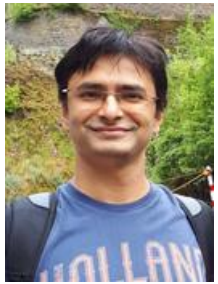
CIKM 2021
1-5 NOVEMBER

Complex Temporal Question Answering on Knowledge Graphs

Zhen Jia[1], Soumajit Pramanik[2], Rishiraj Saha Roy[3], Gerhard Weikum[3]






1. Southwest Jiaotong University, 2. IIT Bhilai, 3. Max Planck Institute for Informatics

西南交通大學 Southwest Jiaotong University

IIT BHILAI

max planck institut informatik

Temporal questions

asking for date

When was Obama born?

explicit

Where did Obama live in 2001?

event

What position was held by Obama during 9/11?

implicit

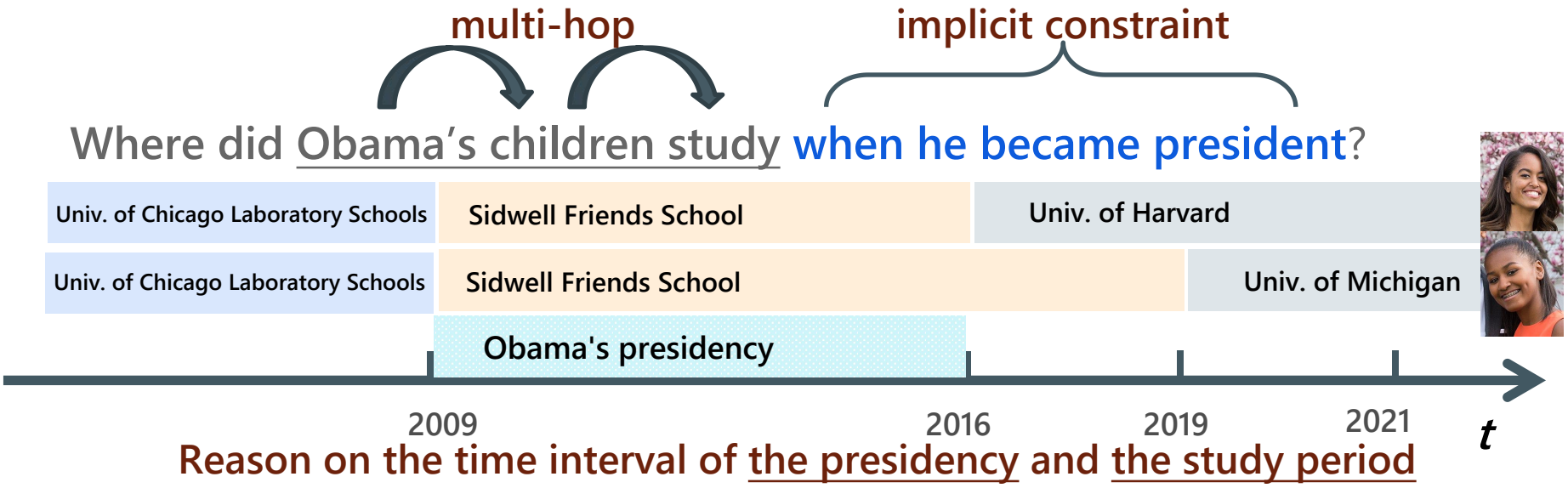
Where did Obama's children study when he became president?

simple

complex

Challenges

- ★ **Explicit, implicit or ordinal temporal** constraints
- ★ **Multi-hop** constraints
- ★ **Identify and reason** on time intervals



Related Work

★ **Rule-based** framework using question decomposition

Jia et al. 2018

★ **Benchmark** containing event-centric questions

Costa et al. 2020

★ Model focusing on **implicit temporal constraints**

Wu et al. 2020

★ Tool plugging **temporal layer** into existing QA system

Saquete et al. 2009

★ KG embeddings-based model on **Temporal KGs**

Saxena et al. 2021

Contributions in EXAQT

★ **EX**plainable **A**nswering of complex **Q**uestions with **T**emporal intent

- 🕓 **End-to-end system** for answering complex temporal questions over KGs
- 🕓 Fine-tuned BERT models to identify **relevant KG facts**
- 🕓 Graph algorithms to compute **compact question subgraphs**
- 🕓 Relational graph convolutional networks (R-GCNs) to predict answers with **time-enhanced mechanisms**

★ **TimeQuestions**: Benchmark with various types of temporal intents

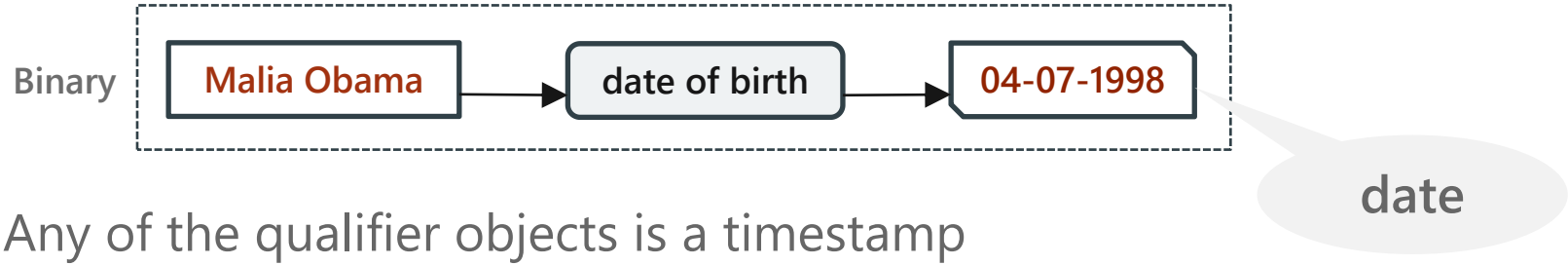
Temporal question

★ A temporal question is one that contains a **temporal expression** or a **temporal signal**, or **whose answer is of temporal nature**.

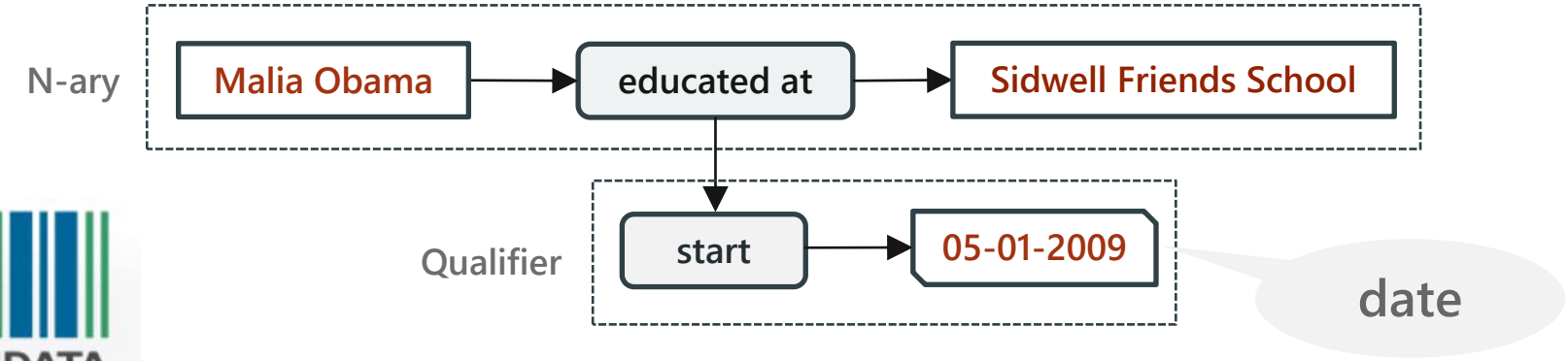
Category	Question	Signal
EXPLICIT	Which movie did Jaco Van Dormael direct **in 2009**?	OVERLAP
IMPLICIT	What club did Cristiano Ronaldo play for **after Manchester United**?	AFTER
	What did Thomas Jefferson do **before he was president**?	BEFORE
ORDINAL	What was the **first** film Julie Andrews starred in?	ORDINAL
TEMP. ANS.	What year did Lakers win their **first** championship?	ORDINAL

Temporal fact

★ Main object is a timestamp



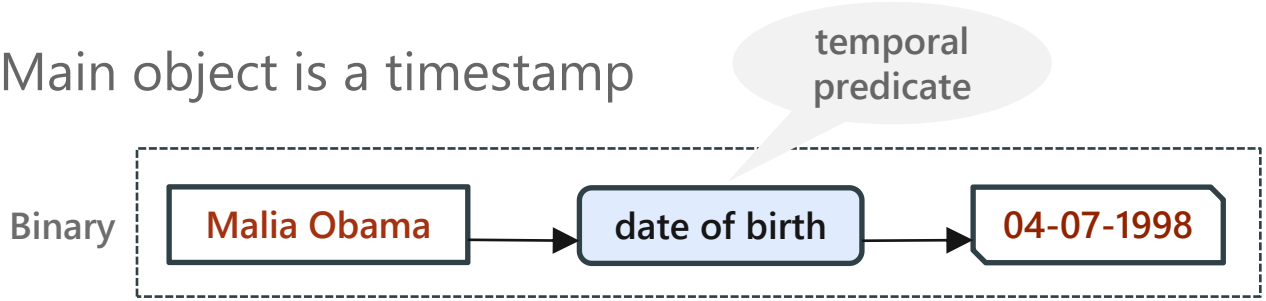
★ Any of the qualifier objects is a timestamp



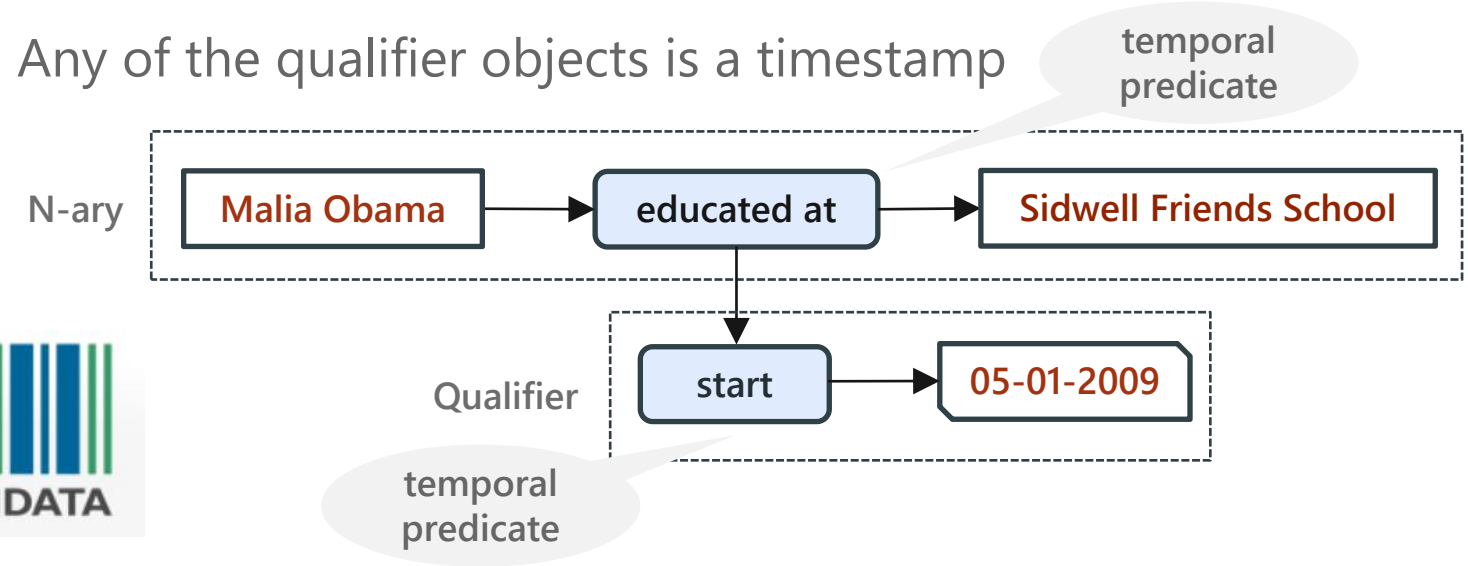
WIKIDATA

Temporal predicate

★ Main object is a timestamp

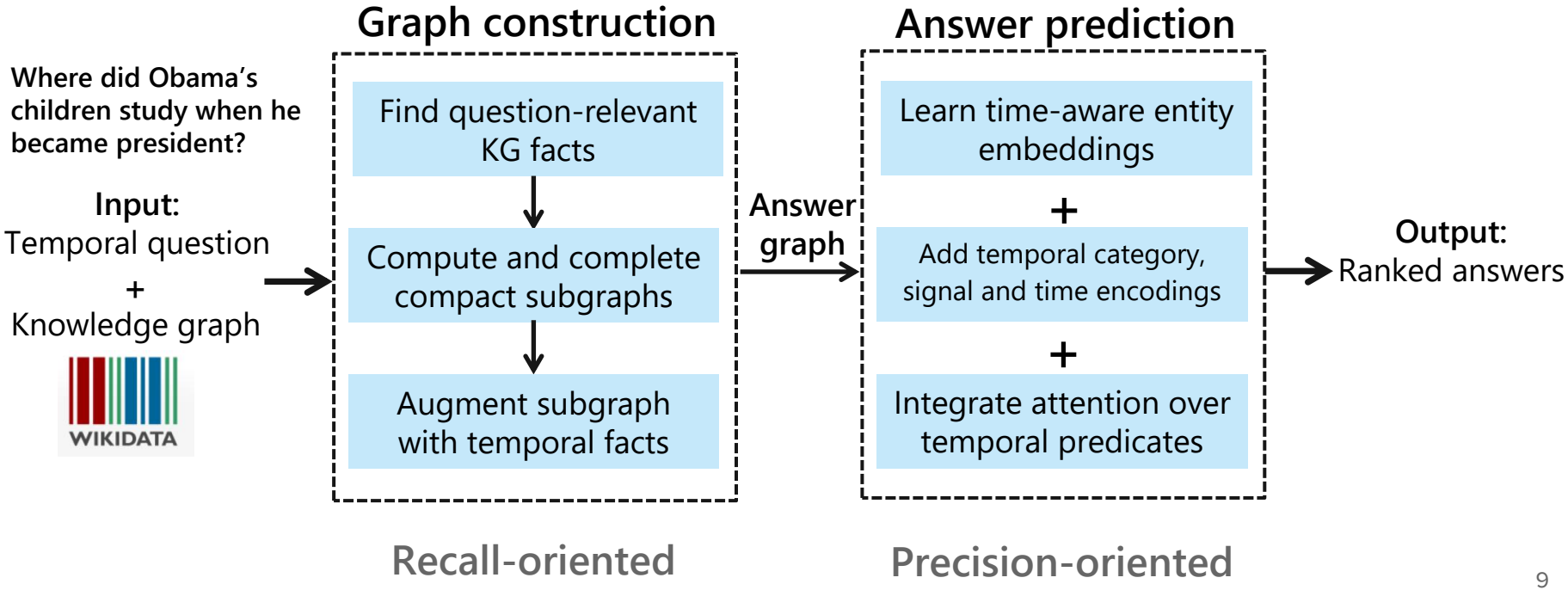


★ Any of the qualifier objects is a timestamp



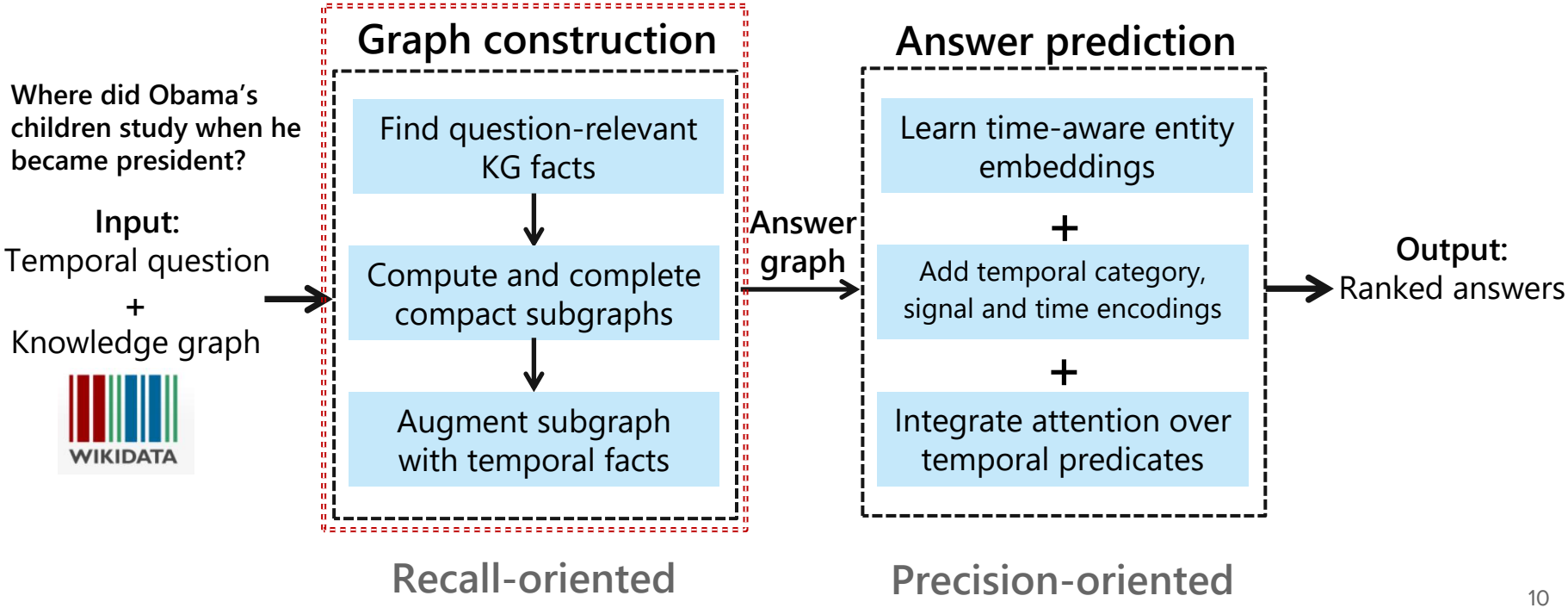
Approach outline

★ Two-stage approach



Approach outline

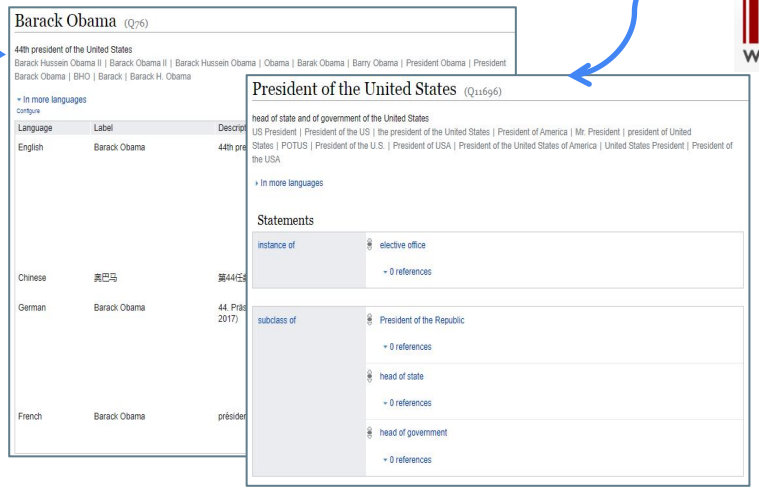
★ Two-stage approach

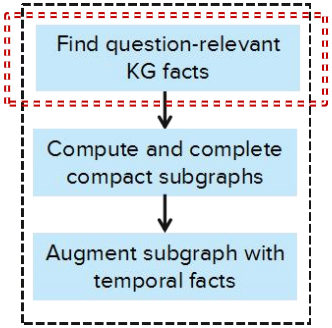


Stage one: Graph construction

Question: Where did **Obama**'s children study when he became **president**?

Use multiple NERD methods to boost answer recall

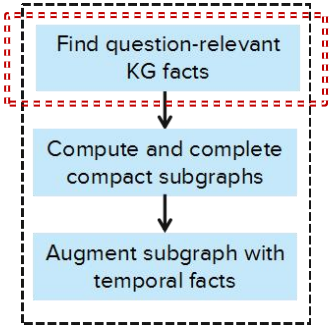


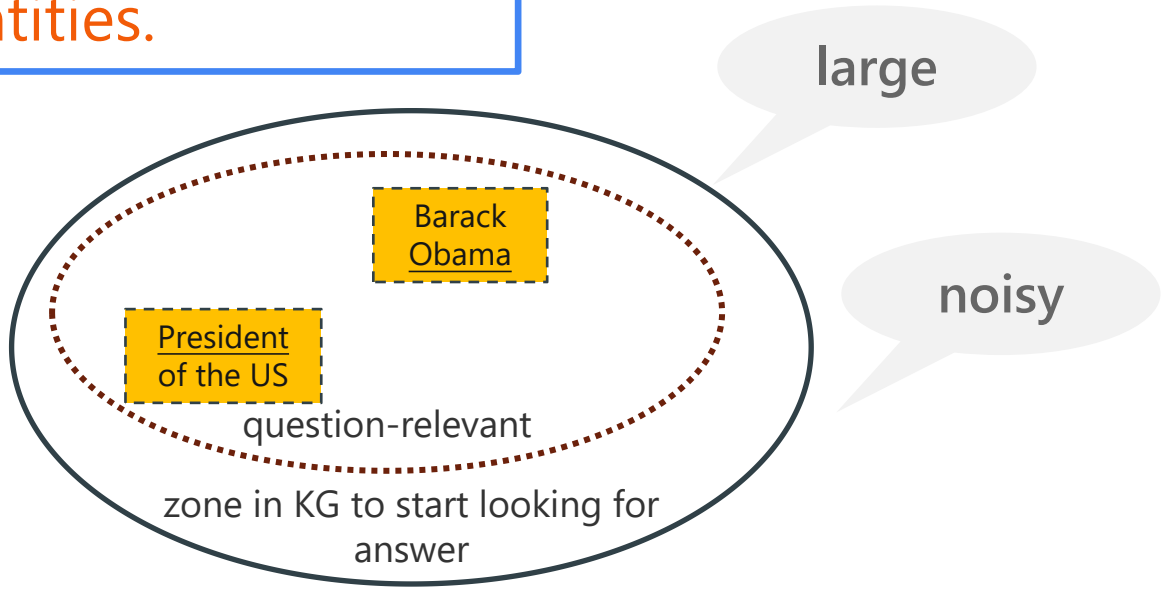


Stage one: Graph construction

Question: Where did **Obama**'s children study when he became **president**?

Retrieve all KG facts of the question entities.





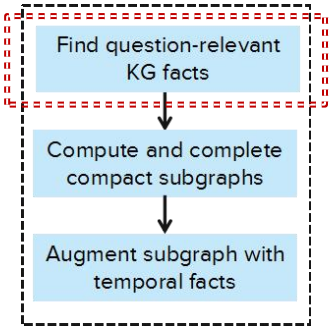
Stage one: Graph construction

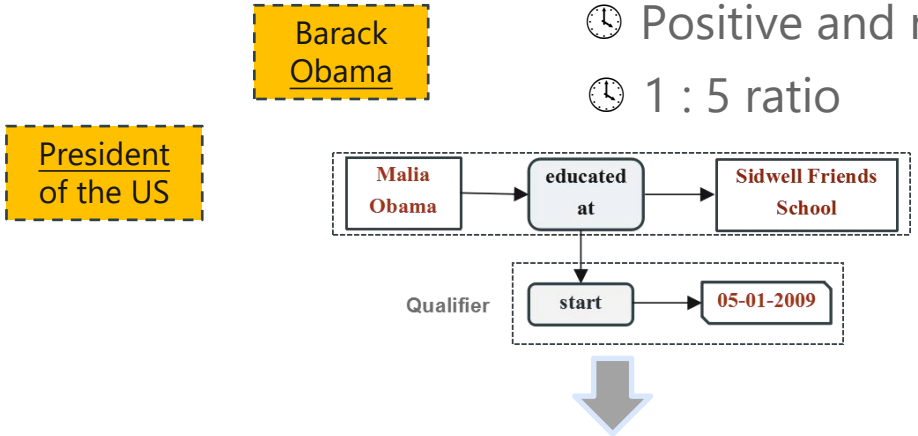
Question: Where did **Obama**'s children study when he became **president**?

Fine tune BERT model to find relevant KG facts

★ Use **distant supervision mechanism** to label training set

- <question, fact> pair
- Positive and negative sample
- 1 : 5 ratio





Malia Obama educated at Sidwell friends School and start 2009-01-05

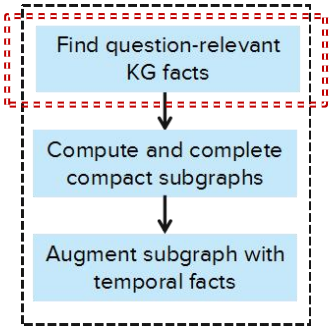
Stage one: Graph construction

Question: Where did **Obama**'s children study when he became **president**?

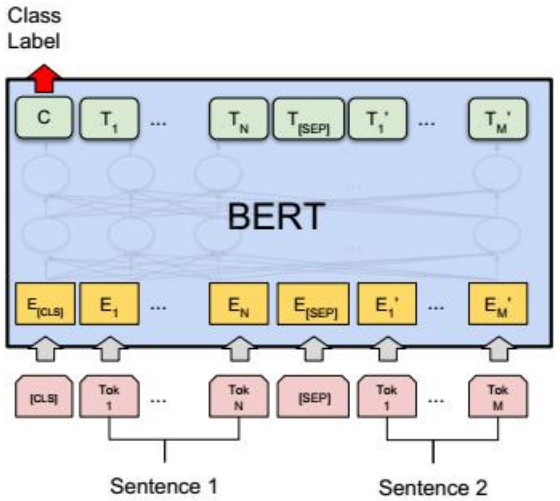
Fine tune BERT model to find relevant KG facts

Barack Obama

President of the US



★ Fine-tune BERT model as a sentence classifier

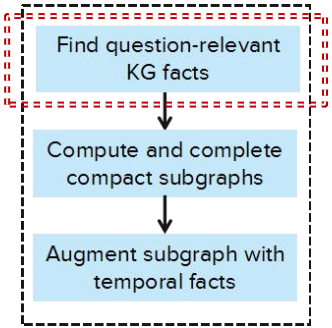


Devlin et al. 2019

Stage one: Graph construction

Question: Where did **Obama**'s children study when he became **president**?

Fine tune BERT model to find relevant KG facts



Barack Obama

President of the US

★ Apply the classifier to sort facts

Rank 1 Barack Obama educated at Harvard Law School and end time ...

Rank 2 Barack Obama educated at State Elementary School Menteng 01 ...

Rank 3 Barack Obama educated at Punahou School start time 1971-01-01 ...

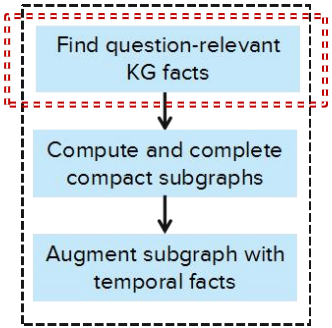
...

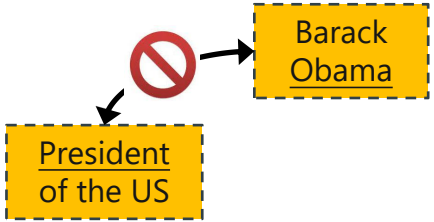
Rank n

Stage one: Graph construction

Question: Where did **Obama**'s children study when he became **president**?

Inject connectivity

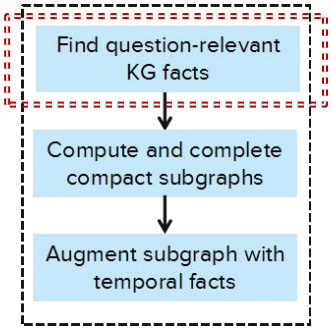


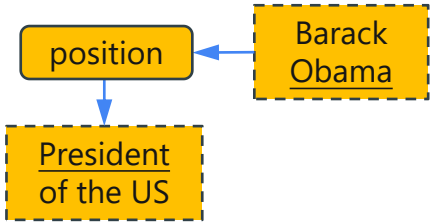


Stage one: Graph construction

Question: Where did **Obama**'s children study when he became **president**?

Inject connectivity



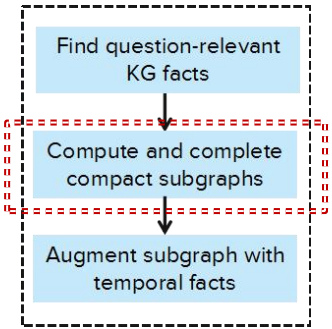


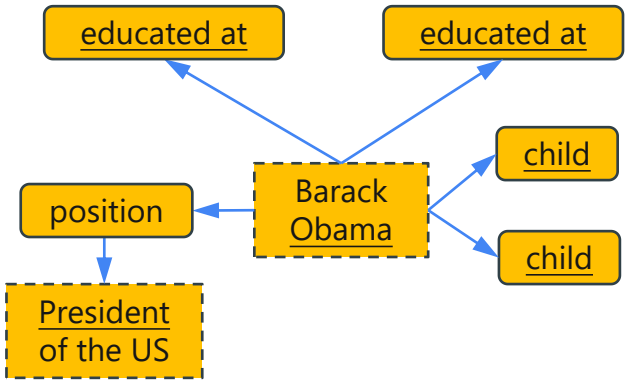
- ★ Compute **shortest path** between pair of entities
- ★ Add the path with the **highest similarity** to answer graph
  - 🕓 Get embeddings from BERT
  - 🕓 Compute cosine similarity

Stage one: Graph construction

Question: Where did **Obama**'s children study when he became **president**?

Compute compact subgraph





★ Group Steiner Trees (GSTs)

🕓 Given

Undirected and weighted **graph**

a subset of **terminals in groups**

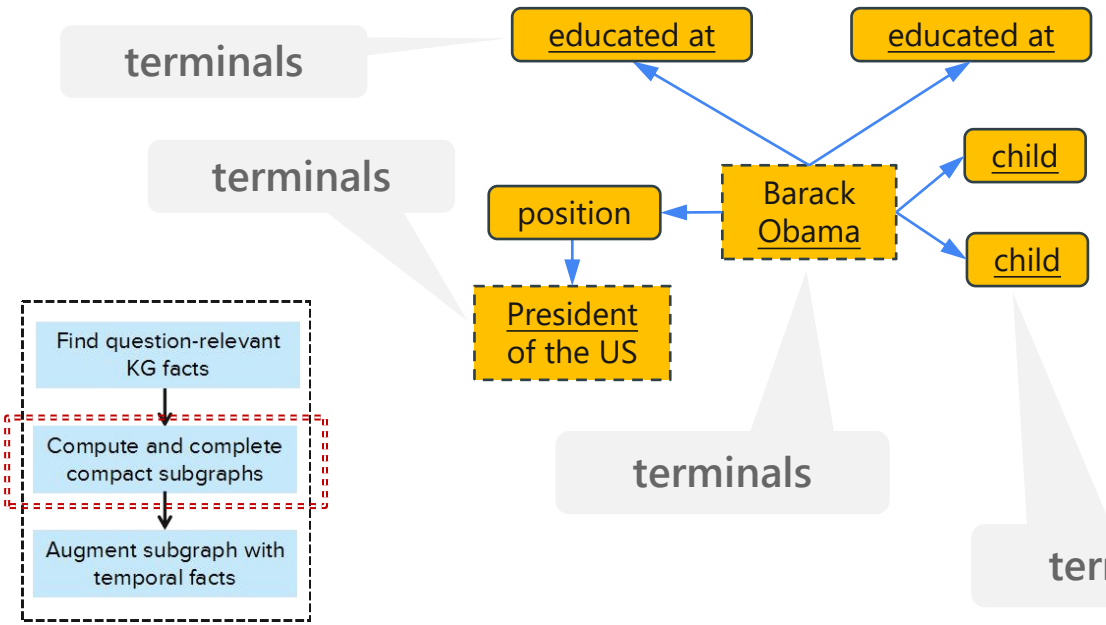
🕓 Find

Minimum cost tree containing at least one terminal from each group

Stage one: Graph construction

Question: Where did **Obama**'s children study when he became **president**?

Compute compact subgraph



★ Group Steiner Trees (GSTs)

🕓 Given

Undirected and weighted **graph**

a subset of **terminals in groups**

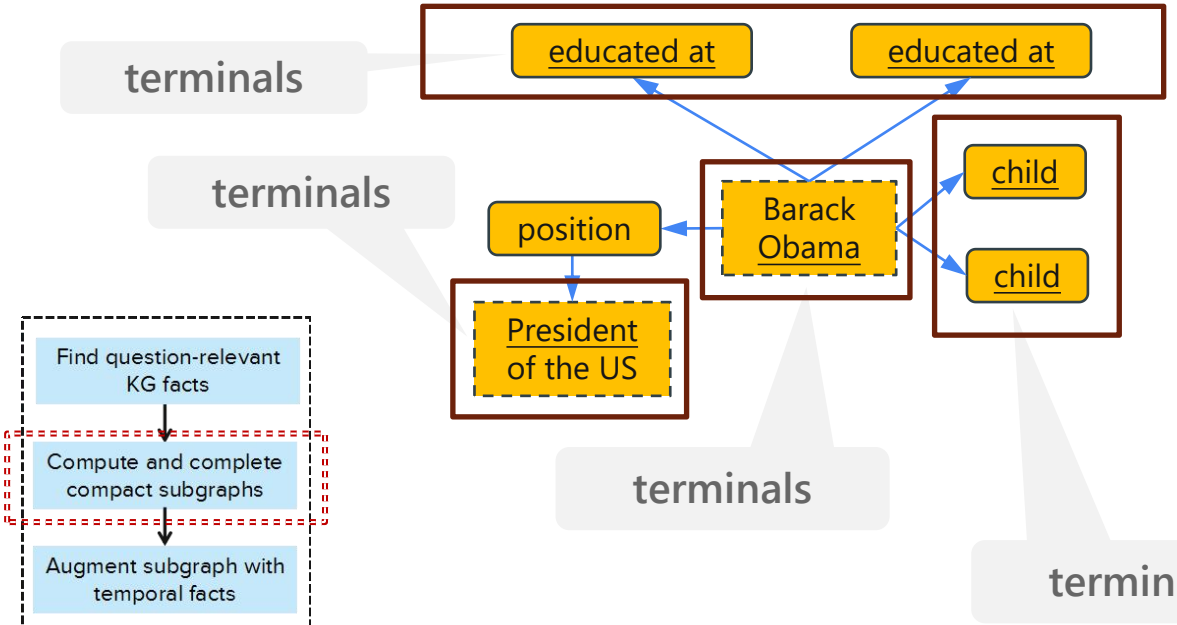
🕓 Find

Minimum cost tree containing at least one terminal from each group

Stage one: Graph construction

Question: Where did **Obama**'s children study when he became **president**?

Compute compact subgraph



★ Group Steiner Trees (GSTs)

🕓 Given

Undirected and weighted **graph**

a subset of **terminals in groups**

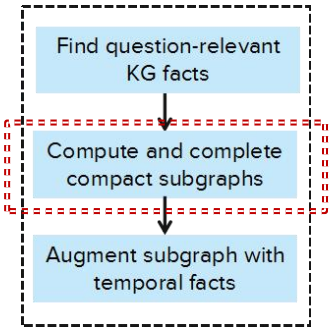
🕓 Find

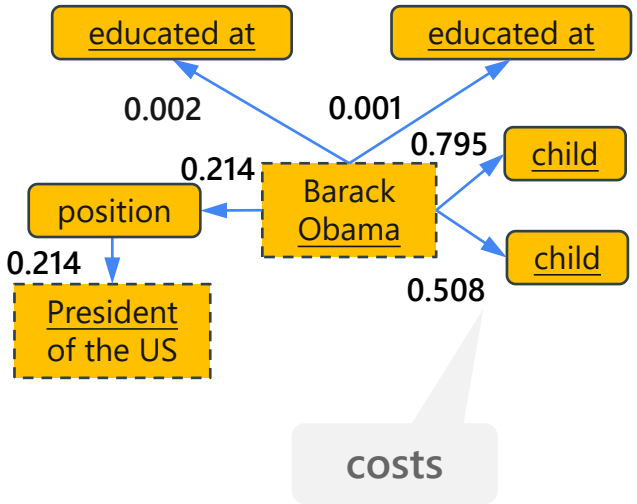
Minimum cost tree containing at least one terminal from each group

Stage one: Graph construction

Question: Where did **Obama**'s children study when he became **president**?

Compute compact subgraph



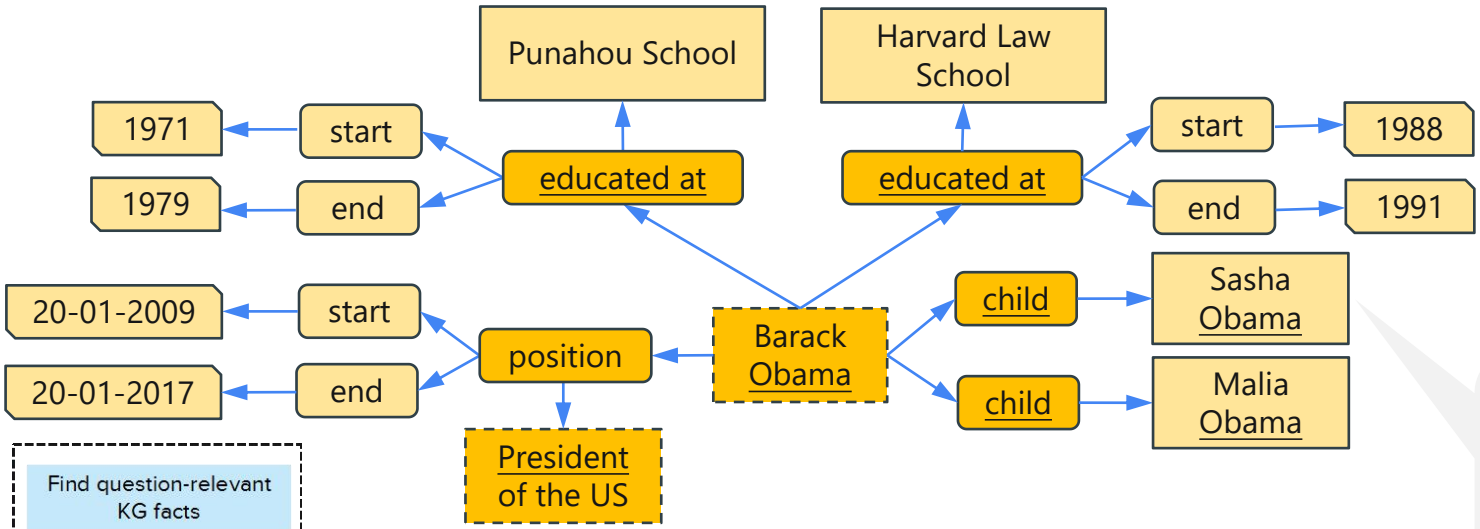


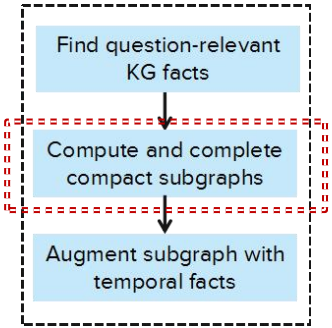
★ Group Steiner Trees (GSTs)

- Cost = 1 - score assigned by the classifier of BERT model
- Method from Ding et al. 2007

Stage one: Graph construction

Question: Where did **Obama**'s children study when he became **president**?



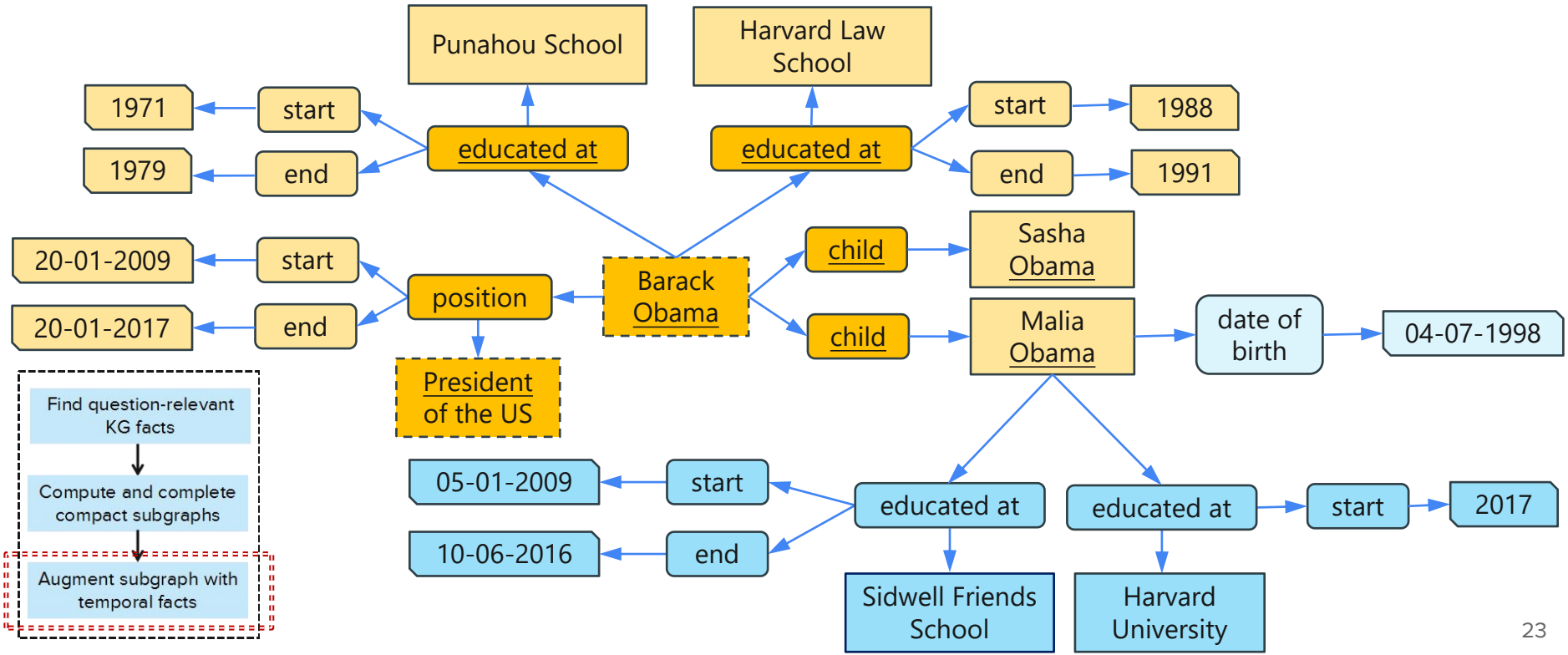


Complete GST compact subgraphs

Facts linked with the predicate terminals in GSTs

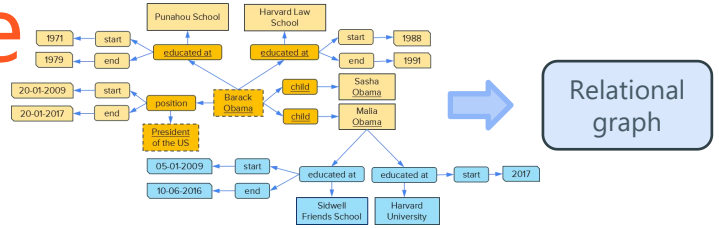
Stage one: Graph construction

Question: Where did **Obama**'s children study when he became **president**?



Approach outline

★ Two-stage approach



Where did Obama's children study when he became president?

Input:
Temporal question
+
Knowledge graph

WIKIDATA

Graph construction

Find question-relevant KG facts

↓

Compute and complete compact subgraphs

↓

Augment subgraph with temporal facts

Answer graph

Answer prediction

Learn time-aware entity embeddings

+

Add temporal category, signal and time encodings

+

Integrate attention over temporal predicates

Output:
Ranked answers

Recall-oriented

Precision-oriented

Stage two: Answer prediction

★ Build upon the KG-only setting of GRAFT-NET (Sun et al. 2018)

★ R-GCN model with multi-pronged mechanisms for temporal QA

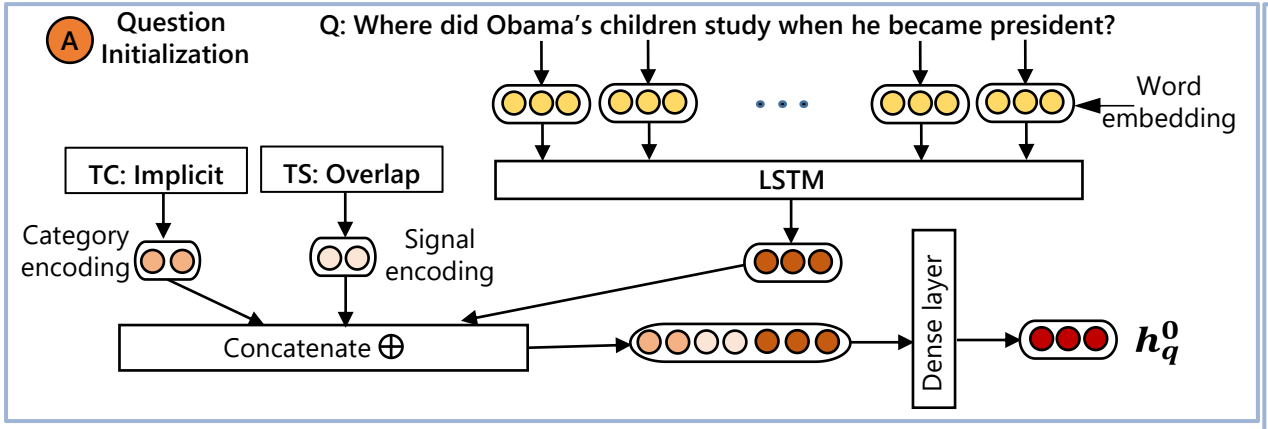
- 🕓 **T**emporal **C**ategory **E**ncoding (**TCE**)
- 🕓 **T**emporal **S**ignal **E**ncoding (**TSE**)

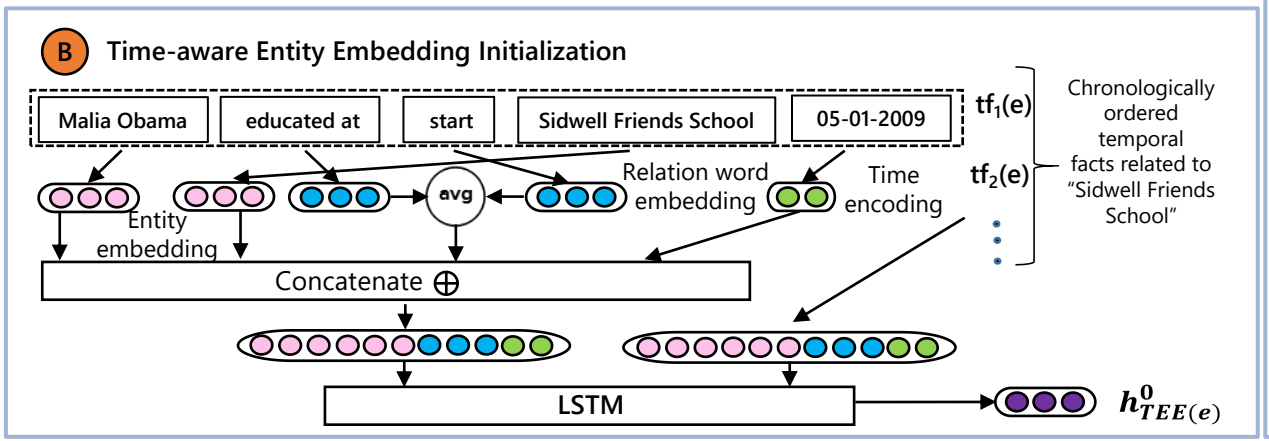
Question representation

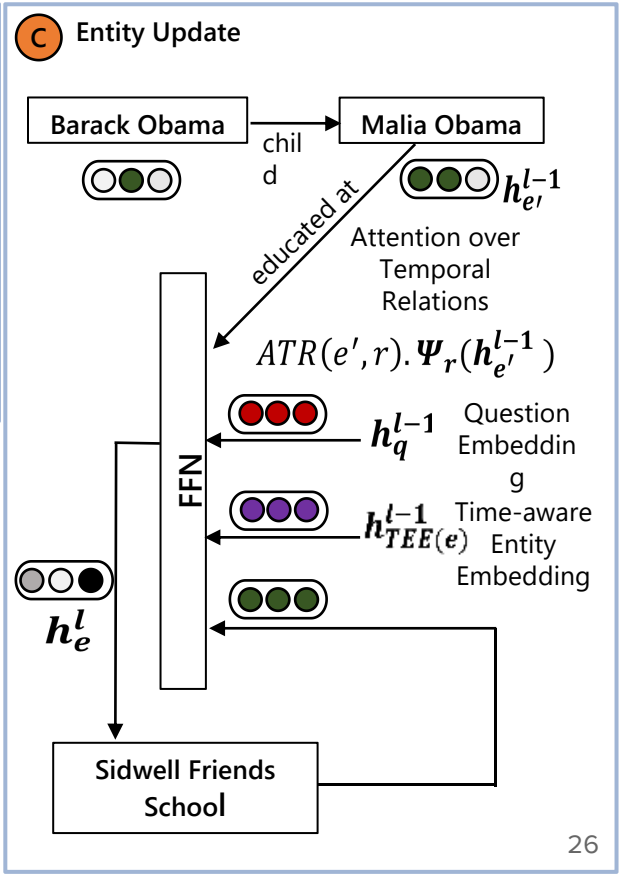
- 🕓 **T**ime **E**ncoding (**TE**)
- 🕓 **T**ime-aware **E**ntity **E**mbedding (**TEE**)
- 🕓 **A**ttention over **T**emporal **R**elation (**ATR**)

Entity representation

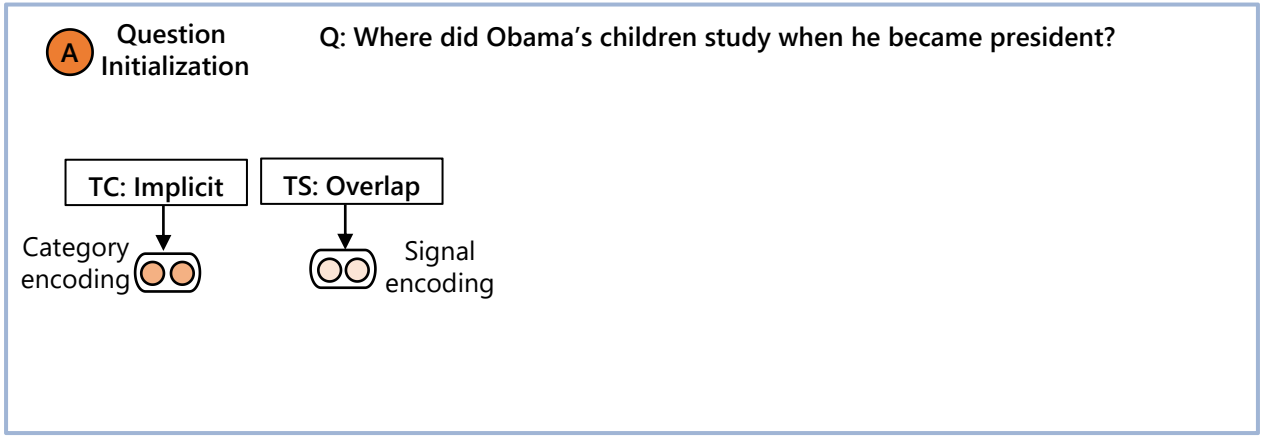
Stage two: Answer prediction







Stage two: Answer prediction



Question embedding initialization

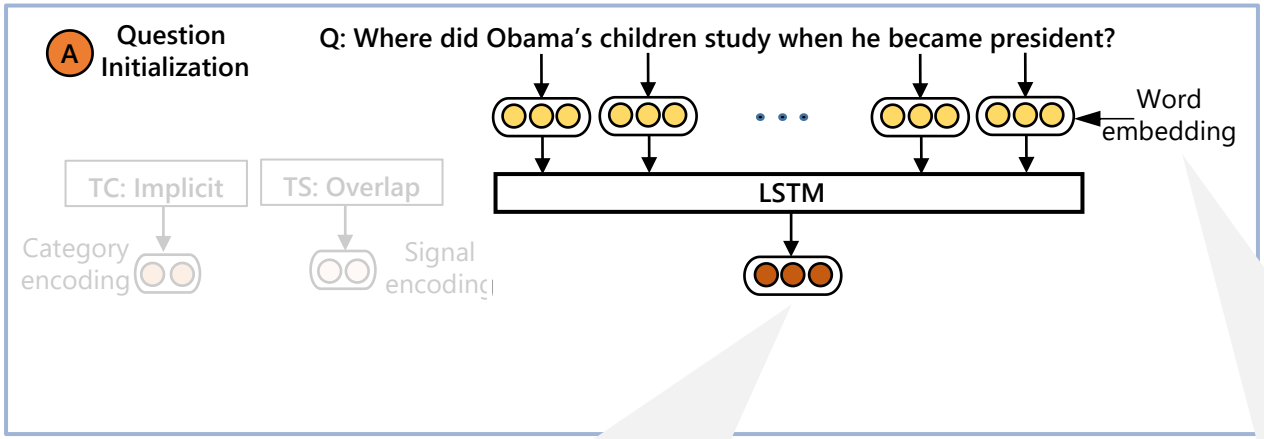
★ Temporal category encoding

- Label categories
  Temporal expression recognition, entity recognition, keywords and POS patterns
- Multi-hot encoding

★ Temporal signal encoding

- Label signals
  dictionary of keywords
- Multi-hot encoding

Stage two: Answer prediction

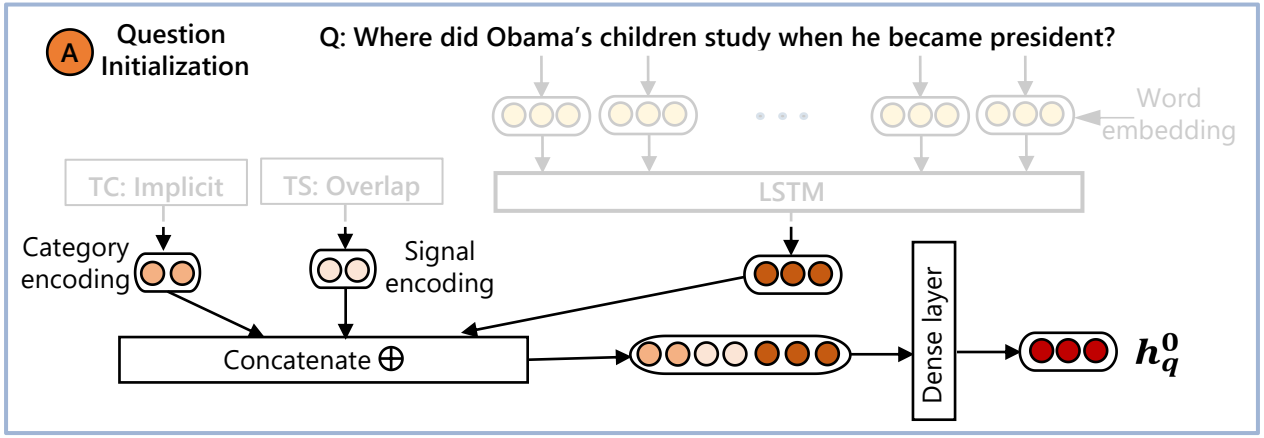


Question embedding initialization

Use LSTM to model the words in the question as a sequence

Pre-trained word embeddings from Wikipedia2Vec (Yamada et al. 2020)

Stage two: Answer prediction



Question embedding initialization

$$h_q^0 = FFN(TCE(q) \oplus TSE(q) \oplus LSTM(w_1, \ldots, w_{|q|}))$$

$$h_q^l = FFN(\sum_{e \in NERD(q)} h_e^{l-1})$$

Update with the embeddings of entities

Question embedding update

Stage two: Answer prediction

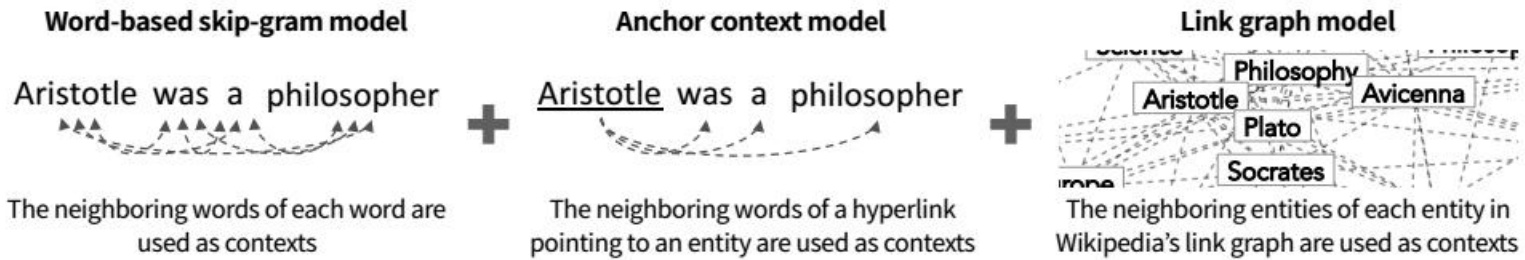
$$h_e^0 = x_e$$

fixed-size pre-trained embeddings from Wikipedia2Vec

Entity embedding initialization

★ Wikipedia2Vec

Yamada et al. 2020



Stage two: Answer prediction

★ Time encoding (similar to position encoding in Vaswani et al. 2017)

- Sinusoidal position encoding
- Provide an **unique encoding**
- Ensure **sequential ordering**

$$TE(k,j)=\begin{cases}\sin(k/10000^{\frac{2i}{d}}), \text{if } j=2i\\ \cos(k/10000^{\frac{2i}{d}}), \text{if } j=2i+1\end{cases}$$

position in time range

vector dimension

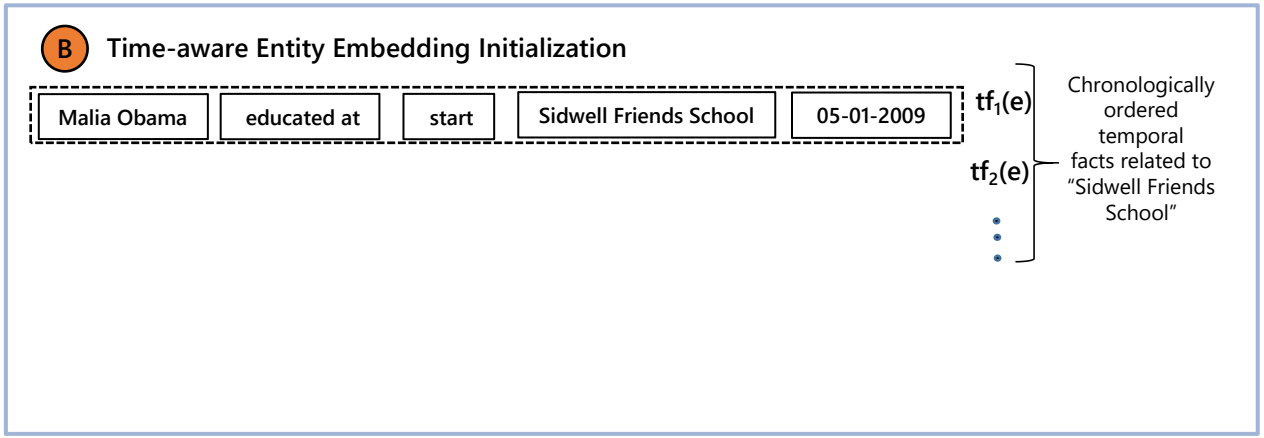
position in vector

Entity embedding update

Stage two: Answer prediction

★ Time-aware entity embedding

- An entity **e** is associated with a set of temporal facts **{tf(e)}**
- The temporal facts of **e** are ordered in a time sequence **{tf_1(e), tf_2(e), ...}**

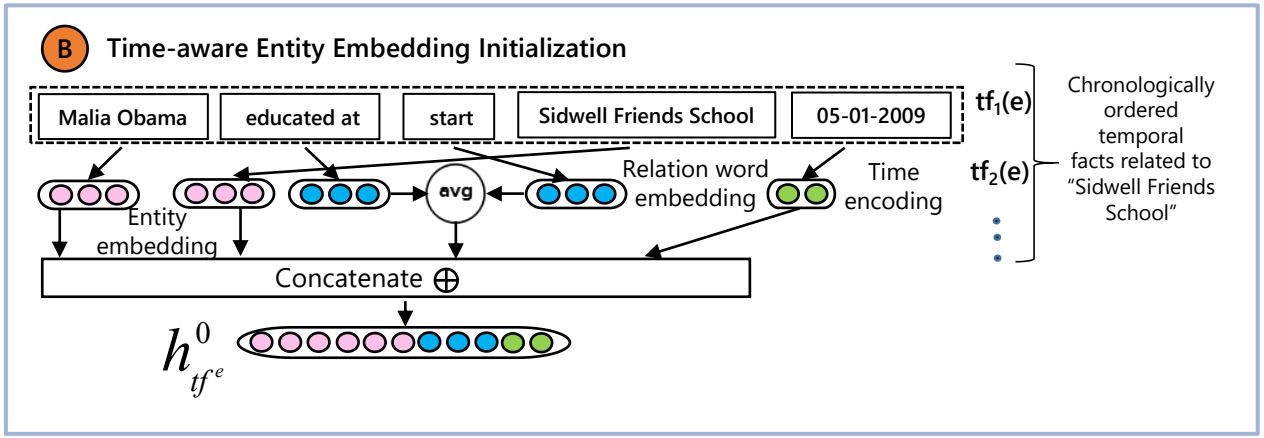


Entity embedding update

Stage two: Answer prediction

★ Time-aware entity embedding

🕓 Encode **tf(e)**



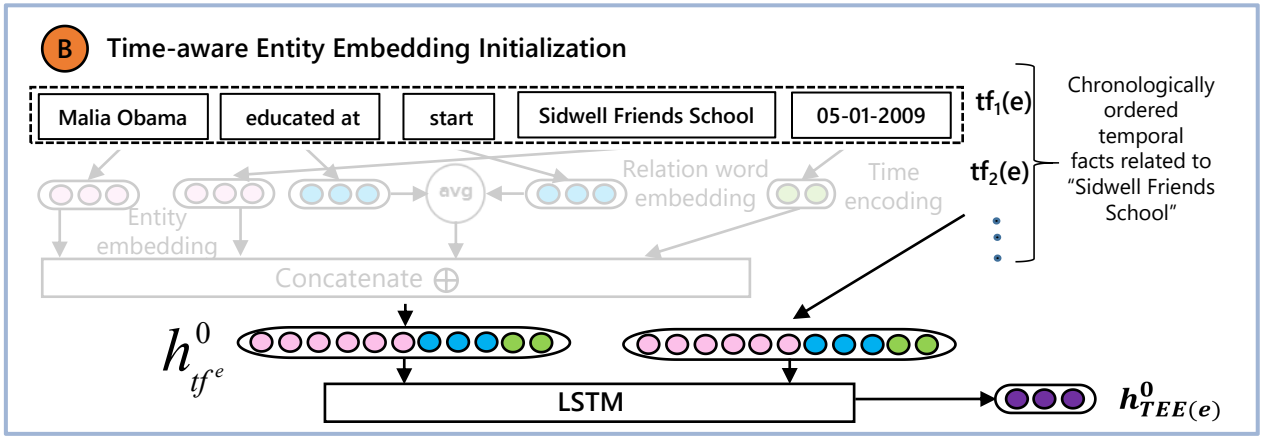
Entity embedding update

Stage two: Answer prediction

★ Time-aware entity embedding

🕓 Use LSTM to model {**$tf_1(e)$**, **$tf_2(e)$**, ...} as a sequence

$$h^0_{TEE(e)} = LSTM(h^0_{tf^e_1}, h^0_{tf^e_2}, ..., h^0_{tf^e_n})$$



Entity embedding update

Stage two: Answer prediction

★ Attention over temporal relation

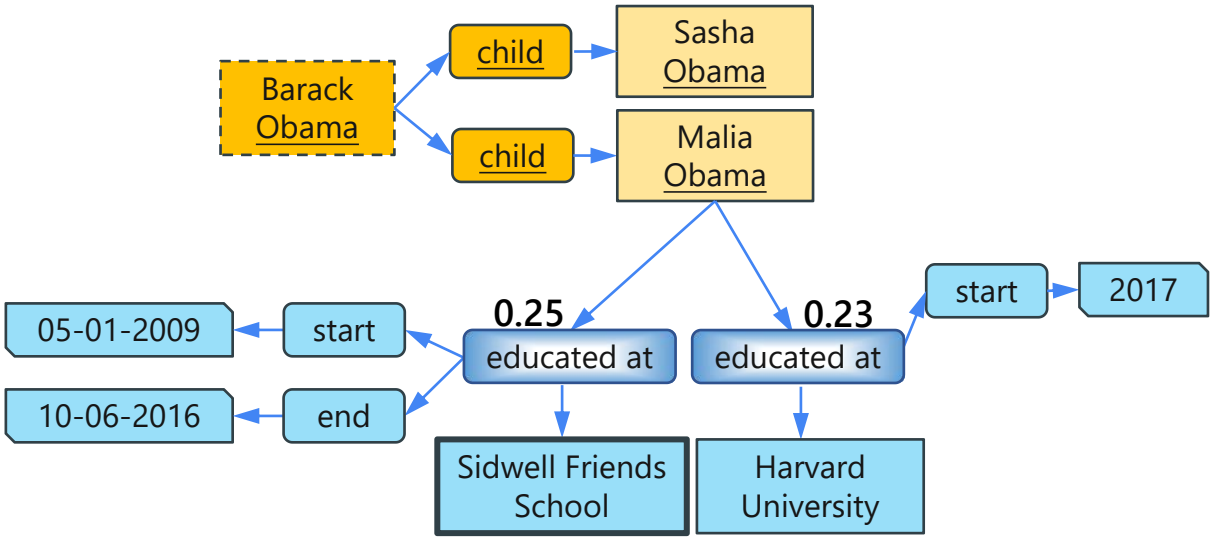
🕓 Distinguish entities with the same relation but having different timestamps

$$ATR(e,r) = softmax(x_r \oplus TE(ts_r)^T h_q^{(l-1)})$$

pre-trained relation embedding

time encoding

question embedding



Entity embedding update

Stage two: Answer prediction

Entity embedding update rule

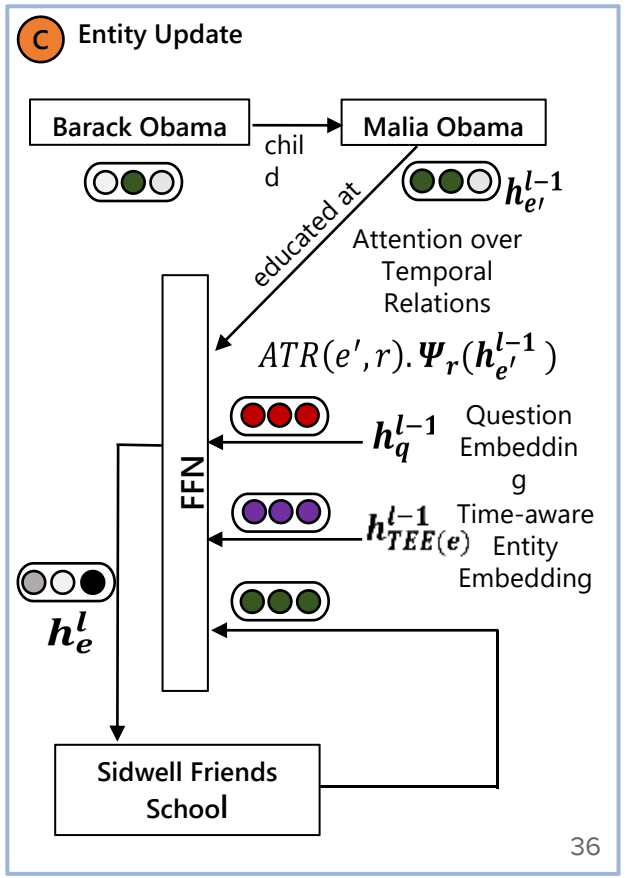
$$h_e^l = FFN\left(\begin{bmatrix} h_e^{l-1} \\ h_q^{l-1} \\ h_{TEE(e)}^{l-1} \\ \sum_r \sum_{e \in nbd_r(e')} (ATR(e',r)\psi_r(h_{e'}^{l-1})) \end{bmatrix}\right)$$

(1) Entity representation

(2) Question representation

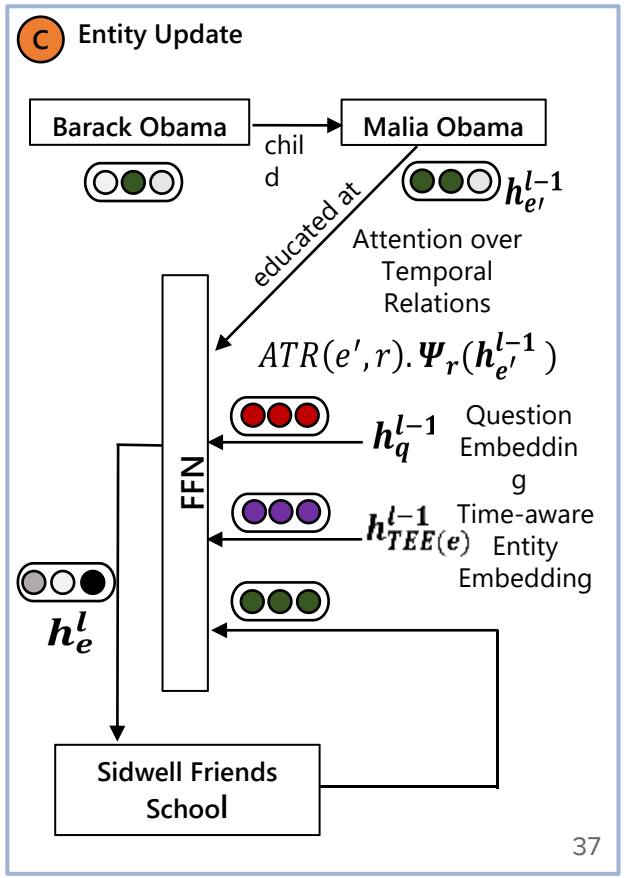
(3) Time-aware entity representation

(4) Aggregate the states from neighbors



Stage two: Answer prediction

$$Pr(e \in \{a\}_q \mid RG_q, q) = \sigma(w^T h_e^l + b)$$



Experiment results: Setup

- ★ **Benchmark**
  - 🕓 TimeQuestions
- ★ **Metrics**
  - 🕓 Presicion@1
  - 🕓 Mean Reciprocal Rank
  - 🕓 Hit@5
- ★ **Baselines**
  - 🕓 UNIQORN (Pramanik et al. 2021)
  - 🕓 GRAFT-Net (Sun et al. 2018)
  - 🕓 PullNet (Sun et al. 2019)

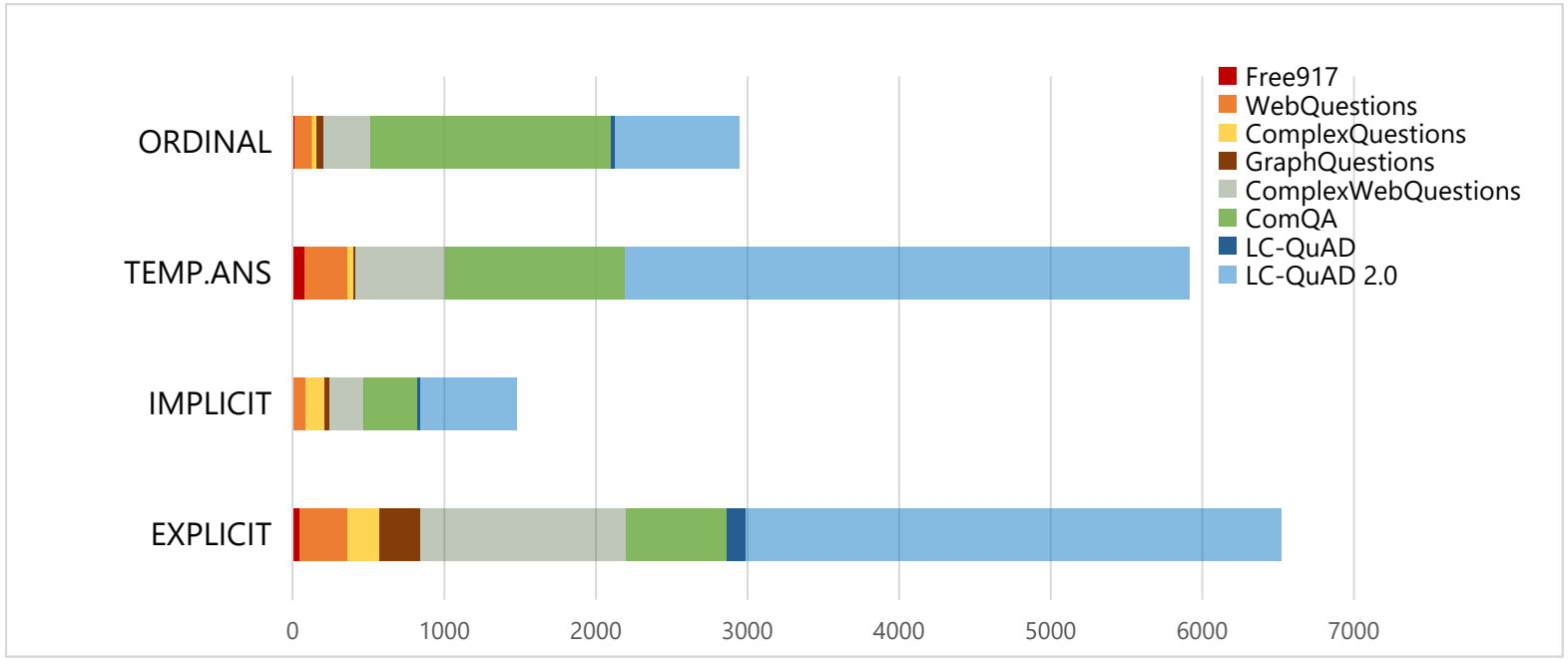
Experiment results: Benchmark

★ Benchmark construction

- Collect temporal questions from **8 popular KG-QA benchmarks**
- Contain **16181** <question, answer> pairs
- Label temporal **categories** and **signals** for each question
- Link answers to **Wikidata** and **Wikipedia**

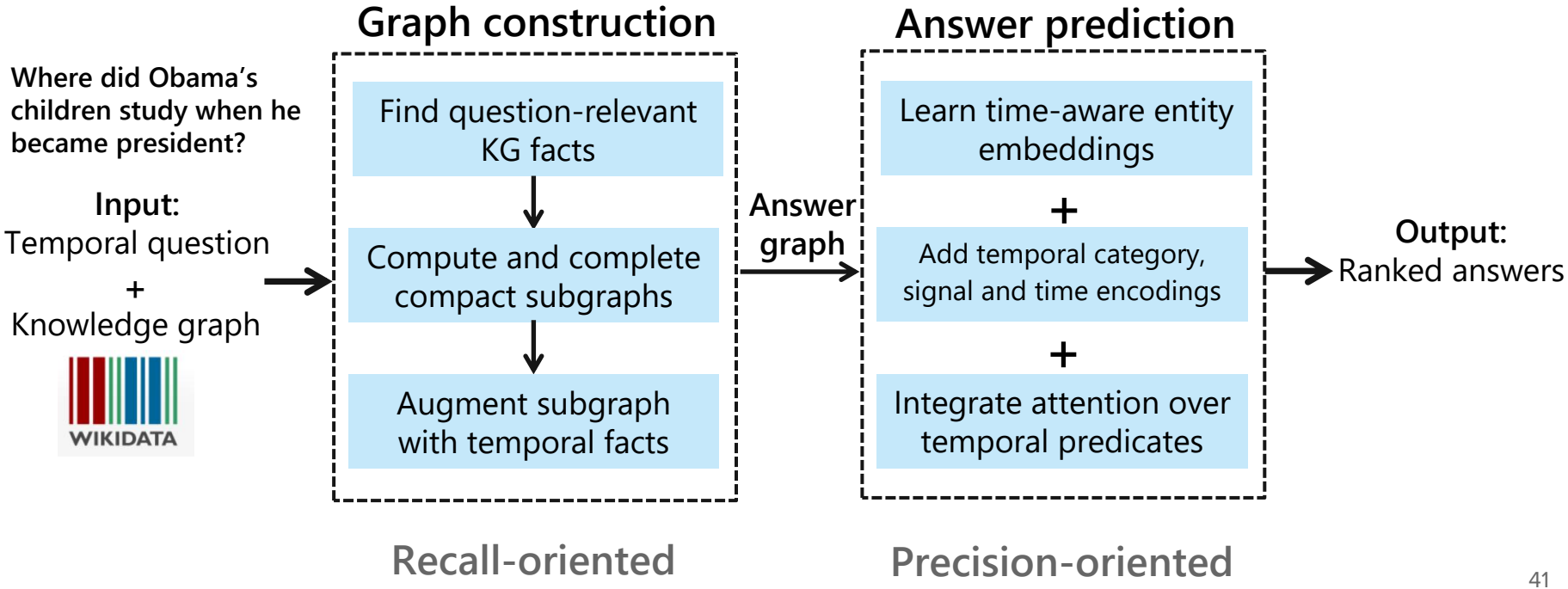
Experiment results: Benchmark

★ Distribution of question categories by source



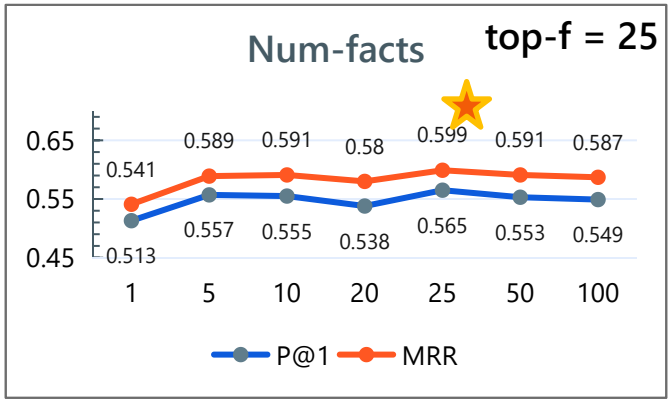
Approach outline

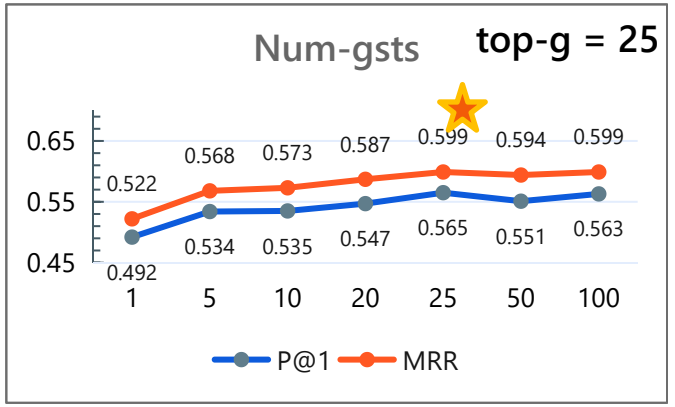
★ Two-stage approach

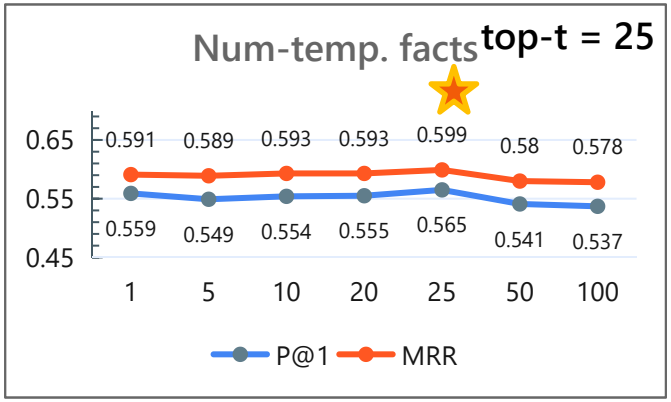


Experiment results: Performance

★ Parameter tuning (S1)





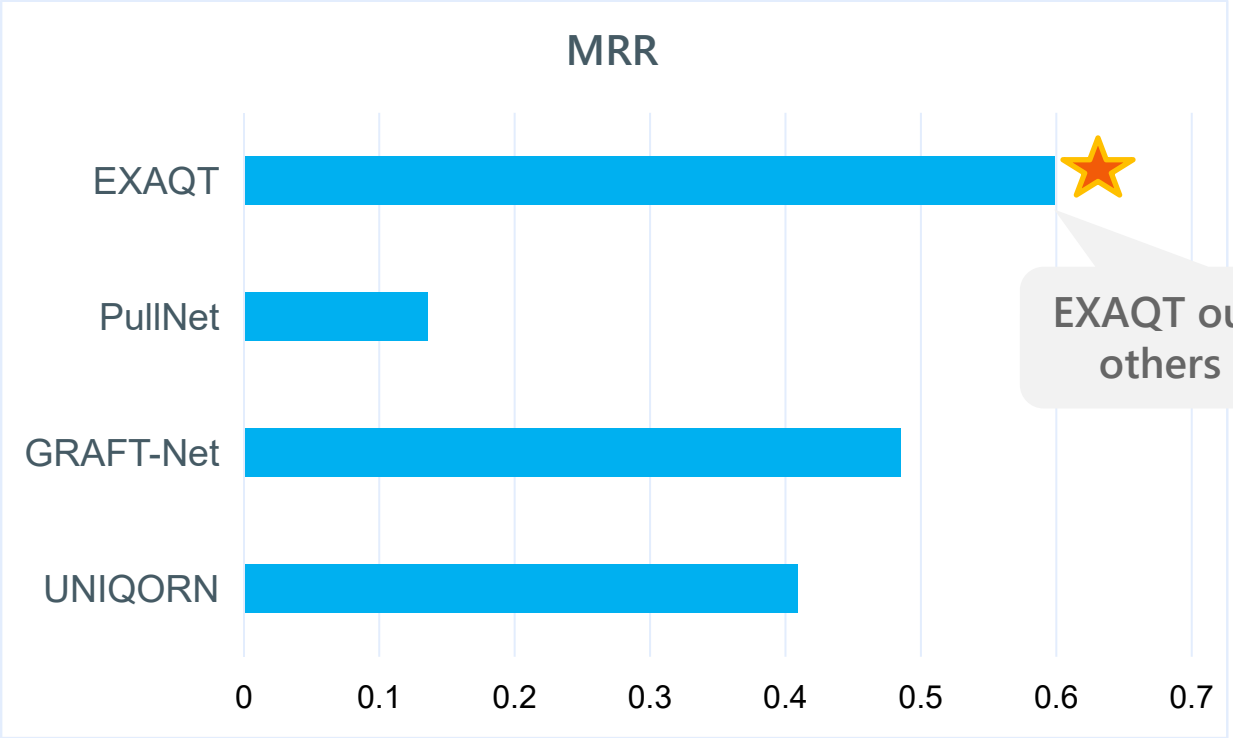


Experiment results: Performance

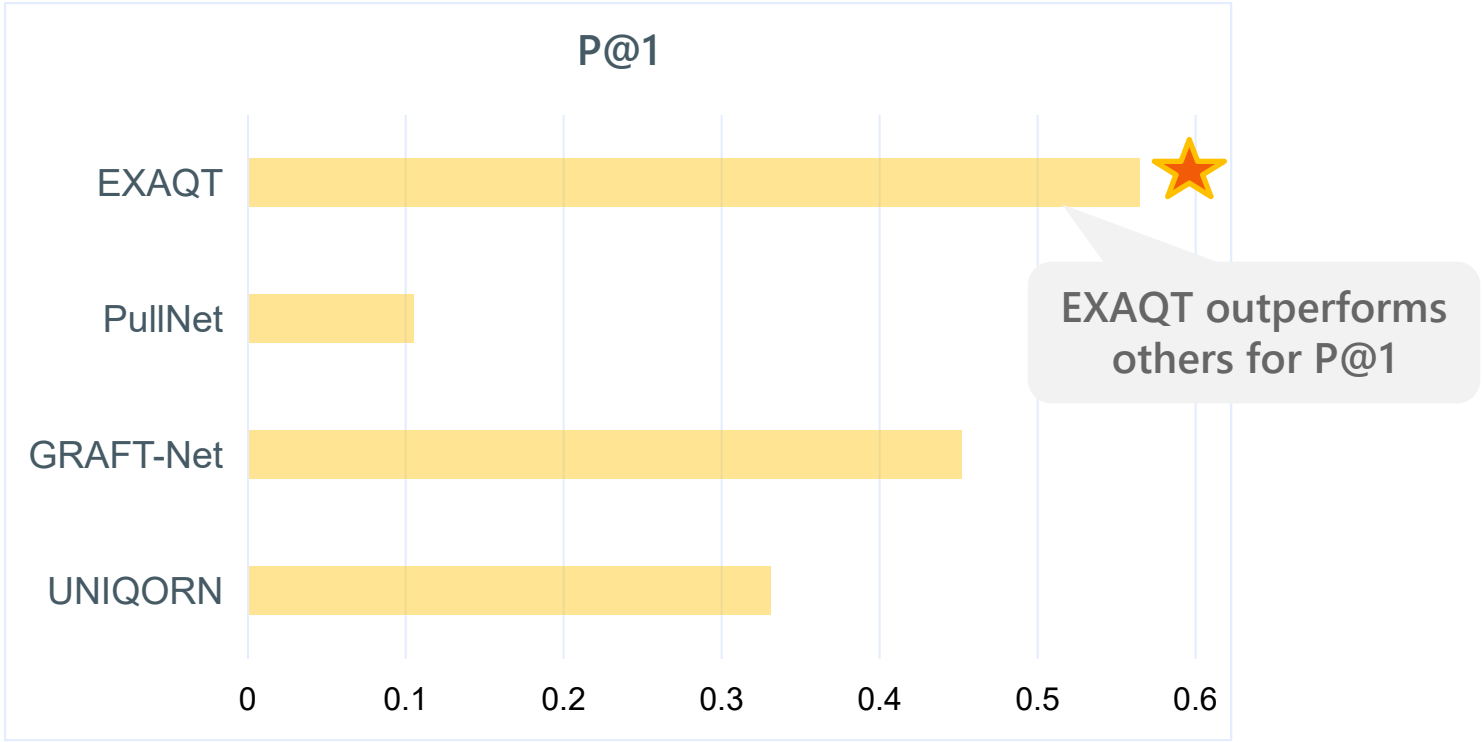
Step in EXAQT pipeline	Recall	#Candidates
All KG facts of **NERD** entities	**0.758**	**2491**
Facts selected by **BERT**	**0.719**	**48**
Shortest paths injected for connectivity	**0.720**	**49**
GSTs on largest component	**0.613**	**13**
Union of GSTs from all components	**0.640**	**14**
Completed GSTs from all components	**0.671**	**21**
Temporal facts added by BERT	**0.724**	**67**

Understanding the recall-oriented stage one

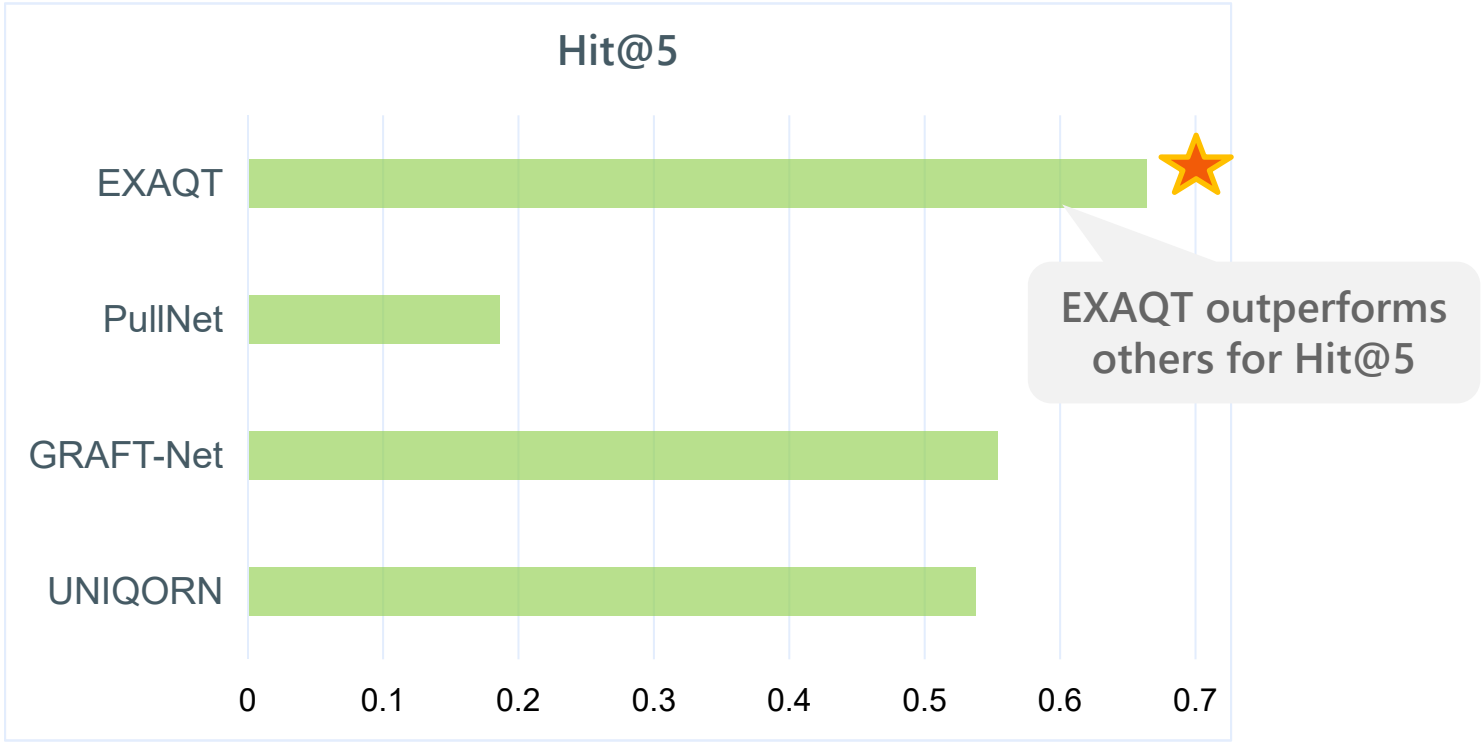
Experiment results: Performance



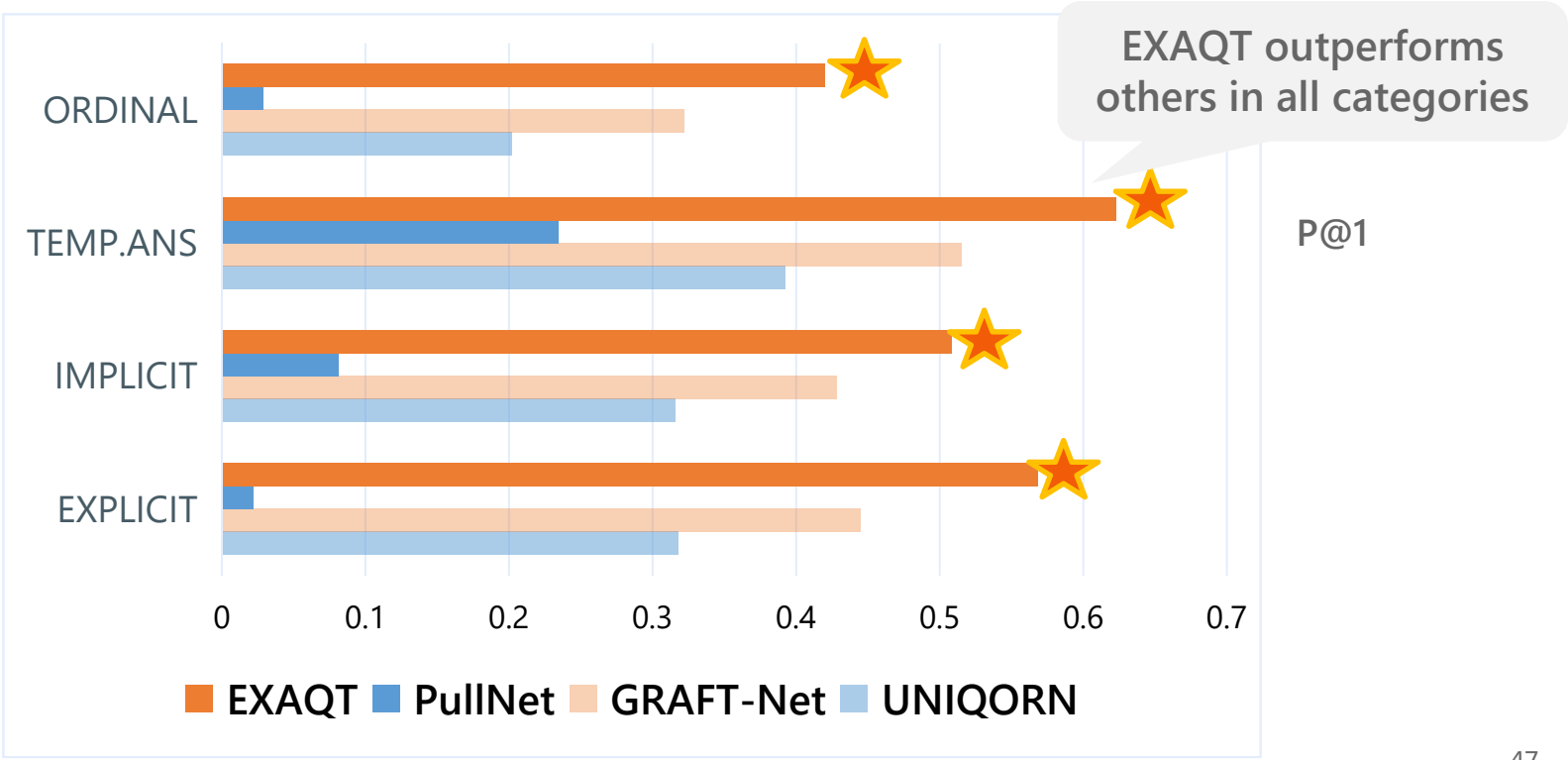
Experiment results: Performance



Experiment results: Performance



Experiment results: Performance



Experiment results: Performance

Category	Overall	EXPLICIT	IMPLICIT	TEMP. ANS.	ORDINAL
EXAQT (Full)	**0.565**	**0.568**	**0.508**	**0.623**	**0.420**
EXAQT without TCE	0.545	0.556	0.481	0.590	0.406
EXAQT without TSE	0.543	0.545	0.465	0.598	0.411
EXAQT without TEE	0.556	0.564	0.475	0.614	0.413
EXAQT without TE	0.553	0.556	0.495	0.613	0.398
EXAQT without ATR	0.534	0.527	0.465	0.594	0.411

Understanding the precision-oriented Stage two

Conclusion

★ **EXAQT**

- **Two-stage approach** for **explainable answering** of **temporal questions** over KGs
- Explainability comes from GSTs, attention and graph visualizations
- Combination of **BERT** classifiers, **GSTs** and **R-GCNs**
- Methods for augmenting components with **temporal** features

★ **TimeQuestions**: benchmark with over 16k temporal questions

Benchmark and demo: https://exaqt.mpi-inf.mpg.de
Code: https://github.com/zhenjia2017/EXAQT

Thank you!