GCoT: Chain-of-Thought Prompt Learning for Graphs

Xingtong Yu, Chang Zhou, Zhongwei Kuai, Xinming Zhang, Yuan Fang

SMU SINGAPORE MANAGEMENT UNIVERSITY

Background

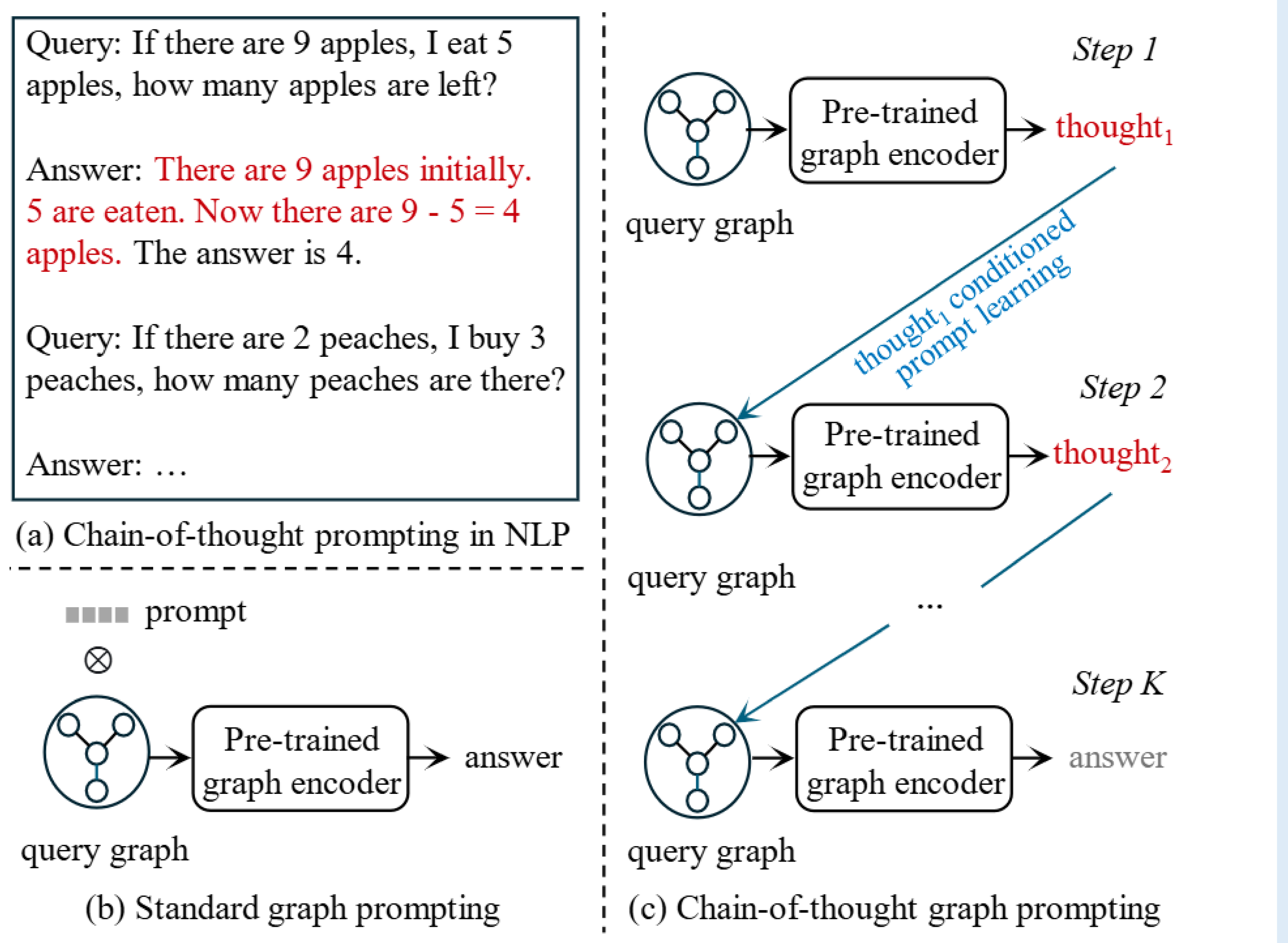

Figure 1: Illustration of chain-of-thought in NLP, standard graph prompts and graph chain-of-thought prompting.

- In NLP, Chain-of-Thought prompting improves reasoning by step-by-step inference.

↓

- Existing text-free graph learning methods produce a "final answer" in a single inference step.

↓

- Would introducing additional inference steps in a CoT style enhance the ability of pre-trained graph models to refine their predictions?

Challenges

- What should be the inference steps and thoughts for a graph task?
- Leveraging "thoughts" to generate prompts for subsequent inference is non-trivial.

Method：GCoT

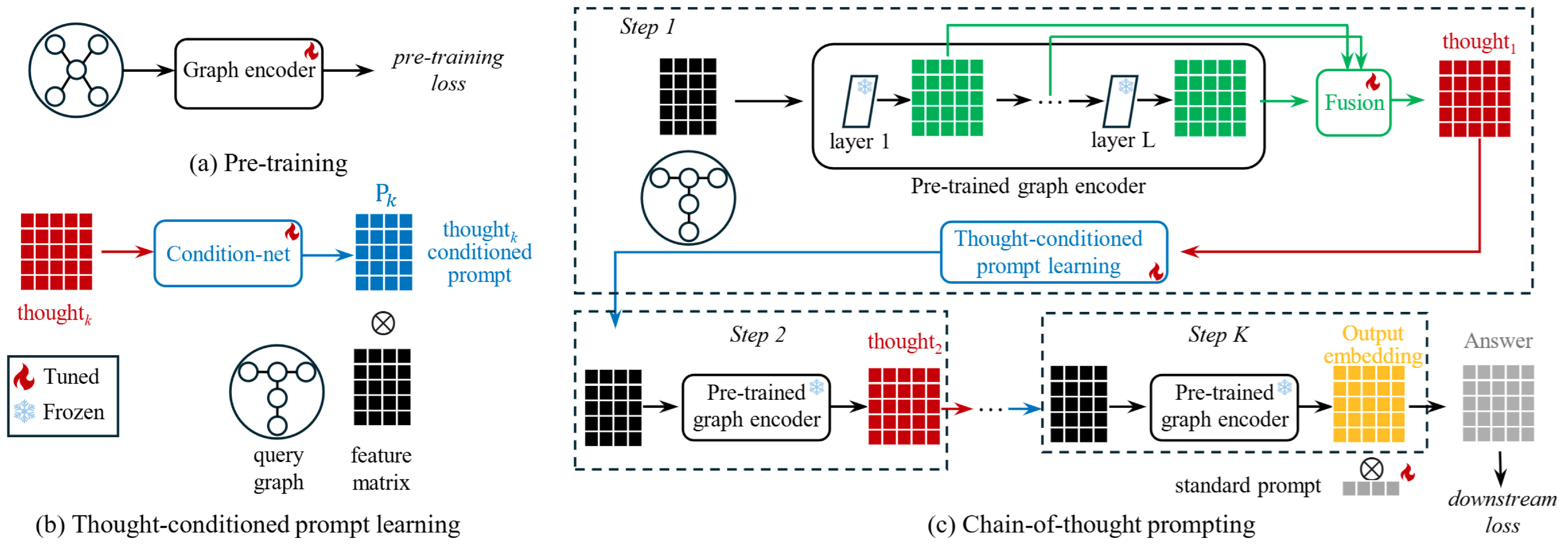

Figure 2: Overall framework of GCoT.

- **GCoT**: Prompt-based inference + Thought construction + Thought-conditioned prompt learning
- **Prompt-based inference** : Feed the graph and prompt into a frozen pre-trained graph encoder to obtain multi-layer node embeddings.
- **Thought construction** : Fuse hidden representations from all layers of the encoder to form a "thought" vector that captures hierarchical topological information.
- **Thought-conditioned prompt learning**: Use a condition-net to generate node-specific prompts based on the current thought, guiding the next inference step.

Results & Experiments

Methods	Node classification: Cora	Citeseer	Pubmed	Photo	Graph classification: MUTAG	COX2	BZR	PROTEINS
GCN	32.50 ± 14.21	26.36 ± 9.03	52.18 ± 8.70	60.18 ± 12.04	43.44 ± 15.14	50.95 ± 23.48	47.25 ± 16.59	40.28 ± 0.03
GAT	31.00 ± 16.22	27.71 ± 8.74	50.02 ± 8.88	51.79 ± 12.85	37.33 ± 10.81	50.58 ± 26.16	46.55 ± 16.57	40.39 ± 0.04
DGI/InfoGraph	54.11 ± 9.60	45.00 ± 9.19	47.46 ± 12.19	58.89 ± 10.97	53.17 ± 17.29	53.82 ± 14.19	49.33 ± 15.11	52.51 ± 10.29
GraphCL	51.96 ± 9.43	43.12 ± 9.61	46.80 ± 9.04	57.78 ± 11.31	54.92 ± 17.09	53.81 ± 14.21	49.73 ± 14.66	53.81 ± 8.97
ProG	50.59 ± 14.64	43.17 ± 8.49	63.07 ± 11.96	66.50 ± 9.46	51.99 ± 4.50	53.45 ± 15.01	53.52 ± 11.97	52.73 ± 6.57
GPF	<u>57.60</u> ± 13.88	43.11 ± 8.80	55.63 ± 10.96	65.29 ± 10.07	56.55 ± 13.95	54.16 ± 14.07	48.65 ± 13.96	53.05 ± 7.62
GPF+	57.42 ± 13.87	43.28 ± 8.82	57.16 ± 10.99	65.07 ± 10.01	<u>56.81</u> ± 12.93	<u>55.24</u> ± 13.29	50.83 ± 19.74	<u>54.58</u> ± 8.70
GraphPrompt	54.25 ± 9.38	<u>45.34</u> ± 10.53	<u>63.11</u> ± 10.01	<u>66.62</u> ± 9.90	55.44 ± 12.56	54.34 ± 14.77	<u>54.59</u> ± 10.52	53.80 ± 7.93
GCoT	**59.67** ± 15.51	**46.21** ± 8.78	**64.43** ± 9.96	**67.16** ± 10.46	**58.75** ± 15.42	**56.26** ± 15.52	**58.03** ± 23.44	**56.24** ± 8.60

- GCoT consistently outperforms all baselines across both node and graph classification tasks, in 1-shot settings.

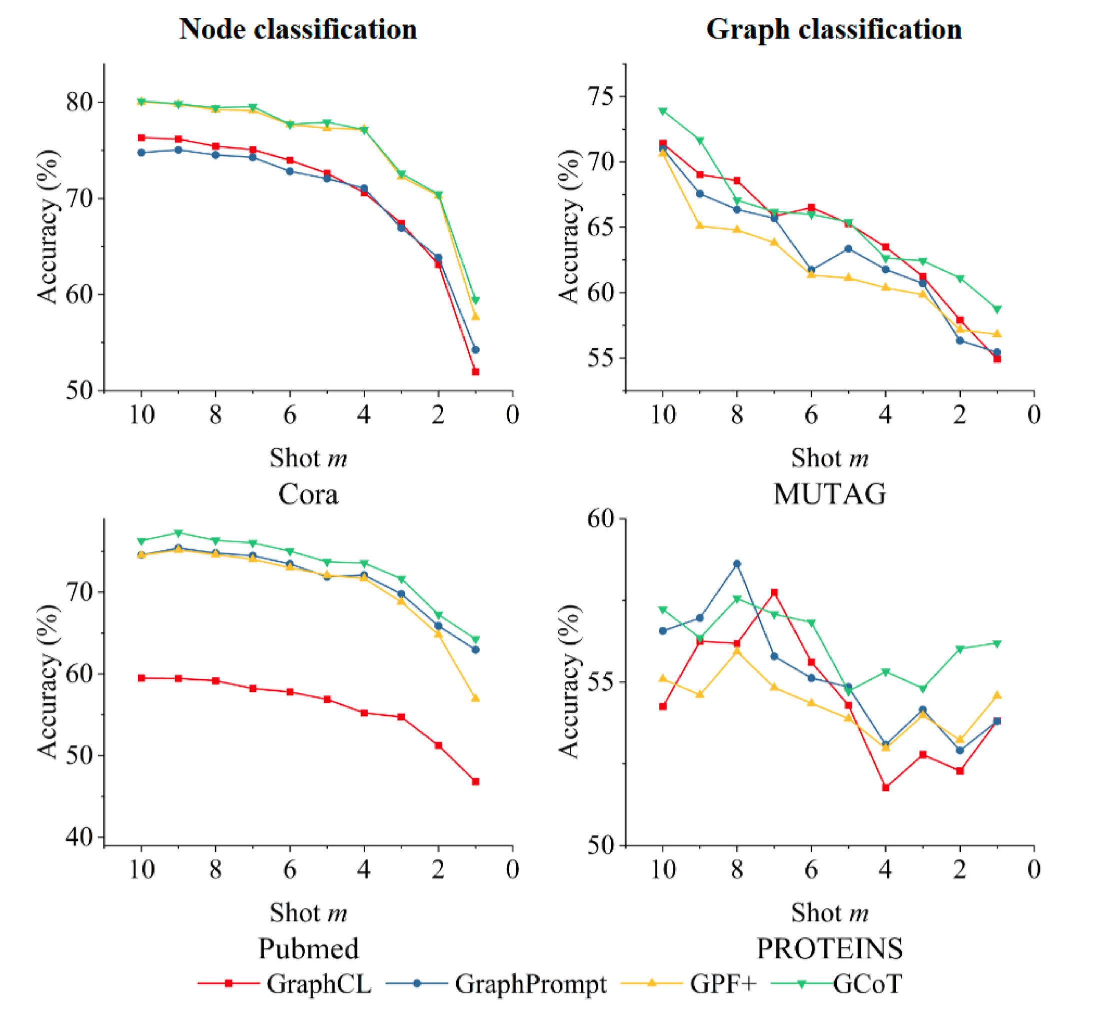

Figure 3: Impact of labeled data size (number of shots) on node and graph classification.

- GCoT consistently outperforms baselines, especially in low-shot settings ($m \leq 5$).
- As m increases, performance improves for all methods, but GCoT still achieves the best or near best results.

Table 3: Ablation study on the effects of key components.

Methods	Node classification: Cora	Pubmed	Graph classification: MUTAG	PROTEINS
GCoT\CoT	56.65±13.97	62.80±10.08	56.49±16.61	53.40±6.66
GCoT-L1	57.18±14.34	63.31±10.05	56.54±14.12	54.71±8.57
GCoT-L2	57.00±14.48	63.20±10.08	57.68±13.84	54.77±8.81
GCoT-L3	57.01±14.66	63.33±10.05	57.85±16.10	56.22±8.45
GCoT	**59.67**±15.51	**64.43**± 9.96	**58.75**±15.42	**56.24**±8.60

- GCoT outperforms all variants, confirming the benefit of step-by-step inference and multi-layer thought fusion.