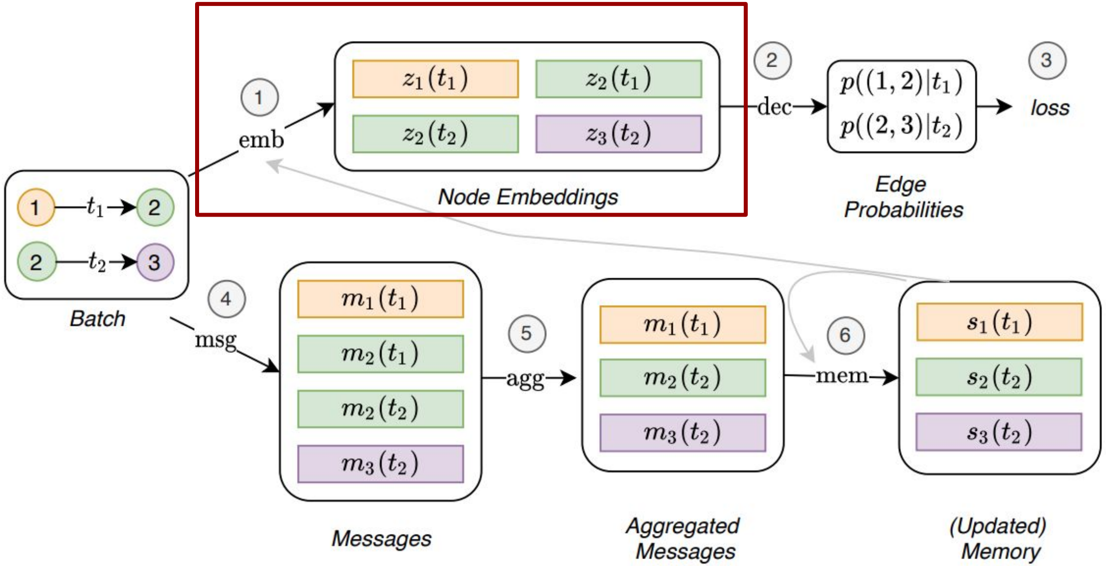
- Replace it. (Update the memory)

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

Temporal Graph Networks for Deep Learning on Dynamic Graphs., ICML workshop 2020

Inductive TKG Reasoning

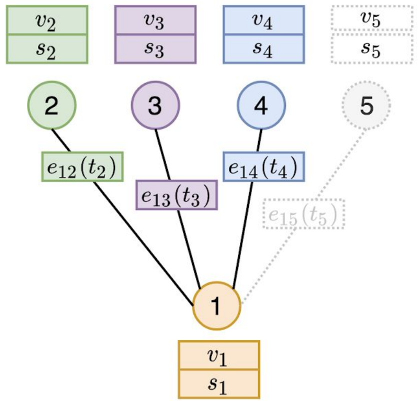
TGN (2020) - Temporal Graph Embedding



Temporal Graph Embedding

- Computes the embedding of a node.

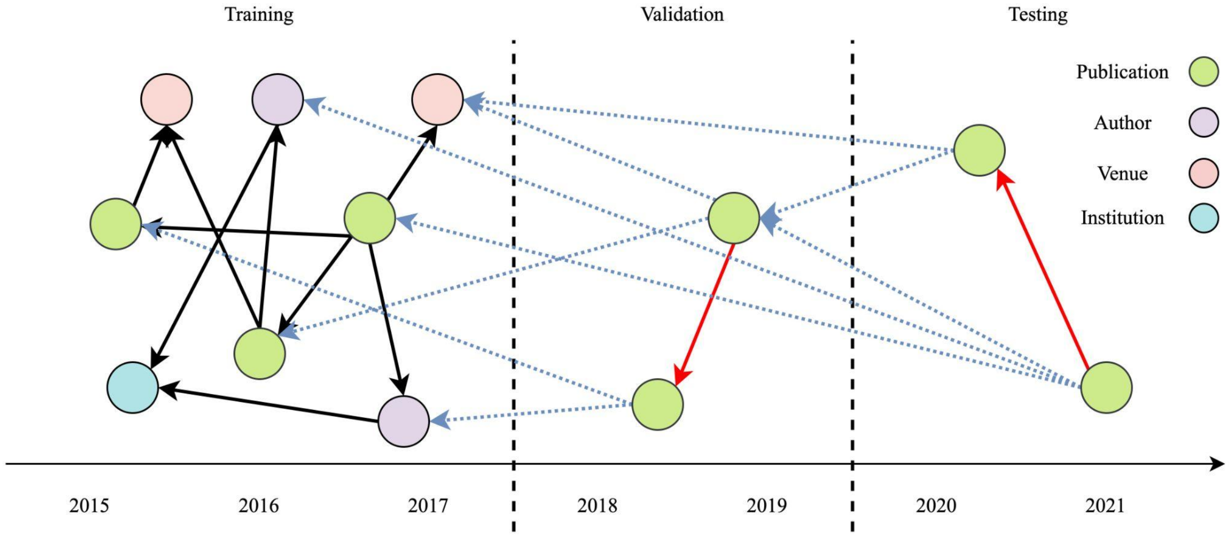
$$\mathbf{z}_i(t) = \text{emb}(i,t) = \sum_{j \in n_i^k([0,t])} h\left(\mathbf{s}_i(t), \mathbf{s}_j(t), \mathbf{e}_{ij}, \mathbf{v}_i(t), \mathbf{v}_j(t)\right)$$



Temporal Graph Networks for Deep Learning on Dynamic Graphs., ICML workshop 2020

Inductive TKG Reasoning

TGN (2020) - Experiments (Future Edge Prediction)



Whether there should be a edge between new nodes or new node and existing node?

Temporal Graph Networks for Deep Learning on Dynamic Graphs., ICML workshop 2020

Inductive TKG Reasoning

TGN (2020) - Experiments (Future Edge Prediction)

	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*	$\mathbf{94.73} \pm 0.2$	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	$\mathbf{96.27} \pm 0.2$	65.79 ± 0.6	60.13 ± 0.6
CTDNE	92.17 ± 0.5	†	91.41 ± 0.3	†	—	†
Jodie	94.62 ± 0.5	$\mathbf{93.11} \pm 0.4$	97.11 ± 0.3	94.36 ± 1.1	$\mathbf{85.20} \pm 2.4$	$\mathbf{79.83} \pm 2.5$
TGAT	$\mathbf{95.34} \pm 0.1$	$\mathbf{93.99} \pm 0.3$	$\mathbf{98.12} \pm 0.2$	$\mathbf{96.62} \pm 0.3$	70.02 ± 0.6	66.35 ± 0.8
DyRep	94.59 ± 0.2	92.05 ± 0.3	$\mathbf{97.98} \pm 0.1$	95.68 ± 0.2	$\mathbf{83.52} \pm 3.0$	$\mathbf{78.38} \pm 4.0$
TGN-attn	$\mathbf{98.46} \pm 0.1$	$\mathbf{97.81} \pm 0.1$	$\mathbf{98.70} \pm 0.1$	$\mathbf{97.55} \pm 0.1$	$\mathbf{94.52} \pm 0.5$	$\mathbf{91.37} \pm 1.1$

Temporal Graph Networks for Deep Learning on Dynamic Graphs., ICML workshop 2020

TKG Reasoning

Challenge 1: Time granularity / Distance

Time granularity:

- Example: month-granularity (Jan 2021 = 1, Feb 2021 = 2, …)
  - Interpolation Question:
    - Trump was no longer president of the US at Jan 19, 2021? → time index = 1
    - Trump was no longer president of the US at Jan 20, 2021? → time index = 1
    - Trump was no longer president of the US at Jan 21, 2021? → time index = 1
    - Can’t capture daily granularity in month granularity setup.
  - **Time granularity needs to be chosen accordingly.**

Time distance:

- Example:
  - going to a vacation for **3 weeks**.
  - Trump was the president of the United states from **2017** to **2021**
- **Need to create multiple quadruples facts within the time range.**

TKG Reasoning

Challenge 2: Sparse graph (Small train data for supervised learning)

Datasets	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{R}\rvert$	$\lvert\mathcal{E}_{train}\rvert$	$\lvert\mathcal{E}_{valid}\rvert$	$\lvert\mathcal{E}_{test}\rvert$	$\lvert\mathcal{E}^{s}\rvert$	$\lvert\mathcal{V}^{s}\rvert$	Time interval
ICEWS18	23,033	256	373,018	45,995	49545	29,774	8,647	24 hours

We have 23,033 node embeddings and 256 relation embeddings to train.

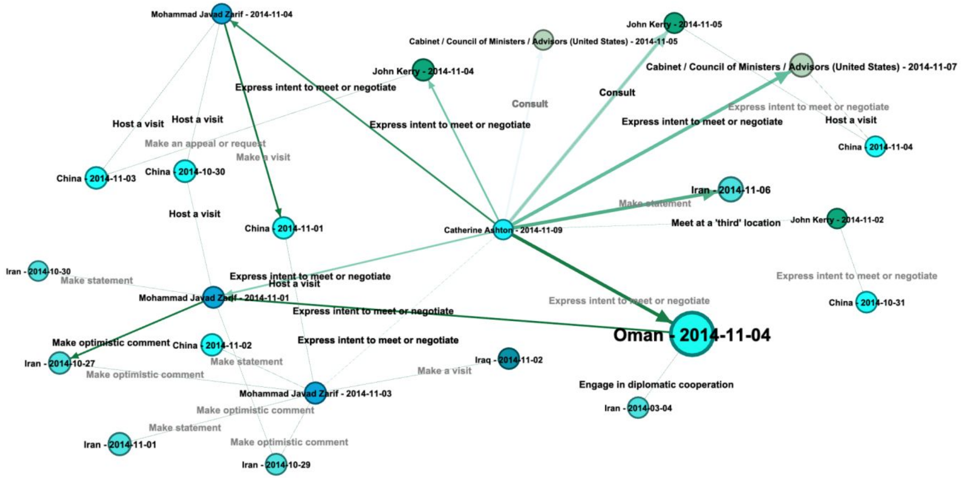
We have 373,018 triples that consists of nodes and relations to train.

Real world…

Dataset	# Entity	# Relation	# Train	# Valid	# Test	# Total
ACLED	243	2 (violence, Battle)	1,778 (2003-11-08 ~ 2021-06-27)	222 (2021-07-02 ~ 2021-10-06)	226 (2021-10-07 ~ 2022-01-27)	2,226
News	555	25	1,005 (2019-02-08 ~ 2021-03-15)	142 (2021-03-16 ~ 2021-03-28)	121 (2021-03-29 ~ 2021-04-16)	1,268
Telegram	522	21	6,450 (2021-01-01 ~ 2021-04-29)	571 (2021-04-30 ~ 2021-05-11)	1,014 (2021-05-12 ~ 2021-05-25)	8,035
Tweets	402	16	438 (2020-02-26 ~ 2020-07-06)	55 (2020-07-07 ~ 2020-07-28)	55 (2020-07-29 ~ 2020-08-26)	548

TKG Reasoning

Challenge 3: Explainability



(Catherine Ashton, Make a visit, ?, 2014-11-09)

Which subgraph contributes more for prediction?

Explainable Subgraph reasoning for Forecasting on Temporal Knowledge Graphs., ICLR 2021

TKG Reasoning

Challenge 3: Explainability

Confidence	Head	Body
0.963	$(E_1, \textit{demonstrate or rally}, E_2, T_4)$	$(E_1, \textit{riot}, E_2, T_1) \wedge (E_2, \textit{make statement}, E_1, T_2) \wedge (E_1, \textit{riot}, E_2, T_3)$
0.818	$(E_1, \textit{share information}, E_2, T_2)$	$(E_1, \textit{express intent to ease sanctions}^{-1}, E_2, T_1)$
0.750	$(E_1, \textit{provide military aid}, E_3, T_3)$	$(E_1, \textit{provide military aid}, E_2, T_1) \wedge (E_2, \textit{intend to protect}^{-1}, E_3, T_2)$
0.570	*(Merkel, consult, Obama, 14/08/09)*	*(Merkel, discuss by telephone, Obama, 14/07/22)*
0.500	*(Merkel, consult, Obama, 14/08/09)*	*(Merkel, express intent to meet, Obama, 14/05/02)* $\wedge$ *(Obama, consult*$^{-1}$*, Merkel, 14/07/18)* $\wedge$ *(Merkel, consult*$^{-1}$*, Obama, 14/07/29)*

(Angela Merkel, Consult, ?, 2014/08/09)

Rules that lead to the correct answer Barack Obama

TLogic: Temporal Logical Rules for Explainable Link Forecasting on Temporal Knowledge Graphs, AAAI 2022
Explainable Subgraph reasoning for Forecasting on Temporal Knowledge Graphs., ICLR 2021

TKG Reasoning

Challenge 4: Difficult to deal with practical forecasting questions

Existing temporal KG methods (Graph Snapshot@t) - *RE-Net*, *CYGNet*

- $G_1 \rightarrow G_2 \rightarrow G_3 \rightarrow \dots \rightarrow \boldsymbol{G_t}$
  - **Which region will be attacked by terrorists next?** **(terrorists, attack, <o>, *t*)**
  - **Who will attack cabo delgado next?** **(<s>, attack, cabo delgado, *t*)**

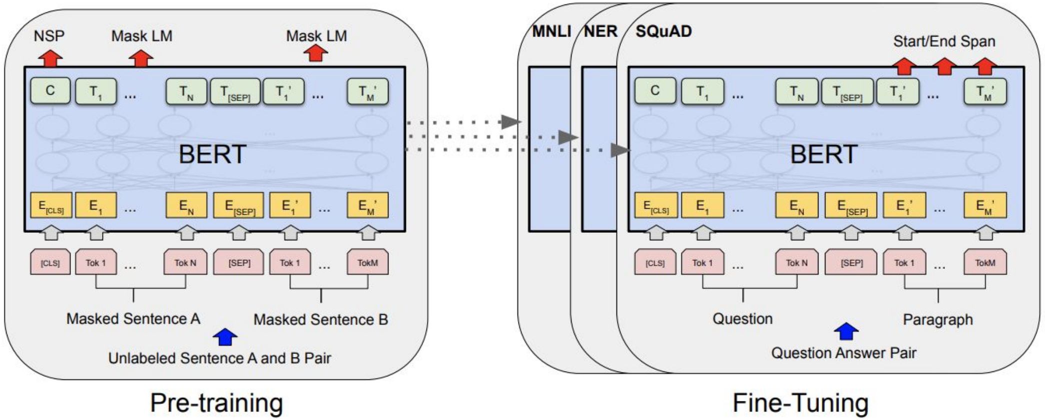
Avoid strong assumption ("Will the attack be in Cabo Delgado")

Our forecasting questions (querying pair / triple)

- **G_t (Conflict = any violence or battle)**
  - **Who are the most likely actors and their actions in cabo delgado?**
    - **Pair-level: (<s>, <p>, cabo delgado, *t*)**
  - **What will happen next in Mozambique?**
    - **Triple-level: (<s>, <p>, <o>, *t*)**

Other research lines of forecasting

Language Models



Original text

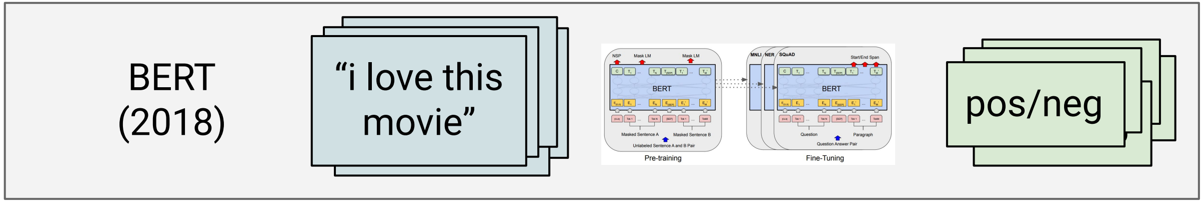
Thank you ~~for inviting~~ me to your party ~~last~~ week.

Inputs

Thank you <X> me to your party <Y> week.

Targets

<X> for inviting <Y> last <Z>



Other research lines of forecasting
Temporal QA data for Language Models

(Entity, Time) Prediction Questions

Reasoning	Example Template
Simple time	When did {head} hold the position of {tail}
Simple entity	Which award did {head} receive in {time}
Before/After	Who was the {tail} {type} {head}
First/Last	When did {head} play their {adj} game
Time join	Who held the position of {tail} during {event}

Example Question
When did Obama hold the position of President of USA
Which award did Brad Pitt receive in 2001
Who was the President of USA before Obama
When did Messi play their first game
Who held the position of President of USA during WWII

(Saxena et al., 2021)
Converting KG -> Natural Language Question

Yes-No Questions

(Sudan, x, x, 2021-08-01)
$\longrightarrow$
Will Sudan host Ramtane on 2021-08-01?

Forecast QA (Jin et al., 2020)
BoolQ (Clark et al., 2019)

Other research lines of forecasting

Forecast QA data for Language Models

	Train	Valid	Test
1-Hop Entity Prediction	252,246	42,991	39,786
2-Hop Entity Prediction	128,810	18,138	16,624
Yes-No	251,537	42,884	39,695
Fact Reasoning	10,103	2694	1859
Total	642,696	106,707	97,964

Fact Reasoning questions

Which of the following statements contributes most to the fact that Envoy (United States) visited China on 2021-08-31?

A. Envoy (United States) expressed the intent to meet or negotiate with China on 2021-08-30.	Answer
B. Envoy (United States) expressed the intent to meet or negotiate with Japan on 2021-08-30.	Hard Negative
C. North Korea criticized or denounced South Korea on 2021-08-22.	Median
D. South Korea had a consolation or a meeting with Tajikistan on 2021-08-19.	Negative

1-hop questions

(Sudan, host, Romatane, 2021-08-01)

-> Who will Sudan host on 2021-08-01?

2-hop questions

(Juan, make a visit, US, 2021-08-03)

(UK, engage in diplomatic cooperation, US, 2021-08-03)

-> Who will visit a country, while UK engages in diplomatic cooperation along with this country on 2021-08-03 ?

Yes-no questions

(Sudan, host, Romatane, 2021-08-01)

-> Will sudan host Romtane on 2021-08-01?

Forecasting Question Answering over Temporal Knowledge Graphs., 2022.08 ArXiv

Other research lines of forecasting

Fine-tune LM

	Train	Valid	Test
1-Hop Entity Prediction	252,246	42,991	39,786
2-Hop Entity Prediction	128,810	18,138	16,624
Yes-No	251,537	42,884	39,695
Fact Reasoning	10,103	2694	1859
Total	642,696	106,707	97,964

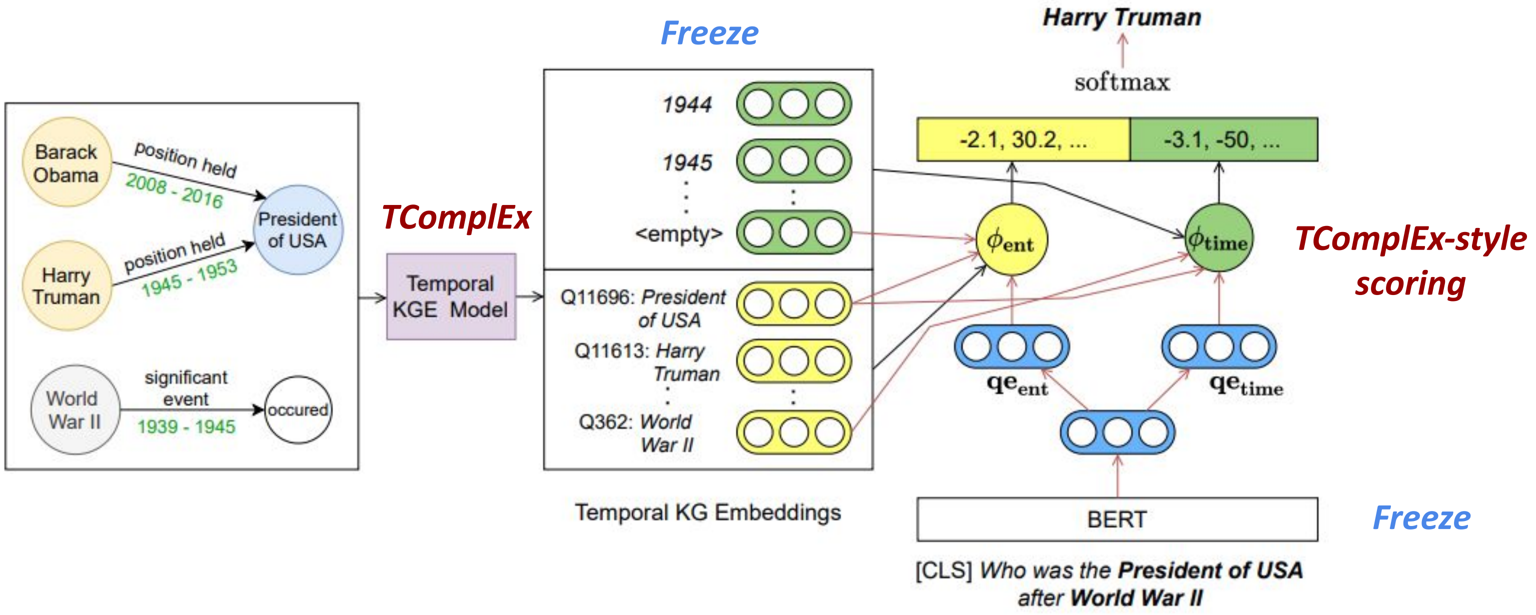
	MRR		
Model	Overall	1-Hop	2-Hop
RoBERTa	0.158	0.166	0.141
BERT	0.260	0.291	0.187

	Accuracy	
Question Type	Yes-No	Fact Reasoning
RoBERTa	0.750	0.514
BERT	0.823	0.567

Fine-tuning LM with such a huge train data can lead a decent performance without time-aware KG (e.g., TComplEX, … etc.)
→ Can be a good option to play with different type of questions.

Other research lines of forecasting

LM+KG



Question Answering over Temporal Knowledge Graph., ACL 2021

Other research lines of forecasting

LM+KG

Model	Hits@1					Hits@10				
	Overall	Question Type		Answer Type		Overall	Question Type		Answer Type	
		Complex	Simple	Entity	Time		Complex	Simple	Entity	Time
BERT	0.071	0.086	0.052	0.077	0.06	0.213	0.205	0.225	0.192	0.253
RoBERTa	0.07	0.086	0.05	0.082	0.048	0.202	0.192	0.215	0.186	0.231
KnowBERT	0.07	0.083	0.051	0.081	0.048	0.201	0.189	0.217	0.185	0.23
T5-3B	0.081	0.073	0.091	0.088	0.067	-	-	-	-	-
EmbedKGQA	0.288	0.286	0.29	0.411	0.057	0.672	0.632	0.725	0.85	0.341
T-EaE-add	0.278	0.257	0.306	0.313	0.213	0.663	0.614	0.729	0.662	0.665
T-EaE-replace	0.288	0.257	0.329	0.318	0.231	0.678	0.623	0.753	0.668	0.698
CRONKGQA	**0.647**	**0.392**	**0.987**	**0.699**	**0.549**	**0.884**	**0.802**	**0.992**	**0.898**	**0.857**

Question Answering over Temporal Knowledge Graph., ACL 2021

Any Questions?