GenTKG: Generative Forecasting on Temporal Knowledge Graph

Ruotong Liao[1,4], Xu Jia[3], Yunpu Ma[1,2], Volker Tresp[1,4]

[1]Ludwig Maximilian University of Munich, [2]Siemens AG, [3]Technical University of Munich, [4]Munich Center for Machine Learning

Email: ruotong.liao@outlook.com

Introduction

Temporal Knowledge Graph (tKG): (e_s, r, e_o, t)

tKG Forecasting: answer test queries: $(e_s, r, ?, t)$ given the history facts before t.

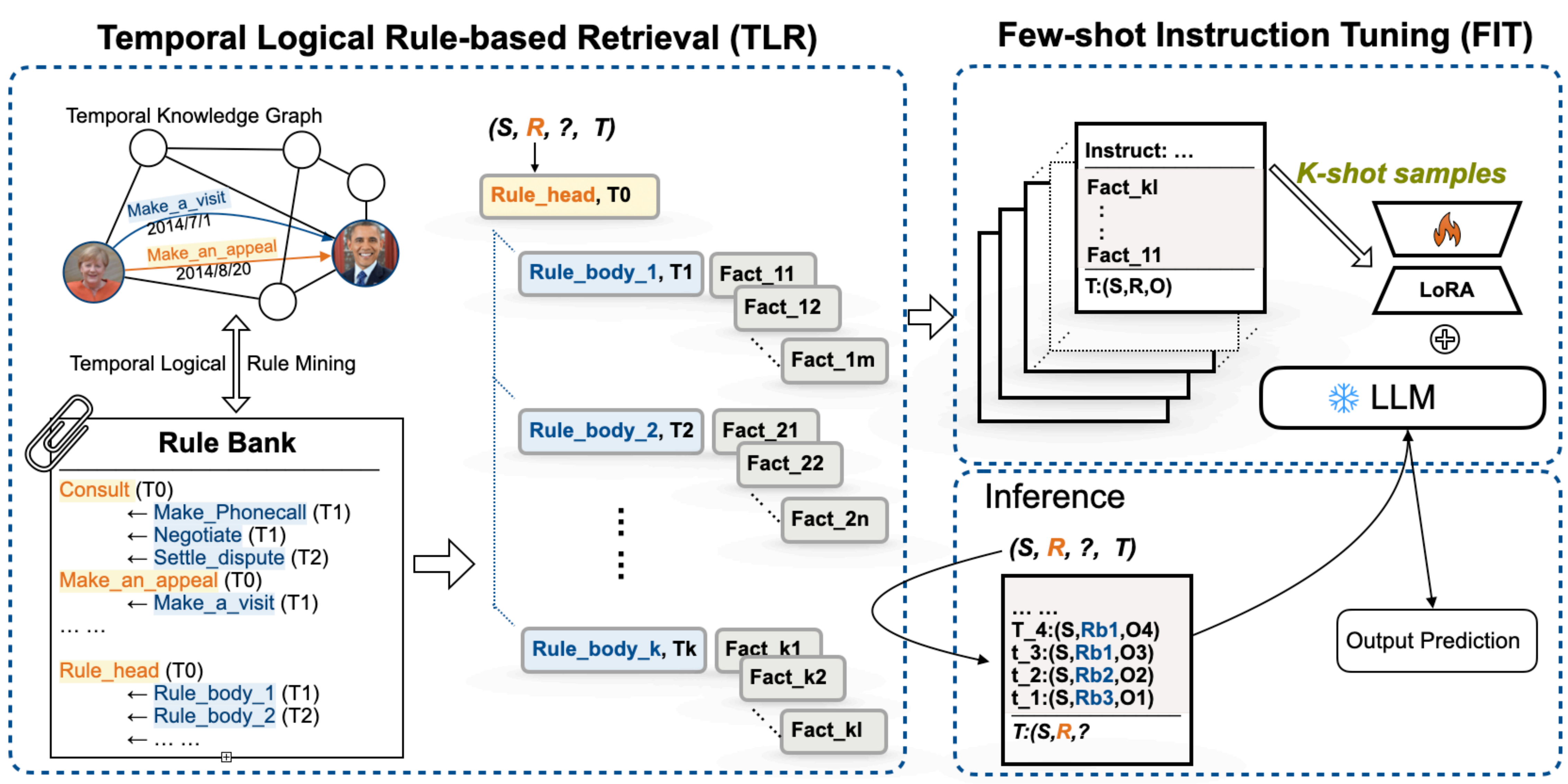
Can pre-trained LLMs understand structured temporal relational data and replace them as the foundation model for temporal relational forecasting?

Challenges:

- Modality challenge between data structures
  - TKG: complex temporal multi-relational graph data
  - LLMs: only process sequential natural language expressions
- Computation challenge
  - Hundreds of thousands of quadruples
  - Enormous costs of fine-tuning LLMs

Solutions: a novel retrieval-augmented generation framework, **GenTKG: Generative Forecasting on Temporal Knowledge Graph**

- Temporal Logical Rule-based Retrieval (TLR) strategy for history facts
  - Enable LLM to comprehend temporal relational data
- Few-shot Parameter-efficient Instruction Tuning (FIT)
  - As few as 0.0027% training data (16 samples) with drastically low computation



Contributions

- **Opening a frontier of generative forecasting on tKG.**
  - The first work that introduces instruction-tuned generative LLM to the tKG domain.
  - A novel retrieval augmented generation Paradigm for tKG forecasting
  - Regardless of the backbone LLM
- **Exceeding performance over conventional methods.**
  - Extensive comparation to embedding-based methods, rule-based method and In-context Learning method with LLM.
- **Task reformulation from data learning to task alignment.**
  - Data-centric learning to task-centric LLM alignment.
  - Aligns LLMs to generative forecasting task on tKG.
- **Generalizability both in-domain and cross-domain without retraining**
  - Cross-domain generalizability: one-time training on a single dataset with exceeding performance on multiple datasets without retraining.
  - In-domain generalizability: training on various partitions of training data of the same dataset with exceeding performance on evaluation.
- **Drastically low computation costs with exceeding performance.**
  - (0.0027%)**16**-shot tuning, comparable results to conventional methods
  - (0.27%)**1024-**shots tuning, outperforming existing methods.

	Datasets	ICEWS14			ICEWS18			GDELT			YAGO		
Method Type	Model	Hits@1	Hits@3	Hits@10	Hits@1	Hits@3	Hits@10	Hits@1	Hits@3	Hits@10	Hits@1	Hits@3	Hits@10
Embedding-based	RE-GCN	0.313	0.473	**0.626**	0.223	0.367	**0.525**	0.084	0.171	0.299	0.468	0.607	0.729
	xERTE	0.330	0.454	0.570	0.209	0.335	0.462	0.112	0.191	0.294	**0.769**	0.787	0.794
	TANGO	0.272	0.408	0.550	0.191	0.318	0.462	0.094	0.189	0.322	0.566	0.651	0.718
	Timetraveler	0.319	0.454	0.575	0.212	0.325	0.439	0.112	0.186	0.285	0.604	0.770	0.831
Rule-based	TLogic	0.332	0.476	0.602	0.204	0.336	0.480	0.113	0.212	0.351	0.638	0.650	0.660
In-Context Learning	GPT-NeoX-20B	0.319	0.439	0.538	0.179	0.297	0.41	0.098	0.165	0.253	0.669	0.787	0.841
	Llama2-7B	0.252	0.427	0.504	0.128	0.272	0.323	0.06	0.164	0.246	0.662	0.760	0.818
GenTKG	GPT-NeoX-20B +TLR	0.35	0.485	0.593	0.205	0.338	0.462	0.156	0.241	0.349	0.681	**0.807**	**0.861**
	Llama2-7B +GenTKG	**0.372**	**0.488**	0.563	0.217	**0.372**	0.443	**0.185**	**0.278**	**0.358**	0.715	0.767	0.786
	Llama2-7B - inductive	-	-	-	**0.234**	0.333	0.409	0.142	0.213	0.281	0.638	0.697	0.762

Table 1: Link prediction results: Hits@1/3/10 (%). The best results among each metric are highlighted in bold and the second bests are underlined. Notably, the TLR denotes only using the retrieval phase in GenTKG

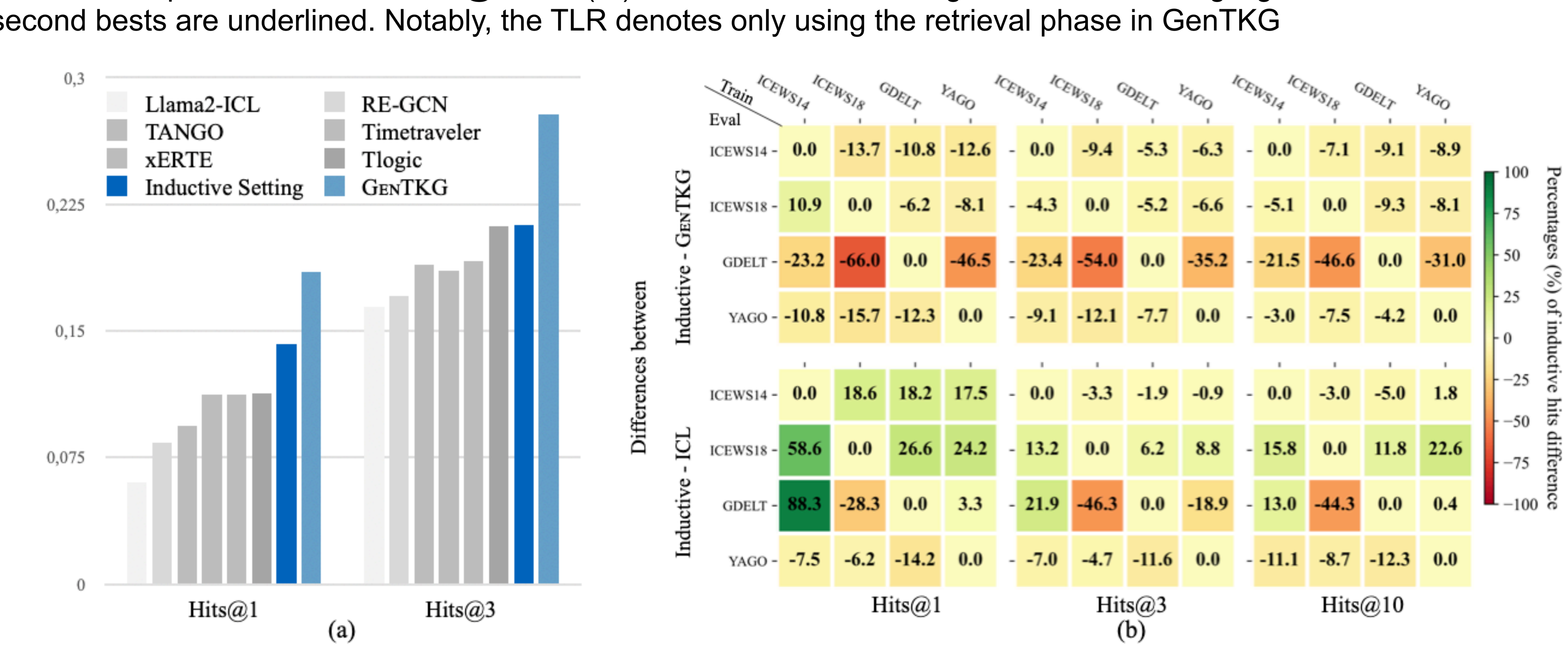


Figure 1. Cross-Domain Inductive Setting.
(a) Single dataset evaluation. All models including GenTKG are trained and evaluated on the GDELT dataset, except that the inductive setting of GenTKG is trained on ICEWS14 and evaluated on GDELT.
(b) Cross-checking. We cross-check the trained LLaMA2 in GenTKG on different training datasets and evaluation datasets. The performance drop compared to the original training setting takes up only small percentages. Even higher performance than ICL can be observed.

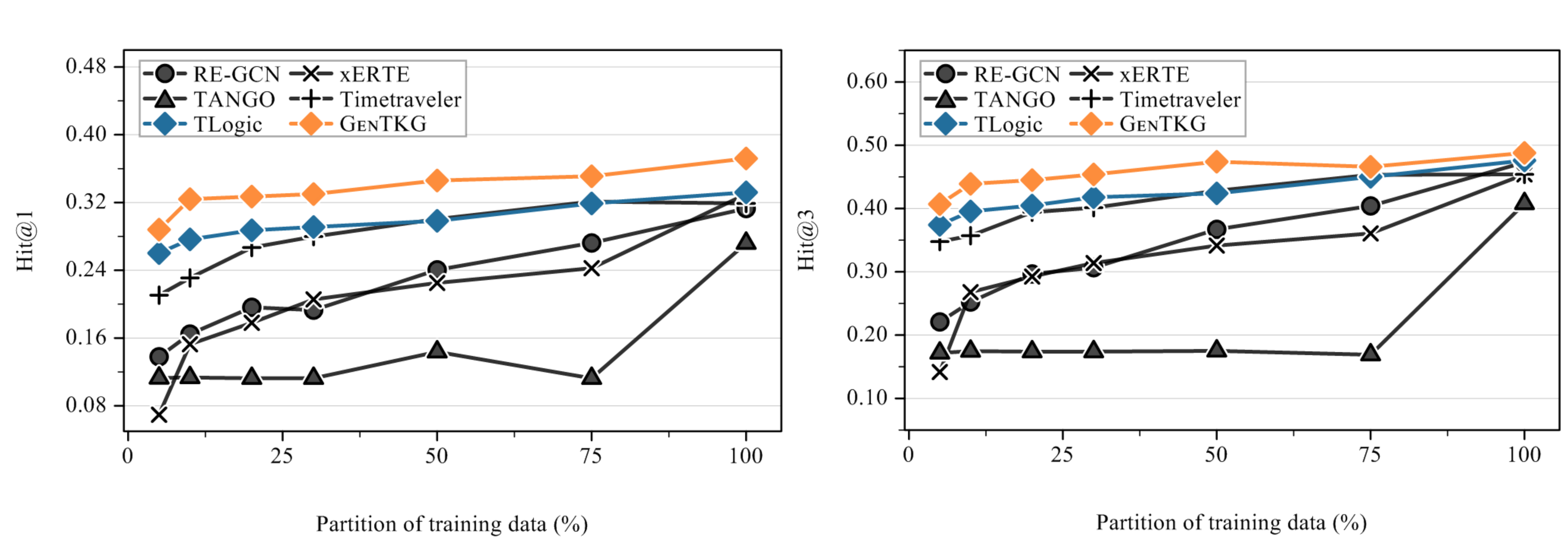


Figure 3. In-domain generalizability. GenTKG generalizes to different training partitions within the same datasets and exceeds conventional methods on all different partitions of training data on ICEWS14.

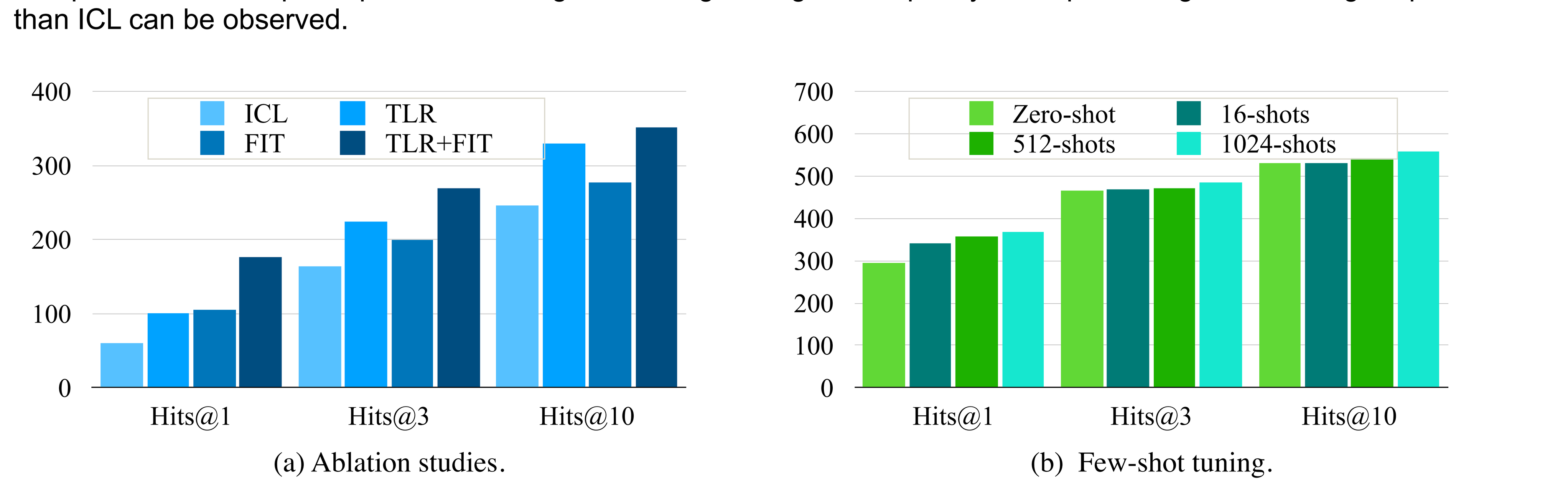


Figure 2. (a) Both TLR and FIT phase in GenTKG framework contributes significantly (b) Increasing few-shot training parameter K improves performance.

References

Dong-Ho Lee, Kian Ahrabian, Woojeong Jin, Fred Morstatter, and Jay Pujara. Temporal knowledge graph forecasting without knowledge using in-context learning. arXiv preprint arXiv:2305.10613, 2023.8.
Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. arXiv preprint arXiv:2106.09685, 2021.
Yushan Liu, Yunpu Ma, Marcel Hildebrandt, Mitchell Joblin, and Volker Tresp. Tlogic: Temporal logical rules for explainable link forecasting on temporal knowledge graphs. In Proceedings of the AAAI conference on artificial intelligence, volume 36, pages 4120–4127, 2022.