TAG-EQA: Text-And-Graph for Event Question Answering via Structured Prompting Strategies

Maithili Kadam and Francis Ferraro, University of Maryland, Baltimore County.

Key idea:

- ➢ Inject causal graph edges to guide reasoning
- ➢ Compare zero shot, few shot, and chain of thought
- ➢ Test text, graphs, and text and graph

Takeaways:

- ➢ Structure helps most in zero shot prompts
- ➢ CoT + graph offers the highest upside
- ➢ Biggest wins on causal and temporal questions

TAG-EQA PIPELINE

DATA PROCESSING

TEXT

Organizers state the two days of music, dancing, and speeches is expected to draw some two million people. But as supporters……..

QUESTION ANSWER SET

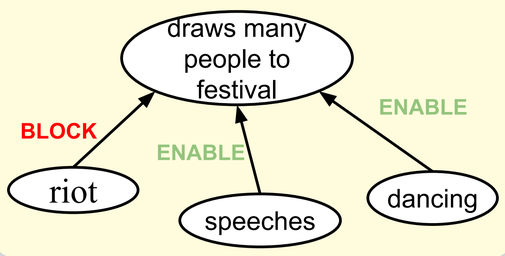
MAIN QUESTION

Did **"gathered"** happen while the organizers made a statement?

QA EXAMPLE

Question: Is "break" events have already finished?
Answer: no
Question: Is "called" events have already finished?
Answer: yes
Question: Is "dancing" events have already finished?
Answer: no

CAUSAL GRAPH



SERIALIZED GRAPH EDGES

- The event "dancing" **enables** the event "draws many people to festival".
- The event "speeches" **enables** the event "draws many people to festival".
- ……..

PROMPT ASSEMBLY

INSTRUCTION

You are provided with text and a causal graph. Analyze both step-by-step to answer the question. Follow the reasoning format below, then provide only "yes" or "no". Respond "unanswerable" if uncertain.

TEXT

Organizers state the two days of music, dancing, and speeches is expected to draw some two million people. But as supporters……

GRAPH

The event "dancing" enables the event "draws many people to festival". The event "speeches" enables the event "draws many people to festival"……..

QA EXAMPLE

Question: Is "break" events have already finished?
Answer: no
Question: Is "called" events have already finished?
Answer: ……

MAIN QUESTION

Did **"gathered"** happen while the organizers made a statement?

Instruction Tuned LLM (T5, QwQ, GPT)

MODEL OUTPUT

REASONING

The event "riot police deployed" blocks the event "protest rally".
The event "speeches" enables the event "draws many people to festival".
….
Therefore the answer is:

ANSWER

YES

RESULTS

RQ 1: Does reasoning (CoT) improve performance over Zero or Few-shot using just text?

Takeaway:
Use Few-shot by default

RQ 2: Are graphs helpful when used alone or combined with text?

Takeaway:
Zero-shot: Graphs > Text.
Few-shot: TAG lifts QwQ & GPT, drops on T5.

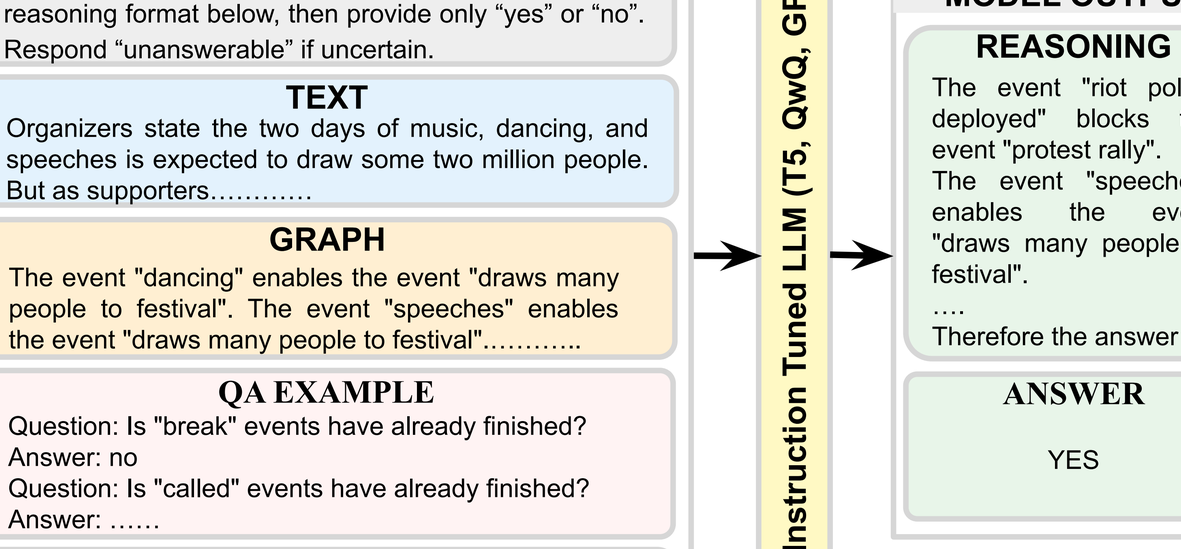
All results are on TORQUESTRA dataset.

Full size data = 53,061. Small size data = 1,024 .

GPT results shown on **small size dataset** only (cost)

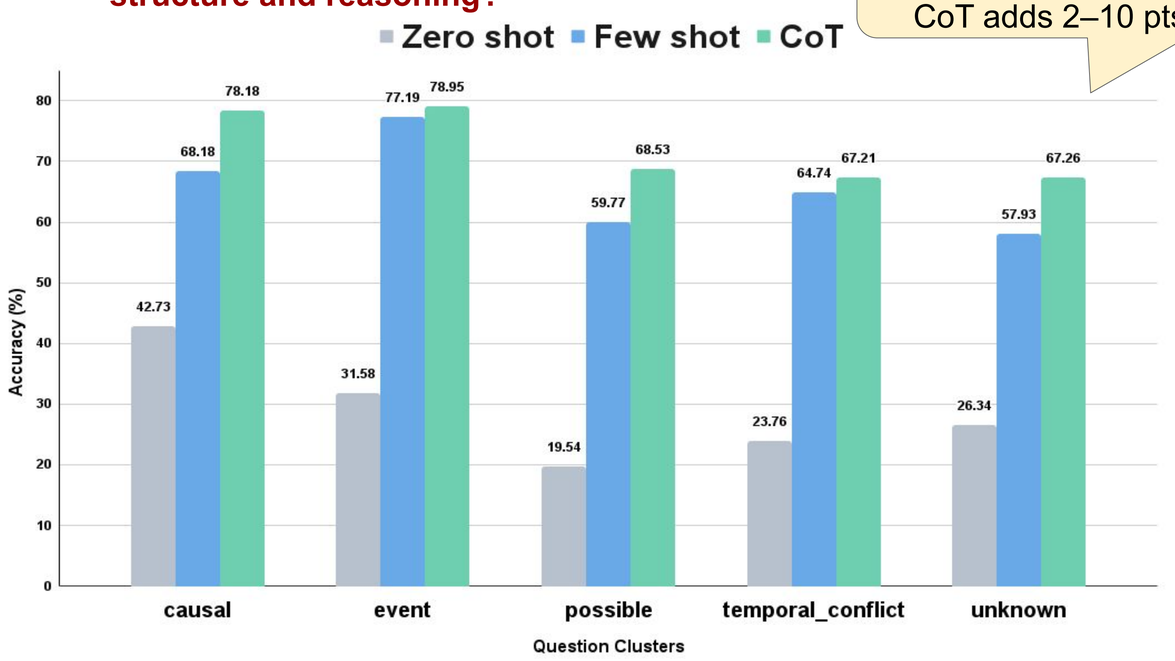
RQ 3: When reasoning is used, does adding graphs help or hurt?

Takeaway:
Few-shot lifts all clusters.
CoT adds 2–10 pts.



RQ 4: Which prompting strategy works best for each model?

MODEL	DATASET	BEST CONFIG	ACCURACY (%)
T5	Full	Few-shot + Text	58.49
QwQ	Full	Few-shot + Text And Graph	79.37
T5	Small	Few-shot + Text	59.47
QwQ	Small	Few-shot + Text	78.32
GPT**	Small	Zero-shot + Text	72.28

RQ 5: Which question types benefit most from structure and reasoning?

Takeaway:
Few-shot lifts all clusters.
CoT adds 2–10 pts.